

**Interaktion gedächtnis- und erklärungs-basierter
Verarbeitungsprozesse bei der pronominalen Auflösung**

Analyse der Effekte von Impliziten Kausalitäts- und Gender-Informationen durch die
Modellierung von Reaktionszeitverteilungen

Dissertation

vorgelegt der Philosophischen Fakultät der Universität zu Köln

von

Sascha Schroeder

A. Theoretischer Teil

1. Interaktion gedächtnis- und erklärungs-basierter Prozesse in der Sprachverarbeitung	1
1.1 Ausgangspunkt: Minimalistische vs. konstruktivistische Inferenztheorie.....	3
1.2 Erweiterung: Gedächtnis- vs. erklärungs-basierte Sprachverarbeitungstheorien.....	11
1.2.1 Gedächtnis-basierte Sprachverarbeitung.....	11
1.2.2 Erklärungs-basierte Sprachverarbeitung.....	21
1.3 Gedächtnis- und erklärungs-basierte Verarbeitung bei der pronominalen Auflösung.....	25
1.3.1 Linguistische Ansätze.....	25
1.3.2 Sprachpsychologische Forschung zur pronominalen Auflösung auf der Satzebene.....	30
1.3.3 Sprachpsychologische Forschung zur pronominalen Auflösung auf der Diskursebene.....	35
1.4 Integration der gedächtnis- und erklärungs-basierten Erklärungsperspektive.....	44
1.4.1 Fehlende Theorienkonkurrenz zwischen gedächtnis- und erklärungs-basierten Sprachverarbeitungstheorien.....	44
1.4.2 Allgemeine Ansätze zur Integration gedächtnis- und erklärungs-basierter Verarbeitungsprozesse.....	48
1.4.3 Interaktion gedächtnis- und erklärungs-basierter Prozesse bei der pronominalen Auflösung.....	51
1.5 Plan der Analyse.....	59
2. Implizite Kausalitäts- und Gender-Informationen als pronominale Auflösungs-Cues.....	61
2.1 Implizite Verbkausalität.....	67
2.1.1 Sozialpsychologie.....	67
2.1.2 Linguistik.....	79
2.1.3 Sprachpsychologie.....	100
2.1.3.1 Übergreifende Ergebnisse zum IC-Effekt.....	102
2.1.3.2 Der IC-Effekt als später Integrations-Effekt.....	105
2.1.3.3 Der IC-Effekt als früher Fokus-Effekt.....	112
2.1.4 Diskussion.....	119

2.2	<i>Gender-Informationen</i>	123
2.2.1	Gender-Informationen als später resolution-Faktor.....	124
2.2.2	Gender-Informationen als früher bonding-Faktor.....	127
2.2.3	Diskussion.....	131
2.3	<i>Interaktion der beiden Cues und Hypothesen</i>	133
3.	Erfassung gedächtnis- und erklärungs-basierter Verarbeitungsprozesse mittels Reaktionszeitverteilungsanalyse	141
3.1	<i>Reaktionszeitmethoden in der Psychologie: Möglichkeiten und Grenzen</i>	141
3.2	<i>Jenseits des Mittelwerts: Modellierung von Reaktionszeitverteilung</i>	149
3.2.1	Probleme der mittelwerts-basierten Auswertung.....	150
3.2.2	Methoden der Reaktionszeitmodellierung.....	156
3.3	<i>Die Weibull-Verteilung als RT-Verteilungsmodell</i>	160
3.3.1	Eigenschaften der Weibull-Verteilung.....	160
3.3.2	Die Ex-Gauss-Verteilung und ihre Beziehung zur Weibull-Verteilung.....	166
3.4	<i>Schätzung der Weibull-Verteilung mittels Hierarchischer Bayes-Verfahren</i>	170
3.5	<i>Evaluation der Modellpassung</i>	174
3.6	<i>Kognitive Interpretation der Parameter der Ex-Gauss- und der Weibull-Verteilung</i>	176
4.	Simulationsstudien	185
4.1	<i>Einleitung</i>	185
4.2	<i>Simulationsstudie I</i>	188
4.2.1	Methode.....	188
4.2.2	Ergebnisse.....	193
4.2.2.1	Konvergenz.....	193
4.2.2.2	Individualparameter.....	196
4.2.2.2.1	Bias.....	197
4.2.2.2.2	RMSE.....	199
4.2.2.2.3	Zusammenfassung.....	202
4.2.2.3	Populationsparameter.....	202
4.2.2.3.1	Bias.....	202
4.2.2.3.2	RMSE.....	204
4.2.2.3.3	Zusammenfassung.....	206

4.3 <i>Simulationsstudie II</i>	207
4.3.1 Methode.....	207
4.3.2 Ergebnisse.....	208
4.3.2.1 Konvergenz.....	208
4.3.2.2 Individualparameter.....	211
4.3.2.2.1 Bias.....	211
4.3.2.2.2 RMSE.....	213
4.3.2.2.3 Zusammenfassung.....	214
4.3.2.3 Populationsparameter.....	214
4.3.2.3.1 Bias.....	215
4.3.2.3.2 RMSE.....	216
4.3.2.3.3 Zusammenfassung.....	217
4.4 <i>Diskussion</i>	218
5. Material	221
5.1 <i>IC-Verben</i>	221
5.1.1 Verbauswahl.....	221
5.1.2 Voruntersuchung I: Kontrolle IC-Bias.....	225
5.2 <i>Stimulussätze</i>	228
5.2.1 Konstruktionsprinzip der Stimulussätze.....	228
5.2.2 Voruntersuchung II: Plausibilität und Stimulusqualität.....	236
B. Empirischer Teil	
6. Experiment Ia	242
6.1 <i>Einleitung</i>	242
6.2 <i>Methode</i>	245
6.2.1 Stichprobenbeschreibung und -bereinigung.....	245
6.2.2 Material.....	246
6.2.3 Untersuchungsablauf.....	247
6.2.4 Design und Abhängige Variablen.....	249

6.3 <i>Ergebnisse</i>	249
6.3.1 Pronominale Auflösungsaufgabe.....	249
6.3.1.1 Akkuratheit: Anteil intendierter pronominaler Auflösungen.....	249
6.3.1.2 Mittlere Reaktionszeit.....	251
6.3.1.3 Reaktionszeitverteilung: Weibull-Parameter.....	253
6.3.1.4 Diskussion pronominale Auflösungsaufgabe.....	258
6.3.2 Verifikationsaufgabe.....	259
6.3.2.1 Akkuratheit: Anteil korrekt verifizierter Aussagen.....	259
6.3.2.2 Mittlere Reaktionszeit.....	260
6.3.2.3 Reaktionszeitverteilung: Weibull-Parameter.....	261
6.3.2.4 Diskussion Verifikationsaufgabe.....	264
6.4 <i>Diskussion Experiment Ia</i>	265
7. Experiment Ib	269
7.1 <i>Einleitung</i>	269
7.2 <i>Methode</i>	272
7.2.1 Stichprobenbeschreibung und -bereinigung.....	272
7.2.2 Material.....	273
7.2.3 Untersuchungsablauf.....	275
7.2.4 Design und Abhängige Variablen.....	275
7.3 <i>Ergebnisse</i>	275
7.3.1 Pronominale Auflösungsaufgabe.....	275
7.3.1.1 Akkuratheit: Anteil intendierter pronominaler Auflösungen.....	275
7.3.1.2 Mittlere Reaktionszeit.....	276
7.3.1.3 Reaktionszeitverteilung: Weibull-Parameter.....	278
7.3.1.4 Diskussion pronominale Auflösungsaufgabe.....	283
7.3.2 Verifikationsaufgabe.....	284
7.3.2.1 Akkuratheit: Anteil korrekt verifizierter Aussagen.....	284
7.3.2.2 Mittlere Reaktionszeit.....	286
7.3.2.3 Reaktionszeitverteilung: Weibull-Parameter.....	288
7.3.2.4 Diskussion Verifikationsaufgabe.....	290
7.4 <i>Diskussion Experiment Ib</i>	291

8. Experiment IIa	297
8.1 <i>Einleitung</i>	297
8.2 <i>Methode</i>	301
8.2.1 Stichprobenbeschreibung und -bereinigung.....	301
8.2.2 Material.....	302
8.2.3 Untersuchungsablauf.....	302
8.2.4 Design und Abhängige Variablen.....	303
8.3 <i>Ergebnisse</i>	304
8.3.1 Satzlesezeiten.....	304
8.3.1.1 Mittlere Lesezeiten.....	304
8.3.1.1.1 Datenbereinigung und Leszeitprofil.....	304
8.3.1.1.2 Ergebnisse auf der Wortebene.....	305
8.3.1.1.3 Ergebnisse auf der Konstituentenebene.....	309
8.3.1.1.4 Reanalyse der „korrekt“ beantworteten Items.....	313
8.3.1.1.5 Zusammenfassung.....	313
8.3.1.2 Reaktionszeitverteilung: Weibull-Parameter.....	314
8.3.1.2.1 Parameterschätzung, Modellgüte und Parameterprofile.....	314
8.3.1.2.2 Ergebnisse auf der Wortebene.....	316
8.3.1.2.3 Ergebnisse auf der Konstituentenebene.....	328
8.3.1.2.4 Reanalyse der „korrekt“ beantworteten Items.....	332
8.3.1.2.5 Zusammenfassung.....	332
8.3.1.3 Diskussion Satzlesezeiten.....	332
8.3.2 Verifikationsaufgabe.....	333
8.3.2.1 Akkuratheit: Anteil „korrekt“ verifizierter Aussagen.....	333
8.3.2.2 Mittlere Reaktionszeiten.....	334
8.3.2.3 Reaktionszeitverteilung: Weibull-Parameter.....	335
8.3.2.4 Diskussion Verifikationsaufgabe.....	337
8.4 <i>Diskussion Experiment IIa</i>	338

9. Experiment IIb	343
9.1 <i>Einleitung</i>	343
9.2 <i>Methode</i>	345
9.2.1 Stichprobenbeschreibung und -bereinigung.....	345
9.2.2 Material, Untersuchungsablauf, Design und Abhängige Variablen	345
9.3 <i>Ergebnisse</i>	346
9.3.1 Satzlesezeiten.....	346
9.3.1.1 Mittlere Lesezeiten.....	346
9.3.1.1.1 Datenbereinigung und Leszeitprofil.....	346
9.3.1.1.2 Ergebnisse auf der Wortebene.....	347
9.3.1.1.3 Ergebnisse auf der Konstituentenebene.....	352
9.3.1.1.4 Reanalyse der „korrekt“ beantworteten Items.....	355
9.3.1.1.5 Zusammenfassung.....	355
9.3.1.2 Reaktionszeitverteilung: Weibull-Parameter.....	355
9.3.1.2.1 Parameterschätzung, Modellgüte und Parameterprofile.....	355
9.3.1.2.2 Ergebnisse auf der Wortebene.....	357
9.3.1.2.3 Ergebnisse auf der Konstituentenebene.....	368
9.3.1.2.4 Reanalyse der „korrekt“ beantworteten Items.....	371
9.3.1.2.5 Zusammenfassung.....	372
9.3.1.3 Diskussion Satzlesezeiten.....	372
9.3.2 Verifikationsaufgabe.....	373
9.3.2.1 Akkuratheit: Anteil „korrekt“ verifizierter Aussagen.....	373
9.3.2.2 Mittlere Reaktionszeiten.....	374
9.3.2.3 Reaktionszeitverteilung: Weibull-Parameter.....	375
9.3.2.4 Diskussion Verifikationsaufgabe.....	377
9.4 <i>Diskussion Experiment IIb</i>	377

10. Zusammenfassung und Diskussion der Ergebnisse	381
<i>10.1 Zusammenfassung und Integration der Ergebnisse aus den Experimenten I und II</i>	383
10.1.1 Gender-Effekte.....	383
10.1.2 IC-Effekte.....	386
10.1.3 Relation und Interaktion der verschiedenen IC-Teileffekte.....	391
<i>10.2 Methodische Reflexion: Analyse von Reaktionszeitverteilungen</i>	396
10.2.1 Stabilität des Verfahrens und Abhängigkeiten in den Parameterschätzungen.....	397
10.2.2 Konstruktvalidität der Parameter.....	403
<i>10.3 Ausblick: Interaktion gedächtnis- und erklärungs-basierter Verarbeitungsprozesse bei der pronominalen Auflösung</i>	407
Literatur	414
Appendix A: Statistische Ausführungen zu Kap. 3	A1
<i>A.1 Location-Scale-Verteilungen: Beispiel Normalverteilung</i>	A1
<i>A.2 Location-Scale-Shape-Verteilungen: Beispiel Weibull-Verteilung</i>	A4
<i>A.3 Momente: Schiefe und Kurtosis</i>	A6
<i>A.4 Methoden zur Schätzung der Weibull-Verteilung</i>	A8
A.4.1 Moment-Methode.....	A8
A.4.2 Maximum Likelihood-Schätzung.....	A8
A.4.3 Bayes-Schätzung.....	A14
<i>A.5 Vergleich verschiedener Verfahren zur Schätzung der Weibull-Verteilung</i>	A18
Appendix B: Stimulusmaterial	B1
Appendix C: Zusatzanalysen zu den Experimenten I und II	C1
<i>C.1 Zusatzanalysen Experiment Ia</i>	C1
<i>C.2 Zusatzanalysen Experiment Ib</i>	C5
<i>C.3 Zusatzanalysen Experiment IIa</i>	C8
<i>C.4 Zusatzanalysen Experiment IIb</i>	C16

Diese Schrift wurde im Wintersemester 2007/08 von der Philosophischen Fakultät der Universität zu Köln als Dissertation angenommen. Die Disputation fand am 5.12.2007 statt.

1. Interaktion gedächtnis- und erklärungs-basierter Prozesse in der Sprachverarbeitung

Sprachverstehen ist ein ausgesprochen komplexes Phänomen, welches viele verschiedene, teilweise sehr heterogene kognitive Prozesse gleichzeitig umfasst und auf vielen Hierarchie-Ebenen angesiedelt ist. Nach Kintsch (2005) können die beteiligten Prozesse in ein allgemeines Spannungsfeld eingeordnet werden, das von eher automatischen, bottom-up-Prozessen (wie sie z.B. für viele Wahrnehmungsphänomene charakteristisch sind) bis hin zu kontrollierten top-down-Prozessen reicht (für die vielleicht das aktive Problemlösen ein prototypisches Beispiel ist). Obwohl ein allgemeiner Konsens darüber besteht, dass beide Arten von Prozessen beim Verstehen von Sprache (wie auch bei den meisten anderen kognitiven Phänomenen) eine Rolle spielen (müssen), kommt es aufgrund der (notorischen) Unbestimmtheit des Verstehensbegriffs sowie der Vielschichtigkeit und Komplexität der am Verstehen von Sprache beteiligten Prozesse meist dazu, dass verschiedene (Teil-)Theorien der Sprachpsychologie hier sehr unterschiedliche Schwerpunktsetzungen vornehmen: Während einige Sprachverstehen eher als einen automatischen Mustererkennungsprozess konzeptualisieren, gehen andere Theorien eher von einer prinzipiellen Analogie zu höheren kognitiven Prozessen, wie z.B. dem aktiven Problemlösen, aus. Prozesse der ersten Art, die (vorläufig) lediglich durch Schlagworte wie *hierarchie-niedrig*, *automatisch*, *bottom-up* etc. charakterisiert werden sollen, werden im Kontext der vorliegenden Arbeit als *gedächtnis-basierte* Sprachverarbeitungsprozesse bezeichnet. Prozesse der zweiten Art, für die sich kontrastiv in der Literatur die Bezeichnung *erklärungs-basiert* etabliert hat (vgl. im Überblick Guéreaud & O'Brien, 2005a), sind hingegen vergleichsweise *hierarchie-hoch*, *kontrolliert* und *top-down*.

In der vorliegenden Arbeit wird davon ausgegangen, dass beide Arten von Prozessen für das Sprachverstehen gleichermaßen relevant sind und deshalb auch gemeinsam von einer umfassenden Theorie des Sprachverstehens berücksichtigt werden sollten. So allgemein formuliert ist diese Aussage relativ unkontrovers: Niemand bestreitet ernsthaft, dass das Verstehen von Sprache sowohl bottom-up- als auch top-down-Anteile enthält. Wenn wir z.B. einen Text lesen, dann sind in diesem immer bestimmte sprachliche Informationen enthalten, die uns vorher so nicht bekannt waren und vielmehr bottom-up dem Text entnommen werden müssen (dies ist ja ein möglicher Grund dafür, warum wir den Text überhaupt lesen wollten). Gleichzeitig wären diese neuen Inhalte jedoch vollkommen unintelligibel, wenn wir nicht schon bestimmte Wissensstrukturen und Verstehensstrategien hätten, die uns ein Verständnis der eintreffenden Informationen ermöglichen oder zumindest erleichtern. Darüber hinaus wird jedoch in der vorliegenden Arbeit die weitergehende These vertreten, dass die beiden Prozess-

klassen nicht nur (unabhängig voneinander) zum Sprachverstehen beitragen, sondern vielmehr eng miteinander verzahnt und aufeinander bezogen sind. Sie ergänzen sich gegenseitig und setzen einander wechselseitig voraus. Auch diese Aussage erscheint zunächst vergleichsweise harmlos, vielleicht sogar trivial. Es wird jedoch noch gezeigt, dass diese Auffassung, wenn man sie einmal konsequent zu Ende denkt, zu einer Konzeption des Verstehensprozesses führt, die sich teilweise erheblich von einer Reihe, aktuell vertretenen Theorien der Sprachverarbeitung unterscheidet. Konkret soll dies in der vorliegenden Arbeit für die pronominale Auflösung gezeigt werden.

Ein harter Indikator dafür, dass eine dynamisch-integrative Verbindung von gedächtnis- und erklärungs-basierten Verarbeitungsprozessen noch nicht selbstverständlich ist, besteht schon allein darin, dass es in der *scientific community* (wenn man von den üblichen Lippenbekenntnisse einmal absieht) eine vergleichsweise stark ausgeprägte „Arbeitsteilung“ gibt, was die beiden beschriebenen Prozessklassen angeht. Dies manifestiert sich zunächst einmal darin, dass sich einzelne Forscher/innen schwerpunktmäßig und isoliert immer nur mit einem der beiden Prozesstypen befassen und den jeweils anderen aus ihrer Betrachtung ausklammern. Wissenschaftstheoretisch bedeutsamer ist jedoch v.a., dass mit dieser unterschiedlichen Schwerpunktsetzung teilweise auch sehr verschiedene Grundauffassungen bezüglich des zu untersuchenden Gegenstandsbereichs verbunden sind, die u.a. bestimmen, was überhaupt die relevanten, zu untersuchenden Phänomene sind, welche Erklärungsstrukturen für diese zur Verfügung stehen und wie der Forschungsprozess konkret abzulaufen hat. Diese holistische Gesamtkonstellation ganz unterschiedlicher determinierender Vorannahmen wird z.B. darin deutlich, dass in der relevanten Literatur nicht nur einfach von gedächtnis- und erklärungs-basierten Verarbeitungsprozessen gesprochen wird, sondern sogar von einem *memory-based approach* (z.B. Gerrig & McKoon, 1998) oder dem sog. *explanation-based view* (z.B. O'Brien & Myers, 1999). Schon an solchen, vergleichsweise oberflächlichen Formulierungen wird deutlich, dass sehr viel mehr auf dem Spiel steht, als einfach nur ein möglichst adäquates Erklärungsmodell für einen bestimmten Objektbereich (nämlich dem des Sprachverstehens) zu finden. Vielmehr scheint es sich bei den beiden Ansätzen um Forschungsprogramme im Herrmannschen Sinne (vgl. z.B. Herrmann, 1976; zusammenfassend: Herrmann, 1994) zu handeln, und zwar mit allen entsprechenden negativen und positiven Heuristiken der Theorieentwicklung (vgl. Lakatos, 1970; im Überblick auch Gethmann, 1995).

Daraus ergibt sich, dass der Versuch einer rein additiven Zusammenführung der Forschungsergebnisse der beiden Strömungen zunächst einmal zum Scheitern verurteilt ist. Vielmehr müssen zunächst die strukturellen Grundannahmen beider Ansätze begrifflich aufgear-

beitet werden, damit sich ihre Ergebnisse nachfolgend sinnvoll integrieren lassen (vgl. auch Groeben, 2003). Eine wissenschaftshistorische Einbettung und konzeptuelle Erläuterung der mit beiden Perspektiven verbundenen Erklärungskonzepte ist entsprechend auch das Hauptanliegen der ersten Hälfte des nachfolgenden Kapitels. Dies soll jedoch nur unter der strukturierenden Zielperspektive einer möglichen Zusammenführung der beiden Ansätze, für die in einem zweiten Schritt dann verschiedene Möglichkeiten vergleichend diskutiert werden. Obwohl die Darstellung damit zunächst noch auf einem vergleichsweise hohen, metatheoretischen Niveau angesiedelt ist (und sein muss), soll möglichst früh der Bezug zum anvisierten Objektbereich der vorliegenden Arbeit, nämlich der pronominalen Auflösung gesucht werden. Einerseits, indem schon bei der einführenden Darstellung der beiden Theoriestränge direkt auf die konkreten theoretischen Implikationen eingegangen wird, die sich aus der Anwendung der beiden Perspektiven auf den Prozess der pronominalen Auflösung ergeben. Andererseits wird abschließend jedoch auch dargestellt, wie sich die Vorstellungen des pronominalen Auflösungsprozesses durch die integrative Zusammenführung der beiden Perspektiven verändern und welche Konsequenzen und Restriktionen dies für den Versuch einer empirischen Untersuchung mit sich bringt.

1.1 Ausgangspunkt: Minimalistische vs. konstruktivistische Inferenztheorie

Ausgangspunkt für die Entstehung des sog. *memory-* und *explanation-based view* ist die Minimalismusdebatte der 1990er Jahre, die dem Umkreis der Inferenzforschung entstammt. Die in dieser Debatte vertretenen Positionen des Minimalismus und des Konstruktivismus können als direkte Vorläufer des gedächtnis- und erklärungs-basierten Ansatzes betrachtet werden, da sich beide Ansätze direkt aus den beiden Positionen entwickelt haben und bis heute sowohl personell als auch argumentativ große Überlappungsbereiche bestehen. Allerdings muss ebenso klar darauf hingewiesen werden, dass der Minimalismus und der *memory-based-view* bzw. der Konstruktivismus und der *explanation-based-view* trotz aller Gemeinsamkeiten und Überschneidungspunkte dennoch unterschiedliche Ansätze sind, die sich sehr wohl voneinander abgrenzen lassen. Denn interessanterweise sind beide Weiterentwicklungen sowohl durch eine Eingrenzung als auch durch eine Ausweitung der Grundaussagen der beiden ursprünglichen Positionen gleichzeitig gekennzeichnet.

Wie erwähnt steht die Minimalismus-Debatte im Kontext der Inferenzforschung. Spätestens seit den einschlägigen Arbeiten von Bartlett (1932/1995) oder Bransford und Franks (1971) gibt es einen prinzipiellen Konsens darüber, dass die menschliche Sprachverarbeitung ein hoch konstruktiver Akt ist, bei dem eintreffende sprachliche Informationen durch zusätzli-

che Informationen angereichert und ergänzt werden (vgl. z.B. Bransford, Barclay & Franks, 1972; für den sprachlichen Bereich speziell vgl. z.B. Groeben & Christmann, 1997; zur kognitiven Konstruktivität des Menschen allgemein vgl. z.B. Groeben, 1997; Erb, 1997). Solche über den reinen Textinhalt hinausgehenden Hinzufügungen und Ergänzungen werden *Inferenzen* genannt, wobei dieser Begriff aufgrund der Heterogenität der mit ihm gemeinten Phänomene meist bewusst sehr offen gehalten wird (vgl. Singer, 1994): Denn in dieser weiten Fassung umfasst der Inferenzbegriff sowohl vergleichsweise einfache, lokale und automatische Hinzufügen und Verknüpfungen (wie z.B. die Auflösung von Anaphern oder das Einfügen einfacher kausaler Konnektoren) als auch umfangreiche, wissensintensive und interindividuell variable Information (wie z.B. instrumentelle oder prädiktive Inferenzen).

Es ist wichtig festzuhalten, dass der Inferenzforschung somit inhärent ein starker top-down-Charakter anhaftet, da Inferenzen ja *per definitionem* etwas sind, was der mentalen Repräsentation eines Textes (oder allgemeiner einer sprachlichen Äußerung) aktiv rezipientenseitig hinzugefügt wird. Wenn man dies allerdings einmal konzidiert hat, dann stellt sich direkt anschließend die Frage, wie weit diese Konstruktivität nun geht. Denn zum einen ist klar, dass (im Normalfall) nicht jede/r Rezipient/in auf der Grundlage ein- und desselben Textes zu vollkommen unterschiedlichen mentalen Repräsentationen desselben kommen wird; es muss also bestimmte restringierende Textinformationen und generelle Strategien zu ihrer Verknüpfung geben, die so etwas wie eine geteilte Kerninterpretation gewährleisten (die darüber hinaus natürlich interindividuell variieren kann). Zum anderen steht die Inferenzforschung traditionell vor dem Problem, genau angeben zu können, welche Inferenzen denn nun bei der Rezeption eines bestimmten Textes gezogen werden und welche nicht. Denn angesichts der Tatsache, dass die Anzahl der möglichen Inferenzen unendlich groß ist, ist von vorneherein klar, dass prinzipiell nicht immer alle Inferenzen konstruiert werden können.

Ein Großteil der frühen Inferenzforschung der 1970er und 1980er Jahre ging jedoch implizit oder explizit von der Annahme aus, dass im Laufe der Textrezeption zunächst einmal die gesamte Menge möglicher Inferenzen zur Verfügung steht, die prinzipiell im Rahmen der gegebenen Ressourcen zur Konstruktion einer sehr reichhaltigen Textrepräsentation genutzt werden. Eine Hauptaufgabe der damaligen Inferenzforschung war entsprechend, angeben zu können, welche Inferenztypen eine höhere Konstruktionswahrscheinlichkeit haben als andere und welche textuellen und rezipientenseitigen Merkmale den Prozess der Inferenzgenerierung beeinflussen (vgl. im Überblick Singer, 1994). So hat sich z.B. herausgestellt, dass bestimmte instrumentelle Inferenzen (wie z.B., dass man Schnee mit einer Schaufel schippt; vgl. Corbett & Doshier, 1978) vergleichsweise selten gebildet werden und eine hohe interindividuelle Va-

riabilität aufweisen, während andere (z.B., dass Bier im nordamerikanischen Raum zur Grundausrüstung für ein Picknick gehört, vgl. Haviland & Clark, 1974) schnell und effizient verfügbar sind.

Vor dem Hintergrund einer solchen, implizit „maximalistischen“ Position ist auch verständlich, warum die sog. *Minimalismus-Hypothese* von McKoon und Ratcliff (1992) ein so großes Aufsehen erregte und zu einer derart starken, bis heute andauernden Debatte führte. Der Kern dieser Position lässt sich durch zwei sehr einfache Aussagen beschreiben:

In the absence of specific, goal-directed strategic processes, inferences of only two kinds are constructed: those that establish locally coherent representations of the parts of a text that are processed concurrently and those that rely on information that is quickly and easily available. (McKoon & Ratcliff, 1992, p. 440).

Diese Position ist wahrlich minimalistisch, und zwar in einem zweifachen Sinne: Zum einen dreht sie die „Bringschuld“ bzw. Begründungspflicht der bisherigen Inferenzforschung radikal um: Man muss nicht mehr begründen, *welche* Inferenzen in der mentalen Repräsentation eines Textes enthalten sind und welche nicht; stattdessen wird nun das Minimum als Referenzpunkt angesetzt. Das heißt, man muss nun begründen, *ob überhaupt* eine Inferenz generiert wurde. Zum anderen geht die minimalistische Inferenztheorie explizit davon aus, dass das Produkt der automatischen Textverarbeitung sehr viel informationsärmer (d.h. minimaler) ist, als vorher (implizit) angenommen wurde. Denn nach Aussage der Theorie werden Inferenzen nur dann automatisch konstruiert, wenn a) ein lokaler Kohärenzbruch vorliegt und b) die notwendigen Informationen zur Konstruktion der Inferenz direkt verfügbar sind.

So einfach die Grundaussage der minimalistischen Inferenztheorie zunächst auch erscheinen mag: Die sich anschließende Debatte hat deutlich gemacht, dass in ihr mehrere, prinzipiell unabhängige Aspekte enthalten sind, die sich durchaus auch voneinander trennen lassen. Die beiden wichtigsten Problemkreise sind a) Annahmen über Umfang und Inhalt des *Inferenzproduktes* und b) Annahmen über Art und Spezifität des *Inferenzprozesses*. In Bezug auf den Produktaspekt scheint der Minimalismus von einer vergleichsweise armen Repräsentation auszugehen, die zudem nur sehr spezielle Inferenztypen enthält; in Bezug auf den Prozessaspekt war in der nachfolgenden Diskussion v.a. die vergleichsweise starke bottom-up-Konzeption dieses Ansatzes dominant. Bevor jedoch auf beide Aspekte näher eingegangen werden soll, ist es empfehlenswert, zunächst die Gegenposition zur minimalistischen Inferenztheorie darzustellen, weil sich beide Problemaspekte besser vergleichend diskutieren lassen.

Denn wie nicht anders zu erwarten, wurde die minimalistische Inferenztheorie von Anfang an stark kritisiert (vgl. z.B. Garnham, 1992; Glenberg & Matthew, 1992; Singer, 1993).

Es lohnt sich jedoch wenig, an dieser Stelle auf die verschiedenen Kritikpunkte einzeln einzugehen (vgl. hierzu z.B. im Überblick Graesser, Millis & Zwaan, 1997; Richter, 2003). Wichtiger ist, dass sich im Laufe der zunächst sehr heterogenen und verstreut vorgebrachten Kritik langsam von verschiedenen Autoren die sog. *konstruktivistische* Inferenztheorie als explizite Gegenposition zum Minimalismus entwickelt wurde, welche die meisten der genannten Kritikpunkte zu bündeln versuchte und zu einem (mehr oder weniger) kohärenten Gesamtentwurf vereinte (vgl. Graesser, Singer & Trabasso, 1994; Singer, Graesser & Trabasso, 1994).

Der emphatische Kern der konstruktivistischen Inferenztheorie findet sich in den drei zentralen Annahmen, die sie macht (vgl. Graesser et al., 1994, pp. 371f.): 1. Der Verstehensprozess dient aktiv den Zielen, die Leser/innen mit bestimmten Texten verbinden (reader goal assumption), 2. Leser/innen konstruieren eine Textrepräsentation, die sowohl lokal als auch global kohärent ist (coherence assumption) und 3. Leser/innen versuchen (sich) zu erklären, warum bestimmte Handlungen und Ereignisse in einem Text vorkommen (explanation assumption). Alle drei Annahmen zusammen konvergieren im Prinzip des sog. *search after meaning* (bzw. *effort for meaning*), d.h. Leser haben das motivationale Bedürfnis, eine möglichst zufrieden stellende Erklärung für die in einem Text genannten Ereignisse zu finden und holen sich die hierfür relevanten Informationen aktiv ein. Damit betont die konstruktivistische Inferenztheorie sehr stark die Zielperspektive des Rezeptionsprozesses. Deshalb werden auch explizit Text- und Lesermerkmale mit in die Theoriebildung einbezogen (sog. Text-Leser-Interaktion sensu Groeben, 1982). Diese werden von der minimalistischen Inferenztheorie vollkommen ausgeblendet.

Insbesondere in Bezug auf den ersten, bereits oben angesprochenen Problemaspekt, nämlich der Frage, wie reichhaltig eigentlich die resultierende Textrepräsentation nach der Inferenzgenerierung ist, nimmt die konstruktivistische Inferenztheorie eine radikal andere Position ein als der Minimalismus. Zum einen besteht schon Uneinigkeit darüber, wie groß der Anteil online generierter Inferenzen ist und welche Klasse von Inferenzen automatisch konstruiert wird. Kern der minimalistischen Position ist, dass nur sehr wenige Inferenzen ohne die Hinzuziehung spezifisch strategischer Prozesse konstruiert werden. Die dominante Gegenstrategie der konstruktivistischen Position besteht hingegen darin nachzuweisen, dass eine größere Teilmenge von Inferenzen online während der Textrezeption generiert wird, als von der minimalistischen Theorie vorhergesagt wird (vgl. z.B. Graesser et al., 1994).

Große Bedeutung kommt dabei der Tatsache zu, dass die beiden Modelle divergierende Vorhersagen dazu machen, welche Informationen Leser/innen abrufen (konstruktivistische Formulierung) bzw. welche Informationen Lesern/innen zur Verfügung stehen (minimalisti-

sche Formulierung), wenn ein Text lokal kohärent ist. Die minimalistische Position macht hier die starke Vorhersage, dass das primäre Ziel der Textverarbeitung die Etablierung lokaler Kohärenz ist und ohne lokalen Kohärenzbruch keine zusätzlichen Informationen abgerufen werden. Ansonsten arbeitet der Rezeptionsprozess lediglich mit den aktuell verfügbaren Informationen, die gerade im Arbeitsgedächtnis aktiv sind. Die konstruktivistische Position geht hingegen davon aus, dass die Inferenzgenerierung sowohl der lokalen als auch der globalen Kohärenzbildung dient. Sie macht deswegen die Vorhersage, dass auch (übergeordnete) Ziele der Protagonisten, kausale Ursacheninformationen für Ereignisse in einer Geschichte und (globale) thematische Inferenzen online konstruiert werden.

Aufgrund dieser klaren Divergenz der Vorhersagen wird mitunter davon ausgegangen, dass sich die beiden Theorien an dieser Stelle in einer Art *experimentum crucis* direkt miteinander konfrontieren lassen (Clifton & Duffy, 2001). Und in der Tat werden die Ergebnisse einiger Experimente häufig so interpretiert, als würden sie die minimalistische Inferenztheorie gefährden, wenn nicht sogar falsifizieren (vgl. im Überblick Singer et al., 1994): So konnten z.B. Suh und Trabasso (1993) zeigen, dass Informationen über noch nicht erfüllte Ziele eines Protagonisten auch dann aktiviert werden, wenn kein lokaler Kohärenzbruch vorliegt. Darüber hinaus haben z.B. Lutz und Radvansky (1997) nachgewiesen, dass Informationen zu nicht-erfüllten Protagonisten-Zielen eher und stärker aktiviert werden, als wenn diese bereits erfüllt worden waren (vgl. auch Albrecht & Myers, 1995).

Obwohl solche und ähnliche Belege sicherlich eindrucksvoll darauf hinweisen, dass Ziel und Motive von Protagonisten wichtige Informationen sind, die von Lesern/innen verfolgt und auch über große Textmengen hinweg noch reaktiviert werden können, ist fraglich, ob damit wirklich die Frage entschieden ist, ob nun die minimalistische oder die konstruktivistische Inferenztheorie in der Realität gilt. Der übergeordnete Grund für diese Skepsis, auf den in Abschnitt 1.4 noch ausführlich eingegangen werden soll, liegt darin, dass zwischen den beiden Positionen m.E. lediglich eine Schein-Dichotomie besteht, da sie sich auf prinzipiell unterschiedliche Phänomenbereiche beziehen. Deswegen sind sie auch nur partiell empirisch miteinander zu konfrontieren und alle Versuche, zwischen den beiden Positionen *entscheiden* zu wollen, sind deswegen von Vorneherein zum Scheitern verurteilt. Ziel sowohl der Theoriebildung als auch der empirischen Forschung sollte eher sein, die komplementären Aspekte der beiden Ansätze herauszuarbeiten und in einem gemeinsamen Rahmenmodell konstruktiv miteinander zu verbinden.

Konkret manifestiert sich die Problematik der Nichtkonfrontierbarkeit der beiden Theorien im vorliegenden Fall darin, dass sämtliche Evidenz, die gegen den ersten Annahmenkern

der minimalistische Position vorliegt, nämlich den Umfang des inferierten Textproduktes, solange inkonklusiv ist, solange keine Entscheidung bezüglich des zweiten oben angesprochenen Problemkreise vorliegt, nämlich welche Prozesse der Inferenzgenerierung von den beiden Theorien überhaupt modelliert werden (sollen). Dieser zweite Annahmestern bezieht sich also auf die Frage, welche Prozesse der Inferenzgenerierung eigentlich zugrunde liegen, d.h. ob für sie eher passiv-automatische bottom-up-Prozesse wie bei der Wahrnehmung verantwortlich sind oder aber aktiv-strategische top-down-Prozesse wie bei der Problemlösung angesetzt werden müssen.

Eng damit verbunden ist die Fragestellung, ob die bei der Inferenzgenerierung eingesetzten Prozesse eigentlich für das Verstehen von Sprache spezifisch oder vielmehr allgemeiner Natur sind. Denn dann lässt sich die Frage sogar noch etwas präziser dahingehend formulieren, wann Textverstehen eigentlich in (allgemeines) Problemlösen übergeht (vgl. Gerrig & O'Brien, 2005). Inwieweit muss eine *sprachpsychologische* Theorie überhaupt Prozesse in ihre Theoriebildung mit einbeziehen, obwohl sie viel eher in eine allgemeine Theorie des Denkens und Problemlösens gehören? Das heißt, letztlich geht es um die gegenstandskonstituierende Fragestellung, für welche Phänomene eine Theorie der Sprachverarbeitung eigentlich (noch) verantwortlich zeichnen muss und ab welchem Punkt sie den Schwarzen Peter weiterreichen kann? Keine Theorie des Sprachverstehens kann alle Prozesse, die irgendwie mit dem Sprachsystem in Verbindung stehen gleichzeitig berücksichtigen (vgl. Groeben & Christmann, 1997). Denn Sprache ist ein für die menschliche Existenz derart zentrales Phänomen, dass dies langfristig darauf hinausläufe, eine eigene „Psychologie in der Psychologie“ zu betreiben und sämtliche kognitiven, motivationalen und emotionalen Problemstellungen noch einmal sprachspezifisch aufzuarbeiten. Alle Theorien des Sprachverstehens müssen also früher oder später sowohl einen Übergangspunkt dafür bestimmen, wann hierarchie-niedrige bottom-up-Prozesse in hierarchie-hohe top-down-Prozesse übergehen, als auch, welche dieser Prozesse eigentlich für das Sprachverstehen spezifisch sind. Beide Grenzziehungen werden von der minimalistischen und der konstruktivistischen Inferenztheorie sehr unterschiedlich vorgenommen; interessanter Weise in jeweils unterschiedliche Richtung.

Die minimalistische Position bezieht sich ganz dezidiert nur auf solche Prozesse, die sie selbst als „automatisch“ bezeichnet. Dieser Begriff wird von McKoon und Ratcliff (1992) ganz im Sinne der klassischen Automtizitäts-Konzeption von Posner und Snyder (1975) bzw. Shiffrin und Schneider (1977) expliziert. Man kann dieses Modell allgemein als unzureichend kritisieren (vgl. z.B. Richter, 2003), da es weder eine homogene Phänomenklasse generiert (vgl. Bargh, 1992), noch dem fließenden Übergang zwischen automatischen und

kontrollierten Prozessen gerecht wird (vgl. Logan, 1988). Darüber hinaus lässt sich gegen die Verwendung des Begriffes bei McKoon und Ratcliff im Speziellen sicherlich der Einwand formulieren, dass die Autoren ihn unpräzise verwenden, lediglich *ex negativo* explizieren und nicht sprachspezifisch konkretisieren (vgl. z.B. Singer et al., 1994). Der grundsätzliche *spirit* der Begriffsverwendung im Minimalismus ist jedoch m.E. dennoch grundsätzlich klar: Er wird im Rahmen einer starken bottom-up-Architektur verwendet und bezieht sich auf schnelle, routinisierte (Gedächtnis-)Prozesse. Diese Vorstellung sollte dann bei der Entwicklung des gedächtnis-basierten Ansatzes der Sprachverarbeitung noch weiter präzisiert werden.

Allerdings ist es dieser Stelle wichtig, darauf hinzuweisen, dass die minimalistische Inferenztheorie explizit keinerlei Aussagen über die strategische Inferenzgenerierung macht („Neither minimalist nor constructionist theories propose models of how strategic, goal-specific inference processing is carried out, and this issue remains on the agenda for future research“, McKoon & Ratcliff, 1992, p. 462); und das, obwohl die minimalistische Position ebenso klar konzediert, dass strategische Prozesse für ein vollständiges Textverstehen notwendig sind (womit sie gegen eines der klassischen Merkmale von Automtizität verstoßen, s.o.) und die von der Theorie beschriebenen Minimalinferenzen häufig nur die Grundlage für spätere strategische Verstehensprozesse bilden:

However, more often than not, readers do have specific goals, especially when learning new information from texts, and so they often engage in strategic processes designed to achieve those goals. The minimalist claim fore these situations is that minimal inferences provide the database for more strategic processes. (McKoon & Ratcliff, 1992, p. 440).

Man kann vollkommen zu Recht kritisieren, dass die minimalistische Inferenztheorie damit einen nur sehr eingeschränkten Erklärungsanspruch besitzt und gewissermaßen nur die Hälfte der Arbeit macht, die man von einer anständigen Sprachverarbeitungstheorie verlangen kann. Wie eingeschränkt der Geltungsbereich der Theorie ist, wird z.B. daran deutlich, dass nach Auffassung der Autoren selbst Pronomen – sonst *das* Paradebeispiel für die Klasse der obligatorischen und automatischen Inferenzen – gerade *nicht* automatisch aufgelöst werden, sondern zumeist strategisch verarbeitet werden müssen (vgl. Greene, McKoon & Ratcliff, 1992). Im Umkehrschluss ließe sich also durchaus die Behauptung aufstellen, dass die sog. minimalistische Inferenztheorie deshalb eigentlich eine verkappte *maximalistische* Theorie ist, und zwar weil sie empirisch gezeigt hat, wie *selten* automatische Verarbeitungsprozesse allein dazu in der Lage sind, das Produkt selbst hochfrequenter Sprachverarbeitungsprozesse adäquat abzubilden.

Im Gegensatz zur minimalistischen Inferenztheorie, welche die Unterscheidung von automatischen und strategischen Verarbeitungsprozessen zum Ausgangspunkt ihrer Überlegungen macht, legt sich die konstruktivistische Inferenztheorie diesbezüglich nicht fest:

It is also beyond the scope of this article to discriminate between those inferences that are automatically versus strategically generated on-line. The distinction between automatic and strategic inferences was central to McKoon and Ratcliff's (1992) *minimalist hypothesis*, but it is not central to the proposed constructionist theory. Instead, the relevant contrast is between on-line and off-line inferences while readers *comprehend* text. (Graesser et al., 1994, p. 372).

Grundsätzlich lässt die konstruktivistische Inferenztheorie also alle Arten von Prozessen bei der Aktualgenese einer Inferenz zu, solange sie nur online vorgenommen werden (im Gegensatz zu Inferenzen, die lediglich offline generiert werden, z.B. in einem sich an die Textrezeption anschließenden Test, vgl. McKoon & Ratcliff, 1998). Da die konstruktivistische Inferenztheorie gleichzeitig nicht spezifiziert, worin denn diese strategischen Prozesse nun genau bestehen und wie sie sich hinreichend präzise formulieren ließen (vgl. Graesser et al., 1994, p. 373), hat der Vorwurf, dass die konstruktivistische Inferenztheorie das Problem der Inferenzgenerierung nicht löst, sondern vielmehr nur umgeht (vgl. Gerrig & O'Brien, 2005), durchaus eine gewisse Berechtigung. Auch hier ist jedoch m.E. der *spirit* wieder hinreichend klar: Die Ebene der Aktualgenese einer Inferenz ist der konstruktivistischen Position prinzipiell eher unwichtig, da sie hier keine präzisen Aussagen machen kann (oder will). Stattdessen ist ihr wichtig, darauf hinzuweisen, dass die aktuellen Verarbeitungsprozesse immer in einen größeren Handlungszusammenhang eingebettet sind, welcher vom Minimalismus ausgeklammert wird: Während sich die minimalistische Position also eher auf die konkreten (automatischen) Verstehensprozesse auf der Mikro-Ebene bezieht, spezifiziert die konstruktivistische Position die übergeordneten Handlungsziele auf der Makro-Ebene und ihre Konsequenzen für den Verstehensprozess. Eine solche Konzeptualisierung führt dann schon fast zwangsläufig zu einer top-down-Architektur, in welcher der Leser als Problemlöser betrachtet wird, der aktiv Informationen zur Befriedigung seiner spezifischen Bedürfnisse sucht.

Auch hinsichtlich der Frage, inwieweit die betrachteten Prozesse eigentlich für die Domäne des Sprachverstehens spezifisch sind, vertreten die beiden Ansätze prinzipiell komplementäre Positionen (was v.a. in der Folgediskussion deutlich wird). Die minimalistische Theorie vertritt hier klar die Auffassung, dass die Sprachverarbeitung auf allgemeinen Prozessen basiert, die immer eingesetzt werden und unabhängig von den einzelnen Lesern/innen und spezifischen Texten sind. Die konstruktivistische Inferenztheorie vertritt hingegen genau die gegenteilige Auffassung, nämlich dass die beteiligten Sprachverarbeitungsprozesse eng auf

die interindividuell variablen Leser-Ziele abgestimmt sind und spezifische Besonderheiten der einzelnen Texte berücksichtigen (vgl. auch Gerrig & O'Brien, 2005).

Die resultierende Kombination der beiden Charakteristika (Minimalismus: automatische, aber unspezifische Prozesse vs. Konstruktivismus: strategische, aber spezifische Prozesse) ist dabei eher verwunderlich. Denn unter Rückgriff auf andere Konzeptualisierungen von Automtizität (wie z.B. der Modularitätsthese von Fodor, 1983), wäre eine umgekehrte Zuordnung plausibler: Eigentlich sollten gerade hoch-automatisierte Prozesse maximal bereichsspezifisch ausfallen, während vergleichsweise hierarchie-hohe Prozesse übergreifend und unspezifisch sind. Hieran kann man m.E. zwei Sachen ersehen: Zum einen, dass bei der Modellierung der hierarchie-niedrigen Prozesse durch den Minimalismus (die sich fast ausschließlich an allgemeinen Gedächtnismodellen orientiert) wichtige automatische, aber sprachspezifische Teilkomponenten (wie z.B. syntaktische Parsing-Prozesse) vollständig ausgeklammert bleiben. Zum anderen wird aber auch deutlich, dass die vermeintliche „Spezifität“ der hierarchie-hohen Prozesse wohl eher daraus resultiert, dass solche Prozesse durch ihre bereichsübergreifende Konzeption flexibler sind.

1.2 Erweiterung: Gedächtnis- vs. erklärungs-basierte Sprachverarbeitungstheorien

1.2.1 Gedächtnis-basierte Sprachverarbeitung

In der anschließenden Debatte über die minimalistische und die konstruktivistische Inferenztheorie hat sich sehr bald herausgestellt, dass beide Theorien in ihrer ursprünglichen Formulierung nicht korrekt sein können (vgl. Gerrig & O'Brien, 2005). So wurde vollkommen berechtigt gegen die minimalistische Position eingewendet, dass ihre starke Annahme, dass Inferenzen nur dann generiert werden, wenn ein lokaler Kohärenzbruch vorliegt, nicht mit den zahlreichen empirischen Belegen vereinbar ist, dass Textinformationen auch dann reaktiviert werden, wenn die lokale Kohärenz eines Textsegmentes gesichert ist (vgl. z.B. Albrecht & O'Brien, 1992). Auch wurden zahlreiche Belege dafür vorgelegt, dass nicht nur lokale, sondern auch globale Textinformationen im Laufe der Inferenzgenerierung genutzt werden (Magliano & Radvansky, 2001).

Die gegenläufige Kritik des Minimalismus an der konstruktivistischen Inferenztheorie ist hingegen v.a. auf der theoretischen Ebene angesiedelt. Abgesehen von dem generellen Vorwurf, dass sie nicht (genau) spezifizieren kann, wie denn die Suche des/der Lesers/in nach globalen Kausalinformationen konkret ablaufen soll, geht es meist darum, dass die (vermeintlich) vom Konstruktivismus postulierten aktiven Suchprozesse kognitionspsychologisch unplausibel sind: Nicht nur, dass eine solche Architektur nicht schnell genug operieren kann, um

den reibungslosen und effizienten online-Charakter der Sprachverarbeitung erklären zu können. Da die Verarbeitungskapazität des verbalen Arbeitsgedächtnisses begrenzt ist (vgl. Just & Carpenter, 1992), stehen nur wenige Ressourcen für solche aufwändigen Verarbeitungsprozesse zur Verfügung. Wenn hingegen Leser/innen wirklich den gesamten Text an jeder Stelle in Bezug auf seine globale und kausale Kohärenz prüfen würden, dann führt dies zu solch unplausiblen Vorhersagen wie z.B., dass die Lesezeit für ein und denselben Satz mit steigender Länge des Textes sukzessive zunehmen müsste, weil ja immer mehr globale Informationen, wie z.B. Ziel, Motive oder Charaktereigenschaften eines Protagonisten etc. vorliegen, die geprüft werden müssen.

Vor dem Hintergrund derartig prinzipieller Einwände gegen beide Positionen ist es nicht verwunderlich, dass der nachfolgende theoretische Entwicklungsprozess zunächst einmal durch eine Liberalisierung der beiden Ausgangspositionen gekennzeichnet war. Das heißt, beide Seiten haben entscheidende Grundannahmen ihrer Theorien modifiziert und so die Möglichkeit zu einer grundsätzlichen Annäherung der zunächst relativ unversöhnlich wirkenden Positionen geschaffen. Interessanterweise war jedoch in beiden Fällen mit dieser theoretischen Umstrukturierung gleichzeitig auch eine Generalisierungstendenz des jeweiligen Erklärungsprinzips verbunden, die einen Konsens wesentlich erschwert. Kurz: Beide Seiten haben sich auf wenige, empirisch besser haltbare (bzw. schwerer zu falsifizierende) Grundannahmen zurückgezogen, diese dann aber noch radikalisiert.

Am deutlichsten wird diese doppelt-gegenläufige Form der Theorieentwicklung beim gedächtnis-basierten Ansatz. Eingeführt wurde der Begriff *memory-based approach* (bzw. allgemeiner: *memory-based text processing*) von McKoon, Gerrig und Greene (1996). Der Ansatz greift einen zentralen Anspruch der konstruktivistischen Position auf, indem er das Prinzip der sog. *readiness* (ein Begriff der m.E. nur schwerlich ins Deutsche zu übersetzen ist) in den Vordergrund seiner Betrachtung stellt (vgl. v.a. Gerrig & McKoon, 1998). Denn das Konzept der *readiness* scheint viele Anforderungen zu erfüllen, die von der konstruktivistischen Position in Bezug auf die allgemeine Funktionalität des textuellen Rezeptionsprozesses und seine Einbindung in höhere Handlungsziele gemacht wurden: Leser/innen rezipieren Texte ja nicht, um irgendwelche minimalen Inferenzen automatisch zu etablieren, sondern vielmehr weil diese Texte eine bestimmte *Relevanz* für sie haben (die allerdings sehr unterschiedlich ausfallen kann und von Zielen wie ‚neue Informationen erhalten‘ über ‚spannende Geschichten lesen‘ bis hin zu ‚eine Vp-Stunden erwerben‘ reichen kann). Von daher scheint die zentrale Funktion gedächtnis-basierter Verarbeitungsprozesse, die als „making ready a range of information of potential relevance to readers’ ongoing understanding of texts“ (Gerrig &

McKoon, 1998, p. 67) expliziert wird, zunächst einmal einen zentralen Kritikpunkt der konstruktivistischen Position aufzugreifen. *Readiness* bezieht sich auf die Tatsache, dass Leser/innen an einer bestimmten Stelle der Textrezeption *bereit* sind, bestimmte Inferenz zu generieren. Die Inferenz liegt also gewissermaßen schon „in der Luft“ und wartet nur darauf, konstruiert zu werden. Allerdings wird sehr bald deutlich, dass McKoon et al. diese *readiness* nicht darauf zurückführen, dass die Inferenz etwa von dem/der Rezipienten/in *antizipiert* worden wäre oder dass die Bereitschaft sonst irgendwie auf eine aktive leserseitige Vorbereitung zurückgeführt werden könnte (so wie man z.B. vor einem Sprung seine Muskeln anspannt etc.): „This claim would impute a goal directedness to this mental activity with which we would be quite uncomfortable.“ (Gerrig & McKoon, 1998, p. 69). Stattdessen ist es vielmehr *der Text*, der den/die Leser/in auf die Inferenz vorbereitet; sie werden also gewissermaßen „bereit gemacht“.

Und gerade diese primär passivische Bedeutung von *readiness* ist der Grund dafür, weshalb dieser Begriff so schwer ins Deutsche zu übersetzen ist: Die zunächst nahe liegende Übersetzung mit dem Wort *Bereitschaft* ist trügerisch. Denn zumindest nach meinem Sprachempfinden ist dieser Begriff im Deutschen eindeutig aktivisch konnotiert, was sich auch darin manifestiert, dass die meisten Synonyme für diesen Begriff das Konzept *Wille* bemühen: Bereitwilligkeit, Entschluss, Willigkeit, Neigung etc. Der komplementäre Begriff der *Vorbereitung* gibt hingegen die elegante aktivisch-passivische Doppelbedeutung des Begriffes im Englischen nicht vollkommen wieder: *Oxfords Advanced Learner's Dictionary* (1989) gibt z.B. als Bedeutungen des Begriffes *readiness* an: 1. state of being ready or prepared, 2. willingness or eagerness, 3. quickness and promptness. Während Bedeutung 1 vergleichsweise neutral ist, spiegelt Bedeutung 2 eindeutig eher den aktivischen und Bedeutung 3 eher den passivischen Anteil des Begriffes wieder. Vor diesem Hintergrund ist die Wortwahl von McKoon et al. sicher geschickt: Der Begriff greift vordergründig ein aktives Prinzip auf, wird dann aber passivisch umgedeutet.

Denn statt einer eigenständigen aktiven Bereitschaft des Lesers geht der gedächtnisbasierte Ansatz davon aus, dass bestimmte Textelemente und Inferenzen wieder verfügbar werden „as a natural consequence of the way in which memory processes work on text representations“ (Gerrig & McKoon, 1998, p. 69.). Das heißt, Inferenzen werden dann (und nur dann) konstruiert, wenn relevante, aber inaktive Informationen des bereits existierenden Diskursmodells und des Weltwissens durch einen allgemeinen Gedächtnisprozesse reaktiviert werden können, und zwar ohne Rücksicht auf die Nützlichkeit oder Relevanz dieser Informationen für den/die Leser/in. Und in seiner starken Form ist der gedächtnisbasierte Ansatz sogar durch die Behauptung charakterisiert, dass *alle* sprachlichen Verstehensprozesse auf solchen allgemeinen Prozessen der Informationsverarbeitung basieren:

The central tenet of memory-based text processing is that comprehension is enabled by information in memory. Whether for text being read of for speech being heard, each new piece of linguistic information is understood in terms of the information that it evokes from memory. (McKoon et al., 1996, p. 919).

Der Mechanismus, auf dem diese schnelle Informationsaktivierung beruhen soll, ist dabei als ein allgemeiner Gedächtnisprozess konzipiert, wie er von verschiedenen modernen Gedächtnistheorien für den Abruf von Informationen aus dem Langzeitspeicher angenommen wird (vgl. Hintzman, 1986; Gillund & Shiffrin, 1984), nämlich ein schneller, passiver Aktivationsprozess, bei dem ein aktuell zu verarbeitender Abruf-Cue im Arbeitsgedächtnis mit allen Informationen im Langzeitspeicher gleichzeitig und parallel in Verbindung tritt. Speziell bezieht sich der gedächtnis-basierte Ansatz auf den Resonanz-Mechanismus, wie er von Ratcliff (1978) spezifiziert wurde: Der Abruf-Cue im Arbeitsgedächtnis sendet ein (ungerichtetes) Signal in den Langzeitspeicher aus (der u.a. die episodische Spur der bisher verarbeiteten Textinformationen, aber auch alle allgemeinen deklarativen Wissensbestände enthält) und alle Informationen im Speicher „resonieren“ in Hinblick auf dieses Signal. Dabei ist das Ausmaß, in dem eine bestimmte Gedächtnisspur vom Cue aktiviert wird, vom Grad der Assoziation zwischen Input-Cue und Gedächtnis-Information abhängig. Die durch den Cue initial aktivierten Elemente aktivieren wiederum andere mit ihnen assoziierte Elemente. Dieser Prozess wiederholt sich zyklisch bis sich ein stabiles Aktivationsmuster herauskristallisiert hat. Die Elemente, die am Ende des Abrufprozesses noch aktiv (bzw. am aktivsten) sind, werden dann ins Arbeitsgedächtnis transferiert und stehen damit zur weiteren Verarbeitung zur Verfügung.

Von entscheidender Wichtigkeit ist, dass der Resonanzprozess drei charakteristische Eigenschaften hat: Er ist a) passiv, b) „dumm“ (engl. *dumb*) und c) unrestringiert (vgl. O'Brien & Gerrig, 2005). Alle drei Eigenschaften bestimmen entscheidend mit, welche Textinformationen wie bearbeitet werden (können). Denn obwohl die Stärke des Abruf-Signals zumindest teilweise als abhängig von seiner Aufmerksamkeitsstärke konzipiert ist (s.u.), ist der Resonanzprozess selbst vollkommen autonom: Alle Gedächtnisspuren resonieren unbegrenzt und allein als Funktion der (semantischen und kontextuellen) Merkmalsüberlappung von Spur und Signal.

Auch wenn der Merkmalsüberlappung in diesem Zusammenhang ein besonders zentraler Stellenwert zukommt, da die einzelnen Gedächtniselemente umso stärker resonieren, je mehr Merkmale sie mit dem Abruf-Cue teilen, so lässt sich dieses allgemeine Prinzip durchaus noch weiter spezifizieren. Dies wurde z.B. von Myers und O'Brien (1998; vgl. auch O'Brien & Myers, 1999) getan, welche die zentralen Einflussvariablen auf den Resonanzprozess präzise ausdifferenzierten. So sind z.B. die *Häufigkeit*, *Ähnlichkeit* (bzw. *Typikalität*) und die (textuelle und zeitliche) *Distanz*, die zwischen Abruf-Cue und Gedächtnis-Elemente liegt, ziemlich nahe liegende Faktoren, welche die Abrufbarkeit einer Gedächtnisspur beeinflussen können (für den Bereich der Brückeninferenzen vgl. z.B. Garrod & Sanford, 1977). In eine

ähnliche Richtung gehen auch die beiden Einflussfaktoren der *Zentralität* (Konzepte, die zentraler für das Thema eines Textes sind, werden schneller abgerufen) und der *Assoziiertheit* zwischen zwei Konzepten (welches sich nicht nur auf die Anzahl der überlappenden Merkmale, sondern auch auf ihre Gewichtung bezieht).

Allerdings ist von entscheidender Bedeutung, dass die Relation von Abruf-Cue und Gedächtnis-Spur nicht isoliert betrachtet werden darf: Der Abruf-Cue aktiviert *alle* Inhalte des Langzeitspeichers und diese sind auch untereinander verbunden. Wichtiger ist deshalb meist das *Merkmalsmuster*, d.h. die *Konstellation* der Merkmale von Abruf-Cue und Gedächtnis-Elementen. So ist z.B. äußerst relevant, ob der Abruf-Cue für eine bestimmte Gedächtnisspur spezifisch ist. Wenn z.B. ein Cue für mehrere Episoden eines Textes gleichzeitig relevant ist, dann ist er wesentlich schwächer, als wenn er ganz gezielt nur ein einziges Konzept des Textes aktiviert. Diese Eigenschaft ist die Grundlage des sog. Distraktor-Effekts (vgl. z.B. Corbett, 1984): Wenn ein zweites, ähnlich stark mit einem Cue assoziiertes Gedächtniselement in einem Text enthalten ist, dann „behindert“ dieses die intendierte Auflösung, weil es automatisch mitaktiviert wird und sich die Signalstärke des Abrufelements dadurch reduziert.

Solche oder ähnliche Kontexteffekte werden sogar noch wichtiger, wenn man bedenkt, dass im Zuge des Resonanzprozesses nicht nur die episodische Repräsentation des bisher verarbeiteten Textes, sondern gleichzeitig auch alle deklarativen Wissensbestände des generellen Weltwissens mitaktiviert werden. Hier ergeben sich dann geradezu unendlich viele Möglichkeiten, wie die einzelnen Gedächtniselemente auf den verschiedenen Merkmalsdimensionen miteinander interagieren können: Wenn ein Abruf-Cue zwar hoch mit einer spezifischen Textepisode assoziiert ist, allerdings auch sehr stark mit Konzepten des allgemeinen Weltwissen, so kann es u.U. passieren, dass der Cue die Textepisode nicht aktivieren kann, weil die Aktivationsmuster miteinander interferieren und sich evtl. sogar gegenseitig auslöschen (vgl. O'Brien & Myers, 1999).

Allerdings beschränkt sich die Liste der Einflussfaktoren nicht nur auf solche vergleichsweise hierarchie-niedrigen Merkmale, sondern kann u.U. auch Merkmalskonstellationen umfassen, welche der vom Konstruktivismus postulierten aktiven Suche schon ziemlich nahe kommen. So konzedieren z.B. manche Vertreter/innen des gedächtnis-basierten Ansatzes explizit (vgl. Gerrig & O'Brien, 2005), dass leserseitige Aufmerksamkeitsprozesse, mit denen bestimmte Textinformationen mehr oder weniger stark fokussiert werden, einen entscheidenden Einfluss auf den Resonanz-Prozess haben können. Denn unter der Voraussetzung, dass fokussierte Elemente eine höhere Cue-Signalstärke haben, können dadurch bestimmte Merkmale u.U. mehr oder weniger stark betont werden. Auch lässt sich die Stabili-

sierungsdynamik des sich ergebenden Aktivationsmuster zumindest partiell „lenken“. Dies ist ein wichtiger Punkt, der bei der Integration der beiden Ansätze eine entscheidende Rolle spielt.

Allerdings besteht nicht nur rezipienten-, sondern auch produzentenseitig die Möglichkeit der Einflussnahme. So können Autoren/innen z.B. durch verschiedene strukturelle Maßnahmen (wie Aktiv- vs. Passivformulierung eines Satzes, Topikalisierung von Kernkonzepten etc.) den Fokus ihrer Rezipienten/innen steuern. Ein weiterer Bereich, der rezipienten- und produzentenseitigen Gestaltungsraum eröffnet, ergibt sich aus der Tatsache, dass Textkonzepte, die wichtiger als andere sind, eine größere Chance zur Reaktivierung haben (vgl. O'Brien & Myers, 1999). Und dies aus dem einfachen Grund, weil in diesem Sinne „wichtige“ Konzepte häufiger in einem Text genannt werden und „tiefer“ in ihm verankert sind. Oberflächenstrukturell wird dies dadurch vermittelt, dass wichtige Konzepte häufiger wiederholt oder vermehrt elaboriert werden.

Insgesamt weist der so spezifizierte Resonanzprozess eine enge konzeptuelle Nähe zum *Construction-Integration-(CI)-Modell* von Kintsch (1988, 1998) auf, unterscheidet sich aber in einigen Punkten auch entscheidend von ihm. Denn auch im CI-Modell stehen bottom-up-Prozesse im Vordergrund, da in der Konstruktionsphase zunächst alle möglichen Inferenzen in die initiale Textrepräsentation integriert werden und einzelne Wörter vollkommen ohne top-down-Kontrolle ihre übergeordneten Propositionen aktivieren, welche wiederum andere Propositionen aktivieren etc. Konflikte, die sich aus dieser etwas chaotischen Anfangsstruktur ergeben, werden im CI-Modell sehr ähnlich wie beim zyklischen Resonanzprozess dann in der anschließenden Integrationsphase gelöst: Diese wird von Kintsch primär als *constraint-satisfaction*-Prozess modelliert, bei dem sich in quasi-konnektionistischer Weise Propositionen und Inferenzen, die zueinander passen, gegenseitig verstärken, während andere gehemmt werden. Auch hier ist Textverstehen also eher ein Wahrnehmungs- denn ein Problemlöseprozess.

Eine mathematische Implementierung des Resonanz-Modells, welche sich sowohl was die Repräsentation als auch die Verknüpfungsstruktur der einzelnen Text- und Gedächtniselemente angeht stark am CI-Modell von Kintsch orientiert, wurde von Myers und O'Brien (1998) vorgelegt. Nachdem ein Element ins aktuelle Arbeitsgedächtnis gelangt ist, wird der Resonanzprozess automatisch und passiv aufgerufen. Alle Gedächtniselemente werden danach direkt und ohne jegliche Kontrolle vom Abruf-Cue angesprochen, lediglich Anzahl und Stärke ihrer Verbindungen beeinflussen den Abruf-Prozess. Jedes Abruf-Element besitzt eine bestimmte Signalstärke, die es auf die verschiedenen zur Verfügung stehenden Gedächtnis-

Elemente aufteilen muss (womit der oben beschriebene Distraktor-Effekt modelliert werden kann). Anders als beim CI-Modell nimmt diese Signalstärke im Laufe der Textverarbeitung kontinuierlich ab, während sich die Verbindungsstärken zwischen den einzelnen Elementen nicht verändern. Das heißt, ähnlich wie beim CI-Modell haben Elemente, die stärker miteinander verbunden sind, eine größere Abrufwahrscheinlichkeit, aber anders als beim CI-Modell wird diese Wahrscheinlichkeit lediglich vom Ausmaß der Interferenz mit anderen resonierenden Elementen bestimmt.

Ein wesentlicher Vorteil des Resonanz-Modells oder des vergleichbaren *landscape models* von van den Broek und Mitarbeitern (s.u.; vgl. Linderholm, Virtue, Tzeng & van den Broek, 2004; van den Broek, Tzeng & Kendeou, 2004) ist, dass die einzelnen Konzepte und Propositionen eines Textes in ihrer Zugänglichkeit dynamisch variieren können und u.U. zu verschiedenen Zeitpunkten unterschiedlich aktiv sind: Ein Konzept, welches in der Satzmitte noch verfügbar ist, hat am Satzende vielleicht schon keinen Einfluss mehr, weil es von anderen Signalen überlagert wird (s. auch Rapp & van den Broek, 2005). Gleichzeitig ist das Modell aber auch so flexibel, dass es eine Reaktivierung von inzwischen deaktivierten Textelementen ermöglicht; vorausgesetzt, es liegt ein hinreichend starker und spezifischer Abruf-Cue vor. Das heißt, die Alles-oder-Nichts-Perspektive, wie sie der Minimalismus-These anhaftete (easily available = aktueller Inhalt des Arbeitsgedächtnis = die letzten ein oder zwei Sätze), wird aufgegeben und gegen einen Mechanismus eingetauscht, welcher der Fluidität der Textverarbeitung gerecht wird und beschreiben kann, wie die Verfügbarkeit verschiedene Information im Laufe der Textverarbeitung zunehmen und auch wieder abnehmen kann (vgl. auch McKoon et al., 1996; Gerrig & McKoon, 2001).

Eine weitere wichtige Eigenschaft des Resonanzprozesses ist, dass er passiv und kontinuierlich ist. Denn dadurch können gedächtnis-basierte Sprachverarbeitungstheorien erklären, warum auch dann (globale) Informationen abgerufen werden, wenn ein Text lokal kohärent ist und somit eigentlich keine Notwendigkeit dafür besteht. Denn wie bereits dargestellt, war dies die vielleicht größte empirische Schwäche der minimalistischen Inferenztheorie (vgl. Singer et al., 1994). Die Spezifizierung des Prinzips der direkten Verfügbarkeit als Resonanzprozess macht deutlich, dass dem/der Leser/in viel größere Wissensmengen zur Verfügung stehen (bzw. durch Reaktivierung zur Verfügung stehen können), als ursprünglich angenommen. Einfache Gedächtnisprozesse wie z.B. der Resonanz-Mechanismus können deswegen u.U. vollkommen ausreichen, um auch viele „höhere“ Verstehensphänomene zu erklären, wie z.B. prädiktive, thematische oder instrumentelle Inferenzen (vgl. Gerrig & O'Brien, 2005): Ein starker Kontext kann zu einer passiven (Re-)Aktivierung von globalen Textinformationen

oder dem allgemeinen Weltwissen führen, welche den Aufbau einer sehr reichhaltigen Textrepräsentation auch mit „minimalen“ Mitteln ermöglichen. Im Gegensatz zur minimalistischen Inferenztheorie geht der gedächtnis-basierte Ansatz also davon aus, dass es keinen prinzipiellen Unterschied zwischen notwendigen und elaborativen Inferenzen gibt (insbesondere bedürfen letztere nicht unbedingt der strategischen Verarbeitung). Beide Inferenztypen werden mittels des gleichen passiven Verarbeitungsprozesses generiert; allerdings hat die Klasse der notwendigen Inferenzen aufgrund der Abrufstruktur des Resonanzprozesses eine höhere Enkodierungswahrscheinlichkeit.

Ein sehr eindrucksvolles Beispiel dafür, dass passive Abrufprozesse in der Tat ein sehr mächtiges Instrument zur Erklärung von „höheren“ Verstehensprozessen sein können, liegt mit den Ergebnissen aus dem sog. *inconsistency paradigm* der Forschergruppe um O'Brien vor (vgl. O'Brien & Albrecht, 1992; Albrecht & O'Brien, 1993). In diesem experimentellen Paradigma werden kleine Geschichten gelesen, die einen bestimmten Zielsatz enthalten (z.B. *She decided to go **outside** and stretch her legs a little.*). Zuvor werden in dem Text entweder Informationen gegeben, die mit diesem Zielsatz konsistent sind (z.B. *As Kim stood **inside** the health club she felt a little sluggish.*) oder nicht (z.B. *As Kim stood **outside** the health club she felt a little sluggish.*). Die Annahme ist, dass Leser/innen inkonsistente Fortsetzung verlangsamt lesen, weil diese aufgrund ihrer Inkonsistenz nur schlecht in die bestehende Diskursrepräsentation integriert werden können und zusätzliche inferentielle Prozesse oder repair-Strategien erfordern. Die entscheidende experimentelle Manipulation bestand nun darin, dass zwischen Ausgangs- und Zielsatz entweder eine längere Passage lag (von ca. 3 Sätzen) oder nicht, sodass die Inkonsistenz entweder lokal oder global war. Die Einfügung war darüber hinaus so konstruiert, dass Einfügung und Zielsatz lokal kohärent waren. Die Vorhersage der minimalistischen Inferenztheorie ist in diesem Fall, dass die Inkonsistenz in der globalen Bedingung nicht erkannt wird, weil der Text lokal kohärent ist und die inkonsistente Information nicht mehr verfügbar sind (weil mehr als 2 Sätze entfernt). Tatsächlich prüften McKoon und Ratcliff (1992, Exp. 2) genau diese Vorhersage in einer sehr ähnlichen Anordnung und fanden keine Aktivierung der distalen Informationen, wenn nicht gleichzeitig eine lokale Inkohärenz vorlag. O'Brien und Albrecht (1992) fanden jedoch auch einen Inkonsistenz-Effekt, wenn der Zielsatz lokal kohärent war (Albrecht & O'Brien, 1993, prüften die gleiche Vorhersage, wenn die Inkonsistenz durch Eigenschaften der Protagonisten vermittelt wurde; mit dem gleichen Ergebnis). Dieser Befund spricht klar dafür, dass entgegen der minimalistischen Annahme ein komplexes Diskursmodell aufgebaut und auf seine globale Kohärenz geprüft wird.

In einer nachfolgenden Serie von Experimenten konnte die gleiche Forschergruppe (vgl. z.B. O' Brien, Rizzella, Albrecht & Halleran, 1998) jedoch zeigen, dass zur Entdeckung der globalen Inkonsistenz nicht unbedingt auf komplexe aktive Suchprozesse im Sinne der konstruktivistischen Theorie zurückgegriffen werden muss. Vielmehr ist der gleiche Effekt auch im Rahmen des gedächtnis-basierten Ansatzes erklärbar, wenn man annimmt, dass der inkonsistente Zielsatz als Abruf-Cue dient, der als Resonanz-Signal alle Gedächtnis-Spuren aktiviert, die mit ihm in Verbindung stehen. Wenn man also zuerst in einer Geschichte gelesen hat, dass die Protagonistin Mary eine strikte Vegetarierin ist und später dann erfährt, dass sie sich in einem Restaurant einen Hamburger bestellt, dann kommt einem das deshalb seltsam vor, weil bei Verarbeitung des inkonsistenten Satzes alle Informationen, die bis jetzt in der Geschichte über Mary vermittelt wurden (passiv) reaktiviert werden ebenso wie generelle Weltwissensbestände darüber, dass strikte Vegetarierinnen kein Fleisch essen (insbesondere keine Hamburger). Die gleichzeitige Aktivierung dieser inkompatiblen Wissensbestände führt dann zu einem lokalen Kohärenzbruch, der beim Lesen entsprechend repariert werden muss (vgl. auch Singer, Halldorson, Lear & Andrusiak, 1992; Singer, 1994).

Um zu überprüfen, ob ein komplexer Prozess der Situationsmodell-Generierung und des aktiven Kohärenz-Monitorings im Sinne des Konstruktivismus notwendig ist oder der Resonanz-Mechanismus alleine ausreicht, wurden weitere Versionen der Geschichte konstruiert, in denen die Inkonsistenz dadurch aufgehoben wurde, dass zwischen den beiden inkonsistenten Sätzen weitere Informationen gegeben wurden, die plausibel machen sollten, warum Mary trotz ihres Vegetariertums in diesem speziellen Fall Fleisch isst. So wurde z.B. in einem Experiment mitgeteilt, dass Mary sich zwar meist vegetarisch ernährt, dies jedoch nicht tut, wenn sie sich mit ihren Freunden in einem Restaurant trifft (Exp. 1 + 2). In einer anderen Version (Exp. 3 + 4) wurde dem/der Leser/in explizit mitgeteilt, dass Marys Vegetarier-Dasein der Vergangenheit angehört und sie sich heute vorzugsweise von Fast Food ernährt. In einem abschließenden Experiment (Exp. 5) wurde den Probanden/innen sogar mitgeteilt, dass Mary niemals Vegetarierin gewesen war, sondern lediglich, dass eine Freundin sie mal spaßeshalber auf einer Party als solche vorgestellt hatte. Das Ergebnis war immer das gleiche: Inkonsistente Sätze wurden langsamer gelesen, selbst wenn sie perfekt mit dem jeweiligen Situationsmodell kompatibel waren. Dieses Befundmuster spricht in der Tat dafür, dass zumindest bei der initialen Verarbeitung ein vergleichsweise unrestringierter und „dummer“ Abrufprozess wirksam ist, der vollkommen unabhängig vom jeweiligen Diskurskontext und den Zielen der Leser einfach alles auf (In)Konsistenz zum aktuellen Input überprüft. Allerdings war es gleichzeitig auch so, dass solche plausibilisierenden Einfügungen zwar nicht den

anfänglichen Inkonsistenz-Effekt beeinflussten, gleichwohl aber ihre spätere Integration (gemessen über einen nachfolgenden Targetsatz).

In diesem Experiment wird die zentrale Strategie des gedächtnis-basierten Ansatzes als Weiterentwicklung der minimalistischen Inferenztheorie noch einmal sehr deutlich: Einerseits konzediert der Ansatz klar, dass die Textrepräsentation sehr viel reichhaltiger ist, als der Minimalismus anfänglich angenommen hat. Gleichzeitig zieht sie sich noch radikaler auf die Aussage zurück, dass auch solche „hohen“, reichhaltigen Textrepräsentationen von relativ basalen und allgemeinen Gedächtnis-Prozessen generiert werden können. Dadurch, dass O'Brien et al. das Konzept der „leicht verfügbaren Informationen“ des Minimalismus im Sinne des Resonanz-Modells explizieren und damit in Reichweite und Erklärungskraft zunächst einmal erheblich aufwerten, machen sie in vielen Fällen eine darüber hinausgehende strategische Verarbeitung überflüssig. So muss z.B. das kausale Gerüst einer Geschichte nicht unbedingt aktiv konstruiert (und überwacht) werden, da eine passive Aktivierung von „schwachen“ syntaktischen und semantischen Basisinformationen hierfür u.U. vollkommen ausreichend ist. Und aus den gleichen Gründen ist eine umfangreiche exhaustive Suche nach den Zielen und Motiven eines Protagonisten häufig gar nicht nötig, weil die meisten der hierfür benötigten Informationen sowieso schon vorhanden sind und sozusagen „umsonst“ geliefert werden.

Mit der Einsicht, dass die Textrepräsentation reichhaltiger und flexibler ist, als ursprünglich angenommen, hat sich der gedächtnis-basierte Ansatz schon einen gewaltigen Schritt der konstruktivistischen Position angenähert. Nicht nur, dass dadurch die zentrale (aber z.T. unberechtigte) Sorge vieler Vertreter/innen des Konstruktivismus, nämlich dass der Minimalismus alle höheren Formen der Textrepräsentation (wie z.B. Situationsmodelle) abschaffen will, beseitigt oder zumindest abgemildert werden konnte. Darüber hinaus schenkt der gedächtnis-basierte Ansatz dem Konstruktivismus sogar einen kognitionspsychologisch plausiblen Mechanismus, mit dem sich auch die Funktionen der Textverarbeitung, die dem Konstruktivismus besonders am Herzen liegen, realisieren lassen.

Diese prinzipielle Annäherung muss man jedoch in meinen Augen damit „bezahlen“, dass sich die gedächtnis-basierten Sprachverarbeitungstheorien fast ausschließlich auf die Klasse der passiv-automatischen Prozesse konzentrieren. Hier sind sie sogar noch viel weitgehendere als der ursprüngliche Minimalismus: Dieser vertrat ganz explizit eine Zwei-Prozess-Theorie mit frühen automatischen Prozessen zur Etablierung der notwendigen Inferenzen, während strategische Prozesse für die Generierung höherer elaborativer Inferenzen zuständig waren. Im gedächtnis-basierten Ansatz gibt es hingegen eine starke Tendenz dahingehend,

eine Ein-Prozess-Theorie zu vertreten, da passive Prozesse ja häufig auch schon allein hinreichend sind, um das Verstehensprodukt zu gewährleisten.

An dieser Stelle können allerdings zwei unterschiedliche Spielarten des gedächtnisbasierten Ansatzes voneinander unterschieden werden: In seiner *starken* Version (wie sie zumindest ansatzweise von Gerrig & McKoon, 2001, für die pronominale Auflösung vertreten wird) geht der Ansatz davon aus, dass gedächtnis-basierte Prozesse sowohl notwendig als auch hinreichend sind, um ein Verstehensprodukt zu generieren, und zwar egal, ob es sich dabei um eine notwendige oder elaborative Inferenz handelt. In seiner *schwachen* Version geht der Ansatz hingegen (nur) davon aus, dass gedächtnis-basierte Prozesse notwendig zum Verstehen sind (vgl. z.B. O'Brien et al., 1998). (Diese Unterscheidung stimmt im Großen und Ganzen mit der von Richards und Singer (2001) überein, allerdings definieren sie die schwachen Ansätze darüber, dass diese gedächtnis-basierte Verarbeitungsprozesse Prozesse allein als hinreichend ansetzen.)

Allerdings wird gerade in letzter Zeit auch bei Vertretern/innen eines eher gedächtnisbasierten Ansatzes klar die Notwendigkeit gesehen (und manchmal sogar die Bereitschaft bekundet!), höhere top-down-Prozesse in ihre Modellbildung mit einzubeziehen (Kintsch, 2005; O'Brien & Myers, 1999; Gerrig & O'Brien, 2005). Ja, im Prinzip basiert ein guter Teil der empirischen Überprüfungsstrategie der gedächtnis-basierten Sprachverarbeitungstheorien gerade darauf, dass es auch strategische Prozesse gibt: So ist z.B. die zentrale Abhängige Variable im Inkonsistenz-Paradigma (und in weiten Teilen der Inferenzforschung) gerade eine Erhöhung der Verarbeitungszeit des Zielsatzes, wenn keine passive Voraktivierung stattfindet (was als erschwerte Integration in das Diskursmodell interpretiert wird).

1.2.2 Erklärungs-basierte Sprachverarbeitung

Die Ausführungen zur Weiterentwicklung der konstruktivistischen Inferenztheorie zum sog. erklärungs-basierten Sprachverarbeitungsansatz werden vergleichsweise kurz ausfallen. Zum einen, weil es hier sehr viel weniger Entwicklung gegeben hat, und zum anderen, weil wesentlich unklarer ist, was überhaupt die gemeinsam vertretenen Kernannahmen der erklärungs-basierten Sprachverarbeitungstheorien sind. Schon McKoon und Ratcliff (1995) hatten in ihrer ersten öffentlichen Replik auf die Kritik seitens der konstruktivistischen Inferenztheorie festgestellt, dass sie sich mit der Veröffentlichung ihrer Minimalismus-These eigentlich gewünscht hätten, dass die konstruktivistische Position den Staffelstab aufnimmt und zu spezifizieren versucht, wie der Prozess der Inferenzgenerierung jenseits der initialen automatischen Anfangsphase weiter aussehen könnte. Das Gegenteil ist leider eingetreten: Die kon-

struktivistische Reaktion war überwiegend destruktiv, und zwar in zweierlei Hinsicht. Zum einen gab es keine konstruktiven Bemühungen, die sich aus der Darstellung der Minimalismusthese ergebenden offenen theoretischen und empirischen Fragen mit entsprechenden Gegenüberlegungen und -experimenten eigenständig und kreativ anzugehen. Stattdessen beschränkte sich der Konstruktivismus auf der theoretischen Ebene darauf, vergleichsweise einseitig, aber dafür umso vehementer die minimalistische Position zu kritisieren. Und auch in Bezug auf die empirische Ebene blieb der Konstruktivismus vergleichsweise reaktiv und begnügte sich (wenn überhaupt) meist mit etwas trotzigem Versuchen, vermeintliche Implikationen des Minimalismus vereinzelt zu widerlegen. Ich habe bereits dargelegt, dass sich diese Diskussion v.a. dadurch auszeichnet, dass beide Positionen grandios aneinander vorbei geredet haben (und aneinander vorbei reden mussten).

Zum anderen kam es im Laufe der Debatte zu einer schleichenden Zersetzung der anfangs noch vergleichsweise geschlossenen konstruktivistischen Position. Statt sich gemeinsam auf ein kohärentes Rahmenprogramm zu einigen, das man der (vermeintlich) starken Position des Minimalismus entgegenhalten konnte, stürzte man sich lieber in (ver)alte(te) Grundsatzstreitigkeiten (z.B. ob prädiktive Inferenzen denn nun online konstruiert werden oder nicht). Von daher bleibt mir wenig anderes übrig, als der Einschätzung von McKoon und Ratcliff (1995, p. 99) zuzustimmen:

However, instead of a unified view, it now appears that constructionism is a collection of informal views that differ on various dimensions. Each view is apparently molded by data collected from whatever methodology the holder of that view believes to reveal underlying text representations.

Beide Merkmale, das der destruktiven Reaktivität und das der internen Zersplitterung, werden perfiderweise schon im Gebrauch des Terminus „erklärungs-basiert“ (*explanation-based*) manifest. Obwohl er eindeutig als Gegenbegriff zum Begriff „gedächtnis-basiert“ gedacht ist, ist er schon rein oberflächensemantisch parallel zu diesem konstruiert und bezieht sich reaktiv auf diesen. Zum anderen wird der Begriff – im Gegensatz zum Begriff „gedächtnis-basiert“, um den sich die Erben des Minimalismus wie um eine Fahne scharen – nicht von allen Vertretern/innen eines solchen Ansatzes einheitlich verwendet und schon gar nicht einheitlich gebraucht. Während ungefähr die Hälfte aller konstruktivistischen Forscher/innen ihre Theoriebildungen als „erklärungs-basiert“ bezeichnet, nennt sich die andere weiterhin „konstruktivistisch“.

Das hat von daher seine Berechtigung, als dass der Begriff „erklärungs-basiert“ äußerst fragwürdig ist: Während die Prozesse des gedächtnis-basierten Ansatzes ja in der Tat auf allgemeinen Gedächtnisprozessen *basieren*, basieren die Prozesse des erklärungs-basierten An-

satzes ja gerade nicht auf Erklärungen, sondern *zielen* vielmehr auf solche *ab*. Von daher wäre eine Bezeichnung wie „erklärungs-orientiert“ m.E. viel sinnvoller und würde auch die von mir präferierte Integrationsperspektive besser vorbereiten: Sprachverarbeitung basiert auf allgemeinen kognitiven Funktionen, benutzt diese aber zur individuellen Sinngenerierung. Da jedoch der Begriff „erklärungs-basiert“ eine Abgrenzung zum historisch vorgeordneten Konstruktivismus erlaubt und bereits in der Literatur etabliert ist, soll er im Folgenden ebenfalls verwendet werden.

Die Grundannahmen des erklärungs-basierten Ansatzes sind nur sehr unscharf umrissen. Auf höchstem Abstraktionsniveau lässt sich vielleicht feststellen (vgl. auch Sanford & Garrod, 2005), dass eigentlich alle Vertreter/innen so etwas wie ein Diskursmodell als vermittelnde Instanz zwischen Text und Welt annehmen (obwohl sie sich dann nicht darüber einig sind, wie dieses Diskursmodell aussieht). Ebenfalls sind sie sich (halbwegs) einig darüber, dass man versucht, einkommende sprachliche Äußerungen so schnell wie möglich auf dieses Diskursmodell und situationsspezifisches Hintergrundwissen zu beziehen (wie z.B. *scenario mapping*). Eine weitere Gemeinsamkeit besteht schließlich darin, dass sie skeptisch sind, ob Aufbau und Aufrechterhaltung eines solchen referentiellen Bezugsmodells allein durch gedächtnis-basierte Prozesse erklärbar sind. Stattdessen geht diese Theriefamilie in Tradition der konstruktivistischen Inferenztheorie davon aus, dass Sprachverstehen auch analytische Prozesse umfassen muss, mit denen Leser/innen z.B. aktive Gründe für textuelle Ereignisse inferieren. Hierfür muss im Laufe der Textrezeption die Kohärenz des bisher gebildeten Diskursmodells ständig überprüft werden. Wird dabei ein Kohärenzbruch entdeckt, dann versuchen Leser/innen, diesen möglichst schnell und adäquat zu reparieren, indem sie z.B. gezielt nach einer zusätzlichen Erklärung suchen oder weitere Textinformationen und Hintergrundwissen (re)aktivieren, wenn sie für ihr Verständnis wichtig sind. Alle diese Verarbeitungsschritte sind jedoch zumindest teilweise als abhängig von bestimmten Lesermerkmalen (Lesemotivation, Arbeitsgedächtniskapazität, Kohärenzstandards etc.) gedacht, die u.a. bestimmen, wie gründlich sie ausfallen.

In Bezug auf die zugrunde liegenden Verarbeitungsprozesse verhalten sich die meisten Vertreter/innen (ähnlich wie im Konstruktivismus) vergleichsweise neutral, auch wenn durch die Betonung von Leserzielen (die jedoch auch meist eher theoretisch, denn empirisch behandelt werden) top-down-Prozesse eine stärkere Rolle spielen. Inzwischen wären jedoch die allermeisten Vertreter/innen eines eher erklärungs-basierten Ansatzes bereit zur Annahme, dass viele automatisiert ablaufen (müssen). Dies ist aber u.a. deshalb kein Problem, weil auch davon ausgegangen werden kann, dass viele zur Kohärenzbildung notwendigen Wissensbe-

stände und die notwendigen Kontrollprozesse vergleichsweise überlernt und schnell verfügbar sind. Grundsätzlich wird diese Zugänglichkeit jedoch von der jeweiligen Domäne abhängig sein und auch interindividuell stark variieren: Während manche Inferenzen den meisten Leser/innen vergleichsweise einfach fallen, sind sie in anderen Fällen vergleichsweise schwer und anstrengend (vgl. Long & Lea, 2005). In diesen Fällen bedarf es dann eher einer kontrolliert-strategischen Verarbeitung, die durch entsprechende motivationale und kognitive Ressourcen gedeckt sein muss. Durch die Einbeziehung solcher Konzepte wie Verfügbarkeit und Zugänglichkeit nähern sich die erklärungs-basierten Ansätze der gedächtnis-basierten Position deutlich an.

Ebenfalls herrscht eigentlich Einigkeit darüber, dass basale Resonanzprozesse die Verarbeitung beeinflussen *können*. Das ist das minimale Zugeständnis, dass die erklärungs-basierte Position machen muss, um Befunden wie denen von O'Brien et al. (1998; s.o.) gerecht zu werden. Hier wurden Textelemente ja sogar dann reaktiviert, wenn diese Reaktivierung *kontraproduktiv* für den Verstehensprozess ist. Wie sich die verschiedenen Prozessklassen jedoch sonst zueinander verhalten, wird sehr unterschiedlich bewertet. Die meisten Autoren (z.B. Sanford & Garrod, 2005; Magliano & Radvansky, 2001) bezweifeln, dass gedächtnis-basierte Prozesse hinreichend sind, um eine bestimmte Inferenz zu ziehen. Andere Autoren (z.B. Richards & Singer, 2001) scheinen jedoch sogar zu bezweifeln, dass automatische Prozesse notwendig sind (vgl. auch Singer & Richards, 2005). Die erste Position, die davon ausgeht, dass gedächtnis-basierte Prozesse nicht hinreichend, erklärungs-basierte aber notwendig sind, möchte ich im Folgenden als *schwache* erklärungs-basierte Position bezeichnen. Die zweite Position, die davon ausgeht, dass gedächtnis-basierte Prozesse weder notwendig, noch hinreichend sind, erklärungs-basierte jedoch sehr wohl, bezeichne ich als *starke* erklärungs-basierte Position.

Auch hier ist eine deutliche Annäherung von Minimalismus und Konstruktivismus kaum zu übersehen, und zwar von beiden Seiten: Während die gedächtnis-basierte Position berücksichtigt, dass das Produkt des Verstehensprozesses sehr reichhaltig sein kann, konzediert die erklärungs-basierte Position, dass häufig auch hierarchie-niedrige Prozesse am Sprachverstehen beteiligt sind. Das grundsätzliche Problem (vgl. Abschnitt 1.4), dass beide Theorien zumindest partiell aneinander vorbeilaufen, bleibt dabei jedoch bestehen: Während der gedächtnis-basierte Ansatz gewissermaßen das Fundament des Sprachverarbeitungsprozesses abdeckt, bezieht sich der erklärungs-basierte Ansatz auf den kognitiv-motivationalen Überbau. Beide Ansätze behandeln also sehr unterschiedliche Phänomene.

1.3 Gedächtnis- und erklärungs-basierte Verarbeitung bei der pronominalen Auflösung

Bevor weiter auf die beiden unterschiedlichen Ansätze der gedächtnis- und erklärungs-basierten Sprachverarbeitung und ihre Verbindung eingegangen wird, ist es sinnvoll, die Diskussion dadurch zu „erden“, dass an einem konkreten Gegenstand gezeigt wird, wie beide Ansätze genau funktionieren und welche Vorhersagen sie machen. Eine Domäne, die sich dabei in besonderer Weise für eine kontrastive Gegenüberstellung der beiden Ansätze eignet, ist die anaphorische Auflösung. Denn Anaphern sind sprachliche Elemente, die keinen eigenen semantischen Gehalt (bzw. keine eigene Referenz) haben, sondern sich vielmehr auf eine bereits im Diskurs vorhandene Entität beziehen, d.h. mit diesem *koreferieren* (sprachliche Elemente, die sich auf den *nachfolgenden* Diskurskontext beziehen, nennt man *Kataphern*; vgl. im Überblick Garnham, 1999). Den Vorgang der Referenzzuweisung nennt man entsprechend anaphorische Auflösung, und es ist relativ leicht vorstellbar, dass an diesem Zuweisungsprozess sowohl primär gedächtnis- als auch eher erklärungs-basierte Mechanismen beteiligt sind.

1.3.1 Linguistische Ansätze

Die alltagspsychologisch sicher typischste Art der Anapher ist sicherlich das Pronomen (das jedoch linguistisch betrachtet häufig deiktisch verwendet wird, vgl. Cornish, 1989). Betrachten wir das Beispiel (1a).

(1a) Peter sah ein, dass Otto gegen ihn war.

Auch Laien ist intuitiv klar, dass das Pronomen *ihn* in Satz (1a) keine „eigene“ Bedeutung hat, sondern sich vielmehr auf eine andere Person beziehen muss. Eine nahe liegende Interpretation des Pronomens ist in diesem Fall, dass *ihn* mit der Person *Peter* koreferiert. Das heißt, dass man (was den semantischen Gehalt, nicht was die Akzeptabilität des Satzes betrifft) im Satz (1a) statt *ihn* auch einfach wieder *Peter* schreiben könnte (1b):

(1b) Peter sah ein, dass Otto gegen Peter war.

Um solche seltsamen Namenswiederholungen zu vermeiden, behilft man sich in der Linguistik häufig mit schlichten Koindizierungen. Zwei sprachliche Elemente, welche die gleiche Referenz haben, bekommen den gleichen Index zugewiesen (vgl. (1c)).

(1c) Peter_i sah ein, dass Otto gegen ihn_i war.

Das Diskursuniversum dieses Satzes (bzw. besser: dieser Interpretation des Satzes) besteht also aus zwei Personen (*Peter* und *Otto*), die in einer bestimmten Relation zueinander stehen

(in diesem Fall, dass Otto gegen Peter ist). Man bemerke jedoch, dass dies nicht die einzige Interpretation des Satzes ist. Nichts „zwingt“ das Pronomen, mit Peter zu koreferieren; stattdessen könnte es sich auch auf eine ganz andere Person beziehen (nennen wir sie *Klaus*). Man könnte also wieder schreiben:

(1d) Peter sah ein, dass Otto gegen Klaus war.

oder allgemein mit Koindizierung:

(1e) Peter_i sah ein, dass Otto gegen ihn_j war.

Allerdings ist es schwer, diese Interpretation zu generieren, wenn kein entsprechender Kontext vorhanden ist, in dem eine Person *j* (*Klaus*) bereits in das Diskursuniversum eingeführt wurde. Es ließe sich jedoch ziemlich leicht ein solcher Kontext etablieren, z.B. durch einen vorhergehenden Satz wie:

(1f) Otto machte Klaus_j ständig Vorwürfe. (Selbst) Peter_i sah ein, dass Otto gegen ihn_j war.

Schon an diesem sehr simplen Beispiel wird deutlich, dass für die pronominale Auflösung sowohl gedächtnis- als auch erklärungs-basierte Prozesse relevant sind: Zum einen wird man geneigt sein, dass Pronomen mit einer Person im Diskursuniversum zu identifizieren, die vergleichsweise stark aktiviert ist. Zum anderen ist allerdings auch klar, dass eine rein gedächtnis-basierte Erklärung kaum hinreichend sein kann, um die Auflösung des Pronomens umfassend zu erklären. Denn es ist wahrscheinlich, dass der Zuweisung des Pronomens mitunter komplexe, erklärungs-basierte Schlussfolgerungsprozesse zugrunde liegen (Welcher Referent ist plausibler? Peter oder Klaus?) und es zudem noch übergreifende Prinzipien gibt (wie z.B. den ökonomischen *constraint*, das Diskursmodell so einfach wie möglich zu halten), welche die Interpretation mit beeinflussen (vgl. Sanford & Garrod, 2005).

An dem kleinen Beispiel wird allerdings auch ersichtlich, dass bestimmte Interpretationen des Satzes *nicht* möglich sind, und zwar aus rein grammatischen Gründen. Die grammatischen Regularitäten der pronominalen Auflösung teilen sich dabei in zwei große Klassen auf: morpho-syntaktische und bindungstheoretische. Zum einen gibt es im Deutschen (als genus-, numerus- und kasus-flektierende Sprache) die generelle morpho-syntaktische Regel, dass Pronomen und Referent genus- und numerus-kongruent sein müssen. Das Pronomen *ihn* in (1a) muss sich also auf eine männliche Einzelperson beziehen, Referenten wie *Maria* oder *die Studierenden* kommen nicht in Frage. Schwieriger ist es, die zweite Klasse von syntaktischen Beschränkungen zu erklären. Man betrachte jedoch den Satz (1g):

(1g) *Peter sah ein, dass Otto_i gegen ihn_i war.

Hier sind nun *Otto* und *ihn* koindiziert. Die semantische Interpretation entspricht also ungefähr der Tatsache, dass *Otto* gegen *sich selbst* ist. Der Punkt ist nun, dass Satz (1g) im Deutschen ungrammatisch ist (was mit einem initialen Sternchen gekennzeichnet ist): *ihn* darf sich nicht auf *Otto* beziehen, sondern nur auf *Peter* (bzw. irgendeine andere Person). Das genaue Gegenteil gilt jedoch im Fall (1h): Die Anapher *sich* muss sich auf *Otto* beziehen und darf nicht mit *Peter* koreferent sein.

(1h) *Peter_i sah ein, dass Otto gegen sich_i war.

In der Tat lässt sich zeigen, dass Personalpronomina (wie *ihn*) und Reflexivpronomina (wie *sich*) eine vollkommen komplementäre Distribution haben (ähnliches gilt auch für den Kontrast von Personalpronomen mit Reziprokpronomina wie *einander*): Überall, wo die einen auftauchen, können die anderen nicht stehen, und umgekehrt. Diese Regularität wird in den beiden Prinzipien A und B der sog. *Bindungstheorie* (*binding theory*) formalisiert (vgl. Chomsky, 1981):

(2) Prinzip A: Reflexiv- und Reziprokpronomina müssen lokal gebunden sein.

(3) Prinzip B: Personalpronomina dürfen lokal nicht gebunden sein.

Darüber hinaus gibt es noch weitere Prinzipien für Spuren und sog. R-Ausdrücke, aber das ist im Moment nicht relevant. Es ergibt sich die Frage, was *lokal* und *gebunden* hier eigentlich genau heißt. Beides kann durch den Begriff des *c-Kommandos* beantwortet werden (den ich hier vergleichsweise umgangssprachlich expliziere, für eine formale Darstellung vgl. z.B. Fanselow & Felix, 1993b). Denn dass zur Explikation der Bindungsrelation die *lineare* Anordnung der Konstituenten alleine nicht ausreichend ist, macht ein Vergleich von (1c) mit (1i) deutlich, weil hier der Nicht-Referent *Otto* vor dem Referenten *Peter* steht:

(1c) Peter_i sah ein, dass Otto gegen ihn_i war.

(1i) Dass Otto gegen ihn_i war, sah Peter_i ein.

Stattdessen führt kein Weg daran vorbei, die Relation *strukturell-hierarchisch* zu formulieren. Und hier ist die Intuition zunächst einmal die, dass das Antezedens eines Pronomens wohl nicht im gleichen Teilsatz wie dieses stehen darf, sondern strukturell *höher* angesiedelt sein muss (in diesem Fall im Hauptsatz, während das Pronomen im Nebensatz steht). Und ganz im Gegenteil muss anscheinend bei Reflexiv- und Reziprokpronomina das Antezedens gerade im gleichen Teilsatz stehen. Genau dies wird nun im Konzept des *c-Kommandos* ausgedrückt (vgl. (4), nach Grewendorf, Hamm & Sternefeld, 1989):

(4) *Begriff des c-Kommandos:*

Ein Knoten X c-kommandiert einen Knoten Y genau dann, wenn die Bedingungen (a) und (b) erfüllt sind:

(a) X und Y nehmen verschiedene Positionen im Strukturbaum eines Satzes ein, sodass keiner der Knoten den anderen dominiert.

(b) Die nächste maximale Projektion, von der X dominiert wird, dominiert auch Y.

Obwohl sich diese Formulierung zunächst noch kompliziert anhört, ist das, was damit gemeint ist, vergleichsweise einfach. Man betrachte den hierarchischen Strukturbaum des Satzes (1c) in Abb. 1.1, wie er (ungefähr) von einem generativen Ansatz (aber auch diversen anderen Syntax-Systemen) angenommen wird (die Darstellung ist stark verkürzt und sehr frei nach Grewendorf et al., 1989; vgl. auch Fanselow & Felix, 1993a; Radford, 1997).

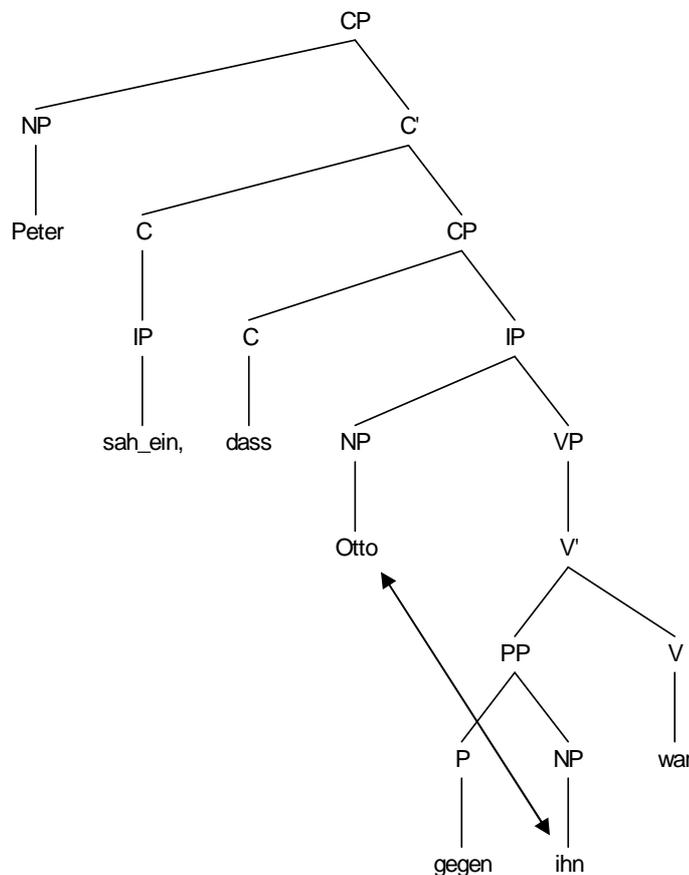


Abb. 1.1: Syntaktischer Strukturbaum für Satz (1). Lokales vs. nicht-lokales c-Kommando.

Es ist eine wesentliche Eigenschaft solcher Syntax-Bäume, hierarchisch strukturiert zu sein: Bestimmte Knoten stehen „tiefer“ im Baum und werden von „höher“ liegenden Knoten dominiert. Eine solche Anordnung haben sehr viele (natürliche wie künstliche) Systeme (wie z.B. auch das Linnische Klassifikationssystem), und im vorliegenden Fall ist es sehr hilfreich, für kurze Zeit den syntaktischen Strukturbaum mal in Analogie zum Stammbau einer Familie zu betrachten: Die einzelnen Wörter und Konstituenten folgen dann in verschiedenen Generationen aufeinander und sind unterschiedlich nah miteinander verwandt. Uns geht es im Mo-

ment um das jüngste Kind der Familie, nämlich das Pronomen *ihn*. Die Relation des c-Kommandos ist in dieser Familien-Analogie gewissermaßen das Onkel-Neffen-Verhältnis. Denn die Definition des c-Kommandos in (4) besagt, dass eine c-kommandierende Konstituente dieses Kommando nicht *direkt* ausüben darf (wodurch alle direkten Vorfahren von *ihn*, die Eltern-PP, die Großeltern-VP etc., schon einmal ausscheiden). Die zweite Bestimmung besagt jedoch zusätzlich, dass die beiden Konstituenten, die in einer c-kommandierenden Relation stehen, von einer gemeinsamen Konstituente dominiert werden müssen, also mindestens einen gemeinsamen Vorfahren haben. Von daher bleiben eigentlich nur noch die Onkel-, Großtanten-, Urgroßonkel-Positionen etc. übrig, die eine Konstituente c-kommandieren können. Und der nächste Onkel von *ihn* ist nun einmal *Otto*, d.h. *Otto* c-kommandiert *ihn* (*Otto* dominiert *ihn* nicht direkt; *Otto* wird jedoch von der gleichen maximalen Projektion dominiert wie *ihn*, nämlich dem Knoten *IP*).

Findigen Lesern/innen wird vielleicht aufgefallen sein, dass das Pronomen *ihn* jedoch nicht nur den „Onkel“ *Otto* hat, sondern auch den „Urgroßonkel“ *Peter*, der entsprechend der obigen Ausführungen ebenfalls *ihn* c-kommandiert. Warum kann *ihn* dann mit *Peter* koreferent sein? Die Antwort auf diese Frage findet sich in der Bestimmung „lokal“ der Formulierung der Bindungsprinzipien A und B in (2) und (3). *Otto* c-kommandiert *ihn* lokal, *Peter* hingegen nicht; deswegen kann *ihn* auch mit *Peter* koreferent sein, nicht jedoch mit *Otto*. „Lokalität“ bestimmt sich dabei dadurch, ob die c-kommandierende Konstituente in der sog. Rektionskategorie des Pronomens enthalten ist oder nicht. Um hier eine weiterführende Definition des Konzepts *Rektionskategorie* zu vermeiden (welches wiederum die Einführung der Begriffe *Rektion*, *m-Kommando* und *vollständiger funktionaler Komplex* notwendig machen würde), sei an dieser Stelle einfach festgestellt, dass ein R-Ausdruck ein Pronomen dann lokal c-kommandiert, wenn er in der gleichen CP wie dieses steht (der/die interessierte Leser/in möge die genaue Ableitung selbst bei Fanselow & Felix, 1993b, nachlesen).

Nicht nur an diesem Beispiel, sondern auch bei vielen anderen wird deutlich, dass es beim Prozess der pronominalen Auflösung also vornehmlich darum geht, einer freien Variable einen bestimmten Wert zuzuweisen und sie dadurch zu binden. Solche Bindungsphänomene sind jedoch in keiner Weise für Pronomen oder Anaphern im Allgemeinen spezifisch. Bekannte Beispiele aus der Semantik, bei der ebenfalls Variablen gebunden werden müssen, sind Quantoren wie *alle*, *einige* etc. Ähnlich wie Anaphern haben solche Quantoren keinen eigenen semantischen Gehalt, sondern dienen eher dazu, den Referenzraum eines anderen sprachlichen Ausdrucks mengentheoretisch zu spezifizieren (*alle Professoren*, *einige Studierende*). Und wenn mehrere Quantoren in einem einzigen Ausdruck verwendet werden, dann

ist auch hier eine Entscheidung darüber notwendig, in welcher Relation diese zueinander stehen. Ein bekanntes Beispiel ist der Satz (5),

(5) Jeder am Lehrstuhl spricht eine Fremdsprache.

der entweder bedeuten kann, dass alle Lehrstuhlmitglieder ein und dieselbe Fremdsprache sprechen (nämlich Englisch), oder aber, dass alle Lehrstuhlmitglieder jeweils eine unterschiedliche Fremdsprache sprechen (Inga Spanisch, David Englisch etc.; vgl. Lohnstein, 1996). Auf der semantischen Tiefenebene (der *logischen Form* sensu Chomsky, 1995) kommen diese Interpretationsunterschiede dadurch zustande, dass im ersten Fall der Existenzquantor im sog. *Skopus* des Allquantors steht und so durch diesen gebunden wird ($\forall x[\text{Lehrstuhlmitglied}'(x) \rightarrow \exists y[\text{Fremdsprache}'(y) \wedge \text{sprechen}'(x,y)]]$), im zweiten Fall jedoch der Allquantor im Skopus des Existenzquantors ($\exists y[\text{Fremdsprache}'(y) \wedge \forall x[\text{Lehrstuhlmitglied}'(x) \rightarrow \text{sprechen}'(x,y)]]$). In ähnlicher Weise ist es möglich, die syntaktisch-semantischen Beschränkungen, welche die pronominalen Interpretation bestimmen, ebenfalls als Skopus-Beschränkungen zu formulieren. Die Relation des c-Kommandos entspricht dann der skopus-höheren Relation etc. Solche und ähnliche Beispiele machen deutlich, dass es eine enge Verbindung zwischen der Bindung von Anaphern und der allgemeinen Interpretation semantischer Ausdrücke gibt. Neuere Ansätze der linguistischen Anapher-Forschung und Bindungstheorie (vgl. z.B. Huang, 2000; Büring, 2005) versuchen deswegen auch, die Problematik der Interpretation von Anaphern nicht allein aufgrund ihrer syntaktisch-strukturellen Regularitäten zu lösen. Stattdessen versuchen sie, im Rahmen einer allgemeinen semantischen Theorie zu explizieren, inwieweit syntaktische Struktur und semantische Interpretation bei der anaphorischen Auflösung ineinander greifen.

1.3.2 Sprachpsychologische Forschung zur pronominalen Auflösung auf der Satzebene

Was jedoch alle linguistischen Theorien – egal ob eher syntaktisch oder eher semantisch orientiert – rundheraus ablehnen, ist die Beantwortung der Frage, wie ein bestimmtes Pronomen denn nun genau aufgelöst wird. Das heißt, sie spezifizieren lediglich die *Bedingungen*, unter denen bestimmte Koreferenzen möglich sind oder nicht, *welche* Variablenbelegung jedoch gewählt wird, sagen sie nicht und wollen das auch nicht sagen („In this book we will usually speak of assignments as contexts, though we will have little to say about just how the assignment comes to reflect an actual context“, Büring, 2005, p. 27). Das heißt, alle linguistischen Sprachtheorien elaborieren zwar, dass das Pronomen *ihn* in Satz (1a) zwar nicht mit *Otto* koreferieren kann, können aber nichts dazu sagen, ob mit *ihn* denn nun Peter gemeint ist oder eine ganz andere Person (außerhalb des aktuellen Kontextes). Diese Frage wird mehr

oder weniger weitergereicht, und zwar an die Sprachpsychologie bzw. die allgemeine Kognitionsforschung (vgl. jedoch die primär linguistischen Ansätze der sog. *Discourse Representation Theory* von Kamp & Reyle, 1993; oder die sog. *Zettelkasten-Semantik* von Heim, 1983).

Dort kümmert man sich in zwei Unterdisziplinen um sie, nämlich einmal im Rahmen der primär auf syntaktische Prozesse bezogenen Forschung auf der Satzebene und einmal im Rahmen der Diskursforschung, bei der Pronomen eher als spezielle Form von Kohäsionsmarkern begriffen und mit anderen anaphorischen Bezügen verglichen werden. In diesem Abschnitt sollen zunächst die wichtigsten Ergebnisse der psycholinguistischen Forschung auf der Satzebene dargestellt werden, im nächsten Abschnitt erfolgt dann eine darüber hinausgehende Einbettung in einen diskurstheoretischen Kontext (vgl. Abschnitt 1.3.3).

Die psycholinguistische Anapher-Forschung auf der Satzebene (vgl. im Überblick z.B. Nicol & Swinney, 2002) betrachtet die pronominale Auflösung v.a. im Kontext des Satz-Parsings und weist zunächst noch eine große konzeptuelle Nähe zur linguistischen Bindungsforschung auf (vgl. Mitchell, 1994). Allerdings ist sie im Vergleich zur rein linguistischen Analyse eher an der online-Analyse der pronominalen Verarbeitung interessiert und damit an der Verarbeitung von Pronomen in *unvollständigen* Sätzen, während die grammatikalische Analyse der Auflösungsbeschränkungen ja meist vom fertigen Satzprodukt ausgeht. Eine implizite Vorstellung ist dabei, dass bei Verarbeitung eines Pronomens dieses eine Suche in der Menge möglicher Antezedents aktiviert. Aus dieser wird dann schließlich ein Antezedens ausgewählt, während die anderen deaktiviert werden (vgl. Abb. 1.2).

Input: Peter sah ein, dass Otto gegen **ihn** war.

1. Phase: Verarbeitung des Pronomens triggert Antezedens-Suche (*ihn*)
2. Phase: Kandidaten-Menge (vorher genannter NPs) wird generiert (*{Peter, Otto}*).
3. Phase: Kandidaten werden aus Antezedens-Menge eliminiert (*{Peter, ~~Otto~~}*).

Abb. 1.2: Phasen der pronominalen Auflösung nach Nicol & Swinney (2002).

Die entscheidenden Fragen innerhalb dieses generellen konzeptuellen Rahmens sind nun z.B., wie schnell die Antezedens-Suche initiiert wird (Phase 1), welche Kandidaten überhaupt in die abzusuchende Kandidaten-Menge gelangen (Phase 2) und nach welchen Kriterien die Eliminierung von Kandidaten vorgenommen wird (Phase 3). Ohne allzu detailliert auf einzelne Ergebnisse einzugehen, sei hier nur das generelle Grundbefundmuster skizziert (vgl. hierzu im Überblick Nicol & Swinney, 2002).

In Bezug auf die erste Fragestellung, nämlich wann die Antezedens-Suche eigentlich initiiert wird, sind die vorliegenden Befunde eher uneinheitlich. Insgesamt liegt Evidenz für die Annahme vor, dass die Suche in der Tat auf dem Pronomen bzw. kurz danach gestartet wird,

aber ebenfalls dafür, dass der Prozess der pronominalen Auflösung mitunter eine ziemlich lange zeitliche Erstreckung hat (vgl. z.B. Ehrlich & Rayner, 1983). Studien, die Aktivierungsmaße als Abhängige Variablen verwenden (*probe recognition* oder lexikalische Entscheidung), deuten auf eine eher frühe Aktivierung der Antezedens-Suche hin, während Lesezeitstudien (*self-paced-reading* oder *eye-tracking*) meist für eine etwas verzögerte Verarbeitung sprechen.

In Bezug auf die zweite und dritte Fragestellung, welche Prinzipien die initiale Generierung des Kandidaten-Sets und den Eliminierungsprozess beeinflussen, herrscht eigentlich Einigkeit darüber, dass folgende Merkmalsgruppen für die pronominale Auflösung relevant sind (vgl. Badecker & Straub, 2002; Nicol & Swinney, 2002):

- Prominenz, Salienz, syntaktische Position: ob der Antezedent als Subjekt oder Objekt des Matrix-Satzes genannt wird; ob Antezedent als Kopf oder Modifizierer einer Phrase fungiert etc.
- Bindungsprinzipien: Restriktionen für Personal- und Reflexiv-/Reziprokpronomen, die sich aus ihrer syntaktischen Relation ergeben: c-Kommando etc. (s.o.).
- Morpho-syntaktische Merkmale: Kategoriale, Genus- und Numerus-Informationen
- Lexikalisch-semantische Merkmale: z.B. Animiertheit, implizite Verbkausalität etc.
- pragmatische Informationen: Plausibilität der Auflösung in Bezug auf Weltwissen etc.

Uneinigkeit besteht allerdings darüber, *wann* die einzelnen Merkmale die pronominale Auflösung beeinflussen. Charakteristisch für diese Linie von Forschung ist die Fragestellung, wann die oben skizzierten Prinzipien der Bindungstheorie den Auflösungsprozess restringieren: Kommen Referenten, die aus bindungstheoretischen Gründen nicht als Antezedens für ein Pronomen in Frage kommen, gar nicht erst in die Kandidaten-Menge? Oder sind sie in der initialen Kandidatenmenge zwar enthalten, müssen dann aber aus ihr eliminiert werden (vgl. z.B. Badecker & Straub, 2002; Nicol & Swinney, 1989; Sturt, 2003a; 2003b)? Das heißt: Ist der Auflösungs-Mechanismus vielleicht „blind“ für strukturell nicht zugängliche Positionen?

In Bezug auf diese Fragestellung ist hinlänglich bekannt, dass kompetente Sprecher/innen in Bezug auf das Verstehensprodukt die Prinzipien der Bindungstheorie meist befolgen (wenn auch nicht ausschließlich anwenden; vgl. z.B. Gordon & Hendrick, 1997; s. auch Harris, Wexler & Holcomb, 2000). Nicol und Swinney (1989) vertraten jedoch die darüber hinausgehende Auffassung, dass bindungstheoretische Restriktionen auch auf der Prozessebene wie ein Filter wirken. Strukturell nicht zugängliche Referenten sollten also im Verlaufe der Satzrezeption überhaupt nicht in die abzusuchende Kandidaten-Menge aufgenommen werden (vgl. auch Clifton, Kennison & Albrecht, 1997). Sie fanden, dass in einem Satz wie (6a), der ein Personalpronomen im Nebensatz enthält, nach der Anapher zwar die beiden Hauptsatz-NPs aktiv sind, nicht jedoch das (strukturell unzugängliche) Nebensatz-Subjekt. In

Satz (6b), der ein Reflexivpronomen enthält, verhält es sich jedoch genau umgekehrt: Hier ist zwar direkt auf dem Pronomen das Konzept *Doktor* aktiv, nicht jedoch die Wörter *Boxer* und *Trainer*. Sie folgern daraus, dass bestimmte Referenten aus bindungstheoretischen Prinzipien in der Tat nicht in die initiale Suchmenge gelangen. Andere Merkmale (wie Genus- und semantische Informationen) dienen dann der weiteren Eingrenzung der Kandidatenmenge.

(6a) Der Boxer sagte dem Trainer, dass der Teamarzt ihn für die Verletzung verantwortlich hält.

(6b) Der Boxer sagte dem Trainer, dass der Teamarzt sich für die Verletzung verantwortlich hält.

Gegen eine solche (starke) Annahme sprechen jedoch jüngere Befunde (Badecker & Straub, 2002, bzw. Sturt, 2003b), die zusätzlich noch Genus-Informationen manipulierten. Denn wenn es stimmt, dass bestimmte Satzpositionen bereits aus rein syntaktischen Gründen vom Auflösungsprozess ganz ausgeschlossen werden, dann sollten Merkmale dieser Konstituenten auch nicht die Auflösung beeinflussen. Entsprechend variierten sie das Geschlecht der nicht-zugänglichen Positionen (vgl. den Kontrast in (7), welcher der Anschaulichkeit halber in Analogie zu (6) konstruiert wurde; Badecker & Straub arbeiteten tatsächlich mit Eigennamen und etwas anderem Stimulusmaterial).

(7a) Der Boxer sagte dem Trainer, dass der Teamarzt ihn für die Verletzung verantwortlich hält.

(7b) Der Boxer sagte dem Trainer, dass die Teamärztin ihn für die Verletzung verantwortlich hält.

(Im Englischen sind Reflexivpronomina – im Gegensatz zum Deutschen – gender-markiert, vgl. *himself* vs. *herself*; deswegen besteht dort die Möglichkeit, auch den analogen Satz (6b) zu überprüfen). Badecker und Straub fanden, dass Sätze, in denen das Gender-Merkmal eines Pronomens mit einem strukturell nicht zugänglichen Referenten übereinstimmte, langsamer gelesen wurden. Im Gegensatz zu Nicol und Swinney (1989) gehen sie deswegen gerade davon aus, dass alle Referenten, die in Bezug auf das Gender-Merkmal passen und hinreichend fokussiert sind, zunächst in die Suchmenge zur anaphorischen Auflösung aufgenommen werden und syntaktische Prinzipien (nur) als nach geordnetes Eliminationskriterium fungieren.

Sturt (2003a, Exp. 1; vgl. auch Sturt, 2003b) geht sogar noch ein Stück weiter. Er konnte zeigen, dass „weiche“ Diskursinformationen (er benutzte stereotypisierte Genderinformationen wie z.B. Arzt = männlich) u.U. sogar die „harten“ Bindungsprinzipien im Verstehensprodukt „überschreiben“ können: Wenn Versuchspersonen Diskursfragmente wie (8) lesen, dann waren sie in einem substanziellen Anteil der Fälle (ca. 30 %) der Meinung, dass Jennifer gespritzt worden war:

(8) Jennifer was pretty worried at the City Hospital. She remembered that the surgeon had pricked herself with a used syringe needle. There should be an investigation soon.

Eine vorhergehende *eye-tracking*-Studie hatte allerdings schon ergeben, dass die Prinzipien der Bindungstheorie sehr wohl und sehr früh wirksam sind, während sich die Diskursinformationen v.a. auf die späten Lesezeitenmaße (second pass reading, Regressionen) auswirkten. Die unterschiedlichen Merkmale sind also unterschiedlich schnell verfügbar. Er interpretiert diesen Befund im Rahmen des *bonding-resolution*-Ansatzes von Garrod und Sanford (1994).

Für die anderen der oben genannten Einflussfaktoren auf die pronominale Auflösung liegen ebenfalls vergleichbare experimentelle Befunde vor, wenn auch diese nicht ganz so umfangreich sind. Zusammenfassend lässt sich vielleicht feststellen, dass vergleichsweise gute Evidenz dafür vorliegt, dass sowohl bindungstheoretische Prinzipien als auch morphosyntaktische Merkmale die pronominale Auflösung früh beeinflussen und die mögliche Kandidatenmenge restringieren. Für die Klasse der lexikalisch-semanticen Merkmale (Belebtheit, Verb-Bias etc.) ist die Evidenz für eine frühe Beeinflussung eher gemischt (vgl. z.B. Hirst & Brill, 1980; Badecker & Straub, 2002). Pragmatische und diskursive Merkmale scheinen zwar einen bedeutenden Einfluss auf das Verstehensprodukt zu haben, werden aber anscheinend vergleichsweise spät verarbeitet (vgl. Sturt, 2003a).

Damit weist die Debatte um die Einflussfaktoren der pronominalen Auflösung auf der Satzebene eine deutliche Strukturparallelität zur allgemeinen Diskussion in der sprachpsychologischen Parsing-Forschung seit ca. 1980 auf: Ebenso wie dort gibt es einen theoretischen Dissens zwischen eher modular-sequenziellen Ansätzen und eher interaktiv-parallelen Modellen (vgl. Pickering & van Gompel, 2006; s. auch Groeben, 2005). Vielleicht ist aber in Bezug auf die pronominale Auflösung die Möglichkeit zu einer integrativen Zusammenführung der zunächst sehr unterschiedlich wirkenden Modelle sogar noch besser: Alle vorliegenden Daten sprechen eigentlich für einen *constraint-satisfaction*-Ansatz (vgl. Trueswell, Tanenhaus & Garnsey, 1994; MacDonald, Pearlmutter & Seidenberg, 1994), bei dem Faktoren der syntaktischen, semantischen und pragmatischen Ebene gleichzeitig und prinzipiell unabhängig voneinander den Auflösungsprozess beeinflussen. Allerdings ist davon auszugehen, dass die einzelnen Einflussfaktoren ein jeweils unterschiedliches Gewicht haben und sowohl zur Aktivierung als auch zur Unterdrückung eines Antezedens führen können (vgl. Gernsbacher, 1989). Um der Vorgeordnetheit der morphologischen und syntaktischen Cues theoretisch gerecht zu werden, ist jedoch kein (starres) initiales Filtermodell notwendig. Stattdessen ist es m.E. vollkommen ausreichend, unterschiedliche zeitliche Verlaufsprofile für die einzelnen Merkmalsgruppen anzunehmen, die damit das Verstehensprodukt in den einzelnen Phasen des Auflösungsprozesses differenziell und selektiv beeinflussen können. Ob ein solcher Ansatz eher im Rahmen eines eher interaktiv-konnektionistischen Ansatzes (im Stile vergleichbarer Modelle

für die Satzverarbeitung, vgl. MacDonald et al., 1994) oder eher durch ein flexibles Filtermodell (wie es z.B. Sturt, 2003b, vertritt) modelliert wird, ist m.E. hingegen zunächst einmal vergleichsweise unerheblich.

1.3.3 Sprachpsychologische Forschung zur pronominalen Auflösung auf der Diskursebene

Parallel zur psycholinguistischen Forschung zur pronominalen Auflösung auf der Satzebene existiert auch eine große Menge an Literatur, welche dieses Phänomen eher im Kontext diskurspsychologischer Fragestellungen betrachtet (vgl. Garrod & Sanford, 1994). Da der zu untersuchende Phänomenbereich beider Forschungsstränge in der Tat genau der gleiche ist und sie sich nur in ihrem theoretischen „framing“ unterscheiden, ist es schon ein wenig erstaunlich, wie wenig Überlappung es zwischen ihnen gibt. Erst in letzter Zeit entstehen langsam Arbeiten, welche versuchen, beide Traditionen miteinander zu verbinden (vgl. z.B. Sturt, 2003a).

Ebenso wie im Bereich der Anapher-Forschung auf der Satzebene, lehnt sich die theoretische Diskussion um die pronominale Auflösung auf der Diskurs-Ebene deutlich an die allgemeinen metatheoretischen Problemkomplexe und Fragestellungen an, die für diesen Bereich charakteristisch sind (vgl. z.B. Kintsch, 1994; Graesser, Millis & Zwaan, 1997; Zwaan & Rapp, 2006). Anstatt Pronomen und ihre Auflösung im Kontext des Satz-Parsings zu sehen und semantische und syntaktische Einflussfaktoren auf den Auflösungsprozess zu identifizieren, werden Pronomen eher als eine basale Form von Inferenz betrachtet (vgl. z.B. Greene et al., 1992) und v.a. auf ihre kohäsive Funktion im Rahmen des Aufbaus einer kohärenten Diskursrepräsentation hingewiesen.

Diskursanalytische betrachtet sind Pronomen zunächst einmal die am häufigsten vorkommende Form von Kohäsionsmarkern, da sie ein explizit-textimmanentes Signal sind, welches die Notwendigkeit mit sich bringt, verschiedene Sätze eines Textes miteinander zu verbinden und zu einer gemeinsamen (Makro-)Repräsentation zu integrieren. Diese Funktion teilen Pronomen mit anderen Formen von Anaphern, allerdings unterscheiden sie sich auch auf mehreren Dimensionen von ihnen. So hat z.B. Ariel (1990; 2004; vgl. bereits Givón, 1983) eine hierarchische Rangreihe verschiedener Formen von (NP-)Anaphern postuliert, die sich hinsichtlich ihres semantischen Gehalts, ihrer diskursiven Präsupponiertheit und ihrer referentiellen Rigidität voneinander unterscheiden (vgl. (9)):

- (9) a) 0-Anaphern: *Caesar_i kam, t_i sah und t_i siegte.*
b) pronominale Anaphern: *Dadurch, dass Caesar_i den Rubikon überquerte, provozierte er_i Rom.*
c) explizite NP-Anaphern: *Alle, die Caesar_i kannten, fürchteten **den Senator_i.***
d) Eigennamen: *Gib Caesar_i, was **Caesar_i** gehört!*

Sog. *0-Anaphern*, worunter u.a. Ellipsen, Spuren und alle koverten Pro-Formen (wie *PRO*, *pro*, vgl. Grewendorf et al., 1989) fallen, haben keine explizite phonologisch-graphematische Realisierung (mehr), weil ihnen diese im Laufe des syntaktischen Derivationsprozesses gewissermaßen abhanden gekommen ist. Sie bedürfen aber sehr wohl (noch) der Interpretation auf der Ebene der Logischen Form: Und an der Tatsache, dass z.B. in dem Beispiel-Satz in (9a) vollkommen klar ist, dass Caesar sowohl kommt als auch siegt (und nicht etwa Asterix), ist ersichtlich, dass auch solche impliziten Anaphern einen korefentiellen Bezug haben (müssen; vgl. McElree & Bever, 1989). *Pronominale Anaphern* machen diesen Bezug eigentlich dann nur noch overt und weisen darüber hinaus einige wenige morpho-syntaktischen Merkmale auf (wie Genus und Numerus), welche ihre Interpretation restringieren. Bei den „höheren“ *NP-Anaphern* ist dies radikal anders: Obwohl diese einen so großen semantischen Eigengehalt haben, dass sie (im Gegensatz zu den Proformen, die koreferieren müssen) als sog. R(eferenzielle)-Ausdrücke auch alleine stehen könnten, beziehen sie sich dennoch anaphorisch auf eine bereits eingeführte Entität im Diskursuniversum: So koreferiert die NP *der Senator* in (9c) z.B. mit *Caesar*, der damit noch weiter bestimmt wird. Allerdings könnte es durchaus auch sein, dass Caesars Bekanntenkreis eine gemeinsame Furcht vor einem anderen Senator teilt (sagen wir: *Brutus*). Dies kann jedoch bei anaphorisch verwendeten *Eigennamen*, die immer genau eine spezifische Entität im Diskursuniversum bezeichnen, dann nicht mehr der Fall sein.

Der Bezug zur Inferenzforschung besteht nun darin, dass in allen skizzierten Fällen die Anapher in irgendeiner Form mit dem jeweiligen Antezedens assoziiert werden muss, entweder über eine direkte Identifizierung oder – falls eine solche scheitert – mittels einer wissensgestützten Brückeninferenz (Haviland & Clark, 1974): So ist bereits die Verknüpfung von Anapher und Antezedens selbst eine (basale) Inferenz, da beide Ausdrücke koindiziert werden müssen (indem z.B. in (9c) den beiden Propositionen $KENNEN(x, y)$ und $FÜRCHTEN(x, z)$ die Variablenbelegung $y = z$ gewählt wird). Darüber hinaus ist jedoch für diese Koindizierung die Konstruktion einer weiteren Proposition notwendig (z.B. $SEIN(Caesar, Senator)$), die der Textrepräsentation ebenfalls hinzugefügt werden muss. Dieses Denkmuster ist sowohl für die frühe Given-New-Forschung charakteristisch (Anaphern sind vor diesem Hintergrund einfach eine spezifische Form von Given-Information, vgl. Clark & Sengul, 1979), hat aber auch die gesamte nachfolgende Inferenzforschung, insbesondere die zur Klasse der wissensgestützten Kausalinferenzen, ganz wesentlich beeinflusst (vgl. z.B. Singer et al., 1992; Noordman, Vonk & Kempff, 1992).

Ariel (1990) hat allerdings darauf hingewiesen, dass sich die verschiedenen Anapherformen auf mehreren Dimensionen voneinander unterscheiden und sich darüber hinaus gegenläufig zueinander verhalten: So nimmt der semantische Gehalt der verschiedenen anaphorischen Formen mit steigender Hierarchiehöhe zwar zu, gleichzeitig jedoch ihre diskursive Bestimmtheit ab: Während 0-Anaphern noch vollständig aus ihrer syntaktischen Position heraus interpretiert werden müssen, ist dies bei pronominalen Anaphern schon nicht mehr der Fall, und mit vollen NP-Formen können sogar neue Entitäten in das Diskursmodell eingeführt werden (d.h. sie müssen nicht zwingend anaphorisch verwendet werden). Bei Eigennamen ist diese Etablierungsfunktion sogar dominant, weshalb ihre anaphorische Verwendung der stilistischen Lizenzierung bedarf (oder zu einer sog. *repeated name penalty* führt, vgl. Gordon et al., 1993). Entsprechend ergeben naturalistische Produktionsstudien auch, dass volle NP-Anaphern und Eigennamen dazu benutzt werden, Charaktere in den Diskurs (wieder) einzuführen, während Proformen dazu dienen, auf einen Charakter, der bereits im Fokus der Erzählung ist, zu verweisen (Marslen-Wilson, Levy & Tyler, 1982).

Die Unterscheidung zwischen den verschiedenen anaphorischen Bezugsformen ist für die folgende Diskussion sehr wesentlich: Denn während man anfänglich (z.B. Haviland & Clark, 1974; Clark & Sengul, 1979) noch dachte, dass man alle anaphorische Bezüge in einem einheitlichen Prozess (nämlich einer seriellen Suche des aktuellen Diskurskontext + evtl. Brückeninferenz) erklären könne, so ist man in der folgenden Diskussion von dieser Vorstellung abgerückt (vgl. z.B. Sanford & Garrod, 1989): Nicht nur in Bezug auf ihre diskursive Funktion, sondern auch in Bezug auf ihre Verarbeitung scheint es einen Unterschied zwischen den hierarchie-niedrigen (0-Anaphern und Pronomina) und den hierarchie-hohen Anaphern (NP-Anaphern und Eigennamen) zu geben. Denn hierarchie-niedrige und -hohe Anaphern unterscheiden sich nicht nur hinsichtlich ihres semantischen Gehaltes und ihrer Fähigkeit, einen spezifischen Diskurs-Referenten identifizieren zu können. Sie divergieren auch bezüglich der Anforderungen, die sie an ihren diskursiven Kontext stellen: Die semantisch leeren („lexikalisch transparenten“ sensu Gernsbacher, 1989), hierarchie-niedrigen Anaphern sind nicht oder nur eingeschränkt dazu in der Lage, selbstständig einen Antezedens aktiv „auszuwählen“, weil sie hierfür semantisch nicht spezifisch genug sind. Stattdessen erhalten sie ihre diskursive Zuweisung passiv aus dem Kontext, d.h. sie sind für ihre Auflösung stärker auf kontextuelle Auflösungs-Cues angewiesen als dies bei den hierarchie-hohen NP-Anaphern der Fall ist.

Ein Teil der zur Auflösung von Pronomen herangezogenen Auflösungsfaktoren ist dabei primär syntaktisch-grammatischer Natur und wurde im Rahmen der pronominalen Auflösungs-forschung auf der Satzebene (vgl. Abschnitt 1.3.2) bereits ausführlich dargestellt. Darüber

hinaus gibt es jedoch auch viele Auflösungs-Cues, die primär *diskursiv* sind, d.h. sie ergeben sich nicht aus einer lokalen Konstellation syntaktischer und semantischer Merkmale, sondern aus der Struktur der bisher aufgebauten Repräsentation des Textes und zusätzlichen Weltwissensbeständen. Zwei bedeutende Bereiche, auf die sich die Forschung zur pronominalen Auflösung auf der Diskursebene konzentriert hat, ist dabei die Zugänglichkeit (*accessability*) des Antezedens und die Kohärenz der finalen Satzstruktur.

In Bezug auf die Zugänglichkeit des Antezedens ist Ariels (1990) *Accessability Theory* einschlägig (vgl. auch Ariel, 2004; Sanders & Gernsbacher, 2004): Pro-Formen als sog. *high-accessability markers* sind die default-Variante des anaphorischen Bezugs und signalisieren eine fortgesetzte Fokussierung des derzeitigen Topik-Referenten, während sog. *low-accessability markers* (NP-Anaphern) einen Topik-Wechsel indizieren (vgl. auch Vonk, Hustinx & Simons, 1992). Die Wahl der jeweiligen anaphorischen Form kovariiert also direkt mit dem kognitiven Status des Referenten im Diskurssystem. (Ariel, 1990, geht sogar soweit zu postulieren, dass syntaktische Prinzipien wie das c-Kommando nur die grammatikalisierte Form solcher Zugänglichkeits-Restriktionen sind). Die verschiedenen Formen des anaphorischen Bezugs sind also eine Möglichkeit, die Aufmerksamkeit der Leser/innen zu steuern und durch den Text zu lenken: Denn die Form der Anapher bestimmt nicht nur, welcher Referent wahrscheinlich gewählt wird (fokussiert vs. nicht-fokussiert), sondern gleichzeitig besteht durch die Wahl der einführenden NP die Möglichkeit, die Aktivierung des Konzepts zu modulieren. So sind z.B. anaphorische Bezüge zu definit eingeführten NPn wahrscheinlicher als zu indefiniten (vgl. Gernsbacher & Shroyer, 1989).

Eine andere prominente Theorie, welche die Zugänglichkeit eines Referenten zum zentralen Auflösungsprinzip macht, ist die sog. *Centering Theory* (Grosz, Joshi & Weinstein, 1995; zusammenfassend: Walker, Joshi & Prince, 1997). Diese hebt allerdings nicht so sehr auf den Produktions-, sondern vielmehr auf den Rezeptionsaspekt ab. Auch hier wird davon ausgegangen, dass Aufmerksamkeitsfokus, inferentielle Komplexität und Form des anaphorischen Bezugs miteinander interagieren und so die Auflösung eines Pronomens determinieren: Entitäten, die im Zentrum der Aufmerksamkeit stehen, werden eher als Referenten ausgewählt als nicht-zentrierte Entitäten. Auch ergibt sich der Kohärenzgrad eines Textes direkt aus der Einhaltung dieses Prinzips.

Es gibt durchaus noch viele andere Theorien, die primär auf das Konzept der Zugänglichkeit abheben (vgl. z.B. Brennan, 1995; Gordon et al., 1993), was unterstreicht, dass solchen aufmerksamkeitsbasierten Faktoren im Laufe der Diskussion ein immer größeres Gewicht zugekommen ist. Weiterhin ist offensichtlich, dass die Aktivierungslogik solcher An-

sätze den Mechanismen der gedächtnis-basierten Sprachverarbeitungstheorien sehr nahe steht. Es ist deswegen auch wenig erstaunlich, dass beide Theriefamilien von den gleichen Einflussfaktoren ausgehen, wie z.B. Merkmalsüberlappung, Elaboration, textuelle Distanz sowie die Anwesenheit anderer potenzieller Antezedenten (vgl. z.B. Myers & O'Brien, 1998). Nach Kehler (1997, vgl. auch Kehler, 2002) haben zugänglichkeits-basierte Theorien zwei entscheidende Vorteile: a) Sie können erklären, warum die spezifische Form des anaphorischen Bezugs selbst einen Einfluss auf den Auflösungsprozess hat; b) Sie können erklären, warum es manchmal zu *garden-path*-ähnlichen Auflösungstendenzen wie in (10) kommt (vgl. Grosz, Joshi & Weinstein, 1995):

- (10)
- a. Terry really goofs sometimes.
 - b. Yesterday was a beautiful day and he was excited about trying out his new sailboat.
 - c. He wanted Tony to join him on a sailing expedition.
 - d. He called him at 6AM.
 - e. **He** was sick and furious at being woken up so early.

Aufmerksamkeitsbezogene Faktoren legen eine Auflösung des Pronomens *he* in (10e) in Richtung *Terry* nahe (er ist der Hauptcharakter der Geschichte und wurde bis jetzt immer – durch die fortgesetzte Verwendung einer pronominalen Anaphern und Subjekt-Positionierung – im Zentrum der Aufmerksamkeit gehalten). Unser Weltwissen sagt uns jedoch, dass *Tony* gemeint sein muss. Initial können die Leser/innen dies jedoch nicht wissen, weshalb sie den Satz (10e) zunächst einmal so verarbeiten, als würde es weiter um *Terry* gehen. Erst später, bei *being woken up so early*, wird klar, dass etwas nicht stimmt und eine Reanalyse des Satzes notwendig ist.

Was solche aufmerksamkeits-getriebenen Theorien allerdings nicht erklären können, ist, wie die finale Interpretation des Satzes denn nun zustande kommt. Dies ist vielmehr die Domäne der sog. kohärenz-getriebenen Theorien. Diese gibt es wiederum in zwei Spielarten: Auf der einen Seite stehen primäre Schlussfolgerungstheorien, die eine starke Überlappung zur Künstlichen Intelligenz (Abk.: KI) haben und meist direkt auf der Ebene der propositionalen Struktur eines Textes operieren; auf der anderen Seite stehen hingegen Ansätze, die eine referenzielle Repräsentation annehmen und den anaphorischen Auflösungsprozess v.a. im Zusammenhang mit dem Aufbau eines Mentalen Modells und der Passung des Pronomens zu schema-, skript- oder szenarien-bezogenen Informationen als ausschlaggebend ansehen.

Prototypisch für die erste Spielart ist der Ansatz von Hobbs (1978, 1979), der explizit den anaphorischen Auflösungsprozess als einen (logischen) Syllogismus betrachtet (vergleichbare Ansätze mit starkem Bezug zur KI-Forschung der 1970er Jahre sind z.B. Winograd, 1972; Charniak, 1972 oder Wilks, 1975). Veranschaulichen lässt sich das Grundprin-

zip solcher Ansätze am besten, wenn man das berühmte Beispiel von Winograd (1972) betrachtet, welches in (11) in der Modifikation von Kehler (2002) wiedergegeben ist (vgl. zum Folgenden Hobbs, 1978, und Kehler, 2002):

(11) The city council denied the demonstrators a permit because...

- a) ...they feared violence.
- b) ...they advocated violence.

Dieses Beispiel eignet sich aus mehreren Gründen gut für die Darstellung der kohärenzgetriebenen syllogistischen Ansätze: a) Offensichtlich liegen kaum aufmerksamkeitsbezogene Cues zur Auflösung des Pronomens *they* vor, b) die beiden Nebensätze sind mit einer kausalen Konjunktion miteinander verbunden (die eine Erklärungsrelation signalisiert) und c) ist dieser Interpretationsunterschied allein auf das Weltwissen zurückzuführen.

Hobbs zufolge muss zur Auflösung des Pronomens nun kein eigener Prozess (z.B. ein Suchprozess etc.) angenommen werden. Vielmehr ergibt sich seine Bedeutung ganz einfach und natürlich aus der allgemeinen Interpretation des Satzes, d.h. bei der Kohärenztablierung. Letztlich ist die pronominale Auflösung also nur ein Epiphänomen eines basaleren Verarbeitungsmechanismus, bei dem Pronomen als freie Variablen aufgefasst werden, die im Laufe eines formallogischen Schlussprozesses gebunden werden. Der Hauptsatz (mit der Struktur VERWEIGERN(Stadtrat, Demonstranten, Erlaubnis)) ist gewissermaßen die Konklusion eines klassischen Syllogismus und gesucht sind die beiden Prämissen, die in ihrer Konjunktion diesen Schluss ermöglichen. Sie müssen jetzt durch eine Art Abduktionsschluss erschlossen werden (wobei die Konjunktion *weil* als Äquivalent zur logischen Implikationsbeziehung fungiert).

(12) Gesetz: FÜRCHTEN(x, Gewalt) \wedge BEFÜRWORDEN(y, Gewalt)
 \wedge ERMÖGLICHEN(Erlaubnis, y, Gewalt) \rightarrow VERWEIGERN(x, y, Erlaubnis)
 Antezedens: FÜRCHTEN(x, Gewalt) bzw. BEFÜRWORDEN(y, Gewalt)

\therefore VERWEIGERN(Stadtrat, Demonstranten, Erlaubnis)

Der entscheidende Punkt ist nun der, dass in den beiden Nebensätzen jeweils eine Antezedens-Bedingung gegeben wird. Ganz im Sinne der deduktiv-nomologischen Erklärung (sensu Hempel & Oppenheim, 1948) bedarf es einer allgemeinen Gesetzmäßigkeit, welche den Schluss vom Antezedens auf die Konklusion logisch rechtfertigt. In diesem Fall besteht dieses „Gesetz“ genau darin, dass man berechtigt ist, jemanden eine Erlaubnis zu verweigern, wenn man selbst Gewalt fürchtet, dieser jemand jedoch Gewalt befürwortet und die Erlaubnis dazu führen würde, dass dieser jemand die Gewalt auch ausüben kann. Wie man jedoch sehen kann, sind die entscheidenden Argumentpositionen dieses Gesetzes zunächst nur mit den frei-

en Variablen *x* und *y* belegt. Erst der Hauptsatz instantiiert eine spezifische Lesart dieses Gesetzes und setzt *x* = *Stadtrat* und *y* = *Demonstranten*; die freie Pronomen-Variable des Nebensatzes wird dann bei Gültigkeit des Gesetzes im Laufe des Schlussprozesses gewissermaßen automatisch mit gebunden.

Der entscheidende Schritt der Interpretation des Satzes ist damit nach Hobbs das Finden des allgemeinen Gesetzes, das gewissermaßen die eigentliche Bedeutung des Satzes abbildet. Die beiden Satzvarianten a) und b) sind hingegen nur unterschiedliche Darstellungsformen dieser geteilten Sinnstruktur. Zentral ist v.a. der Schritt der Bedeutungsgenerierung qua Etablierung einer übergreifenden Erklärungsrelation. Diese ergibt sich dabei in diesem Fall daraus, dass die beiden Nebensätze mit der Konjunktion *weil* verbunden sind; andere Konjunktionen beziehen sich im Hobbschen System auf andere prototypische Relationen: z.B. Kontrast, Parallelität, zeitliche Abfolge etc. Die Parallele zum „need to know“ der konstruktivistischen Inferenztheorie und der Annahme, dass Leser/innen immer Erklärungen für die Textereignisse suchen (vgl. Graesser et al., 1994), ist damit mehr als deutlich („Find a causal chain from the purported cause to the purported effect.“, Hobbs, 1978, p. 328).

Die zweite Spielart kohärenz-getriebener Ansätze bezieht sich auf die Idee, dass Menschen nicht nur über eine abstrakte Beschreibung ihrer Wissensbestände verfügen, sondern diese auch in Form einer referentiellen Sachverhaltsrepräsentation als Mentales Modell verfügbar haben (vgl. Garnham & Oakhill, 1990; im Überblick: Garnham, 2001). Ein Pronomen zu verstehen heißt dann, einen adäquaten Referenten innerhalb dieses Modells zu finden. Im Prinzip unterscheiden sich diese Ansätze nicht fundamental von dem von Hobbs: bei beiden ist die Auflösung eines Pronomens direkt in den übergreifenden Verstehensprozess eingebettet, der als Nebenprodukt die Koreferenz-Relation etabliert. Die Ansätze divergieren lediglich im „Format“, das sie für die zur Interpretation benötigten Wissensbestände ansetzen: Ansätze wie der von Hobbs (1978) gehen davon aus, dass dieses Wissen in Form von elementaren Propositionen in einem mentalen Lexikon gespeichert ist und mit dem inferentiellen Repertoire der Prädikatenlogik kombiniert werden kann. Ansätze, die auf Mentalen Modellen basieren, gehen hingegen von „höheren“ Wissensstrukturen wie Skripten, Schemata und Szenarien aus und nehmen an, dass der aktuelle Text-Input in Bezug auf diese interpretiert wird. In Termini der klassischen Unterscheidung zwischen Textbasis und Situationsmodell (vgl. van Dijk & Kintsch, 1983) findet die Verarbeitung also eher auf der Ebene des Situationsmodells statt.

Ein gutes Beispiel für die unterschiedliche Verarbeitung auf Textbasis- und Situationsmodell-Ebene ist vielleicht die Unterscheidung zwischen *name-* und *role-mapping*, die u.a.

von der Forschergruppe um Garnham gemacht wird (vgl. Oakhill, Garnham & Vonk, 1989; Garnham & Oakhill, 1990; Garnham, Oakhill & Cruttenden, 1992; im Überblick: Garnham, 2001). Betrachten wir hierfür den (etwas blutrünstigen) Satz (12):

(12) Peter ermordete Hans, weil er zuviel wusste.

Nach Rezeption des Satzes wissen wir zunächst einmal nur, dass Peter Hans ermordet hat und dass eine Person zuviel wusste. Darüber hinaus wissen wir, dass dieses Zuviel-Wissen als ursächlich für die Ermordung von Hans durch Peter angesetzt wird. Dadurch dass im Nebensatz nur ein Pronomen und kein Eigename verwendet wird, haben wir keine Informationen darüber, wie die Person heißt, die zuviel weiß, sondern nur Informationen über die *Rolle*, die sie in dem Mentalen Modell erfüllt. Der Prozess des *name-mappings* (bzw. vollständig: *name-to-role-mappings*) bezieht sich nun darauf, dass dieser namenlosen Rolle ein *Name* zugeordnet wird, z.B. *Hans*. Eine solche Zuordnung ist vergleichsweise oberflächlich, weshalb Informationen der propositionalen Textbasis hierfür vollkommen ausreichend sind. So kann z.B. in dem Satz (13) eine solche Zuordnung vollkommen unabhängig von der Bedeutung des Satzes vorgenommen werden; man muss den Satz also nicht wirklich „verstehen“, um das Pronomen aufzulösen, eine Identifikation der Gender-Informationen reicht aus

(13) Petra ermordete Hans, weil er zuviel wusste.

Role-mapping (bzw. auch hier wieder eigentlich: *role-to-role-mapping*) liegt hingegen dann vor, wenn der funktionalen Rolle des Zuviel-Wissens die funktionale Rolle, ermordet zu werden, zugewiesen wird. Eine solche Zuordnung ist modell-interpretativ, weil sie nicht allein durch Oberflächen-Informationen bewerkstelligt werden kann, sondern nur durch Rückgriff auf das zugrunde liegende Situationsmodell. Dies ist ein inferentieller Prozess, der entsprechendes Hintergrundwissen über die Welt (und in diesem Fall über Krimis) voraussetzt.

Vielleicht am konsequentesten wurde diese Verbindung von anaphorischer Auflösung und Situationsmodell-Aktivierung von Sanford und Garrod (1981; 1998) in ihrem *Scenario Mapping and Focus-Ansatz* realisiert (im Folgenden: SMF genannt). Dessen Grundannahme ist, dass alle Bestandteile eines Textes direkt und so früh wie möglich auf das vorhandene Hintergrundwissen der Rezipienten (sowie deren Annahme über die Produzenten) bezogen wird (sog. *primary processing*). Obwohl Sanford und Garrod den Begriff Mentales Modell oder Situationsmodell nicht direkt verwenden, so gehen sie bei ihrer Explikationen der relevanten Wissenseinheiten klar von höheren kognitiven Repräsentationen mit hohem Realitätsgehalt aus. Konkret verstehen sie darunter eine Mischung von skript-, frame- und schema-basierten Wissenseinheiten, deren Zusammenspiel sie unter dem Begriff *Szenario* zusammen-

fassen. Diese Szenarien machen sich an spezifischen Situationen fest und sind im Langzeit-speicher gespeichert.

Insgesamt geht der SMF-Ansatz von vier Wissenssystemen aus, die zur Interpretation eines Textsegmentes zur Verfügung stehen und sich auf zwei Dimensionen einteilen lassen (vgl. Abb. 1.3): Zum einen gibt es eher dynamische Gedächtnissysteme, die wie ein textbezogenes Arbeitsgedächtnis jeweils die aktuell relevanten Textelemente enthalten. Ähnlich wie bei den entsprechenden aufmerksamkeits-basierten Theorien (s.o.) werden diese dynamischen Gedächtnissysteme als *Fokus* bezeichnet. Aufgabe dieses dynamischen Speichers ist es, bestimmte Wissens-elemente des statischen Systems aktuell hervorzuheben und in den Vordergrund zu stellen. Wie der Name *dynamisch* schon impliziert, sind diese Inhalte ausgesprochen flexibel und bilden den Wechsel bzw. Fluss des Geschehens variabel ab.

	Text-basiert	Wissens-basiert
Dynamisch	Expliziter Fokus	Impliziter Fokus
Statisch	Gedächtnis für Diskurs	Szenarien

Abb. 1.3: Die vier Gedächtnissysteme des SMF-Ansatzes.

Auf einer zweiten Dimension variieren die Gedächtnissysteme darin, ob sie eher text- oder eher wissensbasiert sind. Im textbasierten, expliziten Fokus sind die aktuell im Text genannten Entitäten enthalten (also die Charaktere etc.), während im impliziten Fokus die aktuell relevanten Situationsmodell- und Szenarien-Elemente enthalten sind. Sanford und Garrod argumentieren nun dafür, dass diese beiden Gedächtnissysteme zwar getrennt, aber eng aufeinander bezogen sind: Beim Satz (12) sind im expliziten Fokus z.B. bestimmte Entitäten enthalten (*Peter* und *Hans*) und im impliziten Fokus ein Szenario-Skript mit zwei Diskursrollen (das Skript ERMORDEN mit den Rollen *Mörder* und *Ermordeter*). Das eigentlich charakteristische für den SMF-Ansatz ist nun, dass die beiden Gedächtnisrepräsentationen so schnell wie möglich aufeinander bezogen werden, und zwar so, dass die Slots des Skripts im impliziten Fokus jeweils direkt und frühestmöglich mit den aktuellen Aktanten des expliziten Fokus besetzt werden. Dieser Zuordnungsprozess von Aktanten zu Diskursrollen ist dabei ein komplexer, mehrebigiger Integrationsprozess, bei dem low- und high-level-Prozesse miteinander interagieren (*bonding* vs. *resolution*; vgl. Abschnitt 1.4.3).

Damit ergibt sich für den spezifischen Bereich der pronominalen Auflösung ein sehr ähnliches Bild, wie es bereits für die übergeordneten Inferenztheorien insgesamt galt (s. Abschnitt 1.2): Die Auflösung eines Pronomens bzw. die Identifizierung eines adäquaten Antezedenten ist ein ausgesprochen komplexer und vielschichtiger Prozess, der auf mehreren Verständnisebenen gleichzeitig angesiedelt ist. Diese Vielschichtigkeit manifestiert sich konkret

darin, dass anscheinend sehr viele, aber auch sehr unterschiedliche Informationsarten und Auflösungs-Cues vom Sprachsystem für diese Aufgabe herangezogen werden. Einige von diesen, wie z.B. die primär grammatikalischen, sind vergleichsweise rigide und können kaum überschrieben werden. Andere Merkmale, wie z.B. Zugänglichkeits- und Fokus-Informationen, sind hingegen sehr flexibel und beeinflussen den Auflösungsprozess bereits sehr früh. Eine dritte Klasse von Merkmalen, nämlich die pragmatisch-situativen Faktoren, ist zwar sehr mächtig, weil sie die finale Gesamt-Interpretation letztlich determiniert. Allerdings können sich solche Merkmale erst vergleichsweise spät auf den Auflösungsprozess auswirken, weil sie bereits eine abgeschlossene situative Gesamtrepräsentation voraussetzen.

1.4 Integration der gedächtnis- und erklärungs-basierten Erklärungsperspektive

Im bisherigen Kapitel wurde sowohl die minimalistische als auch die konstruktivistische Inferenztheorie und ihre Erweiterungen und Weiterentwicklungen zum sog. gedächtnis- bzw. erklärungs-basierten Ansatz der Sprachverarbeitung aufgearbeitet und die wesentlichen theoretischen Erklärungsmodelle zur pronominalen Auflösung skizziert. In den restlichen Abschnitten geht es darum, die verschiedenen, ausgesprochen heterogenen Auffassungen zumindest soweit zu vereinheitlichen und miteinander zu verbinden, dass sich ein konkretes Prozessmodell ergibt, welches der vorliegenden Arbeit zugrunde gelegt werden kann. Dabei empfiehlt es sich, deduktiv vorzugehen und zunächst einmal auf der allgemeinen Ebene die Schein-Konkurrenz gedächtnis- und erklärungs-basierter Ansätze transparent zu machen und konkrete Ansatzpunkte für eine Verbindung herauszuarbeiten. Diese können dann in einem zweiten Schritt dazu verwendet werden, in Bezug auf den speziellen Gegenstandsbereich der pronominalen Auflösung ein konkretes Prozessmodell zu skizzieren.

1.4.1 Fehlende Theorienkonkurrenz zwischen gedächtnis- und erklärungs-basierten

Sprachverarbeitungstheorien

Gedächtnis-basierte Ansätze gehen davon aus, dass Textelemente durch einen allgemeinen Gedächtnisprozess, wie dem der Aktivierungsausbreitung oder dem der Resonanz, im Verlaufe des Rezeptionsprozesses (re)aktiviert werden. Diese Verarbeitung ist passiv und automatisch, Textverarbeitung ist deshalb im Wesentlichen bottom-up und „umsonst“. Bereits in den vorhergehenden Abschnitten wurde allerdings darauf hingewiesen, dass beim gedächtnis-basierten Sprachverarbeitungsansatz eine starke und eine schwache Version voneinander unterschieden werden muss. Diese unterscheiden sich dahingehend, ob sie solche hierarchie-niedrigen Prozesse (nur) als notwendig zur Generierung einer Inferenz ansetzen oder sogar als hinreichend. Es wurde dafür argumentiert, dass in der Tat einige Evidenz dafür vorliegt, dass

ein Großteil der Textverarbeitung von low-level-Prozessen getragen wird. Die Frage ist allein, ob solche Prozesse auch hinreichend sind, um ein inferentielles Verstehensprodukt gewährleisten zu können. Hierfür ist die Evidenz eher gemischt: Während die minimalistische Inferenztheorie (vgl. McKoon & Ratcliff, 1992) davon ausging, dass die Klasse rein automatisch generierter Inferenzen (sehr) eng begrenzt ist (selbst die pronominale Auflösung gehört ja nicht dazu), gehen die gedächtnis-basierten Sprachverarbeitungstheorien (z.B. Gerrig & McKoon, 2001) davon aus, dass low-level-Prozesse in bestimmten Fällen auch hinreichend sein können, um „höhere“ Verstehensprozesse (wie die anaphorische Auflösung) zu erklären, weil z.B. durch kontextuelle Cues auch entfernte Textinformationen reaktiviert werden können (vgl. z.B. O'Brien et al., 1998). Da gedächtnis- und erklärungs-basierte Verarbeitungsprozesse als dichotom, sich gegenseitig ausschließende Kategorien konzipiert werden, bedeutet das im Gegenzug ebenfalls, dass starke gedächtnis-basierte Ansätze davon ausgehen, dass höhere, erklärungs-basierte Prozesse weder hinreichend (schwache Version), noch notwendig (starke Version) für den Verstehensprozess sind.

Erklärungs-basierte, konstruktivistische Ansätze gehen hingegen davon aus, dass Leser/innen bestimmte Standards und Ziele haben, die sie aktiv im Rezeptionsprozess erfüllen wollen. Hierzu greifen sie u.U. auch auf strategische high-level-Prozesse zurück und die Verarbeitung unterliegt damit prinzipiell einer top-down-Steuerung. Die meisten erklärungs-basierten Ansätze würden deshalb auch davon ausgehen, dass solche high-level-Prozesse notwendig für das Textverstehen sind; die Generalität des Ansatzes macht es darüber hinaus wahrscheinlich, dass sie auch davon ausgehen, dass solche Prozesse auch hinreichend angesetzt werden können. Im Gegenzug wird wieder eine Ausschlussrelation aufgemacht: Auch wenn mitunter konzediert wird, dass hierarchie-niedrige, gedächtnis-basierte Verarbeitungsprozesse notwendig sind, werden sie meist als nicht hinreichend angesehen.

Es wurde bereits vielfach darauf hingewiesen (vgl. z.B. van den Broek et al., 2005), dass die implizite oder explizite Gleichsetzung von gedächtnis-basierten Prozessen mit bottom-up und erklärungs-basierten mit top-down-Prozessen sehr wackelig ist. Auf der einen Seite schließt ein allgemeiner Gedächtnisprozess wie z.B. der Resonanz-Mechanismus nicht aus, dass u.U. auch höhere Wissensstrukturen (wie Schemata, Skripte) mitaktiviert werden. Diese können als habitualisierte top-down-Strukturen ebenfalls den Verstehensprozess mit beeinflussen, auch wenn sie rein passiv aktiviert werden. Auf der anderen Seite ist davon auszugehen, dass auch die search-after-meaning-Prozesse der erklärungs-basierten Ansätze nicht zwangsläufig (nur) strategische top-down-Prozesse sind, da ja bereits in der Zielidee einer adaptiven Suche in Hinblick auf die motivationalen Ziele des Lesers die Berücksichtigung des

aktuellen Textinputs implizit enthalten ist. Von daher ist eher davon auszugehen, dass sowohl gedächtnis- als auch erklärungs-basierte Prozesse eher eine (komplexe) Mischung von bottom-up- und top-down-Anteilen sind, allerdings mit unterschiedlicher Schwerpunktsetzung.

Was gedächtnis- und erklärungs-basierte Theorien stattdessen am meisten voneinander unterscheidet, ist ihre grundsätzliche *Perspektive* auf den Sprachverarbeitungsprozess. Denn vor dem Hintergrund, dass die Verarbeitung von Sprache ein höchst komplexer, holistischer Gesamtprozess ist, bei dem verschiedene Ebenen miteinander interagieren, scheint es eher so, als würden sich beide Theoriefamilien auf unterschiedliche Ebenen des Verstehensprozesses beziehen: Während gedächtnis-basierte Theorien eher auf die konkreten Verarbeitungsprozesse der molekularen Ebene einzelner Texte, manchmal sogar nur einzelner Sätze fokussieren, haben erklärungs-basierte Theorien eine eher molare Orientierung und wollen den Verstehensprozess als Ganzes modellieren, der z.B. auch motivationale Aspekte mit umfasst. Während erstere durch ihre konsequente Ausblendung des kognitiv-motivationalen Überbaus des Verstehensprozess evtl. entscheidende Teilaspekte übersehen, sind letztere auf der Ebene der konkreten Verarbeitungsprozesse vergleichsweise unterspezifiziert.

Damit laufen beide Theoriefamilien jedoch tendenziell aneinander vorbei und stehen deshalb auch nicht – obwohl genau dies insbesondere in der Minimalismus-Debatte immer wieder behauptet wurde – in einer echten Konkurrenzrelation zueinander (vgl. auch Christmann & Groeben, 1999); ihre Kontrastierung ist vielmehr eine falsche Dichotomie (Garnham, 1992). Wissenschaftstheoretisch betrachtet handelt es sich beim sog. gedächtnis- und erklärungs-basierten Ansatz um zwei verschiedene Forschungsprogramme, d.h. als Ansammlung von Theorien, die als Mittel zur Lösung unterschiedlicher objekttheoretischer Problemkomplexe fungieren (vgl. Herrmann, 1976; 1994, im Überblick s. auch Groeben, 1999). Als solche liegen ihnen unterschiedliche Annahmenkerne, aber auch unterschiedliche Zielvorstellungen zugrunde (vgl. auch Westermann, 2000). Der wesentliche identitätsverleihende Annahmekern des gedächtnis-basierten Forschungsprogramms ist dabei eine Festlegung auf die Domäne der aktualgenetischen Textverarbeitung und das zentrale Problem besteht darin, wie diese im Rahmen eines kapazitäts-beschränkten kognitiven System psychologisch plausibel modelliert werden kann. Der Annahmekern des erklärungs-basierten Forschungsprogramms besteht hingegen in der Einbettung des aktualgenetischen Verstehensprozesses in den größeren Rahmen einer psychologischen Handlungstheorie und die zentrale Problemstellung besteht darin, wie sich dieses Einbettungsverhältnis möglichst adäquat und reibungslos erklären lässt (vgl. hierzu auch Christmann & Groeben, 1997; Herrmann, 1995). Wie noch zu zeigen sein wird, haben diese unterschiedlichen Ausgangsbedingungen erhebliche Implikationen für

die jeweils gewählte Theoriestructur und die Art ihrer empirischen Überprüfung (bis hin zum Aufbau der jeweiligen Experimente und dem Einsatz unterschiedlicher Methoden).

Damit stehen gedächtnis- und erklärungs-basierte Sprachverarbeitungstheorien jedoch in keiner (echten) Konkurrenzrelation zueinander. Nach Herrmann (1979, S. 49) liegt diese nur vor, wenn „zwei oder mehrere (nicht ineinander überführbare) Theorien in einer Situation von der Art des Entscheidungsexperimentes stehen, wenn sie also *bei Voraussetzung derselben Meß- bzw. Testtheorien und unter identischen Meß- bzw. Testbedingungen (Randbedingungen) unterschiedliche Daten vorhersagen.*“ Genau dies ist jedoch m.E. bei der Kontrastierung gedächtnis- und erklärungs-basierter Theorien nicht gegeben, denn: „Werden zwei Theorien als Rekonstruktionsmittel für zwei verschiedene Domains oder für verschiedene Domain-Komponenten eines Forschungsprogramms verwendet, so liegt keine Theorienkonkurrenz vor“ (l.c.). Dieser Fall wird von Herrmann vielmehr als „fehlende Theorienkonkurrenz“ bzw. „Pseudokonkurrenz“ bezeichnet (o.c., S. 50).

Vielleicht ist es sinnvoll, an dieser Stelle darauf hinzuweisen, dass die Pseudokonkurrenz der beiden Ansätze m.E. nicht oder zumindest nicht primär dadurch zustande kommt, dass sie mit unterschiedlichem Stimulusmaterial operieren (also v.a. unterschiedliche Testbedingungen aufweisen, vgl. z.B. Groeben, 2005). Auch kann man die Scheinkonkurrenz m.E. nicht einfach dadurch auflösen, dass man in Abhängigkeit von spezifischen Leserzielen mal das eine und mal das andere Modell als zutreffend ansetzt (vgl. z.B. Christmann & Groeben, 1999). Das Problem besteht meiner Auffassung nach wirklich darin, dass beide Theoriefamilien sich auf jeweils unterschiedliche Teilaspekte des Verstehensprozesses beziehen. Gedächtnis-basierte Theorien untersuchen die Wichtigkeit genereller gedächtnis-basierter Prozesse beim Sprachverstehen und siehe da: Sie finden jede Menge Evidenz dafür, dass solche in der Tat beim Sprachverstehen wirksam sind. Erklärungs-basierte Theorien betonen die Notwendigkeit, die basalen Verstehensprozesse in eine übergeordnete Erklärungsstruktur einzubinden; und auch hier findet sich jede Menge Evidenz dafür, dass dies der Fall ist. Ein solches Befundmuster ist jedoch m.E. kein Indiz für Konkurrenz der beiden Ansätze, sondern vielmehr für ihre *Konvergenz*. Das Problem entsteht erst dadurch, dass beide Positionen ihre Befunde übergeneralisieren und im (falschen) Umkehrschluss postulieren, dass die Gegenposition deshalb unrecht hat.

Dabei sind beide Ansätze in ihrer jeweils schwachen Version perfekt miteinander kompatibel: Sowohl gedächtnis- als auch erklärungs-basierte Prozesse sind *notwendige* Bestandteile des Verstehensprozesses, keiner der beiden Prozesstypen ist alleine *hinreichend*. Die Notwendigkeit zur Kombination der beiden Prozesstypen in einem multiplen hinreichenden

Kausalschema ergibt sich dabei dadurch, dass beide tendenziell für unterschiedliche Teilaspekte des Verstehensprozesses relevant sind: gedächtnis-basierte für die aktualgenetische Verarbeitung auf der molekularen Mikro-Ebene, erklärungs-basierte für die zielorientierte Verarbeitung auf der molaren Makro-Ebene. Deswegen sind m.E. auch beide Ansätze komplementär ausgelegt und sollten sich gut in einem gemeinsamen Rahmenprogramm kombinieren lassen. Problematisch ist allenfalls der Übergangsbereich, in dem sich beide Ebenen überschneiden. Für diesen gilt es eine plausible Integrationsperspektive zu entwickeln, die den Ansprüchen beider Positionen gerecht wird.

1.4.2 Allgemeine Ansätze zur Integration gedächtnis- und erklärungs-basierter Verarbeitungsprozesse

Wenn man die eben skizzierte Perspektive wirklich ernst nimmt, dann impliziert dies eine Sichtweise auf das Sprachsystem, in dem beide genannten Prozessarten notwendige Bestandteile des Verstehensprozesses sind: Sowohl die schnellen, aber „dummen“ low-level-Prozesse des gedächtnis-basierten Ansatzes als auch die langsamen, aber „smarten“ high-level-Prozesse des erklärungs-basierten Ansatzes tragen zum Verstehensprodukt bei, wobei sie sich gegenseitig komplementieren. Eine umfassende Theorie des Sprachverstehens sollte deswegen auch nicht nur einfach beide Prozessarten in einem gemeinsamen Rahmen modellieren können, sondern sollte darüber hinaus auch ihre (dynamische) Interaktion abbilden können, und zwar in beide möglichen Richtungen: hemmend oder fördernd. Anstatt also zu fragen, welcher Prozesstyp für das Verstehen von Sprache wichtiger oder angemessener ist, geht es v.a. um die Frage, wie bottom-up-Prozesse das Verstehensprodukt aufbauen und wie top-down-Prozesse diesen Konstruktionsprozess leiten können.

Die Einsicht darin, dass eine Integration beider Ansätze nicht nur möglich, sondern auch notwendig ist, wird in der *scientific community* inzwischen allgemein (ein)gesehen: Im Jahre 2005 widmete die Zeitschrift *Discourse Processes* dieser Fragestellung sogar eine eigene Sonderausgabe (Guéraud & O'Brien, 2005a) und ließ Vertreter/innen sowohl eher gedächtnis- als auch eher erklärungs-basierter Ansätze darüber nachdenken, wie eine solche Integration aussehen könnte. Wenn man sich die dort versammelten Vorschläge genauer betrachtet (und die übliche Mischung von Allgemeinplätzen und Eigenwerbung dabei ignoriert), dann kristallisieren sich in meinen Augen vier hauptsächliche Integrationsperspektiven ab (die weder exhaustiv noch wechselseitig exklusiv sind):

1. Zunächst einmal wird darauf hingewiesen (und damit eine Argumentationsfigur der Minimalismus-Debatte repetiert, vgl. z.B. Garnham, 1993), dass die für den erklärungs-

basierten Ansatz relevanten Prozesse nicht immer zwangsläufig kontrolliert-strategisch sein müssen. Vielmehr wird von erklärungs-basierter Seite konzediert, dass es häufig vergleichsweise hierarchie-niedrige Prozesse sind, welche die lokale und globale Kohärenz gewährleisten. Von der gedächtnis-basierter Seite wird hingegen das Zugeständnis gemacht, dass u.U. auch hoch organisierte Wissensstrukturen (wie Schemata, Skripte etc.) durch den passiven Resonanz-Prozess mitaktiviert werden können, solange diese nur gut gelernt und damit leicht zugänglich sind. So gehen z.B. Sanford und Garrod (2005) in ihrem SMF-Ansatz davon aus, dass auch komplexe Szenarien mit dem Resonanzprozess abgerufen werden können und so sehr schnell im impliziten Fokus zur Verfügung stehen.

2. Einen Schritt weiter gehen bereits Integrationsvorschläge, welche die erklärungs-basierten Prozesse als eine Art übergeordneten Aufmerksamkeitsprozess betrachten, welcher die Verarbeitung auf der gedächtnis-basierten Ebene lenkt. Ähnlich wie in der Aufmerksamkeitsforschung allgemein stehen sich hier Konzeptionen, die eher von einer *Filter*-, und Konzeptionen, die eher von einer *Verstärker*-Wirkung ausgehen, gegenüber. Beiden ist jedoch die Idee gemeinsam, dass die spezifischen Ziele, die Leser haben, darüber entscheiden, welche Textinformationen fokussiert werden. Diese erhalten dann z.B. im Resonanz-Prozess eine höhere Signalstärke, wodurch sich die Textrepräsentation evtl. ändert. Ein guter Kandidat für eine solche Aufmerksamkeitssteuerung qua Fokussierung sind z.B. Ziel- und Motiv-Informationen von Protagonisten. So ist z.B. gut bekannt (vgl. Arnold, 2001), dass GOAL-Informationen von Sprach-Rezipienten/innen stärker fokussiert und eher elaboriert werden. Aufgrund dieser stärkeren Aktivierung werden dann wiederum im Resonanz-Zyklus ebenfalls verstärkt Informationen über den Protagonisten und seine Motive abgerufen etc. Ein solcher, sich gegenseitig verstärkender Fokussierungsprozess könnte z.B. eine im Rahmen des Resonanz-Mechanismus plausible Erklärung dafür liefern, wie Leser/innen aktiv Informationen suchen können, um sich die im Text genannten Ereignisse zu erklären.

3. Wiederum einen Schritt weiter gehen Vorschläge, welche erklärungs-basierte Prozesse nicht nur indirekt qua Fokussierung in den Verstehensprozess integrieren, sondern ihnen eine aktive Rolle zusprechen. Allerdings sind solche höheren Prozesse nicht im „Normalfall“ der Sprachverarbeitung operativ, sondern nur als eine Art *repair*-Mechanismus in dem Falle, dass eine Reanalyse notwendig wird. Das heißt, normalerweise sind schnelle bottom-up-Prozesse vollkommen ausreichend, um das Verstehen zu sichern, wenn diese jedoch versagen, dann wird ein langsamerer, algorithmischer Prozess der Inferenzgenerierung gestartet. Eine solche Notfallsituation liegt z.B. dann vor, wenn die bottom-up-Prozesse nicht genügend Informationen reaktivieren oder diese dem bisherigen Diskursmodell widersprechen (vgl. Riz-

zella et al., 1998). Und als remediale Maßnahmen, die in einer solchen Situation zur Anwendung kommen, werden u.a. die Refokussierung auf andere aktive Elemente im Arbeitsgedächtnis, ein „Neustart“ des Resonanzprozesses, die Anwendung einer aktiven Problemlöse-Strategie, oder aber einfach die Rezeption des nächsten Satzes (bei dem sich der Konflikt evtl. auflöst) genannt. Die Ähnlichkeit dieses Integrationsvorschlages zur Klasse der seriell-modularen Parsing-Modelle auf der Satzebene (wie dem garden-path-Modell von Frazier, 1987, mit einer schnellen syntaktischen Erstverarbeitungs- und langsamen, semantischen Reanalyse-Phase) ist kaum zu übersehen.

4. Eine konsequente Weiterentwicklung des letzten Vorschlags besteht darin, den Einfluss der erklärungs-basierten top-down-Prozesse nicht nur auf eine finale Reanalyse-Phase zu beschränken, sondern sie im Verlauf der gesamten Satz- bzw. Textverarbeitung als eine Art kontinuierliche Kontrollinstanz anzusetzen. Das heißt, die hierarchie-hohen Prozesse, die z.B. dazu dienen, die globale Kohärenz eines Textes zu gewährleisten oder leserziel-abhängig Informationen zu aktivieren, evaluieren die allgemeinen Gedächtnisprozesse der Mikroebene kontinuierlich und initiieren z.B. eine zyklische Wiederholung des Resonanzprozesses bis ein bestimmter Zielzustand erreicht ist (ein gewisser Kohärenzstandard erreicht, ein Antezedens identifiziert etc.) ist. Nach dieser Auffassung besteht zwischen den erklärungs-basierten high-level-Prozessen auf der Makro- und den gedächtnis-basierten low-level-Prozessen auf der Mikro-Ebene eine hierarchische Beziehung, wobei sich erstere immer nur *über* letztere auf den Verstehensprozess auswirken können.

Wie gesagt ist diese Einteilung höchstens als Unterscheidung einiger prototypischer Spielarten der gemeinsamen Modellierung von gedächtnis- und erklärungs-basierten Prozessen mit einem primär heuristischen Wert zu begreifen. Es existieren zahlreiche Mischformen, deren genaue Ausformung auch noch eng vom jeweiligen Objektbereich (z.B. Modellierung von elaborativer Inferenzen vs. anaphorische Auflösung) abhängig ist. Über die Sinnhaftigkeit des einen oder anderen Integrationsvorschlages lässt sich deswegen m.E. auch erst dann ernsthaft diskutieren, wenn man sich ein spezifisches sprachliches Phänomen herausgegriffen hat und versucht, für dieses ganz konkrete Vorschläge zur Verbindung beider Prozessarten zu machen. Dies soll im nächsten Abschnitt für den Objektbereich der pronominalen Auflösung durchgeführt werden.

Allerdings ist schon bei einer oberflächlichen Betrachtung der verschiedenen Integrationsansätze klar, dass im Rahmen einer integrativen Modellierung u.U. Zugeständnisse notwendig sind. So laufen eigentlich alle skizzierten Varianten darauf hinaus, dass (im Gegensatz zur Grundprämisse der gedächtnis-basierten Ansätze) hierarchie-niedrige, allgemeine Ge-

gedächtnisprozesse nicht allein für den Verstehensprozess verantwortlich sein können. Im Gegenteil sind auch hierarchie-hohe Prozesse ein integraler Bestandteil der Sprachverarbeitung, ja bei den meisten Integrationsversuchen sind sie den gedächtnis-basierten Prozessen sogar klar übergeordnet. Auf der anderen Seite macht schon dieser erste Überblick über mögliche Integrationsperspektiven klar, dass erklärungs-basierte Theorien die Tatsache berücksichtigen müssen, dass die dummen, low-level-Prozesse wahrscheinlich den Großteil der Sprachverarbeitung bestreiten und die für die Erklärungsperspektive relevanten Prozesse aus kapazitäts-technischen Gründen eher selten zum Einsatz kommen. Das heißt, hierarchie-hohe, top-down-Prozesse dominieren den Verarbeitungsprozess nie und sind evtl. sogar nur im Problemfall wirksam.

1.4.3 Interaktion gedächtnis- und erklärungs-basierter Verarbeitungsprozesse bei der pronominalen Auflösung

Wie in Abschnitt 1.3 ausführlich dargestellt, nehmen sehr viele Faktoren auf sehr unterschiedlichen Ebenen Einfluss auf die pronominalen Auflösung. Sie reichen von primär syntaktischen Restriktionen (c-Kommando, morpho-syntaktische Kongruenz) über die Zugänglichkeit und Salienz der Referenten bis hin zu pragmatischen Faktoren, die sich auf die Plausibilität der Auflösung beziehen. Darüber hinaus ist zu bedenken, dass Pronomen als semantisch leere Anaphern nur in einem sehr geringen Ausmaß dazu in der Lage sind, aus eigener Kraft einen eindeutigen Referenten festzulegen. Im Gegenteil bekommen sie ihre Referenz größtenteilig vom Kontext zugewiesen, und dies auf eine vergleichsweise passive Art. Prinzipiell ist deswegen für die pronominalen Auflösung der Einfluss hierarchie-niedriger Faktoren als hoch einzuschätzen. Wie entsprechende Korpus- und Produktionsstudien (vgl. Ariel, 1990; Marslen-Wilson et al., 1982) eindrücklich belegen, werden Pronomen in der alltäglichen Kommunikation v.a. dann verwendet, wenn ein Antezedent strukturell klar definiert und hoch verfügbar ist (Salienz, Zentriertheit etc.).

Von daher ist in Bezug auf den konkreten Phänomenbereich der pronominalen Auflösung davon auszugehen, dass gedächtnis- und erklärungs-basierten Verarbeitungsprozessen wirklich unterschiedliche Funktionen und Rollen zukommen, die jedoch aufeinander bezogen werden können. Denn auf höchstem Abstraktionsniveau lässt sich der Prozess der pronominalen Auflösung als ein Verstehensprozess charakterisieren, bei dem einerseits zunächst ein Referent aktiviert werden muss (Informationsaktivierung), bevor dann in einem zweiten Schritt die diskursive Relation zwischen Antezedens und Pronomen bestimmt werden kann (Integration). Von daher scheint sich in diesem Fall eine wirkliche Arbeitsteilung anzubieten:

Während für die Aktivierung des Antezedens primär gedächtnis-basierte low-level-Prozesse relevant sind, ist für die interpretations-etablierende Integration ein erklärungs-basierter Mechanismus entscheidend. Darüber hinaus es ist wahrscheinlich, dass höhere sprachliche Informationsquellen (wie pragmatische oder Diskurs-Informationen) v.a. dann effizient sind, wenn es gilt, zwischen verschiedenen bereits etablierten Strukturen zu *entscheiden*, und nicht, wenn es gilt, diese Strukturen *aufzubauen*. Ähnlich wie für den Bereich der syntaktischen Ambiguitäten auf der Satzebene bzw. für die lexikalischen Ambiguitäten auf der Wortebene kann auch hier festgehalten werden (vgl. z.B. MacDonald et al., 1994): Pragmatische Kontextinformationen wirken sich v.a. dann gewichtend auf den Verarbeitungsprozess aus, wenn überhaupt verschiedene, ungefähr gleich wahrscheinliche Interpretationen zur Verfügung stehen.

Diese Eigenschaften des pronominalen Auflösungsprozesses scheinen für einen prinzipiell zweiphasigen Prozess zu sprechen (vgl. Cook & Myers, 2004): In einer ersten, primär von syntaktischen und basalen Gedächtnisprozessen getragenen Phase wird eine Verbindung zwischen dem Pronomen und dem aktuellen Textinput im Arbeitsgedächtnis etabliert. Mittels eines automatischen Abrufmechanismus werden sowohl episodische Informationen über den bisher verarbeiteten Textinput als auch semantische Informationen über das generelle Weltwissen aktiviert, die auf die dargestellte komplexe Weise via Resonanz mit den gerade aktiven Inhalten des Arbeitsgedächtnisses interagieren. In einer zweiten, primär erklärungs-basierten Phase müssen die aktivierten Informationen mit den aktuellen Inhalten des Arbeitsgedächtnisses integriert werden. Solche Integrationsprozesse sind erheblich langsamer und greifen auf pragmatisches sowie diskursives Kontextwissen zurück (welches ggf. im Laufe der ersten Phase passiv mitaktiviert wurde). Deshalb zeigt sich ihr Einfluss vornehmlich in späten Maßen der Verarbeitung gegen Satzende (spillover-Effekte bei Lesezeiten, Regressionen und *second-pass-reading* beim *eye-tracking* etc.).

Diese zwei-phasige Konzeption wird allerdings schon seit längerer Zeit nicht nur, aber insbesondere für den Bereich der pronominalen Auflösung vertreten, nämlich von Garrod und Sanford mit ihrem sog. *bonding-resolution*-Ansatz (zunächst: *bonding-binding*-Ansatz genannt; vgl. Garrod & Sanford, 1989; 1994; 1999; s. bereits Sanford & Garrod, 1981; Garrod & Sanford, 1983). Dieser ist durch die Beobachtungen motiviert, dass in Satzpaaren wie (14), in denen kein explizites Antezedens für das Pronomen vorhanden ist (sog. „unheralded pronouns“ nach Greene, Gerrig, McKoon & Ratcliff, 1994) der Eindruck einer garden-path-Auflösung entsteht:

- (14) a. John sailed to Ireland.
b. It sank down.

Obwohl durch entsprechende Weltwissensbestände gesichert ist, dass es nicht Irland ist, welches untergeht (sondern z.B. das Boot, mit dem John segelt), wird der Satz initial genau so verstanden (sog. *sounds-like*-Empfindung, vgl. Sanford, Garrod, Lucas & Henderson, 1983). Auch sind die Lesezeiten des zweiten Satzes deutlich erhöht, weil eine vergleichsweise weit reichende (Brücken-)Inferenz generiert werden muss.

Interessanterweise tritt dieser Effekt aber nicht bei Satzpaaren wie (15) auf (vgl. Sanford & Garrod, 2005):

- (15) a. John sailed to the islands.
b. It sank down.

Zwar ist auch hier der Rezeptionsfluss deutlich gestört (es muss ja immer noch inferiert werden, dass mit *it* Johns Boot gemeint ist), aber irgendetwas hat sich geändert. Ein wesentlicher Unterschied zwischen den Satzpaaren (14) und (15) ist, dass für das Pronomen *it* in (14) ein adäquater, d.h. genus- und numerus-kongruenter Referent zur Verfügung steht (*John* scheidet in beiden Sätzen aus), in (15) aufgrund der Plural-Markierung von *the islands* jedoch nicht. Darüber hinaus stellt sich der Effekt ebenfalls nicht ein, wenn auf einen Satz wie in (14) ein Satz folgt, in dem überhaupt kein koreferenzielles, sondern nur ein semantisches leeres (sog. expletives) Pronomen folgt (*It was a nice day.*). Allerdings weiß der Parser bei der Verarbeitung von *it* (noch) nicht, ob es sich um ein referentielles oder nicht-referentielles Pronomen handelt.

Um dieses Befundmuster erklären zu können, gehen Sanford und Garrod (2005) davon aus, dass es in der pronominalen Auflösung eine frühe *bonding*-Phase gibt, in welcher das Pronomen mittels eines schnellen pattern-matching-Prozesses mit einem potenziellen Antezedens eine Verbindung eingeht. Diese etabliert aber nur einen möglichen Bindungspunkt, dem zunächst noch kein semantischer Wert zugewiesen wird. Die eigentliche koreferenzielle Bindung des Pronomens findet erst in einer zweiten, späten *resolution/binding*-Phase statt, wenn sich herausgestellt hat, dass das Pronomen überhaupt eine freie semantische Variable ist, welche der Bindung bedarf. Und eine solche Integration in die bestehende Diskursrepräsentation ist erschwert, wenn kein geeignetes Antezedens zur Verfügung steht, sondern vielmehr inferiert werden muss.

Als Generalisierung der Überlegungen von Garrod und Sanford lässt sich festhalten, dass es also durchaus Evidenz für die Aktivität eines frühen Fit-Mechanismus gibt, der auf der Grundlage eher basaler Merkmale (Fokussiertheit, morpho-syntaktische Merkmale) einige

Entitäten des derzeitigen Diskursmodells sehr schnell als mögliche Antezedenten für ein Pronomen identifiziert und so gewissermaßen zur weiteren Analyse herausfiltert. Obwohl der Fitting-Mechanismus für seine Arbeit durchaus auch auf höhere Wissensstrukturen zurückgreifen kann (wie z.B. ein Segel-Szenario), ist er selbst doch vergleichsweise dumm und arbeitet nur mit semantischen Assoziationen und der syntaktisch-strukturellen Konfigurationen (die allerdings miteinander interagieren können, vgl. z.B. die Ergebnisse von Sturt, 2003a). An dieser Stelle gibt es theoretische Anschlussmöglichkeiten sowohl zu den entsprechenden Theoriebildungen der Linguistik und den psycholinguistischen Prozessmodellen auf der Satzebene als auch zu den gedächtnis-basierten Ansätzen auf der Diskursebene.

Allerdings ist dies nur die eine Hälfte der Geschichte. Denn wie gesagt kann das frühe Fitting nur als eine Art Heuristik dienen, nicht jedoch die korrekte finale Interpretation des Pronomens gewährleisten. Hierfür ist eine abschließende Integration der voraktivierten Referenten in ein adäquates Diskursmodell notwendig, welche den Verstehensakt eigentlich konstituiert. Nach Garnham (1999) handelt es sich dabei um die Identifizierung einer Entität in einer konzeptuellen Struktur, also z.B. einem Personen-token in einem Mentalen Modell (vgl. hierzu auch die linguistische Debatte um die linguistische vs. pragmatische „Kontrolle“ von Anaphern, vgl. Cornish, 1989; und die Unterscheidung von sog. *surface-* und *deep-anaphors* von Hankamer & Sag, 1976; vgl. auch Frazier & Clifton, 2000).

Wie diese zweite Integrationsphase jedoch genau aussieht und welche Prozesse hier eine Rolle spielen, dazu schweigen sich Garrod und Sanford weitgehend aus (vgl. jedoch z.B. Garrod & Sanford, 1985; 1989, und die allgemeinen Annahmen des SMF-Ansatzes von Sanford & Garrod, 1981). Ähnlich gehen auch die meisten anderen Ansätze mit der zweiten Integrationsphase um, die meist nur recht oberflächlich erläutert wird (vgl. z.B. Albrecht & Myers, 1998). Um hier eine weiterführende Differenzierung und Spezifizierung zu erreichen, sind m.E. Überlegungen geeignet, welche von Long und Lea (2005) in Bezug auf eine Neuformulierung des „search after meaning“-Prinzips der konstruktivistischen Inferenztheorie angestellt wurden. Sie haben darauf hingewiesen, dass dieses Prinzip – verstanden als eine aktive, evtl. durch Produktionsregeln vermittelte Suche *nach* Informationen im Langzeitspeicher (*search for*) – nicht brauchbar zur Modellierung der vom Konstruktivismus angenommenen Verstehensprozesse ist. Denn eine solche Abrufprozedur wäre in der Tat mit einigen Befunden des gedächtnis-basierten Ansatzes nur schwer zu vereinbaren (z.B., dass auch dann Informationen abgerufen werden, wenn diese irrelevant oder sogar hinderlich für das Verstehen sind). Stattdessen präferieren sie eine Interpretation der zentralen Erklärungsmaxime des Konstruktivismus, die nicht auf eine aktive Suche, sondern eher als eine schnelle Evaluation

von bereits aktivem Hintergrundwissen hinausläuft (*search of*). Das heißt, höhere Verstehensprozesse beeinflussen nicht direkt die initiale Reaktivierung von Wissensstrukturen auf der Mikroebene, sondern werden im Gegenteil auf das *Produkt* der basalen Aktivierung angewendet. „Search after meaning“ ist also ein sekundärer Prozess auf zweiter Ebene, der die automatisch aktivierten Wissensstrukturen evaluiert. Wenn die derzeit aktiven Wissensstrukturen Informationen enthalten, welche gut in das derzeitige Diskursmodell passen (z.B. weil entsprechende Szenario-Rollen noch frei sind oder sie eine mögliche Ergänzung der derzeitigen Kausalstruktur darstellen), dann stoppt der Evaluationsprozess und die Inferenz wird in das finale Verstehensprodukt integriert. Wenn die Evaluation hingegen keine hinreichende Passung zum derzeitigen Diskursmodell ergibt, dann werden in der Tat allgemeine Problemlöse-Routinen benötigt, welche als repair-Prozess den Konflikt ressourcen-intensiv auflösen müssen und ggf. eine kohärente Interpretation gewährleisten (z.B. mittels der bereits im letzten Abschnitt dargestellten Mechanismen der algorithmischen Verarbeitung). Allerdings ist es durchaus möglich, dass das Pronomen evtl. gar nicht aufgelöst wird, z.B. wenn die Leser/innen nicht über die notwendigen Voraussetzungen für eine solche anstrengende Reanalyse verfügen (vgl. z.B. Greene et al., 1992; Levine, Guzmán & Klin, 2000; Poesio et al., 2006).

Nach dieser Vorstellung ist die zweite Phase der pronominalen Integration also als ein Checking-Prozess aufzufassen, der den Fit der bereits vorliegenden Struktur evaluiert, und zwar sowohl in Bezug auf seine Kompatibilität mit dem aktuell verfügbaren Hintergrundwissen als auch zu früheren Textepisoden. Darüber hinaus ist anzunehmen, dass diese Evaluation mehr oder weniger gründlich ausfallen kann, und zwar in Abhängigkeit von verschiedenen Text- und Lesermerkmalen (wie textueller Strukturierung, der Verfügbarkeit des zur Evaluation notwendigen Hintergrundwissens, der allgemeinen Verarbeitungskapazität und der Leseintention). An dieser Stelle ergeben sich also explizit Anschlussmöglichkeiten an die verschiedenen Theoriebildungen der erklärungs-basierten Ansätze, die Konzepte wie Lesemotivation, Kohärenzbildung und die Suche nach übergreifenden Erklärungsstrukturen in den Vordergrund stellen. In Abhängigkeit vom angezielten Auflösungslevel der finalen Interpretation und der Leichtigkeit, mit der sich diese etablieren lässt, ist davon auszugehen, dass die Integrationsphase sowohl passive als auch aktive Komponenten enthält. Im Gegensatz zur primär passiven Aktivierungsphase dürfte jedoch der aktive Anteil deutlich größer sein. Dies gilt z.B. für ein syllogistisches Schlussverfahren, wie es vom kohärenz-getriebenen Ansatz von Hobbs (1978) angenommen wird, aber auch für die Suche nach einer impliziten Prämisse im Validierungsmodell von Singer et al. (1992). Allerdings ist durch die passiv-automatische Vorbereitung der frühen Aktivierungsphase und die primäre Gestaltung des Integrationsprozesses als

strukturevaluierende und nicht als strukturgenerierende Komponente insgesamt eine schnelle Verarbeitung gewährleistet (s.u.). Dies war ja einer der großen Schwachpunkte der konstruktivistischen Position, die zwar immer wieder auf die Wichtigkeit hierarchie-hoher Prozesse hingewiesen hatte, aber nur selten in der Lage war, präzise anzugeben, wie diese zu implementieren sind.

Es ist wichtig darauf hinzuweisen, dass die meisten derzeit vertretenen Sprachverarbeitungsmodelle nicht oder nur rudimentär mit der skizzierten Integrationsperspektive kompatibel sind. So ist z.B. das CI-Modell von Kintsch (1988, 1998) nur schlecht mit dem skizzierten Modell in Einklang zu bringen. Dies ist eigentlich erstaunlich, weil es von seiner Grundkonzeption – erst Konstruktion, dann Integration – eine parallele Ausgangsstruktur besitzt. Allerdings fehlt in diesem Modell eine eigenständige erklärungs-basierte Komponente, welche die Kraft besitzt, auch aktive top-down-Prozesse und den Einfluss höherer Wissensstrukturen zu modellieren. Das wird direkt klar, wenn man die Integrations-Komponente des CI-Modells genauer betrachtet – also die Stelle, wo in der vorliegenden Konzeption der eigentliche Verbindungspunkt zwischen den beiden Ansätzen zu verorten ist. Das, was dort unter „Integration“ verstanden wird, ist nämlich ganz und gar kein aktiver Evaluationsprozess (vgl. Long & Lea, 2005). Vielmehr werden in der I-Phase des CI-Modells lediglich die vorher schwach wissensgeleitet konstruierten Basis-Propositionen in einem Prozess der zyklischen Aktivierungsmaximierung zu einer gemeinsamen Gesamtrepräsentation verbunden. Damit ist der primäre Integrationsmechanismus auch hier nur der Grad der propositionalen Überlappung verschiedener Wissensstrukturen, woran auch der Vorschlag von Kintsch (2005) nichts ändert, dass im Laufe der komplexen Konstruktionsphase Wissensstrukturen (wie z.B. Skripte) aktiviert werden, die in der nachfolgenden Integrationsphase als Kontrollstrukturen dienen können. Eine solche wissensgeleitete Generierung von Strukturen wäre im vorliegenden Ansatz immer noch der Phase der Informationsaktivierung zuzurechnen, an die sich noch eine weitere Phase der Evaluation anschließen müsste.

Einen Schritt weiter in Richtung auf eine Integration von gedächtnis- und erklärungs-basierten Verstehensprozessen geht das *Landscape-Model* von van den Broek et al. (vgl. z.B. van den Broek et al., 2005). Denn dieses umfasst zwei verschiedene Prozesstypen, nämlich zum einen die sog. *cohort activation* als eher automatische Komponente und zum anderen das sog. *coherence-based retrieval* als eher strategische Komponente. Während der Mechanismus der *cohort activation* weitgehend identisch mit dem bereits ausführlich erläuterten Resonanz-Prozess ist, wäre für eine integrative Modellierung v.a. das *coherence-based retrieval* interessant, welches den konstruktivistischen Teilaspekt des Verstehensprozesses abbilden soll.

Wenn man sich jedoch die Implementierung des Modells (van den Broek et al., 2004) genauer betrachtet, dann stellt sich auch hier heraus, dass bereits bei der Eingabe des Textinputs bekannt sein muss, an welcher Stelle ein referenzieller Rückbezug vorliegt (das Modell erlaubt kausale, referenzielle und sog. ermöglichende Rückbezüge) Das Verstehensprodukt resultiert also gar nicht aus dem Modell, sondern wird von diesem bereits vorausgesetzt. Das einzige, was sich variieren lässt, ist die Abruftrate für die verschiedenen reaktivierten Textelemente, die somit mit einer evtl. differenziellen Gewichtung in die finale Aktivationsmatrix einfließen.

Viel versprechender scheint in dieser Hinsicht das sog. *interpretation-based*-(IPN-) Modell zu sein, welches von Budiu und Anderson (Budiu, 2001; Budiu & Anderson, 2004) im Rahmen der ACT-R-Architektur vorgelegt wurde. Obwohl auch dieses Modell primär auf der semantischen Nähe verschiedener Textelemente basiert (bestimmt durch die korrespondierenden LSA-Scores), ist für die vorliegende Problemstellung v.a. relevant, dass es durch seine ACT-R-Anbindung über eine Produktionskomponente verfügt, welche in der Lage ist, die erforderlichen Such- und Schlussfolgerungsprozesse (die denen von Hobbs ähneln) in Echtzeit durchzuführen. Dies wird z.B. deutlich, wenn man sich die grundsätzliche Struktur des Modells ansieht, die hier in Abb. 1.5 wiedergegeben ist: Es ist sowohl eine eigenständige Such- (*search*) als auch eine Evaluations-Komponente (*Match?*) mit abschließender Integration vorgesehen.

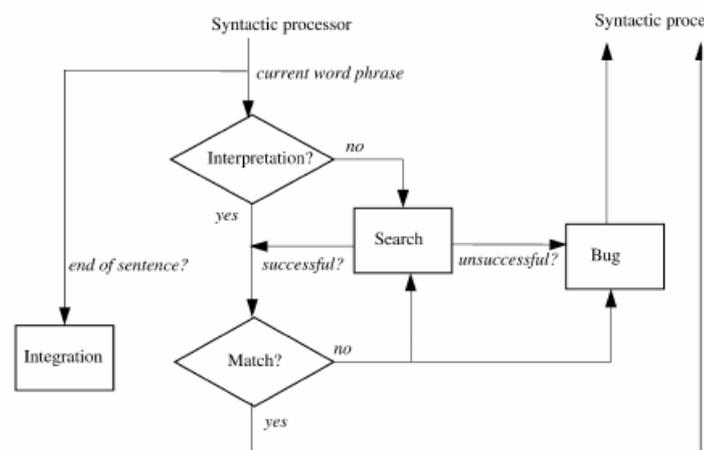


Abb. 1.5: Grundstruktur des semantischen Prozessors im IPN-Modells von Burdiu & Anderson (2004).

Allerdings verfügt das Modell bis jetzt über keine differenzierte syntaktische Komponente (so kann es z.B. keine Nebensätze verarbeiten) und operiert auf der Ebene ganzer Aussagen. Deswegen kann es in seiner vorliegenden Form auch nicht zur Modellierung des pronominalen Auflösungsprozesses herangezogen werden, allerdings liegen Simulationen für Phänomenbereiche vor, die sprachlich auf einer höheren Ebene angesiedelt sind (z.B. die Verarbeitung von Metaphern, vgl. Budiu & Anderson, 2002).

Bevor die den hier vorgelegten Experimenten zugrunde liegende Modellvorstellung der pronominalen Auflösung im nächsten Kapitel noch einmal konkretisiert und in Bezug auf ihre dynamische Prozessstruktur spezifiziert wird, ist es sinnvoll, den derzeitigen Stand der Argumentation noch einmal zusammenzufassen. Ausgehend vom *bonding-resolution*-Ansatz von Garrod und Sanford wird angenommen, dass die pronominale Auflösung notwendigerweise ein zweistufiger Prozess ist. In einer ersten *bonding*-Phase werden in einem schnellen automatischen *pattern-matching*-Prozess verschiedene mögliche Referenten für das Pronomen identifiziert und (evtl. differentiell gewichtete) *bonds* etabliert. Diese Phase ist durch die Interaktion verschiedener Faktoren gekennzeichnet, zu denen neben syntaktisch-strukturellen Restriktionen und aufmerksamkeits- und merkmals-basierten Faktoren explizit auch höhere Wissensstrukturen (wie z.B. Szenarien) gehören, die im Laufe eines passiven Resonanz-Prozesses mit aktiviert werden. In einer zweiten, nachgeordneten Phase werden die bereits etablierten *bonds* in eine kohärente Diskursrepräsentation integriert, wobei diese Integration in Anlehnung an Long und Lea (2005) als ein schneller Evaluationsprozess aufgefasst wird, bei dem die Passung des *bonds* in Bezug zum bereits etablierten Diskursmodell und dem allgemeinen Weltwissen wissenschaftlich gestützt und aktiv überprüft wird. Eine solche Konzeption ist anschlussfähig in mindestens drei Richtungen:

1. Zunächst werden linguistisch-syntaktische Faktoren nicht – wie dies bei den gedächtnis-basierten Modellbildungen der Fall ist – ausgeblendet, sondern explizit als eine mögliche Informationsquelle der frühen *bonding*-Phase mit einbezogen. Bezüglich der relativen Gewichtung und zeitlichen Wirksamkeit syntaktischer constraints ist das Modell offen; prinzipiell ließe sich ein frühes Filtermodell (welches strukturell spezifiziert, welche Referenten überhaupt als *bonds* zur Verfügung stehen) als eine Subphase des *bondings* auffassen.

2. Zu den gedächtnis-basierten Ansätzen weist das hier angenommene Modell ebenfalls eine große Überlappung auf, weil es allgemeinen Gedächtnisprozessen, speziell dem Resonanz-Mechanismus, eine wichtige Rolle zuweist. Sie sind v.a. im Rahmen der frühen *bonding*-Phase operativ und bestimmen damit wesentlich mit, welche Referenten in welcher Gewichtung überhaupt zur Auflösung zur Verfügung stehen. Je klarer und eindeutiger die Lösung am Ende der *bonding*-Phase ist, desto weniger ist der Verstehensprozess auf eine ressourcen-intensive Auflösung in der nachfolgenden Integrationsphase angewiesen.

3. In Bezug auf die erklärungs-basierten Ansätze ist die zugrunde gelegte Konzeption v.a. in zwei Richtungen anschlussfähig: Zunächst stellt sie mit der eigenständigen Integrationsphase eine Realisierung der zentralen Prinzipien des erklärungs-basierten Ansatzes (globale Kohärenzbildung, Erklärungsmotiv) auf der aktualgenetischen Ebene der Satzverarbeitung

bereit. Die Integration selbst wird dabei primär als eine Evaluationsprozedur aufgefasst, welche beim Produkt der gedächtnis-basierten Verarbeitung ansetzt. Sie kann sehr schnell und effizient ausgeführt werden, wenn die Integration durch entsprechend eindeutige Voraktivierung gut vorbereitet ist. Ausmaß und Gründlichkeit der Integration sind jedoch – und das ist der zweite Anknüpfungspunkt zum erklärungs-basierten Ansatz – abhängig von allgemeinen Merkmalen des Text-Leser-Systems, wie Motivation, Rezeptionszielen und Kohärenzstandards.

1.5 Plan der Analyse

Es ist klar, dass die bisher skizzierte Modellvorstellung, welche gedächtnis- und erklärungs-basierte Verarbeitungsprozesse eher grob miteinander in Beziehung setzt, noch keine präzisen Vorhersagen darüber erlaubt, wie und wann ein spezifischer Prozesse auf die pronominale Auflösung einwirkt. Allerdings ist es auch nur als eine Art theoretischer Hintergrund aufzufassen, der noch konkretisiert werden muss. Es ergeben sich aber bereits jetzt einige generelle Konsequenzen für den empirischen Teil dieser Arbeit, weil die skizzierte Konzeption folgende Annahmen zur Interaktion gedächtnis- und erklärungs-basierter Prozesse mit sich bringt:

1. Im Laufe der pronominalen Auflösung sollten sich sowohl passiv-automatische low-level- als auch aktiv-strategische high-level-Prozesse nachweisen lassen.
2. Die beiden Prozesstypen sollten selektiv und differenziell von unterschiedlichen sprachlichen Strukturmerkmalen beeinflusst werden: Strukturmerkmale, welche die Verfügbarkeit und Aktivierungsstärke eines Referenten erhöhen, sollten sich primär in den hierarchie-niedrigen Verarbeitungsprozessen niederschlagen. Strukturmerkmale, welche die Integration eines Referenten in ein Diskursmodell erleichtern, sollten sich hingegen primär in den hierarchie-hohen Verarbeitungsprozessen abbilden.
3. Die beiden Prozessarten sollten sich zeitlich dissoziieren lassen: Hierarchie-niedrige Gedächtnisprozesse sind v.a. in der *bonding*-Phase aktiv und sollten deswegen bereits früh im Laufe des Auflösungsprozesses nachweisbar sein. Hierarchie-hohe Integrationsprozesse sind hingegen v.a. in der *resolution*-Phase wirksam und sollten sich deswegen erst vergleichsweise spät nachweisen lassen.
4. Alle Wissensbestände, die für die Integration benötigt werden, sollten so früh wie möglich abgerufen werden. Das heißt, alle Auflösungs-Cues werden sofort genutzt, allerdings nur dann, wenn sie auch informativ sind.

5. Die beiden Prozessarten sollten dynamisch miteinander interagieren: Wenn gedächtnis-basierte Prozesse die Auflösung eines Pronomens adäquat vorbereitet haben, dann sollten weniger erklärungs-basierte Prozesse zur Auflösung benötigt werden, als wenn dies nicht der Fall ist.

Diese Aussagen befinden sich noch auf einem vergleichsweise hohen theoretischen Niveau und bedürfen der weiteren Spezifizierung, damit sie empirisch überprüfbar werden. Allerdings lassen sich aus ihnen bereits jetzt einige Implikationen dafür ableiten, welcher generelle Untersuchungsrahmen für ihre Überprüfung notwendig ist.

1. Es wird eine textuelle Situation benötigt, in der ein Pronomen aufgelöst werden muss. Dabei ist es wichtig, dass mehrere konkurrierende Antezedenten zur Verfügung stehen, damit es sich wirklich um eine *Auflösung* handelt und nicht einfach nur um eine Identifizierung des Antezedenten.
2. Es werden verschiedene Auflösungs-Cues benötigt, welche in der Lage sind, die verschiedenen Phasen des Auflösungsprozesses und die mit diesen assoziierten Verarbeitungsmechanismen differenziell zu beeinflussen.
3. Auf der Seite der abhängigen Variablen werden Maße benötigt, welche in der Lage sind, gedächtnis-basierte low-level-Prozesse und erklärungs-basierten high-level-Prozesse adäquat und trennscharf voneinander zu erfassen.
4. Es werden mehrere Erhebungszeitpunkte *innerhalb* eines Satzes benötigt, um zu überprüfen, ob die beiden Prozesstypen dynamische miteinander interagieren.

Eine Konkretisierung der ersten beiden Punkte, welche die Manipulation der Unabhängigen Variablen betreffen, wird im nachfolgenden Kapitel 2 erfolgen. Dort werden die notwendigen theoretischen und empirischen Grundlagen des gewählten Textsettings und der eingesetzten Auflösungs-Cues dargestellt. Auf die Punkte 3 und 4, welche die Seite der Abhängigen Variablen betreffen, wird hingegen in Kapitel 3 eingegangen. Dort wird eine reaktionszeit-basierte Methode entwickelt, mit der sich hierarchie-hohe und -niedrige Prozesse sowohl in Bezug auf die Lese- und Bearbeitungszeit ganzer Sätze als auch in Bezug auf wortweise Leseseiten erheben lassen.

2. Implizite Verbkausalität und Gender-Markierung als pronominale Auflösungs-Cues

Es wurde bereits darauf hingewiesen, dass zur Prüfung der in Kap. 1 dargestellten Auffassung, dass gedächtnis- und erklärungs-basierte Prozesse während der pronominalen Auflösung miteinander interagieren, eine sehr spezielle empirische Überprüfungssituation notwendig ist. Denn die zentrale Fragestellung, die sich aus einem solchen Ansatz ergibt, ist die, ob überhaupt beide Prozessarten am pronominalen Auflösungsprozess beteiligt sind und wenn ja, wann und an welcher Stelle sie in ihn eingreifen. Dies lässt sich jedoch empirisch jedoch nur dann feststellen, wenn man überhaupt konkrete sprachliche Faktoren identifizieren kann, die mit einiger Wahrscheinlichkeit in der Lage sind, die beiden Prozesstypen auch (differenziell und selektiv) hervorzurufen. Denn sollte sich ein Pronomen in einer speziellen sprachlichen Konstellation evtl. ausschließlich automatisch auflösen lassen, dann kann man lange nach erklärungs-basierten Prozessen suchen. Und umgekehrt wird vielleicht in einer Diskurs-Konstellation, bei der die pronominale Auflösung zwingend auf höhere Prozesse der Inferenzgenerierung zurückgreifen muss, die automatische Verarbeitung überlagert und kann empirisch nicht mehr erfasst werden.

Allerdings sagt sich dies leichter, als dass sich eine solche Konstellation konkret finden lässt. Denn eine fundamentale theoretische Konsequenz, die sich aus einem integrativen Ansatz der pronominalen Auflösung ergibt, ist ja gerade, dass sich gedächtnis- und erklärungs-basierte Auflösungsfaktoren nicht oder nur sehr schwer voneinander trennen lassen: Beide Prozessarten sind notwendige und essenzielle Bestandteile der pronominalen Auflösung, und das Verstehensprodukt ergibt sich gerade aus ihrer Interaktion. Entsprechend ist auf der empirischen Ebene die Befundlage, ob bei den verschiedenen, in Kap. 1 diskutierten Auflösungs-Cues (syntaktischer, semantischer und pragmatischer Art) eine automatische oder eine strategische Verarbeitung überwiegt, häufig uneinheitlich und unentschieden. Wie so oft in der Psychologie ist es so, dass ungefähr die eine Hälfte der Studien dafür spricht, dass ein bestimmter Auflösungs-Cue (allein) gedächtnis-basiert verarbeitet wird, während die andere Hälfte der Studien gerade zu der Erkenntnis gelangt, dass er (zwingend) der erklärungs-basierten Verarbeitung bedarf. Vor dem Hintergrund der in Kap. 1 skizzierten Auffassung der pronominalen Auflösung überrascht dies allerdings nicht: Wenn unterschiedliche Studien jeweils verschiedene Teilaspekte des pronominalen Auflösungsprozesses fokussieren und methodisch selektiv verschiedene Ebenen des Verstehensproduktes bzw. Phasen des Verstehensprozesses erfassen, dann ist ein solches uneinheitliches Ergebnismuster zu erwarten.

Aufgrund dieser theoretisch wie empirisch schwierigen Ausgangslage erweist es sich als besonderes Glück, dass es bereits eine in der sprachpsychologischen Forschung zur pronominalen Auflösung vergleichsweise gut etablierte Untersuchungskonstellation gibt, welche die geforderten Kriterien hinreichend erfüllt. Es handelt sich dabei um einen Forschungsstrang, der in den 1970er Jahren einsetzte und bis heute kontinuierlich weiterentwickelt wurde. In diesen Studien wird der pronominalen Auflösungsprozess in Abhängigkeit von zwei sehr unterschiedlichen Auflösungsfaktoren untersucht, nämlich dem morpho-syntaktischen Merkmal der Gender-Kongruenz und dem lexikalisch-pragmatischen Faktor der sog. Impliziten Verbkausalität (englisch: *implicit causality*, deswegen im Folgenden als IC abgekürzt). Im Laufe des Kapitels wird dafür argumentiert werden, dass die Gender-Kongruenz und die IC Faktoren sind, welche beide sowohl in den gedächtnis-basierten Teil des Auflösungsprozesses eingreifen als sich auch auf die kohärenz-stiftende erklärungs-basierte Verarbeitung auswirken. Wie noch zu sehen sein wird, lassen sich beide Behauptungen zwar gut aus der bereits vorliegenden Literatur begründen, sie sind aber auch nicht unumstritten.

Bevor jedoch im nachfolgenden Teil des Kapitels eine theoretische Explikation der beiden Auflösungs-Cues erfolgt und eine zusammenfassende Darstellung der empirischen Befundlage gegeben wird, ist es sinnvoll, die „initiale“ Studie dieses Paradigmas (Caramazza, Grober, Garvey & Yates, 1977) etwas näher darzustellen, weil sich alle nachfolgenden Studien mehr oder weniger explizit auf diese beziehen und sich als Modifikationen dieser verstehen lassen.

Ausgangspunkt der Forschung zum sog. IC-Effekte ist eine Beobachtung von Garvey und Caramazza (1974), die zunächst einmal auf einer rein linguistischen Ebene angesiedelt war und auf Überlegungen von Fillmore (1968) und Chafe (1970) basierte (vgl. hierzu auch die linguistische Analyse der IC in Abschnitt 2.1.2): Bestimmte Verben scheinen unterschiedliche kausale Erklärungs- und Attributionsmuster zu implizieren. Dies offenbart sich u.a., wenn man innerhalb eines verbalen Kausalgefüges ein Pronomen einführt und beobachtet, in welche Richtung es wahrscheinlich aufgelöst wird. Betrachten wir z.B. die beiden Satzfragmente in (1):

- (1) a. Peter amüsierte Hans, weil er ...
- b. Peter kritisierte Hans, weil er ...

Garvey und Caramazza beobachteten, dass man im Fall (1a) eher geneigt ist, das Pronomen im Nebensatz als koreferent mit dem Hauptsatzsubjekt zu interpretieren, während man es im Fall (1b) eher als eine Wiederaufnahme des Hauptsatzobjektes auffasst. Entsprechend nannten sie Verben wie in Satz (1a) NP1-Verben und Verben wie in Satz (1b) NP2-Verben.

Dieses konsistent unterschiedliche Auflösungsverhalten ist deshalb bemerkenswert, weil rein grammatikalisch das Pronomen *er* ambig ist und sehr viele semantische und syntaktische Faktoren an der Auflösung beteiligt sind. Da sich jedoch die beiden Satzfragmente nur im Verb unterscheiden, folgerten Garvey und Caramazza, dass es wohl ein verbales Merkmal sein muss, welches für das unterschiedliche Auflösungsverhalten verantwortlich ist. Sie nannten es „Implizite Kausalität“ und nahmen an, dass es sich dabei um so etwas wie lexikalisier-tes Weltwissen handelt: Es ist Teil unseres Weltwissens zu wissen, dass Menschen, die amüsiert sind, meist einen (externen) Grund haben, warum sie amüsiert sind. Und genauso wissen wir, dass man meist einen externen Grund braucht, um einen anderen Menschen kritisieren zu dürfen. Über die Frage, wie sich die IC genau begründen lässt, wurde jedoch später (v.a. in der sozialpsychologischen Forschung) noch heftig debattiert. Im Prinzip ist die Frage bis heute noch nicht endgültig geklärt (vgl. Abschnitt 2.1.1). Im Moment ist es jedoch zunächst einmal egal, *warum* manche Verben implizite Kausalität aufweisen; wichtig ist allein, *dass* sie es tun.

Die Datengrundlage, auf die sich Garvey und Caramazza beriefen, ist eine Studie von Garvey, Caramazza und Yates, (1975), welche mit einer Satzergänzungsaufgabe arbeitete. In einer ersten Studie gaben sie Versuchspersonen (Vpn) eine Reihe von Satzfragmenten wie in (1), die frei ergänzt werden konnten. Im Anschluss an die Untersuchung wurden die Antworten dahingehend kodiert, ob mit dem Pronomen das Hauptsatzsubjekt oder -objekt gemeint war (oder keines von beiden). Sie fanden, dass die 16 von ihnen untersuchten Verben in der Tat einen sehr starken Fortsetzungsdruck erzeugen, da 14 von ihnen einen starken Bias in die vorhergesagte Richtung aufwiesen. In einer zweiten Studie wurde nun untersucht, ob diese Fortsetzungstendenz von verschiedenen anderen linguistischen Faktoren (wie Aktiv- vs. Passivformulierung des Satzes, Negation, Valenz des Verbs, Status der beteiligten Personen etc.) beeinflusst wurde. Hier ergaben sich z.T. komplizierte Interaktionsmuster (so verstärkt z.B. die Passivierung den NP1-, schwächt jedoch den NP2-Bias etc.), die generelle Bias-Tendenz war jedoch in allen Konstellationen immer noch sehr stark ausgeprägt (vgl. auch Grober, Beardsley & Caramazza, 1978, und Caramazza & Gupta, 1979).

Wie bereits dargestellt, nehmen die Autoren als Konsequenz dieses Befundmusters an, dass die von ihnen untersuchten Verben über ein besonderes semantisches Merkmal verfügen. Wichtig ist dabei, dass sie (im Gegensatz noch zur Argumentation in Garvey und Caramazza, 1974) die IC wirklich als ein *Merkmal* im Sinne des linguistischen *feature*-Begriffes auffassen: Ebenso wie Phoneme über ein Merkmal [\pm sonorant] verfügen oder Pronomen über ein Gender-Merkmal wie [\pm weiblich], so verfügen IC-Verben über ein abstraktes Merkmal [\pm

N1], das bestimmt, welcher Referent eine kausale Rolle trägt (wobei die Autoren darauf hinweisen, dass dieses evtl. graduell konzipiert werden muss). Genau genommen halten Caramazza et al. (1977) die IC für ein sog. *transfer-feature*, welches im Verb lokalisiert ist, aber die Interpretation der Satzumgebung mit beeinflusst.

Dieses Denken in Merkmalskonstellationen war in der linguistischen Theoriebildung der 1960er und 1970er ausgesprochen dominant und hat die nachfolgende Diskussion entscheidend mitbestimmt. Der Rückführungsversuch der IC auf ein einziges abstraktes linguistisches Merkmal war jedoch in meinen Augen ausgesprochen unglücklich, weil er in der Folgediskussion zu großer Verwirrung und sehr unterschiedlichen Konzeptualisierungen des Phänomens geführt hat, die u.a. auch die Art der empirischen Überprüfung entscheidend beeinflussten (was wiederum einen Vergleich der Studien erschwerte). Auch wenn eine detaillierte Diskussion noch folgt (vgl. Abschnitt 2.1.3), liegt m.E. das Grundproblem darin, dass die IC nicht einfach nur den kausalen Fokus auf einen der beiden Aktanten legt. Vielmehr werden mit IC-Verben komplexe Weltwissensbestände (z.B. Szenarien oder Skripte) aktiviert, die eine bestimmte kausale Rollenverteilung implizieren. Für eine solche Auffassung spricht, dass Garvey et al. signifikante Interaktionen des IC-Bias mit dem Status der beteiligten Aktanten fanden: Wenn Sätze mit einer „verkehrten“ Rollenverteilung dargeboten wurden (*Der Richter gestand dem Angeklagten...*), dann drehte sich die Bias-Richtung der untersuchten Verben entweder um oder wurde eliminiert. Dies deutet darauf hin, dass das Verb allein wohl kaum für die Auflösungstendenz verantwortlich gemacht werden kann, sondern dass es eher um die Kohärenz bzw. Plausibilität der mit dem IC-Verb implizierten Gesamtsituation geht.

Diese (Fehl-)Interpretation des IC-Phänomens ist umso erstaunlicher, wenn man bedenkt, dass die zentrale Studie von Caramazza et al. (1977) gerade auf der verstehensförderlichen Wirkung einer kohärenten Gesamtinterpretation aufbaut. Schon Garvey und Caramazza (1974) hatten spekuliert, dass anaphorische Sätze, die eine zur IC kongruente Fortsetzung etablieren, wesentlich einfacher und deswegen schneller zu verstehen sind, als wenn explizite Begründung und implizite Verbkausalität divergieren. Genau dies wurde in der Studie von Caramazza et al. (1977) überprüft. Sie kontrastierten Sätze wie in (2) miteinander:

- (2) a. John telephoned Bill because he wanted some information (ic-kongruent).
- b. John telephoned Bill because he withheld some information (ic-inkongruent).

Die Überlegung ist dabei folgende: Durch den Kausalkonnektor werden die Vpn dazu gezwungen, eine kausale Relation zwischen den beiden Aktanten zu etablieren. Wenn die Vpn im Laufe des Auflösungsprozesses in der Tat IC-Informationen (die einen der beiden Aktan-

ten kausal präferieren) genutzt haben sollten, dann muss in inkongruenten Sätzen eine Neu-
zuweisung des Pronomens stattfinden, die zusätzliche Zeit beansprucht.

Wie so häufig war es eine Nebensächlichkei, welche die nachfolgende Forschung ent-
scheidend bestimmte. Caramazza et al. führten nämlich noch eine Kontrollbedingung ein, in
welcher das Pronomen gender-disambiguiert war (vgl. (3)):

- (3) a. John telephoned Sue because he wanted some information (ic-kongruent).
- b. Sue telephoned Bill because he withheld some information (ic-inkongruent).

Der Grund hierfür war, dass die Gender-Kongruenz des Pronomens und einem der beiden
Aktanten in diesem Fall vollkommen ausreicht, um das Pronomen erfolgreich aufzulösen.
Wenn sich der IC-Kongruenz-Effekt auch hier zeigen sollte, dann wäre dies sogar noch stär-
kere Evidenz dafür, dass sich die IC auf den Auflösungsprozess auswirkt.

Als Aufgabe verwendeten Caramazza et al. eine pronominale Auflösungsaufgabe: Die
Vpn bekamen die Sätze mit einem Tachistoskop dargeboten und sollten so schnell wie mög-
lich angeben, welche der beiden im Hauptsatz genannten Personen mit dem Pronomen im
Nebensatz gemeint ist. Das Ergebnismuster, das Caramazza et al. erhielten, ist ausgesprochen
eindeutig und hat sich auch in (fast) allen nachfolgenden Studien zur IC als stabil erwiesen:
Wenn explizite und implizite Begründung übereinstimmen, dann lässt sich das Pronomen viel
schneller auflösen. Wichtig ist weiterhin, dass die beiden Faktoren (IC-Kongruenz und Gen-
der-Informationen) in Bezug auf die Antwortgeschwindigkeit nicht miteinander interagierten:
Sowohl die Gender-Disambiguierung als auch die IC-Kongruenz führten zu dramatisch ver-
kürzten Antwortzeiten und der IC-Effekt war in der gender-disambiguierten Bedingung ähn-
lich stark ausgeprägt wie in der ambigen Bedingung. Die Autoren interpretieren die Ergebnis-
se nun dahingehend, dass IC-Informationen im Laufe des Auflösungsprozesses genutzt wer-
den. Speziell gehen sie davon aus, dass das Pronomen bereits bei seiner Einführung in Rich-
tung des IC-Bias aufgelöst wird und dann am Ende des Satzes ggf. eine Neuuzuweisung statt-
findet.

Der Witz ist jedoch, dass – entgegen der Argumentation von Caramazza et al. – der IC-
Effekt gar nichts mit der Auflösung des Pronomens zu tun haben muss. Er ist genauso gut mit
der Aussage kompatibel, dass es halt einfach nur schwerer ist, schema-inkongruente Informa-
tion in eine kohärente Gesamtinterpretation zu integrieren; wobei in diesem Fall das Schema
durch das IC-Verb instantiiert wird und die Inkongruenz darin besteht, dass bei der expliziten
Auflösung die „falsche“ Rollenzuweisung gewählt wird. Das heißt, evtl. ergibt sich die ver-
längerte Antwortzeit nicht aus einer ressourcen-intensive Neuuzuweisung am Satzende, son-
dern nur durch eine erschwerte Integration der inkongruenten Informationen im Zuge der Ko-

härenzbildung, die für die Interpretation des Pronomens notwendig ist. Eine solche Sichtweise würde im Wesentlichen dem kohärenz-basierten Ansatz von Hobbs (1978) entsprechen, der davon ausgeht, dass die pronominale Auflösung gewissermaßen ein Nebenprodukt des allgemeinen Kohärenzbildungsprozesses ist (vgl. auch Kehler, 2002).

Exakt diese Kontroverse um die genaue Natur des (empirisch unbestreitbar vorhandenen) IC-Effekts – frühe Fokus-Wirkung durch abstraktes Merkmal + Reanalyse vs. späte Integrationsschwierigkeit durch schema-inkompatible Informationen – hat nun die gesamte nachfolgende Diskussion geprägt. Wie noch gezeigt werden soll, gibt es Evidenz für beide Sichtweisen. Im vorliegenden Kontext ist jedoch allein entscheidend, dass beide Ansätze ein erhebliches Ausmaß an erklärungs-basierter Verarbeitung implizieren. Denn dass die finale Satzinterpretation v.a. durch die Plausibilität und Kohärenz der Satzinhalte zum Verbschema bestimmt ist, bestreitet ja niemand. Der Punkt, an dem die beiden Ansätze divergieren, ist allein, ob die pronominale Auflösung bereits früh stattfindet und dann später ein erklärungs-basierter Reanalyse-Prozess notwendig ist oder ob die Referenz des Pronomens überhaupt erst im Zuge der finalen Gesamtinterpretation festgelegt wird.

Vor diesem Hintergrund ist es für die weitere Diskussion wichtig, darauf hinzuweisen, dass der IC-Effekt im Hinblick auf seine Natur seit der Studie von Caramazza et al. fast immer und ausschließlich mit dem Gender-Cue verglichen wird, von dem zumindest anfänglich angenommen wurde, dass er als morpho-syntaktisches Merkmal bereits früh in den Prozess der pronominalen Auflösung eingreift. Während der Gender-Cue bei Caramazza et al. also noch primär eine Kontrollfunktion hatte, hat sich dies im Laufe der theoretischen Diskussion entscheidend geändert und die mit dem Kontrast der beiden Effekte verbundene Frage ist vielmehr, ob der IC-Effekt sich genauso verhält wie der Gender-Effekt oder vielleicht anders erklärt werden muss. Das ist eine ausgesprochen günstige Ausgangslage für die vorliegende Untersuchung! Denn wenn die Kontrastierung des (vermeintlich) hierarchie-hohen IC- mit dem (vermeintlich) hierarchie-niedrigen Gender-Effekt empirisch tragfähig ist, dann können die beiden Cues in der Tat dazu genutzt werden, selektiv gedächtnis- und erklärungs-basierte Verarbeitungsprozesse während der pronominalen Auflösung zu evozieren. Die nachfolgenden Abschnitte dienen dazu, genauer herauszuarbeiten, ob diese Annahme berechtigt ist. Dabei wird sich herausstellen, dass das Bild – wie immer – komplizierter ist, als man zuerst gedacht hat.

2.1 Implizite Verbkausalität

Nachdem das Phänomen der Impliziten Verbkausalität einmal in den wissenschaftlichen Diskurs eingeführt worden war, führte es bald ein sehr illustres Eigenleben, da es von sehr vielen wissenschaftlichen (Sub-)Disziplinen gleichzeitig aufgegriffen wurde. Der Grund hierfür ist, dass Verben, die IC aufweisen, fast ausschließlich *interpersonale Verben* sind (vgl. Fillmore, 1968, 1971), d.h. Verben, die sich auf Ereignisse und Zustände im zwischenmenschlichen Bereich beziehen. Damit sind sie jedoch nicht nur für die sprachpsychologische Forschung interessant, sondern auch für die sozialpsychologische Theoriebildung, in der Überlegungen zur (sozialen) Erklärung spätestens seit den 1950er Jahren eine sehr dominante Stellung einnehmen. Gleichzeitig ist diese Eigenschaft der IC-Verben auch für die Linguistik relevant, weil sie ein semantisches Kriterium liefert, mit dem sich die Besonderheiten und syntaktischen Regularitäten dieser Verbklasse in ein übergreifendes System einordnen lassen. Und schließlich wurde in der sprachpsychologischen Forschung die ursprüngliche Fragestellung von Caramazza et al. weitergeführt, inwieweit die IC eines Verbs als lexikalisch-pragmatisches Merkmal auf den pronominalen Auflösungsprozess einwirkt.

Obwohl alle drei Fragestellungen klar miteinander zusammenhängen und m.E. auch aufeinander bezogen werden sollten, wurde die theoretische Diskussion in den drei genannten Disziplinen nach einer kurzen, anfänglichen Überlappungsphase seit den 1980er Jahren fast vollständig unabhängig voneinander geführt. Dies liegt v.a. daran, dass das objekttheoretische Phänomen der IC in allen drei Disziplinen in jeweils unterschiedliche metatheoretische Diskurse eingebunden wurde, nämlich die Attributionstheorie in der Sozialpsychologie, die Theorie der Argumentstruktur und ihren Einfluss auf die Syntax in der Linguistik sowie die seriell-parallel-Debatte in der Psycholinguistik. Im Gegensatz zum geteilten Interesse am Phänomen an sich, bieten solche fachspezifischen Debatten (zunächst einmal) wesentlich weniger Ansatzpunkte für einen interdisziplinären Austausch, und erst in jüngerer Zeit zeigen sich wieder erste Ansätze der interdisziplinären Verknüpfung der verschiedenen Diskussionsstränge. Die Darstellung soll deswegen auch getrennt für die drei Bereiche erfolgen, wobei durch entsprechende Querverbindungen der jeweilige Bezug und die Relevanz der behandelten Probleme für die anderen Disziplinen aufgezeigt werden soll.

2.1.1 Sozialpsychologie

Wie bereits festgestellt, war der dominante Bezugspunkt der sozialpsychologischen Forschung zur IC zunächst einmal die Tatsache, dass sich in ihr sehr allgemeine menschliche Prinzipien der sozialen Organisation sprachlich zu manifestieren scheinen. Spätestens seit

Heiders (1958) grundlegendem Werk zur zwischenmenschlichen Beziehung spielt der Gedanke, dass Menschen *Erklärungen* für die sie umgebenden Phänomene suchen und ihr Handeln nach diesen ausrichten, eine zentrale Rolle für die sozialpsychologische Theoriebildung. In diesem Zusammenhang sind auch Heiders Untersuchungen zur sog. *phänomenalen Kausalität* zu sehen (vgl. z.B. Heider & Simmel, 1944): Auch kausal nicht zusammenhängende Ereignisse (wie Trickanimationen von Dreiecken und Kreisen, die sich scheinbar „verfolgen“ oder aber „fliehen“) werden von uns kausal interpretiert und unsere Einstellung diesen Objekten gegenüber wird durch die gefundene Erklärung beeinflusst (so werden z.B. Objekte, die andere verfolgen, als „böse“ und „aggressiv“ beschrieben; und wenn man könnte, würde man das „böse Dreieck“ evtl. sogar an einer Verfolgung hindern etc.). Anscheinend ist es eine *conditio humana* und fundamental für fast alle menschlichen Lebensbereiche, Dinge als kausal aufeinander bezogen zu erleben. Wir können einfach nicht anders als uns unsere Umwelt als kausal bedingt vorzustellen (man beachte die Parallele zu den Grundannahmen der konstruktivistischen Inferenztheorie). Obwohl solche Überlegungen in der Philosophie eine lange Vergangenheit haben (z.B. Kant, 1787/1998; Blumenberg, 1979) wurde diese Vorstellung im sog. *epistemologischen Subjektmodell* (Kelly, 1955) vielleicht am explizitesten formuliert. Das epistemologische Subjektmodell geht davon aus, dass „man the scientist“ strukturparallel zum Wissenschaftler (subjektive) Theorien über sich und das Funktionieren seiner Umwelt aufstellt und handlungsleitend verwendet (vgl. auch Groeben & Scheele, 1977; Groeben, 1986; im Überblick: Erb, 1997). Eine stärkere Formalisierung hat dieser Grundgedanke in der Attributionstheorie von Kelley (1967) erhalten, welche versucht, die hinter den kausalen Schlüssen stehenden Prinzipien und Dimensionen systematisch zu explizieren (wofür sie insbesondere die Kovariation von Konsens-, Distinktheits- und Konsistenz-Informationen ansetzt, vgl. im Überblick Heckhausen, 1989).

Der Punkt, an dem die IC nun für die Allgemeine Psychologie und Sozialpsychologie relevant wird, ist der, dass sie eine Möglichkeit darstellt, allein durch eine Beobachtung der „Weisheit der Alltagssprache“ den Menschen gewissermaßen „in den Kopf“ zu schauen und zu beobachten, welche Erklärungsprinzipien für sie relevant sind. So meinen z.B. Brown und Fish (1983a) festgestellt zu haben, dass sich Kelleys Kovariationsprinzipien in der Morphologie des Englischen widerspiegeln: So ist es im Englischen prinzipiell möglich, aus einem Verb derivationell ein Adjektiv zu machen, indem man ein Suffix wie *-able* oder *-ful* anhängt (vgl. im Überblick: Bhatt, 1991). Allerdings gibt es einen bedeutenden Unterschied bezüglich der beiden Suffixe: Eine Ableitung mit dem Suffix *-able* bezieht sich meist auf das Objekt des Verbs, eine Ableitung mit dem Suffix *-ful* hingegen auf das Subjekt. Allerdings scheint nicht

jede Ableitung mit jedem Verb möglich: Ein Verb wie *help* hat z.B. lediglich eine „dispositionale“ Ableitung auf *-ful*, die sich auf die Person bezieht, die hilft. Eine Ableitung, die sich auf das Objekt des Helfen-Aktes bezieht, existiert hingegen nicht (**helpable*). Im Gegenzug lässt das Verb *like* nur eine objekt-bezogene Ableitung zu (*likeable*), nicht jedoch eine subjekt-bezogene (**likeful*). Im Deutschen gibt es keine direkte Entsprechung für den *-ful* vs. *-able*-Kontrast des Englischen. Ansatzweise lässt er sich jedoch für die Adjektiv-Suffixe *-voll* vs. *-bar* beobachten: Adjektiv-Bildung mit *-voll* beziehen sich auf das Subjekt des Verbs, Ableitungen auf *-bar* hingegen auf das Objekt (z.B. *x reizt y*; *reizvoll* → Eigenschaft von *x*, *reizbar* → Eigenschaft von *y*). Und auch im Deutschen ist es so, dass nicht alle Verben mit allen Adjektiv-Suffixen kombinierbar sind. So gibt es z.B. für das Verb *ärgern* nur die Ableitung *ärgerlich* (das sich auf das Subjekt bezieht), nicht aber z.B. das objekt-bezogene **ärgerbar*.

Brown und Fish nehmen nun an, dass das unterschiedliche Verhalten der einzelnen Verben bei der Adjektiv-Bildung darauf zurückgeführt werden kann, dass sie ihrem Subjekt und Objekt jeweils ein unterschiedliches kausales Gewicht zuweisen, und in der Tat ist *helfen* ein NP1-Verb und *mögen* ein NP2-Verb im Sinne von Caramazza et al. Sie gehen jedoch – und das ist entscheidend – noch einen Schritt weiter. Zunächst einmal rekonstruieren sie (in Anlehnung an entsprechende Theoriebildungen aus der Linguistik wie Fillmore, 1968, und Chafe, 1970, vgl. Abschnitt 2.1.2) die den Verben zugrunde liegenden semantischen bzw. thematischen Rollen. Das Konzept der thematischen (θ -)Rolle stammt aus der Argument- und Kasus-Grammatik (im Rahmen der Generativen Syntax werden sie im Rahmen der sog. Theta-Theorie behandelt) und bezieht sich im Wesentlichen darauf, wie viele Argumente ein Verb selektiert und welcher Art diese sind (vgl. z.B. Jackendoff, 1972). So gibt es 0-stellige (Gewitter-)Verben wie *regnen*, 1-stellige (intransitive) Verben wie *schnarchen*, 2-stellige (transitive) Verben wie *schlagen* und 3-stellige (bitransitive) Verben wie *schenken*, die jeweils unterschiedlich viele Aktanten benötigen. Man kann z.B. nicht sagen **Peter schenkte Maria.*, weil hier ein obligatorisches Argument fehlt (man muss *etwas* schenken), aber man kann auch nicht sagen **Peter schnarchte Maria.*, weil hier ein Argument zuviel ist. Mitunter wird die Behauptung vertreten, dass für die Syntax lediglich die *Anzahl* der von einem Verb verwendeten Argumente entscheidend ist (vgl. z.B. Grewendorf et al., 1989). Wie noch gezeigt wird, stimmt das nicht ganz: Mitunter hat auch die *Art* des Arguments einen Einfluss auf die syntaktische Strukturierung eines Satzes. *Welche* θ -Rollen ein Verb verteilt, hängt von seiner Semantik ab (vgl. Levin, 1993; Grimshaw, 1990): *Handlungsverben* verteilen z.B. (im 2-stelligen Fall) die beiden Rollen AGENS und PATIENS, sog. *Zustandsverben* die beiden Rollen STIMULUS und EXPERIENCER. Handlungsverben sind (prototypisch) solche Verben, die einen

(willentlichen Handlungs-)Akt ausdrücken (z.B. *schlagen*). Dabei ist der AGENS der Urheber der Handlung, während der PATIENS vom Handlungsakt affektiert wird. Zustandsverben sind hingegen (prototypisch) nicht-willentliche affektive, motivationale oder kognitive Befindlichkeiten, die vom EXPERIENCER in Reaktion auf den STIMULUS erlebt werden (z.B. *fürchten*).

Es gibt traditionell einige linguistische Testverfahren, um (tendenziell) zu überprüfen, ob ein Verb ein Handlungs- oder ein Zustandsverb ist. Dazu gehören u.a. folgende (vgl. im Überblick Rudolph & Försterling, 1997a): Handlungsverben lassen den Imperativ zu (*Schlag (zu)!*), bei Zustandsverben ist dieser häufig fragwürdig (*?Ekel dich!*), weil man von Außen evozierte Befindlichkeiten nicht willentlich beeinflussen kann; Handlungsverben lassen sich (in den Sprachen, die dies grammatikalisch erlauben, also nicht im Hochdeutschen) aspektuell in der Verlaufsform verwenden, Zustandsverben tendenziell nicht (Approximierung durch die Rheinische Verlaufsform: *Ich bin am fahren* vs. **Ich bin am wissen*); die mit Handlungsverben beschriebenen Ereignisse sind beobachtbar, die von Zustandsverben hingegen nicht; Handlungsverben lassen sich mit *tun* oder *machen* paraphrasieren, Zustandsverben nicht etc.

Brown und Fish beobachten nun, dass zur Klasse der für sie relevanten (IC-)Verben sowohl Handlungs- als auch Zustandverben gehören. Weiterhin nehmen sie an, dass die Handlungsverben unter den IC-Verben ihr kausales Gewicht auf der AGENS-Rolle haben, die syntaktisch immer als grammatisches Subjekt realisiert ist. Alle Handlungsverben sind also ihrer Auffassung nach NP1-Verben im Sinne von Caramazza et al. Bei Zustandsverben können syntaktisches und semantisches Subjekt hingegen divergieren: Bei den sog. Stimulus-Experiencer-(SE-)Verben (wie *beschämen*) fallen beide zusammen, bei den sog. Experiencer-Stimulus-(ES-)Verben (wie *bedauern*) jedoch nicht. (Allerdings gibt es linguistisch gesehen Grund zur Annahme, dass es sich genau andersrum verhält, vgl. Abschnitt 2.1.2). Das kausale Gewicht liegt hier immer auf der STIMULUS-Rolle, egal in welcher syntaktischen Position es realisiert wird. Darüber hinaus behaupten Brown und Fish, dass sich die Adjektiv-Morphologie (des Englischen) nun nach dieser Unterscheidung ausrichtet: SE-Verben lassen nur *-ful*-Ableitungen zu, ES-Verben hingegen nur *-able*-Ableitungen. Handlungs- und Zustandsverben liegen also zwei unterschiedliche, aber auf einer übergeordneten Ebene einheitliche Muster zugrunde, die sowohl die kausale Gewichtung als auch die Adjektiv-Bildung erklären (vgl. Tab. 2.1).

	<i>Handlungsverben</i>	<i>Zustandsverben</i>
Kausales Gewicht	cause(AGENS)	cause(STIMULUS)
Adjektivbildung	nach AGENS	nach STIMULUS

Tab. 2.1: Zuordnung der kausalen Gewichtung und den Prinzipien der Adjektivbildung für Handlungs- und Zustandsverben nach Brown & Fish (1983a).

Die nachfolgende Diskussion hat allerdings gezeigt, dass sich beide Aussagen empirisch nicht halten lassen. Zum einen hat bereits Au (1986) darauf hingewiesen, dass es auch Handlungsverben gibt, die ihr kausales Gewicht auf das Objekt legen. Es gibt also Handlungs-NP2-Verben wie *gratulieren*, die es nach Brown und Fish nicht geben dürfte. Au nennt diese Verbklasse *Action-Patient-Verben* und macht damit klar, dass es sich zwar um Handlungsverben handelt, die jedoch den PATIENS als ursächlich ansetzen. Der tiefere Grund hierfür ist, dass nicht alle Handlungen als (komplett) willentlich verursacht angesetzt werden müssen bzw. können (vgl. auch Groeben, 1986). Vielmehr gibt es auch so etwas wie *reaktive* Handlung, d.h. Handlungen, die man ausführt, weil (weil!) vorher etwas passiert ist, was die Handlung hervorruft (vgl. auch Heider, 1958). *Beschimpfen* ist hierfür vielleicht ein gutes Beispiel: Normalerweise geht man davon aus, dass man jemanden nicht grundlos, spontan und willkürlich beschimpft, sondern (nur) dann, wenn dieser etwas getan hat, was die Beschimpfung rechtfertigt. Auch wenn der Akt des Beschimpfens durchaus volitional ist (ich hätte ja auch etwas anders tun können), ist er in einer gewissen Form von vorherigen Ereignissen abhängig. Genau dem wird nun dadurch Rechnung getragen, dass die kausale Rolle nicht dem (reaktiven) Handlungssubjekt zugesprochen wird, sondern dem bereits zuvor handelnden Handlungsobjekt. Rudolph und Försterling (1997a) haben diesen Verben deshalb recht passend als sog. *Agent-Evocator-(AE-)*Verben bezeichnet und grenzen sie von den „normalen“ Handlungsverben ab, die sie *Agent-Patient-(AP-)*Verben nennen. Diese Unterteilung führt also zu einer revidierten Klassifikation, in der es zwei unabhängige Dimensionen gibt, nämlich Handlungs- vs. Zustandsverben einerseits und NP1- vs. NP2-Verben andererseits (vgl. Tab. 2.2). Rudolph und Försterling nennen diese Einteilung die sog. *Revised Action-State Distinction (ASD)* und zeigen darüber hinaus, dass diese in allen relevanten Punkten mit dem in der Sozialpsychologie gut bekannten *Linguistic Category Model* von Semin und Fiedler (1988) identisch ist.

	<i>Handlungsverben</i>	<i>Zustandsverben</i>
NP1	Agent-Patient	Stimulus-Experiencer
NP2	Agent-Evocator	Experiencer-Stimulus

Tab. 2.2: *Revised Action-State-Distinction* (ASD) nach Rudolph & Försterling (1997a).

Aber auch die zweite zentrale Behauptung von Brown und Fish (1983a), nämlich dass die morphologischen Prinzipien der Adjektiv-Derivation von der kausalen Struktur des Verbs abhängig sind, hat sich (so) als nicht haltbar erwiesen. In einer (nach linguistischen Maßstäben recht informell durchgeführten) Korpusanalyse fanden Hoffman und Tchir (1990), dass 90% der Aktionsverben (des Englischen) eine Adjektivableitung in Richtung des Subjekts (also zum AGENS) zulassen (z.B. *combative*), jedoch auch 25% eine zum PATIENS (z.B. *commendable*; man beachte, dass beide Kategorien nicht exklusiv sind). Bei den Zustandsverben gibt es 75% Adjektivbildungen zur STIMULUS-Rolle (z.B. *thrilling*), aber auch 50% zum EXPERIENCER (z.B. *obsessive*). Obwohl es also durchaus eine Tendenz in Richtung der von Brown und Fish vertretenen Hypothese gibt, fällt diese bei Weitem nicht so eindeutig aus, wie von diesen behauptet wird (vgl. auch Brown & Fish, 1983b). Darüber hinaus haben Rudolph und Försterling (1997a) darauf hingewiesen, dass alle Studien, die den Verbtyp (im Sinne des ASD-Systems) und die Adjektivbildung (Subjekt vs. Objekt) unabhängig voneinander variiert haben (vgl. z.B. Semin & Marsman, 1994; Greene & McKoon, 1995), keinen nennenswerten unabhängigen Einfluss des Derivationsprinzips finden konnten (mit der Ausnahme von Hoffman & Tchir, 1990).

Der eigentliche Grund, warum die allgemein-sozialpsychologische Forschung für die vorliegende Arbeit so interessant ist, besteht jedoch darin, dass sie versucht, den IC-Effekt in eine allgemeine psychologische Theoriebildung einzubinden. Dies ist m.E. sehr wichtig, weil man sonst nicht versteht, was der eigentliche „Motor“ des Effekts ist, und ohne dieses Verständnis ist es sehr schwer, eine angemessene Erklärungsstruktur auch auf der psycholinguistischen Prozessebene zu formulieren. Schon Brown und Fish (1983a) haben zur Erklärung ihrer Ergebnisse zwei basale Kausalschemata für Handlungs- und Zustandsverben postuliert (vgl. auch Abelson & Kanouse, 1966; McArthur, 1972): *Handlungen* werden, weil als willentlich verursacht gesehen, primär dispositional erklärt, d.h. unter Rückgriff auf aktuelle oder stabile Eigenschaften des Handelnden selbst, die bei anderen Menschen so nicht vorhanden sein müssen. *Zustände* werden hingegen primär als von äußeren Umständen verursacht betrachtet. Es gibt also bestimmte Eigenschaften des Stimulus, die zu diesem Verhalten (generell bei allen Personen) führen, der Erlebende selbst hat daran nur einen relativ geringen An-

teil. Diese Unterscheidung korrespondiert ungefähr mit der Skinners (1938) zwischen operantem und respondentem Verhalten und scheint ein allgemeiner Bestandteil der *theory of mind* der meisten Menschen und auch unserer Gesellschaft insgesamt zu sein (z.B. beruht unsere Strafgesetzgebung zu großen Teilen auf dieser Unterscheidung; vgl. z.B. Perner, 1991). Andere Konzepte, die in der psychologischen Literatur sehr ähnlich verwendet werden, sind z.B. „presupposed responsibility“ (Fillmore, 1971), das Intentions-Konzept Heiders (Heider, 1958) und die Initiator-Reactor-Unterscheidung von Osgood (1970) sowie (schon etwas entfernter) auch die Theorie der persönlichen Verursachung (deCharms, 1968) oder Rotters (1954) Unterscheidung zwischen internaler und externaler Kontrolle. An dieser Stelle ließe sich ebenfalls noch ein ganzes Kapitel anschließen bezüglich der Rolle, welche diese Unterscheidung in der Analytischen Philosophie des Geistes zur Erklärung solcher Phänomene wie Intentionalität und Subjektivität spielt (vgl. einführend Bieri, 1981; Beckermann, 2000). Ich will jedoch an dieser Stelle nicht weiter auf die philosophischen und psychologischen Grundlagen eingehen, sondern stelle schlichtweg fest, dass das Denken in verschiedenen Zuständlichkeits- und Verhaltens-Kategorien, mit denen unterschiedliche Formen der Ursachenzuschreibung verknüpft sind, für das menschliche (Alltags-) Denken *fundamental* ist (aber auch für den psychologischen Diskurs, vgl. Groeben, 1986).

Dieses generelle Prinzip scheint deshalb auch mehr oder weniger allen sozialpsychologischen Erklärungsansätzen der IC zugrunde zu liegen, allerdings wird es (meist) noch weiter ausdifferenziert. Rudolph und Försterling (1997) haben in ihrer Metaanalyse zum (sozialpsychologischen) IC-Effekt insgesamt vier unterschiedliche Hypothesen zur Erklärung des IC-Effekts unterschieden. Ihnen allen ist ein einheitlicher Rahmen gemeinsam, in welchem ein Verb mit IC am Anfang einer Erklärungskette steht und das konkrete Attributionsverhalten am Ende. Die einzelnen Theorien divergieren hingegen bezüglich des von ihnen angesetzten mediierenden Erklärungsmechanismus (vgl. Abb. 2.1).

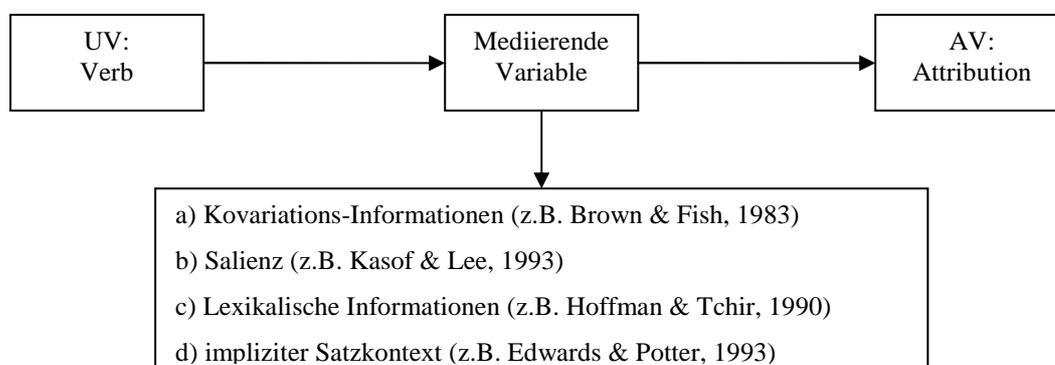


Abb. 2.1: Erklärungskonzepte für den Impliziten Kausalitäts-Effekt nach Rudolph & Försterling (1997a).

Die sog. Kovariations-Hypothese ist sicherlich die am häufigsten vertretene Hypothese. Ihr zufolge wird das verbale Attributionsverhalten von genau den gleichen Informationen gesteuert wie das menschliche Attributionsverhalten überhaupt. Nach der Attributionstheorie von Kelley (1967) sind für die Ursachenzuschreibungen von Menschen insbesondere Informationen über den Konsensus (über Personen), die Distinktheit (über Entitäten) und die Konsistenz (über Zeitpunkte bzw. Modalitäten) des Verhaltens entscheidend (vgl. Heckhausen, 1989; Kap. 12; da Konsistenz-Informationen für die Erklärung des IC-Effekts praktisch keine Rolle spielen, werden sie der Einfachheit halber im Folgenden weggelassen): Wenn ich wissen will, warum Peter Otto *bewundert*, dann brauche ich Informationen darüber, a) ob Peter grundsätzlich alle Personen bewundert (niedrige Distinktheit) oder nur Otto (hohe Distinktheit) und b) ob alle Personen Otto bewundern (hoher Konsensus) oder nur Peter (niedriger Konsensus). Nach Kelleys sog. Kovariationsprinzip attribuieren Menschen nun ein Verhalten auf die Personen selbst, wenn niedriger Konsensus (nur Peter mag Otto) und niedrige Distinktheit vorliegt (Peter mag überhaupt sehr viele Menschen). Wenn jedoch der umgekehrte Fall eintritt und hoher Konsensus (Alle mögen Otto) und hohe Distinktheit vorliegen (Peter mag sonst niemanden), dann wird das Verhalten auf den Stimulus Otto attribuiert. Die Kovariations-Hypothese besagt nun, dass in IC-Verben (ohne Kontext!) aufgrund bereits bestehender Erfahrungen, implizite, sozusagen sprachlich hart-verdrahtete Kovariations-Informationen enthalten sind, welche dazu führen, dass entweder das Subjekt oder das Objekt als kausal verantwortlich gesehen wird (vgl. Tab. 2.3).

UV	Mediierende Variable		AV
Verb	Kovariations-Informationen		Attribution
	Konsensus	Distinktheit	
AP	Niedrig	niedrig	Person (Subjekt)
AE	Hoch	Hoch	Entität (Objekt)
SE	Niedrig	Niedrig	Person (Subjekt)
ES	Hoch	Hoch	Entität (Objekt)

Tab. 2.3: Kovariations-Schema für die verschiedenen Verbtypen des ASD-Systems.

Die zentrale Strategie, diese Hypothese zu überprüfen, besteht seit ihrer expliziten Formulierung durch Brown und Fish (1983a) darin, Vpn Satzfragmente wie *Peter bewundert Otto* zu geben und sie die Konsensus- und Distinktheit beurteilen zu lassen. Dabei findet sich immer wieder und sehr dominant das in Tab. 2.3 beschriebene Muster. Auch verschwinden alle Unterschiede zwischen den Attributionsmustern der vier Verbtypen vollständig, wenn man die entsprechenden Kovariations-Informationen auspartialisiert.

Die Salienz-Hypothese (Kasof & Lee, 1993) ist ebenfalls attributionstheoretisch begründet. Allerdings geht sie nicht davon aus, dass Kovariations-Informationen an sich für die verbale Rollenverteilung verantwortlich sind, sondern stattdessen ein basaler Prozess der Salienz: Verben evozieren mentale Repräsentationen, in denen verschiedene Aktanten unterschiedlich salient sind. Es wird nun auf die Person attribuiert, die salienter ist, also mehr Aufmerksamkeit auf sich zieht (von daher ähnelt das Erklärungsmuster den aufmerksamkeitsbasierten Theorien der pronominalen Auflösung). Zur Testung der Hypothese wurden einerseits Salienz-Ratings herangezogen, die Salienz einzelner Aktanten wurde jedoch andererseits auch experimentell manipuliert (z.B. durch Passivierung). Die Effekte sind schwach, sprechen aber grundsätzlich dafür, dass die Salienz einen Einfluss auf die Attribution hat. Allerdings haben Rudolph und Försterling (1997a) darauf hingewiesen, dass es sich beim Salienz-Konzept im Wesentlichen um eine Reformulierung des Kovariations-Begriffs handelt (so führt z.B. niedriger Konsensus oder hohe Distinktheit im Kelley-Schema zwangsläufig auch zu einer höheren Salienz). Das Konzept bringe deswegen auch keinen essenziellen Erklärungsmehrwert mit sich. Als wichtiger Punkt für die spätere Diskussion ist allerdings festzuhalten: Erklärungs-basierte Kovariations- und aufmerksamkeits-basierte Salienz-Informationen arbeiten meist nicht *gegen-*, sondern vielmehr *miteinander*, denn das, was für eine gute Erklärung wichtig ist, ist auch hervorgehoben, Unwichtiges tritt hingegen in den Hintergrund.

Die lexikalische Hypothese (die v.a. von Hoffman & Tchir, 1990, vertreten wird) basiert wesentlich auf der Beobachtung, dass die unterschiedlichen Verbtypen des ASD-Systems unterschiedliche Adjektiv-Bildungen haben (zum Objekt vs. zum Subjekt, s.o.). Im Gegensatz z.B. zu Brown und Fish (1983a) führen sie diese Tatsache jedoch nicht darauf zurück, dass mit den unterschiedlichen Verben verschiedene Kovariations-Informationen verbunden sind. Stattdessen gehen sie davon aus, dass die IC eines Verbs eine lexikalisch kodierte Information ist, welche sowohl die Adjektiv-Bildung als auch die nachfolgende Attribution determiniert. Sie nehmen an, dass bei Verarbeitung eines IC-Verbs immer seine Adjektiv-Form mitaktiviert wird (oder diese Verbindung bereits habitualisiert ist). Das Adjektiv trägt wiederum die attributive Präferenz als lexikalisches Merkmal. Anstatt also ein allgemeines kognitives Prinzip (Kovariations-Informationen oder Salienz) zu bemühen, wird der IC-Effekt allein dadurch erklärt, dass Verben unterschiedliche Muster der Adjektiv-Bildung haben. Die Autoren vertreten damit eine (starke) sprachdeterministische Position. Genau bei diesem Punkt liegt dann auch die zentrale Schwäche des Ansatzes: Denn abgesehen von der Tatsache, dass die empirische Evidenz für eine starke, eigenständige Wirkung der Adjektiv-Morphologie ausgespro-

chen schwach ist (s.o.), ist die Stipulation eines lexikalischen Merkmals auf der theoretischen Ebene keine wirkliche Erklärung für den IC-Effekt: Ähnlich wie bei der merkmals-basierten Interpretation der IC als reines lexikalisches Merkmal bei Caramazza et al. wird das Problem einfach nur verlagert. Weder wird es in eine allgemeine Theorie der menschlichen Geistes eingebunden (wo es zweifelsohne hingehört!), noch kann der Ansatz etwas darüber aussagen, wie und warum sich das morphologische System gerade so entwickelt hat, wie es sich entwickelt hat und warum die Unterschiede in der Adjektiv-Morphologie so auffällig mit allgemeinen semantischen Merkmalen der Verben kovariieren.

Die Satzkontext-Hypothese (vgl. v.a. Edwards & Potter, 1993; s. auch Au, 1986, und Fiedler & Semin, 1988) besagt nun schließlich, dass IC-Verben nicht – wie die Kovariations-Hypothese dies annimmt – isoliert irgendwelche Konsensus- und Distinktheits-Informationen aktivieren, sondern dass durch sie vielmehr komplexe kausale Szenarien getriggert werden: Anstatt einen Satz wie *Otto belügt Peter* isoliert zu verarbeiten, konstruieren Menschen für diese Aussage einen impliziten Satzkontext, in dem der Akt des Belügens nicht dekontextualisiert, sondern vielmehr in eine Ereigniskette eingebunden ist. Es gibt Ereignisse, die der Handlung vorausgehen und diese hat selbst wiederum Konsequenzen. Am deutlichsten hat diese Position vielleicht Au (1986, p. 119) in Rückgriff auf Fillmore (1977) formuliert:

Understanding a word [...] can be thought as activating a scene or a cluster of scenes and then focussing to a certain part of the scene(s). [...] The findings of the present and other research on interpersonal verbs suggest that the scenes concerning the causes and consequences of interpersonal events can readily be activated in the process of understanding these verbs.

Die Zuweisung der kausalen Rolle ist nach dieser Sichtweise eher eine Frage der „Interpunktion“ der Ereignisse: Je nachdem, ob man den fokalen Satz eher als Konsequenz einer vorhergehenden Handlung ansieht oder als Ausgangspunkt für ein nachfolgendes Ereignis, wird eine unterschiedliche Verantwortlichkeitszuschreibung vorgenommen:

Whenever we use a sentence in a text or in natural discourse, it is embedded in a context of antecedent and consequent sentences or in a context of referential behaviors. These either precede or are anticipated to follow the focal sentence. Granting that this context is necessary to understand the full meaning of all but the simplest sentences, then people may be expected to infer or imagine the temporal sentence context when no context is explicitly provided. Thus the argument advanced here is that the causal impact of verbs somehow reflects an implicit sentence context – that is, the kind of behaviors that are inferred as preceding and/or following the behaviour stated in a focused sentence. (Fiedler & Semin, 1988, p. 28).

Für diese Hypothese spricht, dass Vpn nicht nur sehr schnell eine passende (rückwärts gewandte) kausale Attribution für ein Satzereignis finden, sondern genauso gut darin sind, mögliche Konsequenzen des Ereignisses zu generieren (*Otto mag Peter, deshalb...*; vgl. z.B.

Au, 1986; Edwards & Potter, 1993; Fiedler & Semin, 1988). Dieser Befund ist sprachpsychologisch v.a. deshalb interessant, weil parallel hierzu der Nachweis gelungen ist, dass es nicht nur eine „implicit causality“ in der Verbverarbeitung gibt, sondern auch eine „implicit consequentially“ (Stewart, Pickering & Sanford, 1998, Garnham, 2001). Wenn man Vpn bittet, keine rückwärts-, sondern vielmehr vorwärtsgerichtete Satzfragmente wie *Weil Otto Peter mag, ...* zu ergänzen, dann zeigen sie ein Ergänzungsverhalten, welches (prinzipiell) gegenläufig zum eigentlichen IC-Bias des Verbs ist (vgl. Crinean & Garnham, 2006): So wie die AGENS-Rolle bei Handlungsverben eher als kausal verantwortlich für den Handlungsakt angesehen wird, so wird gleichzeitig angenommen, dass der Handlungsakt eher Konsequenzen für den PATIENS hat. Und in ähnlicher Weise wird bei Zustandsverben der STIMULUS zwar als ursächlich für den Zustand des EXPERIENCER angesehen (weshalb er bei der rückwärtsgerichteten Begründung eher thematisiert wird), allerdings wird ebenfalls angenommen, dass der EXPERIENCER im Folge-turn auf den STIMULUS reagieren wird (weshalb er bei einer vorwärtsgerichteten Ergänzung eher thematisiert wird).

Rudolph und Försterling (1997a) äußern sich in ihrer Metaanalyse eher kritisch zur Satzkontext-Hypothese, jedoch ohne ihre Ablehnung wirklich begründen zu können. Zu ihren Einwänden zählen eher pauschale Bedenken wie z.B., dass dieser Ansatz ein zusätzliches Erklärungskonzept (nämlich das eines impliziten vorhergehenden Ereignisses) einführt, während andere Theorien der IC lediglich dispositionale Ursachen ansetzen. Auch wenden sie ein, dass ein vorhergehendes Ereignis ja nicht zwangsläufig auch die Ursache eines Folgeereignisses sein muss. Rudolph und Försterling präferieren hingegen sehr klar die Kovariations-Hypothese als den Ansatz mit der besten Evidenz, der größten Erklärungskraft und den geringsten Voraussetzungen. Allerdings sind sie da m.E. ein wenig vorschnell. Nicht nur, dass die von ihnen angeführten Gegenargumente in meinen Augen keine wirklichen Nachteile der Satzkontext-Hypothese sind, sondern eher Vorteile. Darüber hinaus unterschätzen sie das Erklärungspotenzial des Ansatzes erheblich: Er liefert nicht nur eine plausible Einbettung der IC in eine umfassende allgemeinspsychologische Theoriebildung, sondern kann darüber hinaus auch einige Dinge erklären, die mit der Kovariations-Hypothese alleine unerklärlich sind. Dazu gehören u.a. eine schlüssige Erklärung für die Diskrepanz zwischen den AP- und den AE-Verben sowie eine konsequente Weiterentwicklung in Richtung auf die impliziten Konsequenzen einer Handlung, die sich als ein sehr robustes Phänomen erwiesen hat (vgl. Crinean & Garnham, 2006).

Allerdings ist es m.E. sowieso überflüssig, sich darüber zu streiten, ob nun die Kovariations- oder die Satzkontext-Hypothese angemessener zur Erklärung des Phänomens der IC ist.

Denn meiner Meinung nach sind alle vier von Rudolph und Försterling genannten Hypothesen miteinander kombinierbar, wenn man sie einmal von ihren (vollkommen unnötigen) Überbeziehungen und Verabsolutierungen befreit hat. Denn es ist in der Tat plausibel anzunehmen, dass mit bestimmten interpersonalen Verben durch entsprechende Sozialisation auf relativ basaler, evtl. bereits lexikalischer Ebene so etwas wie prototypische Szenarien oder Skripte verbunden sind (lexikalische Hypothese), die eine bestimmte Handlungs-Folge-Sequenz abbilden und mit denen ohne explizit gegebenen Kontext eine bestimmte *default*-Interpunktion verbunden ist (Satzkontext-Hypothese). So kann man sich z.B. relativ leicht vorstellen, dass eine Person X eine Person Y lobt und dass das nicht grundlos geschieht, sondern in Reaktion auf ein vorhergehendes Ereignis, z.B. weil sie gerade eine gute Note bekommen hat. Aber selbstverständlich sind mit dieser Verursachungsrelation auch die entsprechenden Kovariations-Informationen verbunden, die das vorhergehende Ereignis überhaupt erst zu einem adäquaten Grund machen (Kovariations-Hypothese): So gehe ich doch gerade davon aus, dass sich das inferierte Ereignis des Gute-Noten-Bekommens (zumindest) durch niedrigen Konsens auszeichnet, sonst gäbe es ja gar keinen Anlass für das Lob. Weiterhin ist es m.E. überaus wahrscheinlich, dass die gerade gelobte Person auch die salienteste im Szenario ist, weil sich der ganze Akt des Gute-Noten-Bekommens und Nachher-gelobt-Werdens schließlich nur um sie dreht (Salienz-Hypothese).

Im Gegensatz zu Rudolph und Försterling schätze ich also das Konvergenz-Potenzial der verschiedenen allgemein- und sozialpsychologischen Theorien als recht hoch ein. Ich würde es deshalb auch sehr begrüßen, wenn die zukünftige sozialpsychologische Forschung eine solche integrative Theorie der IC entwickeln und genauer ausdifferenzieren würde. Meine Aufgabe kann das hier allerdings nicht sein und muss es auch nicht. Denn für die in dieser Arbeit behandelte Fragestellung ist nur wichtig, dass der psychologische Mechanismus, auf dem der IC-Effekt basiert, ein sehr allgemeiner, basaler und evtl. hoch automatisierter Prozess ist, da die Beurteilung von Verursachungs- und Verantwortungsrelationen gewissermaßen zum Standardrepertoire unseres alltäglichen Lebens gehört. Deswegen ist auch davon auszugehen, dass die mit solchen Einschätzungen verbundenen Informationsbestände und Verarbeitungsprozesse sehr schnell verfügbar sind und ressourcen-schonend eingesetzt werden können. Daraus jedoch zu schließen, dass hierfür benötigten Wissensbestände simpel und die beteiligten Prozesse lediglich elementar sind, wäre jedoch ein fundamentaler Fehler. Denn die Diskussion um die sozialpsychologische Erklärung des IC-Effekts hat ja gerade gezeigt, dass zur Etablierung einer erfolgreichen Erklärungsrelation eine hochkomplexe Ereigniskette konstruiert werden muss und u.U. sehr spezifische Objekt-Informationen zu bewerten sind.

2.1.2 Linguistik

Es gibt vergleichsweise wenig linguistische Literatur, die sich direkt mit dem IC-Effekt befasst. Stattdessen wird das Phänomen meist im Rahmen größerer und fundamentalerer Fragestellungen besprochen. Das sind dann im linguistischen Kontext zunächst einmal Fragen zur Verbklassifikation und -semantik (Welche Verbtypen müssen unterschieden werden? Mittels welcher Kriterien lassen sie sich voneinander abgrenzen?), aber auch die Fragestellung, in welchem Verhältnis die Verbsemantik zur übergreifenden Gesamtbedeutung eines Satzes steht (Welche semantischen Restriktionen sind mit bestimmten Verbtypen verbunden?). In einem weiteren Sinne – und für die vorliegende Arbeit relevanter – sind damit aber auch alle Aspekte angesprochen, die sich in Bezug auf die sog. Syntax-Semantik-Interface-Problematik ergeben. Denn die semantische Struktur eines Verbs spiegelt sich in der syntaktischen Struktur des Satzes wider. Allerdings ist aber die Semantik bestimmter Verben und die Konsequenzen, die sie für die syntaktische Strukturierung eines Satzes hat, seit Jahr und Tag *der Punkt*, über den sich die verschiedene Grammatiktheorien (z.B. generative vs. lexikalistische Ansätze) streiten (vgl. im Überblick Newmeyer, 1996). Denn letztendlich geht es dabei um die alte Frage, inwieweit die Sprachverarbeitung (syntaktischer Strukturaufbau und semantische Interpretation) und das generelle Kognitionssystem (z.B. bezüglich allgemeiner Attributionsschemata oder Weltwissensszenarien) zusammenwirken.

Fangen wir mit den einfachen Dingen an. Im letzten Abschnitt wurde bereits ausführlich das Prinzip der Argumentstruktur und -selektion dargestellt. Verschiedene Verben verlangen unterschiedlich viele Argumente und verteilen unterschiedliche θ -Rollen (wie AGENS, PATIENS etc.). Die Literatur zur IC konzentriert sich dabei fast ausschließlich auf zweistellige Verben und arbeitet zusätzlich mit der Unterscheidung von Handlungs- und Zustandsverben (die jeweils unterschiedliche θ -Rollen verteilen, s.o.). Das alles sind jedoch zunächst einmal Beschreibungen allein auf der semantischen Ebene: So gehört es halt zur Bedeutung des Verbs *lieben*, dass es eine Entität impliziert, die den Zustand der Liebe erlebt (EXPERIENCER), und eine entgegen gesetzte Entität, welche diesen Zustand auslöst (STIMULUS). Es wird angenommen, dass im mentalen Lexikon kodiert ist, welche Verben welche Rollen selektieren (der Lexikon-Eintrag könnte also z.B. so aussehen: [‘lieben’ \rightarrow x(EXP, STIM)]). Die Frage ist nun, welche Konsequenzen dies für die Syntax des Satzes hat. Irgendwie muss sich diese Bedeutung in der Struktur des Satzes ja abbilden. Hierfür ist z.B. im Rahmen der Generativen Grammatik die sog. θ -Theorie zuständig. Sie versucht, alle für die Syntax wesentlichen Elemente der Semantik „auszulesen“ und für den syntaktischen Strukturaufbau bereitzustellen (sie ist deswegen eine sog. Interface-Theorie). Es wird angenommen, dass für die *Syntax* nur

die Anzahl der Rollen (das sog. *Theta-Raster*) relevant ist, nicht ihre Art (d.h., ob es auch die „richtigen“ Rollen sind, wird nicht von der syntaktischen, sondern der semantischen Komponente des Sprachsystems überprüft). Im Wesentlichen wird dann eine 1:1-Relation zwischen thematischen Rollen und syntaktischen Argumenten etabliert: Das heißt, jedes Argument darf nur eine θ -Rolle bekommen und jede θ -Rolle darf nur einem Argument zugewiesen werden (selbstverständlich ist das alles wesentlich komplizierter, weil es noch fakultative Adjunkte gibt und Verben unterschiedliche syntaktische Kategorien selegieren etc., vgl. z.B. Fanselow & Felix, 1993b). Die θ -Theorie umgeht also das Problem der Syntax-Semantik-Korrespondenz bzw. lagert es aus der syntaktischen Komponente aus. Nachdem das Theta-Raster einmal etabliert ist, wird der syntaktische Teil der Derivation dann von anderen Teiltheorien der Syntax (wie der Rektions- und der Kasus-Theorie) übernommen und für die Wohlgeformtheit der Interpretation ist die Semantik zuständig.

Damit bleiben aber noch ein paar (entscheidende!) Fragen offen. Die Frage, die für den IC-Effekt am wichtigsten ist, besteht darin, welche thematischen Rollen in welcher syntaktischen Position realisiert werden, weil dies letztendlich darüber entscheidet, ob ein Verb ein NP1- oder ein NP2-Verb ist. Hierzu macht die θ -Theorie keine Aussage. Stattdessen wird dieses Problem in der linguistischen Literatur meist unter dem Schlagwort des sog. *argument linking* diskutiert (vgl. im Überblick Butt, 2006). Denn die Reihenfolge der Realisierung der einzelnen thematischen Rollen ist (im Englischen und im Deutschen, andere Sprachen, z.B. das Japanische, sind hier freier) nicht beliebig, wie ein Vergleich von (4a) und (4b) zeigt:

- (4) a. Peter_{AGENS} schlägt Hans_{PATIENS}.
 b. * Hans_{PATIENS} schlägt Peter_{AGENS}.

Das heißt, wir bekommen einfach keine Lesart hin, in welcher Hans (als der Geschlagene) in der grammatikalischen Subjekt-Position steht, dort *muss* der AGENS Peter stehen. Solche und ähnliche Beispiele haben dazu geführt, dass in der linguistischen Literatur eine Reihe sog. *Linking Hierarchien* postuliert wurden (über die heftig gestritten wird, vgl. z.B. van Valin & LaPolla, 1997, vs. Jackendoff, 1990), in denen die Abfolge der *syntaktischen* Realisierung eines Arguments in Abhängigkeit von seiner *semantischen* Rolle geregelt wird. Allgemeine Einigkeit besteht darin, dass die AGENS-Rolle (so vorhanden) immer als Satzsubjekt kodiert wird, ein unmittelbarer Mitspieler (z.B. ein PATIENS) meist als direktes Objekt. Über dieses basale Prinzip hinaus divergieren die einzelnen Positionen dann allerdings. Ein Art Minimalkonsens ist jedoch eine Abfolge die ungefähr so aussieht (hier in der Version von Grimshaw, 1990; vgl. auch Jackendoff, 1972):

- (5) AGENT → EXPERIENCER → GOAL/SOURCE/LOCATION → THEME (Grimshaw, 1990, p. 8).

Das heißt, wenn ein Verb eine AGENS-Rolle vergibt, dann wird diese immer als erste realisiert, wenn dies nicht der Fall ist, dann wird der EXPERIENCER als Subjekt realisiert etc. (die Rolle THEME ist bei Grimshaw, 1990, eine Art Sammelbezeichnung für sehr unterschiedliche Rollen: im vorliegenden Fall ist allein wichtig, dass damit sowohl die PATIENS- als auch die STIMULUS-Rolle gemeint ist).

Für die Klasse der Handlungsverben, die eine AGENS- und eine PATIENS-Rolle vergeben, stellt die oben dargestellte Hierarchie kein Problem dar. Sie erklärt vielmehr gut, warum der AGENS immer in der Subjekt-Position realisiert ist. Das eigentliche Problem für die Handlungsverben besteht vielmehr darin zu erklären, warum das kausale Gewicht bei ihnen dennoch manchmal auf dem Objekt liegt, nämlich bei den AE-Verben. Hierfür wird dann eine Sondererklärung notwendig werden (s.u.), die dieses „sonderbare“ Verhalten dieser Verbklasse begründet. Im Moment ist die semantische Kausalrelation zwischen den Rollen jedoch (noch) nicht relevant, da sich die in (5) dargestellte Hierarchie einzig und allein auf die syntaktische Argumentposition bezieht.

Stattdessen konzentriert sich die linguistische Debatte allein auf die Zustandsverben (die dort PSYCH-Verben heißen): Denn während die Klasse der ES-Verben der spezifizierten kanonischen Reihenfolge vollkommen entspricht, wird sie von den SE-Verben verletzt. Solche Verben sollte es also gar nicht geben. Konfrontiert mit dieser unschönen Situation hat eine linguistische Theorie mehrere Möglichkeiten:

- a) Sie kann die in (5) spezifizierte Hierarchie aufgeben und wird damit wieder den linguistischen Daten gerecht. Sie gibt damit aber eine Regularität auf, die sich in sehr vielen anderen Fällen durchaus bewährt hat und erhebliche Erklärungsvorteile z.B. für den Spracherwerb mit sich bringt. Denn wenn die Hierarchie gilt, dann muss lediglich eine einzige Regel gelernt werden; wenn nicht, dann muss für jedes einzelne Verb ein eigener Lexikoneintrag erworben werden, der spezifiziert, welche thematische Rolle in welcher syntaktischen Position realisiert wird. Das Problem wird also auf lexikalischer Ebene gelöst.
- b) Sie kann die in (5) spezifizierte Hierarchie als eine allgemeine Regel ansetzen und die Klasse der SE-Verben als Ausnahmen extra definieren (wie man das ja auch für unregelmäßige Verben im Fremdsprachunterricht kennt). Ein solches Vorgehen hat den Vorteil, dass wenigstens nur für die Ausnahmen ein Mehraufwand notwendig ist, der Großteil der Fälle jedoch durch die allgemeine Regel abgedeckt wird. Ein Nachteil ist, dass eine solche Theorie nicht angeben kann, warum die Ausnahmen

systematisch nur bei den PSYCH-Verben auftreten (nicht jedoch z.B. bei den Handlungsverben).

- c) Sie kann versuchen, die irregulären SE-Verben durch die Einführung einer Zusatzregel zu erklären. Das läuft dann meist darauf hinaus, dass man argumentiert, dass die SE-Verben „eigentlich“ gar keine wirklichen SE-Verben sind, sondern verkappte ES-Verben. Auf lexikalischer Ebene gibt es also nur einen einzigen Verbtyp (alle PSYCH-Verben sind ES-Verben!), und bei manchen von ihnen wird halt die STIMULUS-Rolle aufgrund anderer (unabhängiger) Prinzipien oberflächen-strukturell in der Subjekt-Position realisiert.

Die meisten linguistischen Theorien (und zwar vollkommen unabhängig von ihrer jeweiligen theoretischen Ausrichtung, z.B. generativ vs. kasusgrammatisch etc.) entscheiden sich für die Möglichkeit c) (vgl. jedoch neuerdings Jackendoff, 2007). Die verschiedenen Ansätze unterscheiden sich jedoch erheblich darin, welche Prinzipien sie zur Erklärung der Tatsache heranziehen, dass einige PSYCH-Verben auf der Satzoberfläche die Abfolge SE haben. Einschlägig sind hier v.a. die Ansätze von Belletti und Rizzi (1988), von Grimshaw (1990) und von Dowty (1991) (vgl. zusammenfassend und vergleichend: Pittner, 1994).

Die Analyse der SE-Verben (sog. *frighten*-Klasse) durch Belletti und Rizzi (1988) ist generativ orientiert und basiert auf den früheren Analysen dieser Verbkategorie durch Fillmore (1968) bzw. Postal (1970). Entsprechend der eben skizzierten grundsätzlichen Argumentationslogik nehmen sie an, dass SE-Verben auf Tiefenstruktur ES-Verben sind: Während ES-Verben ganz normal basisgeneriert werden, gehen sie davon aus, dass den SE-Verben eigentlich eine Doppelobjekt-Struktur zugrunde liegt, in welcher die STIMULUS-NP das direkte Objekt des Verbs ist und die EXPERIENCER-NP das indirekte (vgl. Abb. 2.2).

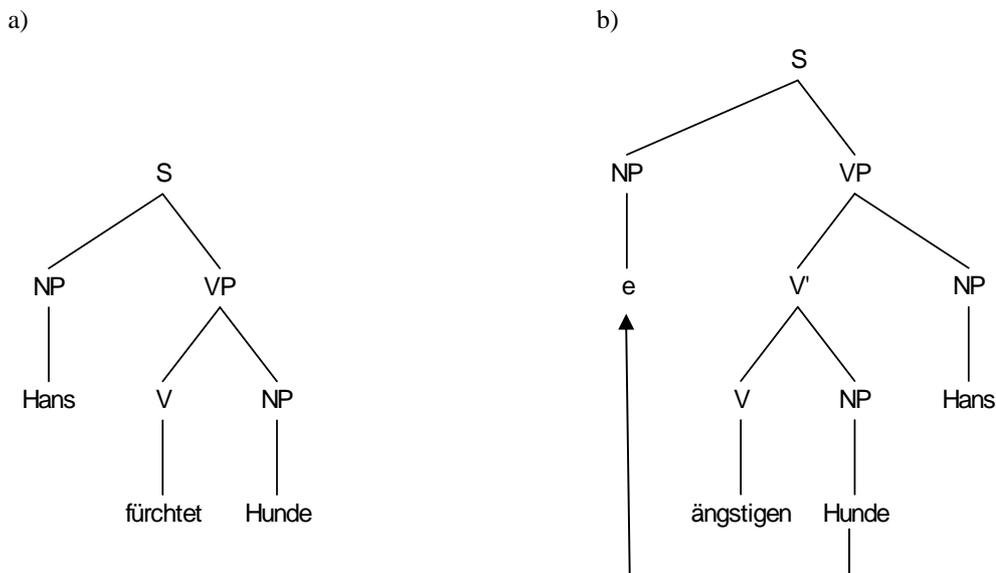


Abb. 2.2: Derivation a) der ES- und b) der SE-Verben nach Belletti & Rizzi (1988).

Obwohl bei beiden Verbklassen also die Stimulus-NP direkt mit dem Verb mit generiert wird, erhalten SE-Verben erst im Laufe der Derivation ihre spezifische Oberflächenstruktur, indem die EXPERIENCER-NP in die freie SpecS der linken Satzklammer bewegt wird (also in die Subjekt-Position).

Belletti und Rizzi führen für ihre Hypothese eine Menge von Material aus dem Italienischen an. Allerdings sind die von ihnen behandelten Phänomene recht technisch und ohne sehr gute Vorkenntnisse der Generativen Grammatiktheorie wahrscheinlich auch unverständlich (weshalb ich sie dem/der Leser/in an dieser Stelle ersparen möchte). Es sei nur kurz angedeutet, dass Belletti und Rizzi sowohl Evidenz dafür vorbringen, dass

- a) das Subjekt in SE-Konstruktionen bestimmte Eigenschaften besitzt, die sonst nur sog. abgeleitete (d.h. bewegte) Subjekte haben (sie können z.B. im Italienischen nicht reflexiv klitisiert werden, erlauben keine pro-Deletion, können nicht in kausativen Einbettungen vorkommen und können nur adjektivisch, nicht verbal passiviert werden);
- b) dem Objekt in SE-Konstruktionen bestimmte Eigenschaften fehlen, die man bei normalen direkten Objekten erwarten würde (u.a. derivationale Transparenz, Block von Extraktion etc.).

Auf einen weiteren Punkt möchte ich allerdings doch etwas näher eingehen, weil er für die vorliegende Problematik direkt relevant ist: Die von Belletti und Rizzi vorgeschlagene Struktur kann das Phänomen erklären, dass SE-Verben sich „unnormale“ in Bezug auf die üblichen Gesetze der Bindungstheorie verhalten. Speziell erlauben sie Konstruktionen wie in (6a), die nach dem Prinzip A der Bindungstheorie eigentlich nicht möglich sind (vgl. 6b; nach Belletti & Rizzi, 1988, p. 312; vgl. auch Kap. 1).

- (6) a. Questi pettegolezzi su di se preoccupano Gianni più di ogni altra cosa
 [Die Gerüchte über sich_i ängstigten Hans_i mehr als alles andere.]
- b. *Questi pettegolezzi su di se descrivono Gianni meglio di ogni biografia ufficiale.
 [Die Gerüchte über sich_i beschreiben Hans_i besser als die offizielle Biographie.]

Prinzip A der Bindungstheorie besagt, dass Reflexiva wie *se/sich* in (6) lokal gebunden sein müssen (d.h. in diesem Fall innerhalb der Konstituente [Die Gerüchte über sich]). Das ist im Fall (6b) nicht gegeben, da *Hans* dort nicht vorkommt. Entsprechend ist der Satz auch ungrammatisch, man will klar ein Pronomen haben: *Die Gerüchte über ihn*.... In (6a) ist diese Struktur im Italienischen jedoch grammatisch (und im Deutschen würde ich sie zwar als ungrammatisch, aber immerhin als akzeptabler als (6b) betrachten). Der Punkt ist nun, dass die syntaktische SE-Struktur wie in Abb. 2.2 b) diesen Kontrast erklären kann: Der STIMULUS

wird innerhalb der VP basisgeneriert und wird deswegen auf D-Struktur auch vom EXPERIENCER c-kommandiert und lokal gebunden. Erst nach der Bindung wird er in die Subjekt-Position bewegt. Da bei Handlungsverben wie *beschreiben* der AGENS direkt in SpecS basisgeneriert wird, gibt es diese Möglichkeit hier nicht.

Auch wenn die Analyse der SE-Verben von Belletti und Rizzi (1988) somit einige wünschenswerte Eigenschaften aufweist, so haben mehrere Autoren/innen (z.B. Grimshaw, 1990; Pittner, 1994) darauf hingewiesen, dass sie allein noch nicht hinreichend sein kann. Denn das Hauptproblem dieser Theorie ist, dass sie postulieren muss, dass die SE-Verben lexikalisch über einen inhärenten Akkustativ-Kasus verfügen, weshalb nicht der EXPERIENCER in die Subjekt-Position bewegt wird, sondern der STIMULUS. Letztendlich ist also die SE-Stellung doch nur eine lexikalische Eigenschaft dieser Verbklasse und folgt keiner generellen Regel. Von daher ist sie eigentlich nur eine – syntaktische anspruchsvolle – Reformulierung des Ausgangsproblems, nicht aber seine Lösung.

Grimshaw (1990) geht deswegen auch ganz anders vor. Sie sucht die Lösung des Problems nicht mehr (allein) auf der syntaktischen, sondern vielmehr auf der semantischen Ebene. Zunächst einmal bringt sie noch weitere Evidenz dafür vor, dass in der kanonischen Hierarchie (5) in der Tat der EXPERIENCER über dem STIMULUS steht und nicht umgekehrt. Ein m.E. besonders schöner Beleg ist dabei, dass die Objekte von ES-Verben ein Kompositum mit ihrem Verb bilden können, die Objekte von SE-Verben jedoch nicht:

- (7) a. Mensch fürchtet Gott. → Ein gottesfürchtiger Mensch.
- b. Gott ängstigt Mensch. → *Ein mensch-ängstigender Gott.

Nach Grimshaw liegt dies daran, dass bei der Derivation die thematische Hierarchie gewissermaßen von unten nach oben abgearbeitet werden muss. Die niedrigste Rolle muss also zuerst vergeben werden, erst dann können sukzessiv die höheren verwendet werden. Wenn das ES-Verb *fürchten* also ein Kompositum mit seinem STIMULUS-Objekt bildet, dann ist diese Rolle verbraucht; danach kann es seine EXPERIENCER-Rolle noch an das Subjekt abgeben. Beim SE-Verb *ängstigen* ist das Objekt jedoch bereits der EXPERIENCER und die Reihenfolge der thematischen Rollen wäre damit verletzt. Aufgrund des irregulären Verhaltens der SE-Verben bringt es ihrer Meinung nach auch nichts, für die beiden Verbklassen einfach nur unterschiedliche thematische Rollen anzunehmen (vgl. z.B. Pesetsky, 1987; 1995: CAUSE OF EMOTION vs. TARGET OF EMOTION etc.), die dann einen eigenen Platz in der Hierarchie bekommen. Denn dies würde dazu führen, die irregulären SE-Verben gewissermaßen künstlich zu „über-regularisieren“, wodurch man der Tatsache nicht mehr gerecht wird, dass sie sich in

vielen Fällen (z.B. bei der Nominalisierung, s.o.) seltsam verhalten. Einfach nur so zu tun, als wäre alles in Ordnung, löst alleine auch noch nicht das Problem...

Im Gegensatz zu Belletti und Rizzi greift sie jedoch nicht auf eine lexikalisch-syntaktische Stipulation zurück, sondern führt zusätzlich zu der thematischen noch eine zweite semantische Dimension ein, die mit der thematischen Hierarchie interagiert und deren Kombination dann die oberflächenstrukturelle Abfolge der Rollen determiniert. Diese zweite Dimension ist nun für meine Fragestellung besonders relevant, weil sie gleich zwei Aspekte aufgreift, die sowieso schon in der Luft liegen: So beobachtet sie, dass sich die SE-Verben von den ES-Verben dadurch unterscheiden, dass sie a) eine kausative Lesart haben und b) sich aspektuell von ihnen abgrenzen, da sie nicht stativ sind.

Die Konzepte *Aspekt* und *Aktionsart* (wobei Aspekt die grammatikalisierte Variante der Aktionsart ist) beziehen sich auf die temporale Struktur, die mit einem Verb verbunden ist (vgl. z.B. Lohnstein, 1996). Denn die mit einem Verb beschriebenen Ereignisse können sich in ihrem zeitlichen Verlauf erheblich voneinander unterscheiden: Eine Aussage wie *Er ist blond* beschreibt z.B. einen vergleichsweise lang andauernden Zustand, das Verb *platzen* ist hingegen eher punktuell. Diese Eigenschaft bezieht sich auf die Dauer der Handlung, die entweder *durativ* oder *nicht-durativ* sein kann [\pm DUR]. Darüber hinaus können die von einem Verb beschriebenen Aktivitäten zielgerichtet sein und einen natürlichen Endpunkt haben (z.B. *sterben*) oder aber nicht (z.B. *wissen*). Diese Eigenschaft bezieht sich auf das Ziel der Handlung, die entweder *resultativ* (auch: *telisch*, *perfektiv*) sein kann oder nicht [\pm RES]. Eine verbreitete Klassifikation von Aktionsarten ist das sog. Vendler-Dowty-System (nach Vendler, 1967, mit Modifikation durch Dowty, 1979). Es unterscheidet zwischen *states* (andauernden Zustände: *wissen*, *lieben* etc.), *activities* (ungerichtete, länger andauernde Handlungen: *laufen*, *tanzen* etc.), *accomplishments* (zielbezogene Handlungen: *umziehen*, *besteigen* etc.) und *achievements* (punktuelle zielbezogene Ereignisse: *entdecken*, *erreichen* etc.). Wenn man noch reine Eigenschaften hinzunimmt und diese von eher situativen Prozessen abgrenzt, dann ergibt sich nach Ehrich (1992) das folgende Klassifikationssystem (vgl. Abb. 2.3):

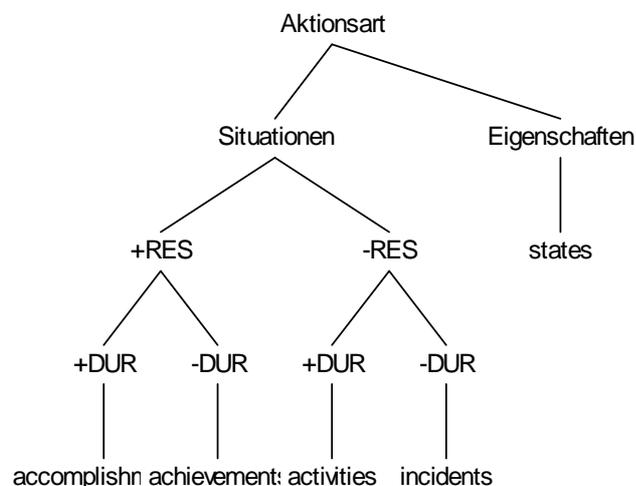


Abb. 2.3: Aktionsartklassifikation nach Ehrich (1992).

Nach Grimshaw (1990) hängen nun beide Punkte (Kausativität und Nicht-Stativität \approx Nicht-Resultativa + *states*) eng miteinander zusammen. Sie bezieht sich dabei auf Überlegungen von Pustejovsky (1988; 1991), der – ähnlich wie Moens und Steedman (1988) – für jedes Verb eine eigene Ereignisstruktur annimmt, die sich aus den Aktionsarten der einzelnen Ereignisse zusammensetzt. So schlägt Pustejovsky z.B. für die Klasse der *accomplishments* (bzw. auch der *achievements*, die sich von diesen ja nur durch ihre Durativität unterscheiden) folgende Ereignisstruktur vor:

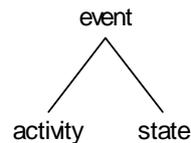


Abb. 2.4: Ereignisstruktur von accomplishment-Verben nach Pustejovsky (1988).

Das heißt, ein *accomplishment*-Ereignis ist ein *komplexes* Ereignis, das sich aus mehreren Teilereignissen zusammensetzt. Im Falle eines Verbs wie *umziehen* involviert das eine (bzw. meist leider ja sogar sehr viele) Aktivität(en), wie Kisten packen, einladen etc., die jedoch irgendwann in einen (End-)Zustand münden, nämlich irgendwo zu *wohnen*. Man bemerke, dass die beiden Teilereignisse kausal miteinander verknüpft sind: Der Akt des *Umziehens* führt zum Zustand des *Wohnens*. Grimshaw (1990) nimmt deswegen an, dass ein Argument immer genau dann kausativ interpretiert wird, wenn es (überwiegend) am *ersten* Teilereignis beteiligt ist. Dieses System arbeitet recht gut für die Klasse der Handlungsverben (natürlich ohne AE-Verben), die sich häufig als *accomplishments* rekonstruieren lassen. Aber auch durative und intransitive Verben werden damit erfasst. Denn wie Abb. 2.5 zeigt, lassen sich auch einfache Aktivitäten und Zustände in das Ereignisraster von Pustejovsky bringen:

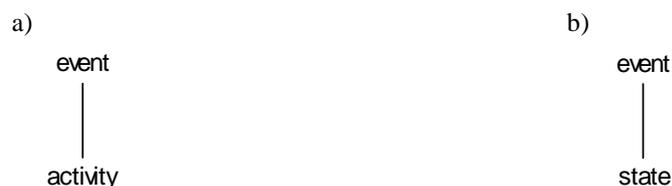


Abb. 2.5: Ereignisstruktur von activities und states-Verben nach Pustejovsky (1988).

Ein einfaches intransitives Verb wie *arbeiten* gehört zur Aktionsartklasse der Aktivitäten (-RES, +DUR) und hat nur ein einziges Teilereignis, nämlich sich selbst. Interessant ist jedoch v.a. was das System mit den Zustandsverben macht: Für die Klasse der SE-Verben, die ja – wie Grimshaw behauptet, nicht belegt! – immer eine resultative Lesart haben, führt die Erklärung zum gewünschten Ergebnis: ein SE-Verb wie *verärgern* ist ein *accomplishment*-Verb (+RES, +DUR) bzw. evtl. auch ein *achievement*-Verb (+RES, -DUR). Auf jeden Fall

lässt es sich durch seine Resultativität als ein komplexes Ereignis interpretieren, bei dem der STIMULUS eine Aktivität vollzieht, die dann zu einem Zustand beim EXPERIENCER führt. Deswegen ist der STIMULUS hier auch kausativ. Problematisch sind (was Grimshaw fairerweise auch selbst eingesteht) in diesem System gerade die ES-Verben. Sie müssen nach Grimshaw zwingend nicht-resultativ interpretiert werden, und beziehen sich dann schlicht auf ein Zustands-Ereignis, das eine Struktur wie in Abb. 2.5 b) hat, d.h. der Zustand ist wiederum sein eigenes Teilereignis und der EXPERIENCER, nicht der STIMULUS erhält die kausative Interpretation.

Bevor wir uns jedoch den mit diesem Ansatz verbundenen Problemen zuwenden, ist es sinnvoll, das System *in sake of the argument* vorläufig als überzeugend anzusetzen und uns anzuschauen, wo Grimshaw uns mit ihm hinführt. Denn der ganze Bogen über die aspektuellen Unterschiede zwischen den ES- und den SE-Verben war nur die Ouvertüre, der Hauptakt folgt erst noch. Was sie nämlich eigentlich erreichen wollte, ist ein unabhängiges Prinzip dafür zu finden, warum bei SE-Verben der STIMULUS hierarchie-konträr als Subjekt realisiert wird. Dieses Ziel erreicht sie nun dadurch, indem sie zusätzlich zur thematischen noch eine kausale Hierarchie ansetzt, welche die Subjekt-Realisierung steuert: Subjekt wird die Rolle, welche das Merkmal CAUSE trägt, also das kausative Element in der Ereignisstruktur. Beide Repräsentationsebenen (die hierarchische und die thematische) tragen also in Interaktion miteinander zur semantischen Interpretation des Satzes bei.

Durch diese meist zusammen, manchmal aber auch gegeneinander arbeitende Doppelhierarchie erhält Grimshaw zunächst einmal das gewünschte Ergebnis: AP-Verben (wie immer wird die AE-Klasse hier zunächst ausgeklammert!) haben eine Argument-Abfolge, die sowohl der thematischen als auch der kausalen Hierarchie entspricht, ES-Verben ihrer Analyse zufolge auch. Im einen Fall wird deswegen der AGENS, im anderen der EXPERIENCER Subjekt. Darüber hinaus sind nun auch die SE-Verben vorhersagekonform, weil die beiden Hierarchien in diesem Fall zwar divergieren (nach der thematischen müsste der EXPERIENCER, nach der kausalen der STIMULUS Subjekt werden), die kausale Hierarchie aber vorgeordnet ist. Gleichzeitig liefert diese Divergenz der beiden Hierarchien nun aber auch einen plausiblen Grund dafür, warum sich die SE-Verben in Bezug auf so viele linguistische Phänomene recht seltsam verhalten: Da bei ihnen die beiden Hierarchien auseinander fallen, haben sie kein maximal prominentes Element und haben damit kein *externes Argument*. Daraus folgen dann direkt viele weitere Eigenschaften dieser Verbklasse wie z.B., warum sie keinen verbalen Passiv bilden (vgl. Belletti & Rizzi, 1988): Da sie überhaupt kein externes Argument haben, kann dieses im Laufe der Passivierung auch nicht unterdrückt werden.

Ich muss sagen, dass diese Theorie ihren Charme hat: Sie erklärt die problematischen Fälle, kann aber auch begründen, warum diese problematisch sind. Gleichzeitig geht sie theoretisch m.E. definitiv in die richtige Richtung: Sie verknüpft syntaktische und semantische Überlegungen und führt (endlich!) die Kausalrelation in Kombination mit thematischen Kriterien als zentrales Erklärungs-konstrukt im Kontext eines situativ-ereignis-bezogenen Interpretationsrahmens ein. Allerdings muss man sich fragen, ob der Preis hierfür nicht ein wenig hoch ist. Abgesehen davon, dass die empirische Fundierung in Bezug auf die aspektuellen Unterschiede zwischen den ES- und SE-Verben m.E. ausgesprochen dürftig ist (s.u.), liegt das Hauptproblem bei ihrer kausalen Analyse der Verben der ES-Klasse. Zum einen ist die Begründung theoretisch post-hoc und lückenhaft: „There is no well-established independent analysis of their [= der Verben der ES-Klasse] aspectual structure that would guarantee that the Experiencer will be aspectually more prominent than the Theme. [...] so for present purposes we must simply stipulate it” (Grimshaw, 1990, p. 27ff.). Zum anderen – und das ist entscheidender – widerspricht das Ergebnis ihrer Analyse der ES-Verben sämtlichen Ergebnissen der IC-Forschung! Dort findet man ja gerade, dass der STIMULUS, und nicht der EXPERIENCER für kausal verantwortlich gehalten wird. Vor diesem Hintergrund erscheint Grimshaws Analyse für die SE-Verben auch zunächst einmal adäquat; die der ES-Verben dient jedoch lediglich dem Zweck, die anfangs postulierte thematische Ausgangshierarchie trotz ihrer Probleme noch weiter aufrechterhalten zu können und gleichzeitig noch ein Abgrenzungskriterium für das Sonderverhalten der SE-Verben zu haben. Allerdings gibt es auch durchaus linguistische Evidenz dafür, an ihrer Analyse der ES-Verben zu zweifeln: So sind sie z.B. cross-linguistisch nicht stabil und eine ganze Subklasse der ES-Verben (die sog. *please*-Klasse) hat überhaupt keine resultative Lesart, sondern ist klar stativ (vgl. Grimshaw, 1990, p. 29; s. auch Härtl, 1999).

Auch wenn Grimshaws Ansatz aus den genannten Gründen in seiner jetzigen Form als nicht-akzeptabel zurückgewiesen werden muss, so ist der Grundgedanke doch sehr begrüßenswert. Er soll deswegen auch im Folgenden als Ausgangspunkt für eine Explikation der thematischen und ereignisstrukturellen Grundlagen der PSYCH-Verben genutzt werden, die vielleicht in der Lage ist, den IC-Effekt besser zu verstehen. Die beiden Hauptaufgaben, die sich einem dabei stellen, sind a) die etwas brüchige Angewiesenheit allein auf die Aspekt-Informationen zu minimieren und b) eine kausale Lesart für die ES-Verben zu etablieren, die zwar einerseits die generelle thematische Rollenhierarchie nicht gefährdet, andererseits aber auch dem linguistischen Sonderstatus der SE-Verben gerecht wird. Dabei ist auffällig, dass Aufgabe b) dem sowieso noch ausstehenden Problem ähnelt, eine adäquate Erklärung für die

Klasse der AE-Verben zu finden. Zunächst einmal soll jedoch Aufgabe a) angegangen werden, für die sich der dritte eingangs erwähnte Erklärungsansatz von Dowty (1991) als hilfreich erweist, auf den deshalb auch als nächstes eingegangen werden soll.

Der sog. *Protorollen-Ansatz* von Dowty (1991) ist eine Theorie, die mit sog. generalisierten semantischen Rollen arbeitet (vgl. Kailuweit, 2004). Er bricht mit dem klassischen Konzept der thematischen Rollen und reduziert die Vielzahl der in der Literatur verwendeten Rollen (AGENS, GOAL, THEME etc.) auf lediglich zwei Protorollen, nämlich (Proto-)AGENS und (Proto-)PATIENS. Ob ein Verbargument die Funktion des AGENS oder des PATIENS hat, wird dabei vom Vorliegen verschiedener, auf einer niedrigeren Hierarchieebene angesiedelten Basiseigenschaften abhängig gemacht. In der Version der Theorie von Primus (1999a; 1999b) gibt es folgende Basis-AGENS-Merkmale: a) x *kontrolliert* etwas (CONTROL(x,...)), b) x *verursacht* etwas (CAUSE(x,...)), c) x *bewegt* etwas (MOVE(x, ...)), d) x *erlebt* etwas (EXPER(x,...)) oder e) x *besitzt* etwas (POSSESS(x,...)) (für eine genaue Definition der im Rahmen der Theorie verwendeten Basiseigenschaften vgl. Klein & Kutscher, 2005). Während Dowty (1991) parallel hierzu auch eine Reihe von Basis-PATIENS-Eigenschaften ansetzt (x wird bewegt, wird ins Leben gerufen etc.), geht Primus vereinfachend davon aus, dass eine Entität PATIENS-Eigenschaften hat, wenn sie als Argument eines AGENS-Ausdrucks vorkommt. Wichtig sind jedoch v.a. zwei Dinge: Zunächst wird die Eigenschaft, etwas zu erleben (also letztlich die EXPERIENCER-Rolle), als eine aktive, proto-agentive Eigenschaft aufgefasst. Während der EXPERIENCER bei Grimshaw noch eng mit der PATIENS-Rolle verwandt war, ist er hier Bestandteil der AGENS-Rolle. Der zweite wichtige Aspekt ist, dass AGENS und PATIENS nicht mehr klar und eindeutig voneinander abgrenzbare Rollen sind, sondern probabilistische Clusterbegriffe. Ein Satz-Argument ist nicht *der* AGENS oder *der* PATIENS, sondern kann *mehr oder weniger* viele AGENS- oder PATIENS-Eigenschaften haben.

Welche und wie viele AGENS- bzw. PATIENS-Eigenschaften ein Argument letztendlich hat, ergibt sich aus seiner Semantik. Ähnlich wie bei McCawleys (1968) berühmter Analyse des Wortes *kill* als [CAUSE [BECOME [NOT [ALIVE]]]] wird ein Verb in verschiedene, in Bezug auf die Protorollen-Analyse relevante Einzelaspekte zerlegt (aber keine vollständige Rückführung der Wortbedeutung auf semantische Primitiva versucht, vgl. hierzu z.B. Jackendoff, 1990). So erhält z.B. das (AP-)Verb *töten* die Protorollen-Struktur $\forall x \forall y [\text{töten}'(x,y) \rightarrow \text{CAUSE}(x, \text{BECOME}(\text{DEAD}(y))) \wedge \text{MOVE}(x,y)]$, d.h. es wird angenommen, dass es sowohl eine Verursachungs-Komponente enthält (CAUSE) als auch eine volitionale Bewegungs-Komponente (MOVE). Da das Argument x zwei prototypische AGENS-Eigenschaften hat (CAUSE, MOVE), das Argument y jedoch eine prototypische PATIENS-Eigenschaft (es ist Argument von BECOME),

ist x in dem Satz das Argument mit den maximalen AGENS-Eigenschaften (A_{\max}) und y das Argument mit den maximalen PATIENS-Eigenschaften (P_{\max}). In diesem Fall ist also eindeutig festgelegt, wer AGENS- und wer PATIENS ist. Und da auch Dowty davon ausgeht, dass der PROTO-AGENS syntaktisch in der Subjektposition steht, der PROTO-PATIENS jedoch als Objekt realisiert wird, ergibt sich für diese Verbklasse eine korrekte Vorhersage.

Allerdings ist es durchaus auch möglich, dass die verschiedenen Argumente gleichzeitig AGENS - und PATIENS-Eigenschaften akkumulieren und so eine Konkurrenz um die Protorolle entsteht. Genau dies ist nach Dowty bei den PSYCH-Verben der Fall: Bei ihnen *erlebt* das eine Argument etwas, was eine prototypische AGENS-Eigenschaft ist, und das andere Argument *verursacht* dieses Erleben, was ebenfalls eine prototypische AGENS-Eigenschaft ist. Beide Argumente haben also gleich viele AGENS-Eigenschaften und konkurrieren damit um die Subjektposition. Diese Beobachtung alleine ist schon einmal viel Wert: Sie kann nämlich erklären, warum die Gruppe der PSYCH-Verben eine so große Konstruktionsvielfalt aufweist. Bei ihnen sind die beiden Argumente prinzipiell gleichwertig und kleine Unterschiede bezüglich der restlichen Bedeutung eines Verbs entscheiden darüber, ob es ein ES- oder SE-Verb ist (z.B., welcher der Aktanten die Eigenschaft BELEBT hat, vgl. Härtl, 1999; Wegner, 1999). Um jetzt noch darüber hinaus zu erklären, warum die ES-Verben mit dem EXPERIENCER in der Subjektposition realisiert werden, greift Dowty wieder auf die Beobachtung zurück, dass ES-Verben stativ sind, während SE-Verben häufig eine inchoative Lesart haben, also resultativ im Sinne des oben dargestellten Aktionsart-Systems sind. Ähnlich wie Grimshaw behauptet er, dass SE-Verben nicht einfach nur *Zustände* sind, sondern vielmehr *Prozesse*, d.h. sie implizieren eine Zustandsveränderung auf Seiten des EXPERIENCERS. Und dies verschiebt nun das Rollengewicht entscheidend: Denn Gegenstand einer Veränderung zu sein, ist eine prototypische PATIENS-Eigenschaft und damit ist der EXPERIENCER der einzige Kandidat, der für diese Rolle in Frage kommt. Dies ist formal noch einmal in (8) dargestellt.

- (8) a) $\forall x \forall y [\text{fürchten}'(x,y) \rightarrow \text{CAUSE}(y, \text{EXP}(x, \text{ANGST}'))]$
 b) $\forall x \forall y [\text{ängstigen}'(x,y) \rightarrow \text{CAUSE}(x, (\text{BECOME}(y, \text{EXP}(y, \text{ANGST}'))))]$

In (8a) ist y der STIMULUS des ES-Verbs *fürchten* (der Gefürchtete), x hingegen der EXPERIENCER (der Fürchtende). y verursacht – wie bereits oben ausführlich dargestellt – das Gefühl der Angst in x ; und da y etwas verursacht und x gleichzeitig etwas erlebt, haben beide AGENS-Eigenschaften. In (8b) ist y hingegen der EXPERIENCER des SE-Verbs *ängstigen* (der Geängstigte) und x der ängstigende STIMULUS. Gleichzeitig ist *ängstigen* (nach Dowty) inchoativ zu lesen, d.h. im Gegensatz zu *fürchten* (welches eher einen länger anhaltenden Zustand bezeichnet) ist es als ein perfektiver Vorgang aufzufassen, der den EXPERIENCER von einem

Zustand der Nicht-Angst in einen der Angst überführt. Dies ist formal durch den zusätzlichen BECOME-Operator dargestellt. *y* ist Argument dieses Operators und hat damit eine PATIENS-Eigenschaft mehr.

Das ist ein recht schönes Ergebnis! Die aspektuelle Dimension ist nicht mehr relevant für die Zuweisung der kausativen Bedeutung. Diese wird vielmehr unabhängig durch rein semantische Kriterien festgelegt. Dies führt dazu, dass beide Verbtypen unabhängig von ihrer Ereignisstruktur kausativ interpretierbar sind (d.h. einen CAUSE-Operator involvieren). Darüber hinaus sind – im Gegensatz zu Grimshaws Analyse – die Argumente bei beiden Verbtypen nun auch mit den (aus psychologischer Sicht) „richtigen“ Rollen verbunden: Es ist immer der STIMULUS, der vom CAUSE-Operator gebunden wird, d.h. dieser Ansatz macht die richtige Vorhersage, dass der STIMULUS als das kausal verantwortliche Element angesehen wird. Bei den ES-Verben musste Grimshaw vollkommen unmotiviert hierfür den EXPERIENCER ansetzen. Ebenfalls schön ist, dass die Beziehung zwischen EXPERIENCER und STIMULUS m.E. psychologisch adäquater realisiert ist: Nicht nur, dass klargestellt wird, dass das Konzept der Intentionalität und Agentivität kein einheitliches Konzept ist, sondern sehr viele unterschiedliche Facetten umfasst. Darüber hinaus stellt es in Rechnung, dass die Beziehung zwischen Welt und Mensch reziprok ist, und zwar in dem Sinne, dass das Erleben ein aktiver Konstruktionsprozess ist (dem deswegen vollkommen zurecht AGENS-Eigenschaften zukommen), gleichzeitig aber auch durch die vorliegenden Weltzustände (im konkreten Fall durch den „Aufforderungscharakter“ des STIMULUS) restringiert wird.

Man könnte also eigentlich festhalten, dass das Problem der PSYCH-Verben damit in einem aus psychologischer Perspektive adäquaten Sinne von der Linguistik gelöst wurde und dass die Ansätze beider Disziplinen konvergieren. Allerdings gibt es zwei Gründe, warum der von Dowty skizzierte Protorollen-Ansatz vielleicht noch weiter ausdifferenziert werden sollte.

Zum einen hat insbesondere die Gruppe um Primus (2002, 2004; vgl. zusammenfassend: Klein & Kutscher, 2005) recht umfangreiches Material dafür zusammengetragen, das darauf hinweist, dass eine Analyse der PSYCH-Verben, die zu ihrer Differenzierung auf aspektuelle und kausative Unterschiede zurückgreift, zumindest im Deutschen kaum gelingen kann (was am vergleichsweise arm ausgeprägten Aspekt-System des Deutschen liegen mag). Insbesondere zeigt sich für die deutschen PSYCH-Verben die sowohl für die Argumentation von Grimshaw als auch Dowty zentrale stativ-inchoativ-Diskrepanz nicht oder nur sehr schwach. Nicht nur, dass die Klasse der SE-Verben aspektuell uneinheitlich ist (ungefähr die Hälfte von ihnen passiert übliche inchoativ-Tests wie Progressivbildung, Zeitspannenadverbial; vgl. Klein & Kutscher, 2005), und auch nicht immer eine streng kausative Interpretation haben

(z.B. erlauben nicht alle eine kausale INSTRUMENT-Ergänzung: *Die Theorie beeindruckt/*interessiert mich durch ihre Neuheit.*; man kann höchstens nicht-instrumental sagen: *Die Theorie interessiert mich wegen/aufgrund ihre Neuheit.*, vgl. Klein & Kutscher, 2005, p. 9); Darüber hinaus greift Downtys Prinzip der Kasusselektion überhaupt nicht für die Klasse der SE-Verben, bei denen der EXPERIENCER nicht im Akkusativ, sondern im Dativ steht. Primus et al. ziehen daraus den Schluss, bei ihrer Modellierung (ihnen geht es speziell um das Problem der Kasuszuweisung von PSYCH-Verben) nicht auf ereignisstrukturelle Daten zurückzugreifen, obwohl sie sonst (mit einigen Modifikationen) im Grundprinzip Downtys Ansatz folgen. Stattdessen argumentieren sie aus einer sprachhistorischen Perspektive damit, dass die Kasusselektion der PSYCH-Verben gar nicht von ihrer „psychologischen“ Interpretation (bei der ein EXPERIENCER einen STIMULUS mental repräsentiert) abhängig ist. Vielmehr stellen sie fest, dass die meisten PSYCH-Verben entweder gleichzeitig oder in der sprachhistorischen Vergangenheit eine parallele „physikalische“ Lesart haben, auf die sich die „psychologische“ metaphorisch bezieht oder aus der sie sich entwickelt hat (s. hierzu auch Wegener, 1999). Bei Verben wie *durchdrehen*, *platzen* etc. ist diese physikalische Lesart auch heute noch transparent, bei anderen jedoch nicht mehr (wie z.B. dem Verb *vergessen*, das sich jedoch aus dem volitionalen *loslassen*, *nicht länger festhalten* entwickelt hat). Indem sie sich auf das Prinzip der Lexikonökonomie berufen (Habe für ein Wort nicht mehrere mentale Lexikon-Einträge!) können sie mittels eines optimalitätstheoretischen (OT-)Präferenzalgorithmus (vgl. Kager, 1999), der ähnlich wie die *constraint-based*-Modellen der Sprachpsychologie mehrere unabhängige Kasusverteilungsprinzipien parallel gegeneinander abwägt, mit hoher Erfolgsquote die Kasusverteilung der PSYCH-Verben des Deutschen vorhersagen. Dies deshalb, weil in der physikalischen Lesart das erste Argument meist vermehrt AGENS-Eigenschaften hat. Im Prinzip wird damit kasustheoretisch die Gruppe der ES-Verben zu AP-Verben gemacht.

Ein gutes, aber unappetitliches Beispiel ist hier das Verb *ankotzen*, welches üblicherweise in seiner psychologischen Lesart im Sinne von *nerven*, *anwidern* etc. gebraucht wird. Gleichzeitig hat das Verb unbestreitbar aber auch eine zweite, physikalische Lesart, bei der eine Person über eine andere erbricht. Der Punkt ist, dass es wesentlich leichter ist, eine agentive Lesart für letztere zu finden. Besonders offensichtlich ist dies bei Verben wie *begreifen* und *verstehen* (vgl. Wegener, 1999).

Obwohl sowohl die theoretische Herleitung als auch die optimalitäts-theoretische Implementierung des Ansatzes sehr bewundernswert sind, sind beide Aspekte jedoch für die Modellierung des IC-Effekts problematisch. So wird z.B. das ES-Verb *fühlen* zwar weiterhin in seiner psychologischen Lesart (EXP(x,y) → *x fühlt y*) interpretiert, aber der EXPERIENCER gelangt nur dadurch in die richtige Satzposition, dass man die wesentlich komplexere physi-

kalische SE-Lesart (z.B. *Puls fühlen*: CTRL(x,s) \wedge PHYS(x) \wedge PHYS(x,y) \wedge EXP(x,y)) zugrunde legt. Ein weiterer kritischer Aspekt ist, dass viele für eine Analyse der PSYCH-Verben wichtige Punkte, von diesem Ansatz zunächst nicht mehr erfasst werden. Ohne Zusatzannahmen können Primus et al. nicht mehr erklären, warum die meisten SE-Verben keinen verbalen Passiv bilden oder sich in Bezug auf die Bindungstheorie atypisch verhalten (s.o.). Um solche und ähnliche Phänomene zu erklären, müssen Primus et al. (z.B. Primus, 2004) auf ein Äquivalent der in (5) spezifizierten Argument-Hierarchie zurückgreifen: Nur bei ES-Verben stimmen Kasus-Vergabe (der EXPERIENCER steht im Nominativ) und Rollen-Hierarchie (EXPERIENCER > STIMULUS) überein, bei den SE-Verben konfliktieren beide Prinzipien jedoch miteinander. Durch diese Diskrepanz wird dann ähnlich wie bei Grimshaw das atypische Verhalten der SE-Verben erklärt.

Es ist interessant, dass Croft (1991, 1993) eine alternative Erklärung für das divergente Verhalten der PSYCH-Verben vorgeschlagen hat, welche besser mit den im letzten Abschnitt skizzierten sozialpsychologischen Theorien kompatibel ist. Croft (1993) kombiniert in seiner Analyse der PSYCH-Verben gewissermaßen Überlegungen von Grimshaw und Dowty, obwohl er auf beide gar nicht direkt eingeht (in der Tat entwickelte er den Ansatz bereits vor ihnen). Ausgangspunkt ist wiederum die Beobachtung, dass a) gerade die PSYCH-Verben eine ausgesprochene cross-linguistische, aber auch intersprachliche Variabilität aufweisen und b) die SE-Verben zu einer kausativen Lesart tendieren (wofür Croft als vergleichender Sprachwissenschaftler u.a. Evidenz aus Sprachen anführt, die – wie die Atztekensprachen Lakhota und Nahtual – explizite, morphologisch realisierte Kausalaffixe besitzen). Ähnlich wie Grimshaw stellt er nun einen Bezug zur Ereignisstruktur der mit den Verben beschriebenen Ereignisse her, begründet diese allerdings nicht (allein) über die Aktionsart der Verben, sondern – ähnlich wie Dowty – über die thematischen Rollen. Speziell werden die verschiedenen thematischen Rollen über ihre Position in der Ereignisstruktur (*causal-aspectual structure*) überhaupt erst definiert. Mit Talmy (1976) unterscheidet er vier unterschiedliche Kausalrelationen: a) *volitional causation* (ein Agent beeinflusst aktiv und intentional ein physikalisches Objekt), b) *physical causation* (ein physikalisches Objekt affektiert ein anderes physikalisches Objekt), c) *inductive causation* (eine Person beeinflusst eine andere Person) und d) *affective causation* (ein physikalisches Objekt beeinflusst eine Person). Dadurch wird es möglich, auch komplexe Kausalgefüge gewissermaßen in kleinere kausale Einheiten zu zerlegen. So dekomponiert er z.B. den Satz *Peter öffnete die Kokosnuss für Maria mit dem Hammer* in: *Peter* (vol) \rightarrow *Hammer* (phys) \rightarrow *Kokosnuss* \rightarrow (become) \rightarrow *offen* \rightarrow (aff) *Maria*. Peter führt also einen Akt der volitionalen Verursachung auf den Hammer aus, der (physikalisch) die Kokosnuss dazu

bringt, sich zu öffnen, was wiederum Maria (affektiv) beeinflusst (man beachte, dass die lineare Abfolge im Satz und die logische Abfolge in der Kette divergieren). Ereignisse bestehen nach seiner Auffassung also immer aus komplexen Kausalketten und Verben können einen bestimmten Ausschnitt aus einer Kausalkette spezifizieren. Die thematische Rolle (und auch die Kasuszuweisung) für eine Entität ergibt sich aus der relativen Stellung innerhalb dieser Kette: Die erste Position, die von einem Verb obligat verlangt wird, wird das Subjekt, während die letzte Position ein Objekt wird.

Da dieses System offensichtlich zunächst einmal nur für asymmetrische Verben funktioniert, muss sich Croft für die symmetrischen (wie *küssen*, (*miteinander*) *tanzen* etc.) etwas einfallen lassen. Er nimmt an, dass auch sie innerhalb des dominanten Kausalschemas interpretiert werden müssen. Da sie jedoch nicht eindeutig gerichtet sind, müssen sie als sich wechselseitig beeinflussend interpretiert werden. Wie bei allen symmetrischen, akausalen Strukturen, lässt sich jedoch auch hier eine Figur-Grund-Unterscheidung vornehmen, d.h. die Entität, der gerade mehr Aufmerksamkeit geschenkt wird, wird in den kausalen Vordergrund geschoben, während die andere im Hintergrund bleibt.

In Bezug auf die Klasse der PSYCH-Verben nimmt er nun an, dass die kausativen bzw. inchoativen Verben bzw. auch alle Verben, die sich auf mentale Aktivitäten beziehen ((*darüber*) *nachdenken* etc.), sich sequenziell als Prozesse interpretieren lassen, in denen ein STIMULUS kausal auf einen EXPERIENCER einwirkt. Für die Klasse der stativen mentalen Zustandsverben setzt er hingegen eine asymmetrische Beziehung an. Das liegt daran, dass er ihre Beziehung – ähnlich wie Dowty – als zwei gegenläufige, aber gleichzeitig ablaufende Prozesse ansieht: Der STIMULUS wirkt auf den EXPERIENCER ein, aber der EXPERIENCER „erlebt“ gleichzeitig den STIMULUS (konkret arbeitet Croft mit dem kognitiven Prozess der Aufmerksamkeitszuwendung). STIMULUS und EXPERIENCER sind also gleichzeitig Anfangs- und Endpunkt der kausalen Kette. Genau aus diesem Grund sagt Croft für diese Verbgruppe auch die meisten Variationsunterschiede voraus. Weiterhin können unterschiedliche Verben innerhalb des symmetrischen Gefüges verschiedene Figur-Grund-Gliederungen vornehmen. Dies wirkt sich wiederum auf die Interpretation aus: Wenn ein Element als Anfangspunkt der Kette interpretiert wird, dann erhält es ein stärkeres kausales Gewicht. Und in der Tat gibt es ja leichte Bedeutungsunterschiede in solchen Minimalpaaren wie *ängstigen* vs. *fürchten*. Beide Verben haben einen ziemlich ähnlichen semantischen Gehalt, aber nehmen eine unterschiedliche Figur-Grund-Gliederung vor: Bei *ängstigen* wird auf den STIMULUS fokussiert, bei *fürchten* auf den EXPERIENCER.

Man bemerke zunächst einmal, dass sich hier alle relevanten Argumentationsfiguren wieder finden lassen, die sowohl in der diskurspsychologischen Forschung im Allgemeinen als auch in der Forschung zur pronominalen Auflösung im Speziellen thematisiert werden. Zum einen die kausal-basierte Analyse der Konstellation verschiedener Aktanten, zum anderen aber auch die aufmerksamkeits-basierte Hervorhebung bestimmter Aktanten und Interpunktion von symmetrischen Beziehungsmustern. Aber auch intratheoretisch ist der Ansatz von Croft anschlussfähig zu den anderen bis skizzierten linguistischen Theorien, obwohl er sich gleichzeitig auch von ihnen abgrenzt. Zum einen nimmt er explizit eine ereignisstrukturelle Repräsentation an, in der die verschiedenen mit einem Verb bzw. Satzsegment bezeichneten Teilelemente miteinander verbunden sind. Anders als z.B. bei Grimshaw wird jedoch kein prinzipieller Unterschied zwischen verschiedenen Verbtypen gemacht (Zustände = ein Unterereignis; accomplishments = zwei kausal verknüpfte Unterereignisse etc.). Vielmehr bestehen alle mit PSYCH-Verben bezeichneten Ereignisse aus (mindestens) zwei Subereignissen, die kausal miteinander verknüpft sind. Im Prinzip sind beide damit vom Aktionsart-Typ „Prozess“, da auch die (symmetrischen) Zustände als reziproke Prozessrelationen interpretiert werden. Dies wird dadurch ermöglicht, dass ähnlich wie bei Dowty ein einheitlicher Verursachungsbegriff aufgegeben wird. „Verursachung“ ist für Croft ein Komplexbegriff, der viele Teilelemente enthalten kann. Entsprechend können gleichzeitig auch mehrere Kausalrelationen aktiv sein, bei den Zustandsverben sogar in unterschiedliche Richtungen. Bei den kausativen Inchoativa (die ungefähr mit den SE-Verben gleichzusetzen sind) gibt es nach Croft lediglich eine Kausalrelation vom STIMULUS zum EXPERIENCER. Für die nicht-kausativen Stativa (die ungefähr den ES-Verben entsprechen) gibt es zwei Relationen, eine kausale vom STIMULUS zum EXPERIENCER und eine attentionale vom EXPERIENCER zum STIMULUS. In jedem Fall gibt es also eine kausale Verursachungsrelation des STIMULUS, aber auch der EXPERIENCER kann am Anfangspunkt der Kausalkette stehen. Abb. 2.6 fasst dies noch einmal zusammen:

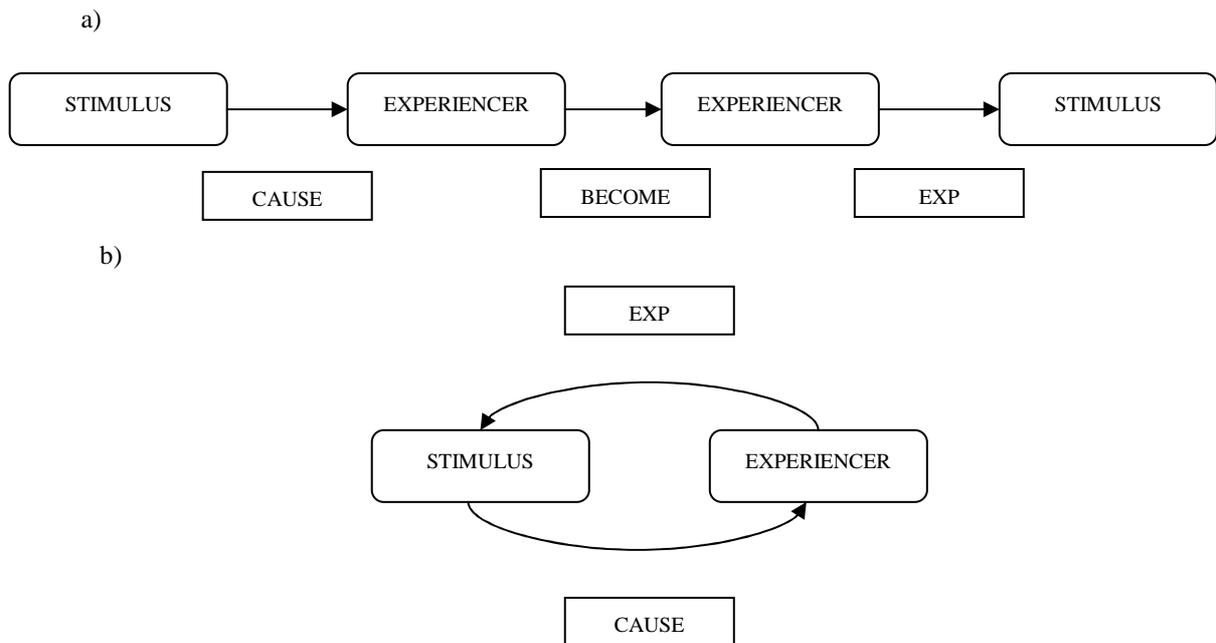


Abb. 2.6: Ereignisstruktur von a) mentalen Prozess- und b) mentalen Zustandsverben nach Croft (1993).

In der Notation von Dowty bzw. Primus entspricht Abb. a) exakt dem Ausdruck für die inchoativen SE-Verben ($\exists x \exists y [\text{CAUSE}(x, (\text{BECOME}(y, \text{EXP}(y,x))))]$), während Abb. b) ungefähr dem Ausdruck $\exists x \exists y [\text{CAUSE}(x,y) \wedge \text{EXP}(y,x)]$ entsprechen würde. Das heißt, ein SE-Verb wie *gefallen* würde ungefähr analysiert als ‚Es gibt ein x, das verursacht, das y in den Zustand gerät, dass y x mag‘, während ein ES-Verb wie *mögen* analysiert würde als ‚Es gibt ein x, für das gilt, das es y affektiv beeinflusst und y beachtet es‘.

Der skizzierte Ansatz scheint also zunächst einmal alle erforderlichen Anforderungen zu erfüllen, die für eine psychologisch adäquate Erklärung der PSYCH-Verben notwendig sind. Dazu gehört v.a., die kausale Struktur so auszurichten, dass das kausale Gewicht auf dem STIMULUS liegt. Als zentrales Erklärungsprinzip ist dabei in Anschluss an Croft (1993) die rezipientenseitige Rekonstruktion einer zugrunde liegenden Ereignissequenz anzusetzen, die von verschiedenen Verben unterschiedlich interpunktiert wird (vgl. Abb. 2.7).

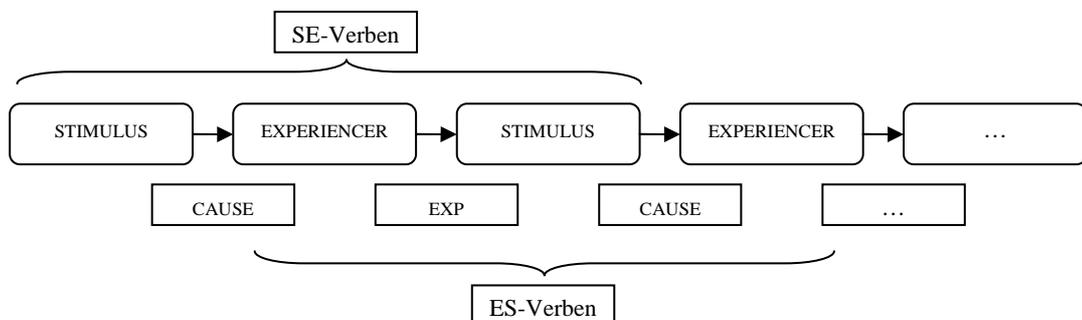


Abb. 2.7: Unterschiedliche Interpunktion der zugrunde liegenden Ereignisstruktur durch SE- und ES-Verben.

Ein weiterer Vorteil des Ansatzes besteht darin, dass er eine direkte Anschlussmöglichkeit zu den im letzten Abschnitt erläuterten sozialpsychologischen Erklärungsansätzen bietet und damit auch eine direkte Einbeziehung der – von der linguistischen Literatur nicht behandelten! – Klasse der AE-Verben erlaubt. Denn deren Erklärung wurde bis jetzt ja immer vertagt, und tatsächlich setzt ihre Existenz zunächst einmal die meisten linguistischen Theorien unter Druck: Wie kann es sein, dass dort ein Aktant klar agentivisch als Satzsubjekt ausgedrückt wird, dieser aber dennoch nicht als kausal verantwortlich angesehen wird?

Der einzige halbwegs gescheite Erklärungsansatz, der zur Erklärung der IC bei AE-Verben vorgebracht wurde, stammt m.E. von Crinean und Garnham (2006). Dass der Ansatz eigentlich in Bezug auf eine vereinheitlichende Erklärung sowohl des Impliziten Kausalitäts- als auch des Impliziten Konsequentialitätseffekts (s. Abschnitt 2.1) formuliert wurde, ist dabei durchaus charakteristisch. Crinean und Garnham nehmen explizit Bezug zum Ereignisstruktur-Modell von Moens und Steedman (1988), das Ereignisse eingebettet sieht in eine kausale Erklärungskette, die drei Elemente umfasst: a) initiale Prozesse, die dem Ereignis vorangehen, b) das Ereignis selbst und c) Konsequenzen, die dem Ereignis nachfolgen. In Anschluss an Stevenson et al. (1994) nehmen sie an, dass bestimmte thematische Rollen mit bestimmten Teilereignissen assoziiert sind. Bei Transfer-Verben (mit einer SOURCE-GOAL-Struktur) ist die thematische Rolle SOURCE mit dem initialen Teilereignis verbunden, die GOAL-Rolle jedoch mit dem finalen.

Eine solche dynamische Ereignisstruktur, die bestimmte thematische Rollen mit der zeitlich und kausal geordneten Ablauflogik bestimmter Teilereignisse korreliert, kann erfolgreich für den Impliziten Kausalitäts-/Konsequenzialitäts-Kontrast der PSYCH-Verben herangezogen werden: Der STIMULUS ist quasi *per definitionem* der Aktant, der einen Zustand im EXPERIENCER hervorruft, entsprechend ist er mit dem ersten Teilereignis assoziiert und damit kausal vorgeordnet. Entsprechend weisen psychische Zustandsverben einen IC-Bias in Richtung des STIMULUS auf. Der EXPERIENCER ist hingegen *per definitionem* die Partei, welche die Konsequenzen des psychischen Zustands erfährt, deswegen haben psychische Zustandsverben einen Impliziten Konsequenzialitäts-Bias in Richtung auf den EXPERIENCER. (Man bemerke übrigens, dass sowohl Stevenson et al. als auch Crinean und Garnham keine Aussagen dazu machen, warum bei bestimmten Verben die beiden Rollen oberflächenstrukturell in unterschiedlichen Positionen realisiert werden; wie für sprachpsychologische Arbeiten typisch, setzen sie dies schlichtweg voraus und halten irgendwelche lexikalischen Eigenschaften der Verben hierfür verantwortlich).

Für die Klasse Handlungsverben gibt es jetzt allerdings ein Problem. Diese sollte mit der gleichen Logik wie bei den Zustandsverben immer einen IC-Bias zum AGENS aufweisen (der *per definitionem* einen Zustand beim PATIENS verursacht) und einen Impliziten Konsequenzialitäts-Bias in Richtung auf den PATIENS. Für die Klasse der AP-Verben stimmt dies ja auch, allerdings nicht für die Klasse der AE-Verben. Diese weisen empirisch einen IC-Bias in Richtung des PATIENS auf und einen Konsequenzialitäts-Bias ebenfalls in Richtung des PATIENS (vgl. die Reanalyse der IC-Verben bei Crinean & Garnham, 2006). Das heißt, sie tanzen gleich zweimal aus der Reihe: Weder zeigen sie das typische CAUSE(AGENS)-Muster, noch gibt es bei ihnen eine komplementäre Beziehung zwischen (impliziter) Kausalität und Konsequenzialität.

Crinean und Garnham führen dieses anormale Verhalten der AE-Verben nun darauf zurück, dass diese semantisch eine verbale Mischklasse sind, die sowohl psychologische Zustandsanteile als auch Elemente von Handlungsverben hat. Speziell behaupten sie, dass der AGENS dieser Verben in der Tat auf eine (präsupponierte) Handlung des PATIENS reagiert, weshalb er auch EXPERIENCER-Anteile hat. Das heißt, der PATIENS ist gleichzeitig ein STIMULUS und der AGENS ist gleichzeitig ein EXPERIENCER. Unter der Voraussetzung, dass die Kausalzuschreibung vom Zustands- und die Konsequenzialitätszuschreibung vom Handlungselement des Verbs gesteuert wird, kommen sie zum gewünschten Ergebnis: In Bezug auf die Frage, warum die Handlung (z.B. Peter beschuldigt Hans.) zustande gekommen ist, ist der PATIENS als STIMULUS für die Handlung des AGENS verantwortlich (Peter denkt, dass Hans ihn bestohlen hat), aber da es sich ja um eine *Re-Aktion* handelt, hat die Handlung des AGENS wahrscheinlich direkte Konsequenzen für den PATIENS (z.B. muss Hans sich anschließend rechtfertigen).

Diese Analyse hat entscheidende Vorteile, insbesondere weil sie ebenfalls auf einer ereignisstrukturellen Analyse basiert und konsequent den STIMULUS als Verursacher ansetzt. Aufgrund dieser konzeptuellen Gemeinsamkeiten weist die Analyse der AE-Verben durch Crinean und Garnham auch eine offensichtliche Parallele zur oben vorgestellten Analyse der ES-Verben auf. Darüber hinaus könnte sie erklären, warum bei den AE-Verben der AGENS als Subjekt realisiert wird: In der Terminologie von Dowty ist er nicht nur Verursacher einer Handlung (CAUSE), sondern (implizit) auch EXPERIENCER eines Zustands (EXP). Da ihm damit mehr proto-agentivische Eigenschaften zukommen, wird er als A_{\max} in der Subjektposition realisiert.

Allerdings ist der Vorschlag von Crinean und Garnham auch nicht unproblematisch. Das liegt m.E. v.a. daran, dass AE-Verben doch zunächst einmal Handlungsverben sind, d.h.

semantisch *Handlungen* und aspektuell *Aktivitäten*; im Gegensatz dazu sind ES-Verben semantisch und aspektuell (psychische) *Zustände*. Dafür sprechen alle einschlägigen Testverfahren, insbesondere Verlaufsformtests und die Kombination mit Zeitrahmen-Adverbialen (s.o.). So können z.B. AE-Verben im Gegensatz zu ES-Verben meist nicht mit einem Zeitspannenadverbial stehen (*Peter vergaß/*beglückwünschte Maria innerhalb einer Woche.*), ES-Verben jedoch im Gegensatz zu AE-Verben normalerweise nicht mit einer prospektiven Referenzzeit-Adverbial (*In einer Stunde wird Peter Maria *lieben/beglückwünschen.*).

Wenn somit auf der einen Seite schon Zweifel daran bestehen, dass dem AGENS der AE-Verben eine EXPERIENCER-Funktion zukommt, so ist es auf der anderen Seite auch problematisch anzunehmen, dass der PATIENS eine STIMULUS-Funktion hat. Zwar halte ich es für sehr wahrscheinlich, dass es Mischformen von psychischen Zustands- und Handlungsprädikaten gibt. In der Tat deuten viele AE-Verben wie *verhöhnern, entschuldigen, belächeln* etc. darauf hin, dass sie einen emotional-mentalen Zustand beim AGENS präsupponieren. Allerdings denke ich ebenfalls, dass dieser emotionale Zustand bei den AE-Verben nicht hinreichend ist, um die Handlung zu lizenzieren. AE-Verben wie *kritisieren, loben, helfen* etc. implizieren nicht nur eine Bewertung des Gegenübers, sondern das Gegenüber muss auch etwas *getan* haben, was diese Reaktion rechtfertigt. Für den STIMULUS eines ES-Verbs ist es hingegen vollkommen ausreichend, dass er einfach nur existiert. So kann ich jemanden *lieben*, nur weil er da ist und gerade auch dann, wenn er *nichts* tut. Das alleine rechtfertigt jedoch noch nicht, dass ich ihn *lobe* oder ihm *danke*, hierfür muss er eine Handlung vollzogen haben.

Wichtig ist jedoch allein, dass m.E. die Interpretation der Handlungsverben ziemlich ähnlich wie für die Zustandsverben vorgenommen werden kann: Das heißt, sie sind beide in eine implizite Handlungssequenz eingebettet, und die beiden Verbtypen interpunktieren diese Kette jeweils unterschiedlich: *AP-Verben* setzen den AGENS als initialen Punkt der Kette an, der klar und eindeutig einen PATIENS affiziert. Die Ereigniskette ist ansonsten vergleichsweise unterspezifiziert, lässt sich aber leicht inferenziell erweitern. So könnte man z.B. erwarten, dass der PATIENS im nächsten *turn* wiederum auf den AGENS reagiert etc. Bei den *AE-Verben* beginnt die Verursachungskette mit einer (impliziten) Handlung des PATIENS und die nachfolgende manifeste Handlung des AGENS ist eine Reaktion hierauf. Unklar ist allerdings, welche Relation zwischen PATIENS und AGENS genau besteht. Eine reine EXPERIENCER-Relation wie bei den Zustandsverben scheint mir nicht angemessen zu sein (s.o.). Evtl. kann hier auf die Unterscheidung von Talmy (1976) zwischen *volitional* und *inductive causation* zurückgegriffen werden (s.o.):

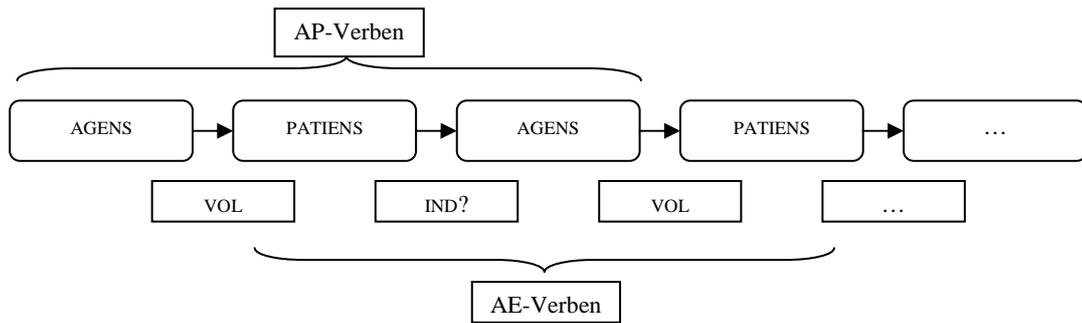


Abb. 2.8: Unterschiedliche Interpunktionsstruktur der zugrunde liegenden Ereignisstruktur durch AP- und AE-Verben.

Ein Vorteil dieser parallelen Strukturierung der Zustands- und der Handlungsverben ist, dass sich aus ihr die gängigen impliziten Kausalitäts- und Konsequenzalitäts-Urteile für die verschiedenen Verbtypen herleiten lassen: Wenn man eine Person nach den kausalen *Ursachen* einer Handlung fragt, dann wird die gerade existente Ereignisstruktur „abgesucht“ und das Element genannt, welches am Anfang einer Verursachungsrelation steht. Eine Frage nach den *Konsequenzen* einer Handlung führt hingegen zu einer Suche nach dem letzten Element der Kausalkette.

Wohlgermerkt erhebt diese nur grob skizzierte Strukturierung jedoch nicht den Anspruch, eine adäquate Erklärung für alle relevanten linguistischen Sachverhalte zu liefern. Ganz im Gegenteil ist eher davon auszugehen, dass dies sicherlich *nicht* kann, u.a. weil bei dem gewählten Auflösungslevel (z.B. ohne weitere Dekomposition der verschiedenen Protollen-Dimensionen) nicht mehr klar ist, wie die Oberflächenstruktur des Satzes etabliert wird. Die vorgestellte Rekonstruktion des IC-Effekts als Interpunktionsstruktur einer impliziten Ereignisstruktur soll vielmehr nur einen bestimmten heuristischen Wert haben und durch ihre (prinzipielle, nicht Detail-)Struktur in der Lage sein, verschiedene linguistische Strukturebenen miteinander zu verbinden. Damit kann sie als konzeptuelles Gerüst zum Verständnis des Phänomens der IC dienen und zur Strukturierung der nachfolgenden Ausführungen eingesetzt werden (vgl. Abschnitt 2.1.4). Gleichzeitig ist sie anschlussfähig zur psycholinguistischen Literatur zur Verarbeitung thematischer Verbinformationen (vgl. z.B. Ferretti, McRae & Hatherall, 2001; Altmann & Kamide, 1999; Knoeferle & Crocker, in press)

2.1.3 Sprachpsychologie

IC ist sprachpsychologisch gesehen ein vergleichsweise gut erforschtes Phänomen. Seit der initialen Studie von Caramazza und Garvey (1974) wurden bisher ca. 25 psycholinguistische Studien im engeren Sinne veröffentlicht, die sich diesem Thema in Bezug auf die pronominale Auflösung hauptsächlich oder randständig widmen. Dabei wurden nur solche Studien

berücksichtigt, welche online-Maße erhoben haben und sich mit dem konkreten Verstehensprozess beschäftigen. Hinzu kommt ungefähr noch einmal die gleiche Menge von Untersuchungen, welche die IC nicht unbedingt aus einer primär sprachpsychologischen Perspektive betrachten oder lediglich offline-Maße zum Verstehensprodukt erheben, worunter ich z.B. alle in Abschnitt 2.1.1 besprochenen sozialpsychologischen Arbeiten zähle (wovon ca. 23 in der Metaanalyse von Rudolph & Försterling, 1997a, besprochen werden). Dabei sind sogar nur solche Veröffentlichungen gezählt, die in gut zugänglichen und weit bekannten sprach- und kognitionspsychologischen Zeitschriften veröffentlicht wurden. Wenn man auf die kaum überschaubare Ebene der Forschungsberichte, Qualifikationsarbeiten und der lokal bzw. unveröffentlichten Studien geht, dann erhöht sich diese Zahl sicherlich beträchtlich (die Suchabfrage „implicit causality“ bei der Datenbank PSYCH-INFO erbringt ca. 50 Treffer, eine Schlagwortsuche bei gängigen wissenschaftsorientierten Internet-Suchmaschinen wie *Google Scholar* ca. 300 Treffer).

Überraschend ist dabei v.a. die Kontinuität, mit welcher dieses Thema behandelt wird. Die sprachpsychologische Forschung ist ja sonst nicht unanfällig für bestimmte Moden, in welchen bestimmte Phänomene schlagartig „in“ sind, aber ebenso plötzlich wieder aus der Diskussion verschwinden. Für die IC scheint hingegen ein vergleichsweise konstantes Interesse vorzuliegen, was v.a. daran liegt, dass das Phänomen der IC für alle zentralen theoretischen Grundfragestellungen des psycholinguistischen Diskurses der letzten Jahrzehnte (seriell vs. parallel, gedächtnis- vs. erklärungsbasiert etc.) direkt relevant ist. Diese lange zeitliche Erstreckung von über 30 Jahren Forschung und die mit ihr verbundenen, teilweise dramatischen Veränderung der zugrunde gelegten Begrifflichkeiten und theoretischen Vorstellungen erschweren den Vergleich der verschiedenen Studien erheblich. Das gleiche gilt in Bezug auf die Heterogenität der eingesetzten Methoden, die sehr unterschiedliche Aufgabentypen, abhängige Variablen und technische Erhebungsmöglichkeiten umfassen – die Spannbreite reicht dabei vom Tachistoskop bei Caramazza et al. (1977) bis hin zu modernen hoch-technisierten neuropsychologischen Studien wie bei Nieuwland und van Berkum (2006).

Die entscheidende, die Diskussion um den IC-Effekt geradezu beherrschende Frage war, ob es sich bei ihr um einen frühen, material-getriebenen *bottom-up*-Effekt handelt oder vielmehr um einen späten, kohärenz-basierten *top-down*-Effekt. Dabei ist auffällig, dass sich alle bisher vorgelegten Studien in ungefähr zwei gleich große Lager einteilen lassen: Während die einen ihre theoretischen Überlegungen v.a. darauf ausrichten, wie die IC eines Verbs automatisch mit frühen syntaktischen und allgemeinen Sprachverarbeitungsprozessen interagiert (und entsprechend ihre Methoden dieser Perspektive anpassen), zielt das andere Lager eher

darauf ab, den Zusammenhang der IC zu übergreifenden diskursiven Phänomenen wie Referenzetablierung, Plausibilitätsabfrage und Kohärenzbildung herauszuarbeiten. Im Folgenden sollen deswegen (in den Abschnitten 2.1.3.2 und 2.1.3.3) die zentralen Ergebnisse dieser beiden theoretischen und empirischen Strömungen getrennt, aber kontrastiv gegeneinander gestellt werden. Zuvor (Abschnitt 2.1.3.1) sollen jedoch noch solche Arbeiten und Ergebnisse diskutiert werden, die einen vergleichsweise übergreifenden Charakter haben und helfen, den Geltungsbereich des Phänomens besser einzugrenzen.

2.1.3.1 Übergreifende Ergebnisse zum IC-Effekt

Der bisher dargestellte „allgemeine“ IC-Effekt, der insbesondere in der sozialpsychologischen und linguistischen Diskussion im Vordergrund steht, bezieht sich auf die Tatsache, dass bei bestimmten Verben einem der Argumente eine größere kausale Verantwortung zugeschrieben wird. Im Gegensatz hierzu geht es bei dem „speziellen“, sprachpsychologischen IC-Effekt ganz dezidiert um die pronominale Auflösung. In der Tat ist auffällig, dass das doch eigentlich sehr allgemein konzeptualisierte Merkmal der IC psycholinguistisch fast *nur* im Kontext der pronominalen Auflösungsforschung untersucht wird (bis auf ein paar, sehr wenige Ausnahmen, vgl. z.B. Burgess et al., 1997).

Der pronominale IC-Effekt, um den es im Folgenden ausschließlich gehen wird, bezieht sich auf die Tatsache, dass ein Pronomen, welches in einen kausalen Nebensatz eingebettet ist, leichter aufzulösen ist, wenn die im Nebensatz genannte *explizite* Begründung mit der *impliziten* Kausalitätsrichtung des Verbs kongruent ist (vgl. schon Caramazza et al., 1977). Das heißt, prototypisch ist der Kontrast wie in (9) (s.o.):

- (9) a. Peter betreute Hans, weil er keine Verwandten mehr hatte. (IC-kongruent).
b. Peter betreute Hans, weil er der einzige Verwandte war. (IC-inkongruent).

Es kann wirklich gar nicht ausdrücklich genug darauf hingewiesen werden, dass dies *nicht* der bisher besprochene allgemeine IC-Effekt ist, sondern eigentlich nur ein Reflex von diesem. Denn was hier variiert wird, ist nicht die IC des Verbs selbst (welche vielmehr vorausgesetzt und konstant gehalten wird), sondern die *Kongruenz* der expliziten mit der impliziten Begründung. Wollte man den IC-Effekt *direkt* erfassen, dann müsste man vielmehr die explizite Begründung konstant halten und die implizite Kausalität variieren, die sich dann auf das Verstehensprodukt auswirken sollte.

Egal auf welche Art manipuliert wird: Das charakteristische Ergebnis des Vergleichs der Sätze a) und b) in Paaren wie in (9) ist, dass Satz a) wesentlich leichter zu verarbeiten ist, d.h. er wird schneller und akkurater verarbeitet. Dieser Befund lässt sich über eine erstaunli-

che Vielfalt von sprach- und kognitionspsychologischen Aufgabentypen und Abhängige Variablen generalisieren: Personen können in solchen Sätzen den Referenten schneller benennen (sowohl in *naming*- als auch in *button-release*-Aufgaben; Caramazza et al., 1977; Ehrlich, 1980), sie lesen die entsprechenden Sätze schneller (egal, ob die Lesezeit selbst-gesteuert auf der Satz-, Teilsatz-, Phrasen- oder Wort-Ebene erfasst wird, vgl. Stewart et al., 2000; Koorneef & van Berkum, 2006, oder mittels *eye-tracking*; vgl. z.B. Vonk 1984; 1985a), sie können schneller seine Akzeptabilität bzw. Grammatikalität beurteilen (Garnham et al., 1992) und sie können schneller ic-kongruente Fortsetzungen formulieren etc. (Rigalleau et al., 2007). Darüber hinaus ist der ic-kompatible Referent bei *probe-recognition*-Aufgaben im Vergleich zum Nicht-Referenten aktiver und in neuropsychologischen Studien, die mit ereigniskorrelierten Potentialen arbeiten, rufen ic-inkompatible Referenten stabil einen P600-Effekt hervor, der eine syntaktische Anomalie indizieren soll (van Berkum, Koorneef, Otten & Niewland, 2007). Aber auch in offline-Maßen lässt sich der Effekt nachweisen, und zwar sowohl durch Verifikations- und Rekognitionsaufgaben als auch durch inhaltliche Verständnisfragen. Allerdings nehmen solche offline-Methoden in der sprachpsychologischen Forschung zum IC-Effekt eine eher untergeordnete Stellung ein. Denn charakteristisch für die sprachpsychologische Forschung ist ihre Ausrichtung auf die Prozessperspektive der Sprachrezeption. Dies erklärt, warum fast ausschließlich *zeitbezogene* AVn eingesetzt werden, wie Lesezeiten, *eye-tracking*-Daten oder ereigniskorrelierte Potenziale. Die sozialpsychologische Forschung konzentriert sich hingegen eher auf die Seite des Verstehensproduktes und die zum Verständnis eines Satzes notwendigen Informationsbestände.

Der IC-Effekt ist aber auch sehr robust in Bezug auf verschiedene *Text- und Sprachmerkmale*. Er findet sich in so unterschiedlichen Kontexten wie einfachen und doppelten Nebensatz-Konstruktionen (Vonk, 1985b; McKoon et al., 1993), aber auch über Satzgrenzen hinweg (Hudson, Tanenhaus & Dell, 1986). Er funktioniert bei transitiven und bitransitiven Verben (Garnham & Oakhill, 1985; intransitive scheiden wegen der fehlenden Ambiguität des Pronomens aus) und sowohl wenn die Sätze isoliert dargeboten werden als auch wenn sie in einen diskursiven Kontext eingebettet sind (der von 3-Satz-Kontexten bis hin zu kleinen Geschichten reichen kann; vgl. Garnham et al., 1992). Auch ließ sich der Effekt bis jetzt in so unterschiedlichen Sprachen nachweisen wie dem (germanischen) Englisch, Niederländisch und Deutsch, dem (romanischen) Französisch, Spanisch und Italienisch, aber z.B. auch im Persischen. In Bezug auf die *rezipientenseitigen Lesermerkmale* findet er sich sowohl bei Erwachsenen als auch bei Kindern (Au, 1986; Corrigan & Stevenson, 1994), und selbst bei verschiedenen neurologischen Patientengruppen mit schwersten Sprachstörungen (vgl. Leonard,

Waters & Caplan, 1997). Auch ist er weitgehend unabhängig von der jeweils verwendeten Untersuchungsinstruktion (die von normalen Lesen bis zu sehr spezifischen Aufgaben reichen kann; vgl. im Überblick Garnham et al., 1992).

Kurz: Der Effekt ist stabil und findet sich fast immer. Allerdings lassen sich auch charakteristische Bedingungen angeben, in denen er sich *nicht* finden lässt. Eine zentrale Rolle spielte hier in der Diskussion sicherlich zunächst einmal die Frage, ob der Effekt auch in Anwesenheit anderer pronominaler Auflösungs-Cues wirksam ist, v.a. wenn das Pronomen gleichzeitig durch Gender-Informationen disambiguiert wird. Wie bereits die Ergebnisse von Caramazza et al. gezeigt haben (und die nachfolgende Diskussion noch einmal bestätigen wird), ist dies meist der Fall (s. Abschnitt 2.3). Eine Kontextvariable, die jedoch für den IC-Effekt ganz entscheidend ist, ist die Art der *Konjunktion*, welche die beiden Nebensätze miteinander verbindet. So konnten u.a. Ehrlich (1980), aber auch Vonk (1984; 1985a) nachweisen, dass der Effekt wirklich nur bei der kausalen Konjunktion *weil* auftritt, bei adversativen Konjunktionen wie *aber*, *obwohl* etc. jedoch nicht, ebenso nicht bei nebenordnenden Konjunktionen wie *und* bzw. *oder* (vgl. auch McKoon et al., 1993). Wie in den letzten Abschnitten ausführlich besprochen dreht sich der Effekt bei prospektiven Konjunktionen wie *deshalb*, *so* etc. sogar um und wird zur impliziten Konsequenzialität (Stewart et al., 1998; Crinean & Garnham, 2006). Eine weitere wichtige Eigenschaft des IC-Effekts ist, dass er wirklich nur mit *Pronomen* funktioniert, nicht jedoch mit anderen Arten von Anaphern wie Eigennamen oder definiten NPn (vgl. u.a. Hudson et al., 1986; Garnham & Oakhill, 1985; vgl. jedoch Stewart et al., 2000).

Darüber hinaus ist bemerkenswert, dass der IC-Effekt bei den vier theoretisch relevanten Verbgruppen (AP-, AE-, SE- und ES-Verben) nicht gleich stark ausfällt. Obwohl er sich prinzipiell bei allen nachweisen lässt, herrscht eigentlich Einigkeit darüber, dass er für die Klasse der (mentalen) Zustandsverben sehr viel stärker ausfällt als für die (behavioralen) Handlungsverben. Dies lässt sich durch den skizzierten thematischen Sonderstatus dieser Verbgruppe (mit ihrer prinzipiellen Ausgeglichenheit in der Protorollen-Zuweisung) erklären. Ungeklärt ist hingegen die Frage, ob der Effekt in Bezug auf die *Kausalitätsrichtung* (NP1: AP, SE vs. NP2: AE, ES) unterschiedlich ausfällt. Einige Studien berichten hier einen leichten Verarbeitungsvorteil hin zum oberflächenstrukturellen Subjekt, der zu einem generellen Auflösungs-Bias in Richtung der NP1-Verben führt; andere Studien berichten hingegen einen leichten Verarbeitungsvorteil für die NP2-Verben. Allerdings lassen sich solche generelle Auflösungstendenzen sowieso nur sehr bedingt auf die Verbkausalität zurückführen, weil die beiden Aktanten durch ihre Position im Satz mit anderen auflösungsrelevanten Faktoren kon-

fundiert sind (z.B. *primacy vs. recency*). Wahrscheinlich sind es jeweils sehr spezifische Konstellationen der Untersuchungsanordnung, des Stimulusmaterials und der eingesetzten Aufgabe bzw. Instruktion, die in den einzelnen Untersuchungen zu einem (meist nur leichten) Verarbeitungsvorteil für eine der beiden Gruppen führen. Dennoch ist auffällig, dass gerade die Klasse der SE-Verben im Vergleich zu den ES-Verben häufig benachteiligt erscheint: Zwar ist das Satzsubjekt zwar stärker aktiviert (*first mention advantage*), aber insgesamt brauchen Menschen länger, um solche Sätze zu verarbeiten und machen bei ihnen auch mehr Fehler (vgl. z.B. McDonald & MacWhinney, 1995; Long & de Ley, 2000). Es sei daran erinnert, dass diese Verbklasse sich auch in der linguistischen Rekonstruktion als problematisch erwiesen hat, da sie viele linguistisch-strukturelle Anomalien aufweist (kein Passiv- und Progressiv-Bildung etc., vgl. Abschnitt 2.1.2).

Als Fazit lässt sich festhalten, dass es bezüglich der Existenz des IC-Effekts kaum eine ernsthafte Diskussion geben kann. Entsprechend hat sich die ganze sprachpsychologische Forschung zum IC-Effekt auch eher auf die Frage konzentriert, welcher *Art* der IC-Effekt ist, d.h. wo und wann er in den Sprachverarbeitungsprozess eingreift. Wie bereits angedeutet, gibt es einerseits Vertreter/innen, die ihn eher als einen frühen, lexikalischen Effekt ansehen, der bottom-up die pronominale Auflösung beeinflusst. Die Gegenposition zu einer solchen gedächtnis-basierten Erklärung geht hingegen davon aus, dass für den IC-Effekt eher hierarchiehohe, top-down-Prozesse verantwortlich sind. Dieser Auffassung zufolge wird bei der pronominalen Auflösung unter Rückgriff auf komplexe Weltwissenbestände erklärungs-basiert nach einer möglichst plausiblen Auflösung für das Pronomen gesucht. Und diese Suche ist wesentlich einfacher, wenn die explizite und die implizite Begründung für ein Satzereignis übereinstimmen.

2.1.3.2 Der IC-Effekt als später Integrations-Effekt

Die Kontroverse, ob der IC-Effekt ein früher lexikalischer oder ein später integrativer Effekt ist, wurde bereits sehr früh von Ehrlich (1980) angestoßen. Nach Ehrlich geht die lexikalische Position in Anschluss an Caramazza et al. davon aus, dass die IC ein inhärentes Merkmal des Verbs ist, welches das kausale Gewicht entweder dem Subjekt oder Objekt zuweist. Dadurch kommt es im Rezeptionsprozess zu einer „Prädisponierung“ für eine bestimmte pronominale Auflösung. Die Vorstellung ist dabei ungefähr die, dass z.B. mit dem Verb *blame* in (10) ein bestimmter attentionaler Vorteil für das Objekt verbunden ist, weil dieses durch die IC bereits voraktiviert ist:

(10) John blamed Bill because he spilt the coffee.

Die kohärenz-basierte Gegenposition geht hingegen davon aus, dass Leser/innen ihr Weltwissen in Kombination mit der Semantik des Verbs dazu benutzen, eine möglichst sinnvolle Auflösung für das Pronomen zu finden: „We know from general knowledge that spilling coffee is usually bad, and from lexical knowledge that Bill doing something bad is a reason for John to blame Bill.“ (Ehrlich, 1980, p. 248). Das heißt, die Auflösungstendenz basiert nicht (allein) auf der lexikalischen Voraktivierung einer der beiden Aktanten, sondern hier ist eine (komplexe) Inferenz notwendig, welche die globale Kohärenz des Satzes gewährleistet.

Obwohl die von Caramazza et al. berichtete Ergebnisse mit beiden Interpretationen kompatibel sind, ist bezeichnend, dass zunächst einmal die *diskursive*, nicht die lexikalische Interpretation des IC-Effekts die Diskussion bestimmt hat. Schon Ehrlich (1980) führte den Effekt auf die (späte) Wirksamkeit von Weltwissensbeständen im Sprachverarbeitungsprozess zurück, weil der Effekt offensichtlich von der verwendeten Konjunktion abhängig ist. Auch Vonk (1984) ging davon aus, dass der IC-Effekt diskursiv-pragmatischer Natur ist. Er kommt dadurch zustande, dass im Laufe der pronominalen Auflösung Weltzustände abgerufen bzw. konstruiert werden, die mit den lexikalischen Informationen des Verbs übereinstimmen müssen. Sie vergleicht den IC-Effekt hier explizit mit einem anderen bekannten Phänomen, das bei der pronominalen Auflösung relevant ist, nämlich dem sog. *experiencer-constraint* (Springston, 1976; vgl. (11)).

- (11) a. Hans sagte Peter_i, dass er_i Maria langweilt.
b. *Hans sagte Peter_i, dass Maria ihn_i langweilt.

In (11) ist keine IC involviert, das Verb *sagen* impliziert normalerweise keine besondere kausale Präferenz für einen der beiden Aktanten. Das Pronomen *er* in (11a) ist deswegen auch perfekt mit beiden Hauptsatz-Referenten kompatibel: Hans kann Peter sowohl sagen, dass er selbst Maria langweilt, aber auch, dass Peter sie langweilt. In Satz (11b) ist hingegen nur noch die eine Lesart möglich: Da *langweilen* ein subjektiver Gefühlszustand ist, wäre es pragmatisch seltsam, wenn Hans Peter über seine (Peters) Gefühle aus der Außenperspektive informieren würde. Diese Interpretation bräuchte zumindest eine starke diskursive Unterstützung.

Vonk (1985a) interpretiert den IC-Effekt nun ziemlich ähnlich wie den *experiencer-constraint*: Das Verb schafft durch seine IC einen kontextuellen Rahmen, in dem bestimmte Auflösungsvarianten präferiert werden. Ähnlich wie Hobbs (1978) geht sie davon aus, dass im Laufe der pronominalen Auflösung ein formallogischer Schluss vorgenommen wird, der auf den Verbinformationen und den Informationen des Nebensatzes als Prämissen aufbaut. Wenn wir also hören, dass jemand einen anderen *beschuldigt* und gleichzeitig wissen, dass

einer von den beiden *Kaffee verschüttet* hat, dann werden diese beiden Informationen kombiniert und der richtige Referent hieraus abgeleitet, also *inferiert* (vgl. (12)):

- (12) a. KAFFEE VERSCHÜTTEN(x) → BESCHULDIGEN(y,x)
- b. BESCHULDIGEN(John,Bill)
- c. KAFFEE VERSCHÜTTEN(?) → Bill!

Entscheidend für die Auflösung sind also nicht die Verbinformationen selbst; in der Tat sind sie im Rahmen einer solchen Konzeption für sich genommen vollkommen nutzlos. Die pronominale Auflösungs-Inferenz kann erst durch die *Kombination* von Hauptsatz- und Nebensatzprädikat erfolgen. Und da zum Zeitpunkt der Verarbeitung des Pronomens noch nicht alle relevanten Informationen vorhanden sind, muss mit der Auflösung bis zum Satzende gewartet werden. Dort wird dann eine Auflösung des Pronomens derart vorgenommen, dass eine adäquate und kohärente Gesamtinterpretation des Satzes in Bezug auf das vorliegende Weltwissen resultiert. IC-Informationen werden also erst am Satzende genutzt, um beide Teilsätze miteinander zu integrieren.

Der entscheidende Punkt ist, dass sich dieser Auffassung zufolge der IC-Effekt allein aus der (In)Kongruenz von Verbinformation und Nebensatzinformation ergibt. Denn – wie in (12) demonstriert – müssen pronominale Schlüsse unter Rückgriff auf eine allgemeine Gesetzmäßigkeit (in diesem Fall, dass Kaffee verschütten Beschuldigungen rechtfertigt) vorgenommen werden (deduktive Variante) oder zumindest muss diese im Laufe des Auflösungsprozesses rekonstruiert werden können (induktive Variante). In Fällen, in denen die IC- und Nebensatzverb-Informationen übereinstimmen, ist eine solche (Re)Konstruktion wesentlich leichter möglich. Hier kann auf gut verfügbare Weltinformationen zurückgegriffen werden (Kaffee verschütten ist ein guter und schnell verfügbarer Grund für den BESCHULDIGEN-Akt). In einer inkongruenten Bedingung (z.B. *John beschuldigt Bill, weil er schlechte Laune hat.*) ist dies hingegen nicht so leicht oder nur durch inferenzielle Zusatzaktivitäten möglich.

Diese theoretische Konzeption hat – worauf v.a. Long und de Ley (2000) hingewiesen haben – eine enge konzeptuelle Nähe zum sog. *Connective Integration Model* von Millis und Just (1994). Dieses geht davon aus, dass die Satzverarbeitung nur lokal inkrementell ist und an den Teilsatzgrenzen abschließende Integrationsprozesse vorgenommen werden. Der Inhalt der Teilsätze wird dann aus dem Buffer des Arbeitsgedächtnisses in einen Zwischenspeicher transferiert. Dieses Modell setzt die v.a. am Anfang der sprachpsychologischen Forschung dominante Vorstellung fort (vgl. z.B. Kintsch, 1974), nach der die Sprachverarbeitung propositionsweise vorgenommen wird. Die Propositionen werden zunächst separat verarbeitet, müssen jedoch dann im Folgenden integriert werden. Nach Millis und Just fungieren nun Satz-

konnektoren (wie die Konjunktion *weil*) als Cues zur Integration verschiedener Teilinformationen. Das heißt, die einzelnen Teilsätze werden jeweils für sich inkrementell verarbeitet, können jedoch am Satzende wieder miteinander kombiniert werden (sofern eine entsprechende Konjunktion vorhanden ist). In Bezug auf den IC-Effekt bedeutet das, dass im Hauptsatz zunächst eine Proposition verarbeitet wird (z.B. *BESCHULDIGEN(y,x)*), die jedoch an der Hauptsatzgrenze in den Zwischenspeicher übertragen wird, um Ressourcen zur Verarbeitung des anschließenden Nebensatzes zu haben. In diesem wird wiederum die entsprechende Proposition (*KAFFEE VERSCHÜTTEN(x)*) für sich verarbeitet. Erst am Satzende dient der Konnektor *BECAUSE* als Hinweisreiz, dass die beiden Propositionen kombiniert werden müssen, und zwar in einer Implikationsrelation (\rightarrow). Das heißt, an dieser Stelle wird die erste Proposition reaktiviert und die notwendigen Inferenzen und Elaborationen vorgenommen, die ungefähr so ablaufen, wie in (12) skizziert.

In Bezug auf die Fragestellung, warum ic-inkonsistente Fortführungen schwerer zu verarbeiten sind, haben übrigens in jüngerer Zeit Guerry, Gimens, Caplan und Rigalleau (2007) ein cleveres Experiment durchgeführt. Sie nahmen dabei die – jedem, der einmal im IC-Paradigma gearbeitet hat, leidvoll bekannte! – Tatsache zum Ausgangspunkt, dass es (sehr) viel schwieriger ist, ic-inkompatible als -kompatible Fortsetzungen zu generieren. Während einem meistens sehr schnell plausible Fortsetzungen einfallen, die mit dem Verb-Bias übereinstimmen, braucht man hierfür bei den inkongruenten sehr viel länger. Guerry et al. prüften dies in einem Produktionsexperiment, in dem gemessen wurde, wie lange Personen brauchen, um verschiedene Fortsetzungen zu IC-Verben zu formulieren. Das heißt, Stimulussätze waren Sätze wie in (13), in denen die jeweiligen Nebensatzanfänge bereits vorgegeben waren:

- (13) a. Peter beschuldigte Linda, weil sie.... (ic-kompatibel)
 b. Linda beschuldigte Peter, weil sie... (ic-inkompatibel)

Ihre Versuchspersonen brauchten in der Tat viel länger für die ic-inkompatiblen als für die -kompatiblen Fortführungen. Menschen erscheint es also in der Tat plausibel, dass jemand einen anderen beschuldigt, weil dieser etwas Beschuldigungswertes getan hat. Die Befindlichkeit des Beschuldigenden selbst ist hingegen kein guter Grund für einen Beschuldigungs-Akt. Zur Begründung dieses Effekts ließen sich im Prinzip alle Ausführungen wiederholen, die aus der sozialpsychologischen Forschung zur IC bereits bekannt sind (z.B. Kovariationsinformationen etc.; vgl. Abschnitt 2.1.1). Wichtig ist jedoch im Moment allein das Grundprinzip, nämlich dass man durch den *weil*-Konnektor zu einer bestimmten Attribution gezwungen wird (Person vs. Objekt) und es unterschiedlich schwer ist, die für die gewünschte Attribution

benötigten Informationsbestände (insbesondere Konsensus- und Distinktheitsinformationen) abzurufen bzw. aufzubauen.

Darüber hinaus ist interessant, dass Guerry et al. bei der inhaltlichen Inspektion der generierten Fortführungen entdeckten, dass viele ihrer Vpn bei den ic-kompatiblen Fortsetzungen auf den „Trick“ zurückgriffen, den vom Verb-Bias präferierten Aktanten als Objekt wieder in die Satzstruktur einzuführen, z.B. wie in (14):

(14) Linda kritisierte Peter, weil sie **seine** Arbeit unzureichend fand.

Hier wird offensichtlich doch wieder ein Grund eingeführt, der das Gewicht der Erklärung auf den ic-kompatiblen Aktanten legt. Der ic-inkompatible Aktant nimmt diesen Zustand hingegen lediglich wahr und *beurteilt* ihn. Dies verleitete Guerry et al. zu der These, dass der IC-Effekt evtl. dadurch zustande kommt, dass bei ic-inkompatiblen Fortsetzungen nicht nur eine, sondern vielmehr *zwei* Inferenzen notwendig sind: Zum einen muss man einen Zustand *Lindas* inferieren, der ihre Kritik rechtfertigt, und zum anderen muss noch ein weiteres Ereignis inferiert werden, welches wiederum *Lindas Zustand* rechtfertigt. Im Rahmen der ereignisstrukturellen Rekonstruktion des IC-Effekts (vgl. Abschnitt 2.1.2) bedeutet das, dass der thematische Rahmen eines (in diesem Fall AE-)Verbs verlangt, dass die Kausalkette mit dem Objekt begonnen werden muss. Das bereits „vor-aufgelöste“ Pronomen in (13) erzwingt jedoch einen Subjekt-turn. Deswegen ist man gezwungen, entweder einen Subjekt-turn zu finden, der kausal so „stark“ ist, dass der verlangte initiale Objekt-turn konsensuell vorausgesetzt werden kann, oder aber man muss diesen explizit in die Ereignisstruktur einbauen. Experimentelle Evidenz zu dieser Hypothese steht jedoch meinem Wissen nach noch aus.

Eine solche Vorstellung schafft ideale Voraussetzungen für eine Verbindung der frühen, auf propositionalen Strukturen und aussagenlogischen Schlussprozessen basierenden Ansätze (wie z.B. der von Hobbs, 1978) und Ansätzen, die eher referenziell orientierten sind und mit Konstrukten wie Mentalen Modellen oder Szenarien operieren. Eine Rekonstruktionsvariante des IC-Effekts, welche explizit auf eine referenziell-modellorientierte Konzeption zurückgreift, wurde z.B. von der Gruppe um Garnham vorgelegt (Garnham & Oakhill, 1985, Garnham et al., 1992; Garnham et al., 1996; im Überblick: Garnham, 2001). Der Ansatz macht sich dabei die Tatsache zu nutze, dass kausale Verknüpfungen ja nicht nur bei der pronominalen Auflösung eine Rolle spielen, sondern für das Sprachverstehen allgemein zentral sind. Einen Text zu „verstehen“ bedeutet im Prinzip, die mit ihm beschriebenen Situationen, also *Verknüpfungen von Ereignissen*, adäquat zu rekonstruieren. Sätze, die eine pronominalen Auflösung verlangen, werden also nicht isoliert verstanden, sondern immer in Bezug auf komplexe Diskursmodelle oder Szenen, welche die Abfolge und Verbindung von Einzeler-

eignissen strukturieren. Die IC ist dann lediglich ein Faktor, der zur Strukturierung solcher Szenen beiträgt. Für die pronominale Auflösung selbst ist für Garnham et al. das Prinzip des sog. *mapping* wichtig, dass sie im Wesentlichen aus Sanford und Garrods *Scenario-Mapping-and-Focus* Ansatz übernehmen. Ein Pronomen aufzulösen, bedeutet dann v.a., ihm einen Referenten zuzuweisen. Es muss also auf eine bereits etablierte Diskursentität *bezogen* werden. Wie in Kap. 1 erläutert, unterscheiden Garnham et al. zwischen einem primär oberflächenstrukturellen *role-to-name-mapping* und einem tiefenstrukturellen, modellbezogenen *role-to-role-mapping*. Während die Auflösung eines Pronomens allein aufgrund seiner Geschlechtszugehörigkeit (Gender-Cue) lediglich ein *name-mapping* verlangt, ist im Falle der IC eine *role-mapping* erforderlich. Das heißt, das Pronomen muss nicht nur morpho-syntaktisch zum Referenten passen, sondern muss seine *Funktion* oder *Rolle* in dem zugrunde liegenden Szenario übernehmen (können). Zur Herstellung solcher Relationen sind u.U. umfangreiche inferenzielle Prozesse notwendig, die zwingend auf das aktuelle Diskursmodell und andere Weltwissensbestände Bezug nehmen. Die erschwerte Verarbeitung von ic-inkongruenten Fortsetzungen resultiert dann daraus, dass sie sich schwerer in das Szenario integrieren lassen. Ic-inkongruente Auflösungen sind gewissermaßen *skript-inkompatibel* und lassen sich ohne kohärenzstiftende Zusatzinferenzen nicht verarbeiten.

Garnham et al. (1996) kontrastieren dieses Verständnis des IC-Effekts (sog. *integration hypothesis*) explizit mit der zuerst von Caramazza et al. (1977) vertretenen Auffassung, die den IC-Effekt eher als einen frühen Aufmerksamkeits-Effekt betrachtet (sog. *focusing hypothesis*). Die Integrations-Hypothese macht die Vorhersage, dass ic-inkompatible Fortsetzungen erst dann die Verarbeitung behindern, wenn die skript-inkompatiblen Informationen ins aktuelle Diskursmodell integriert werden müssen. Der IC-Effekt sollte deswegen erst am Ende des Satzes auftreten, wenn es zu einem Abgleich der impliziten Kausalitätsinformationen mit der expliziten, aber inkongruenten Begründung des Nebensatzes kommt. In diesem Fall werden dann rückwärts-gerichtete Inferenzen notwendig, welche die globale Kohärenz des Szenarios gewährleisten. Garnham et al. gehen deswegen davon aus, dass der IC-Effekt nur auftritt, wenn zum einen überhaupt eine ereignisstrukturelle Verbindung der beiden Teilsätze notwendig ist (durch einen *weil*-Konnektor). Wie besprochen, gibt es in der Tat Evidenz dafür, dass der *weil*-Konnektor für den Effekt entscheidend ist. Zum anderen sollte das aufwändige *role-mapping* vermieden werden, wenn gleichzeitig ein Gender-Cue vorhanden ist, der eine Auflösung aufgrund von *name-mapping* gewährleistet. Im Rahmen mehrerer *probe-recognition*-Experimente legen Garnham et al. Daten dafür vor, dass der IC-Effekt nur dann auftritt, wenn der Satz nicht gleichzeitig gender-disambiguiert ist. Allerdings war dies bei

Caramazza et al. (1977) gerade nicht der Fall und die Unabhängigkeit der beiden Effekte wurde auch von vielen anderen Studien bestätigt (z.B. Vonk, 1984; 1985a, 1985b). Aparenterweise hatten Garnham et al. (1992) selbst zuvor genau die gegenteilige These vertreten, nämlich dass der IC-Effekt sogar *nur* dann auftreten sollte, wenn der Satz gleichzeitig genderdisambiguiert ist. Das Thema der Interaktion der beiden Cues ist jedoch kompliziert und soll deswegen in Abschnitt 2.3 noch separat diskutiert werden.

Stewart et al. (2000) versuchten ebenfalls, die Integrations- und die Fokussierungshypothese gegeneinander zu testen. Hierzu führten sie eine Reihe von *self-paced-reading*-Experimenten durch. Der Nebensatz wurde dabei in zwei Teilfragmente unterteilt und es wurde geprüft, ob sich der IC-Effekt im ersten oder im zweiten Fragment fand. Es war immer der zweite. Darüber hinaus versuchten sie, durch eine Reihe experimenteller Manipulationen plausibel zu machen, dass es sich beim IC-Effekt in der Tat um einen späten Integrationseffekt handelt. So behaupteten sie, dass der IC-Effekt, wenn er wirklich fokus-basiert sei, nicht bei NP-Anaphern auftreten sollte. Sie legten jedoch Evidenz dafür vor, dass es ihn auch für höhere Anaphern gibt (allerdings sind sie damit die einzigen, die dies bisher nachgewiesen haben, im Gegensatz z.B. zu Garnham & Oakhill, 1985; Hudson et al., 1986; McDonald & MacWhinney, 1995). Darüber hinaus sagten sie (ganz im Gegensatz zu Garnham et al.) voraus, dass ein fokus-basierter IC-Effekt gerade mit dem Gender-Cue interagieren sollte. Sie konnten jedoch keine Interaktion der Faktoren nachweisen. Und schließlich konnten Stewart et al. noch nachweisen, dass der IC-Effekt nur bzw. stärker auftrat, wenn nach dem Lesen sog. „deep processing-questions“ gestellt wurden, d.h. Fragen, die eine Auflösung des Pronomens verlangten. Auch dies spricht ihrer Auffassung zufolge dafür, dass es sich beim IC-Effekt um einen integrations- und nicht einen fokus-basierten Effekt handelt.

Die Ergebnisse von Garnham et al. (1996) und Stewart et al. (2000) sind jedoch nicht singulär. Insgesamt sprechen eigentlich fast alle Studien, die Aussagen über den zeitlichen Verlauf des IC-Effekts zulassen, dafür, dass der IC-Effekt satzfinal anzutreffen ist. Unter anderem liegen hierzu Daten aus *eye-tracking*- (Vonk, 1984, 1985b), *probe-recognition*- (Garnham et al., 1996) und *self-paced-reading*-Experimenten (Stewart et al., 2000) vor. Viele Autoren/innen kommen deswegen zu dem Schluss, dass es sich beim IC-Effekt um einen reinen Integrationseffekt handelt. Am deutlichsten wurde diese Position von Stewart et al. (2000, p. 438) formuliert, die behaupten, dass

[...] implicit causality information is employed by the language processor during integration. Implicit causality does not focus the processor on the potential referent consistent with the direction of the implicit causality bias. Rather, it facilitates the process of combining clauses to produce a coherent semantic representation.

Allerdings muss einschränkend hervorgehoben werden, dass die Existenz eines späten Integrationseffektes die Fokussierungshypothese an sich zunächst einmal nicht gefährdet (vgl. Koornneef & van Berkum, 2006). Ansätze, die von einem frühen aufmerksamkeits-basierten Effekt ausgehen, sind ebenfalls kompatibel mit der Aussage, dass es im weiteren Verlauf der Satzverarbeitung zu Integrationsproblemen kommt. Sie gehen geradezu davon aus, dass aufgrund der frühen Fehlaulösung des Pronomens im späteren Satzverlauf eine semantische Reanalyse notwendig ist, die zu erhöhten Verarbeitungszeiten führt (s.u.). Kritischer für die Fokussierungshypothese ist hingegen, dass sich in den genannten Studien keine frühen Effekte der IC finden ließen. Im nächsten Abschnitt sollen jedoch Ansätze und Studien vorgestellt werden, welche dafür sprechen, dass es unter bestimmten Umständen auch frühe IC-Effekte gibt. Allerdings soll ebenfalls gezeigt werden, dass sich die Interpretation dieser Effekte im Laufe der Zeit sehr verändert hat. Während sie anfänglich als rein aufmerksamkeits-basierter Fokussierungsprozesse verstanden wurden, interpretiert man sie heutzutage eher als Reflex früher Prädiktionsprozesse.

2.1.3.2 Der IC-Effekt als früher Fokus-Effekt

Es wurde bereits darauf hingewiesen, dass die genaue Natur des IC-Effekts in seiner initialen Formulierung durch Caramazza et al. (1977) noch eher unklar war. So argumentierten Caramazza und Garvey (1974) – den Zeichen ihrer Zeit entsprechend – für eine rein lexikalische Fundierung des Effekts. Sie behaupteten jedoch nur relativ pauschal, dass die IC als ein abstraktes Verbmerkmal dafür verantwortlich ist, dass einer der beiden Aktanten stärker voraktiviert wird. Eine weitergehende Präzisierung hat diese lexikalische Sichtweise des IC-Effekts dann von Hudson et al. (1986) erhalten, die ihm konkret im Kontext der *centering theory* interpretieren (Grosz et al., 1985; vgl. auch Kap. 1.3.3). Das heißt, der IC-Effekt wurde nun als ein früher Fokussierungseffekt rekonzeptualisiert, der das diskursive Aufmerksamkeits-Zentrum beeinflusst, auf das sich Pronomen bevorzugt beziehen. Der IC-Effekt kommt dann dadurch zustande, dass eine pronominale Reanalyse bei Verarbeitung der inkonsistenten Fortsetzung notwendig ist. Im Prinzip handelt es sich bei dieser Interpretation um eine direkte Übertragung des bekannten garden-path-Modells der syntaktischen Ambiguitäten auf das Phänomen der pronominalen Auflösung.

Exakt diese Auffassung wurde dann von McKoon, Greene & Ratcliff (1993) im Rahmen ihrer minimalistischen, gedächtnis-basierten Inferenztheorie noch weiter ausgebaut (vgl. auch Greene & McKoon, 1995). Wie in Kap. 1 ausführlich dargestellt ging es der Gruppe um McKoon und Ratcliff ja v.a. darum, exakt herauszuarbeiten, welche Inferenzen automatisch

gezogen werden können und welche der strategischen Verarbeitung bedürfen. Ein erstaunliches Ergebnis ihrer umfangreichen empirischen Arbeiten war dabei, dass Pronomen gerade *nicht* zur Klasse der minimalen Inferenzen gehören, da sie nur unter strategischen Bedingungen aufgelöst werden (vgl. Greene et al., 1992). Sie gehen dabei von einem dynamisch flexiblen Diskursmodell aus, in dem verschiedene Entitäten unterschiedliche Grade der Zugänglichkeit besitzen. Die Zugänglichkeit eines Aktanten ist dabei von verschiedenen syntaktischen, semantischen und pragmatischen Faktoren abhängig. Entsprechend der Grundannahmen des gedächtnis-orientierten Ansatzes der pronominalen Auflösung fassen sie Pronomen nicht als sprachliche Elemente auf, welche einer „Auflösung“ im Sinne einer gesteuerten Suchabfrage bedürfen. Vielmehr gehen sie in ihrem „pronoun-as-cue-framework“ davon aus, dass die semantischen Merkmale des Pronomens einen schnellen, parallel-automatischen Resonanz-Prozess auslösen. Dieser reaktiviert zumeist eine eindeutige Diskursentität, die als potentieller Referent zur Verfügung steht. Wenn hingegen kein oder mehrere Referenten am Ende des Resonanz-Prozesses aktiviert sind, dann wird das Pronomen entweder nicht aufgelöst oder es bedarf der strategischen Reanalyse. Der entscheidende Punkt ist nun, dass McKoon et al. bei „normalen“ Verben fanden, dass Pronomen im Gegensatz zu NP-Anaphern von sich aus nicht in der Lage sind, einen eindeutigen Referenten zu etablieren. Erstaunlicherweise funktioniert es bei IC-Verben aber sehr wohl: In solchen Sätzen können Pronomen auch unter automatischen Verarbeitungsbedingungen aufgelöst werden, und zwar sowohl, wenn der Referent mit der IC des Verbs übereinstimmt, als auch, wenn er *ic-inkompatibel* ist. Das heißt, in dem von McKoon et al. verwendeten Stimulusmaterial (vgl. (15)) wurde sowohl das Pronomen in Satz 3a) als auch das in 3b) aufgelöst:

- (15)
1. James and Debbie were working on a political campaign together.
 2. They were both planning on pursuing careers in politics.
 3. James infuriated Debbie because
 - a) he leaked important information to the press.
 - b) she had to write all the speeches.

Zum genauen Mechanismus, welcher der Auflösung zugrunde liegt, äußern sich McKoon et al. jedoch nur sehr unklar. Grundsätzlich lehnen sie die traditionelle Unterscheidung von Handlungs- und Zustandsverben ab und beziehen sich auf die Initiating-Reaction-Klassifikation von Osgood (1970): Bestimmte Aktanten initiieren eine Handlungsabfolge und andere reagieren hierauf. Die Einführung eines kausalen Konnektors leitet nun eine Suche nach einer Erklärung für das initiierte Teilereignis ein. Im Gegensatz zur klar fokusbasierten Analyse des IC-Effekts bei Hudson et al. (1986) gehen sie also explizit davon aus, dass im Laufe der pronominalen Auflösung eine Erklärungsrelation etabliert werden muss, die

Die Grundstruktur des Materials entspricht dabei vollkommen dem seit Caramazza et al. üblichen Muster. *Amaze* ist ein SE-Verb, das seinen Kausalitätsbias auf das Satzsubjekt legt. Die Logik der Studie war nun die, dass man an den mit Zahlen markierten Stellen des Satzes (der auditorisch dargeboten wurde) entweder den Namen des ic-kompatiblen Referenten *Gary* oder des ic-inkompatiblen *Alan* visuell darbot (crossmodales *probe-recognition*-Paradigma). Die entscheidende Frage war, ob am Zeitpunkt T2 auf den Namen *Gary* schneller reagiert wird als auf den Namen *Alan*. Man bemerke, dass zu diesem Zeitpunkt die Informationen des Nebensatzes noch gar nicht relevant sind. Entsprechend wurde die *Kongruenz* vom impliziten Verb-Bias zur expliziten Nebensatz-Begründung auch gar nicht manipuliert: Alle Sätze waren also in der Terminologie von Caramazza et al. „kongruent“. Da sich der IC-Effekt bei McDonald und MacWhinney allein auf einen (relativen) Aktivierungsvorteil des fokussierten Referenten bezieht, werde ich ihn im Folgenden als *IC-Fokus-Effekt* bezeichnen. Der (ursprüngliche) IC-Effekt von Caramazza et al. (um den es auch in den meisten anderen Studien geht), bezieht sich hingegen auf die Kongruenz von expliziter und impliziter Begründung; ich nenne ihn deswegen im Folgenden *IC-Kongruenz-Effekt*.

Sowohl McKoon et al. als auch McDonald und MacWhinney vertreten also die Auffassung, dass die IC eines Verbs zu einer frühen Fokussierung des kausal verantwortlichen Referenten führt. Allerdings ist bei beiden Ansätzen bereits eine deutliche Tendenz vorhanden, den gedächtnis- und aufmerksamkeits-basierten IC-Fokus-Effekt in eine größere erklärungs-basierte Perspektive einzubetten. Long und de Ley (2000), die ebenfalls davon ausgehen, dass dem IC-Effekt eine frühe Fokus-Wirkung des Verbs zugrunde liegt, werden sogar noch deutlicher. Sie beziehen sich auf die in Kap. 1 ausführlich beschriebene *Scenario-Focus-Mapping*-Theorie von Sanford und Garrod (1998), die davon ausgeht, dass Informationen, die sich gerade im expliziten Fokus befinden, unmittelbar auf komplexe Weltwissensbestände im impliziten Fokus bezogen (*gemappt*) werden. Vollkommen parallel gehen sie davon aus, dass IC-Verben so etwas wie kleine Mini-Szenarien sind, die direkt vom Verb instantiiert werden. Sie stellen direkt szenarien-typische Informationen bereit, auf die im Laufe der folgenden Satzverarbeitung zurückgegriffen werden kann, um weitere Informationen zu verarbeiten.

Allerdings weisen Long und de Ley auf eine wichtige Randbedingung hin: Szenario-Informationen können nur dann abgerufen und in die Satzstruktur integriert werden, wenn sie hinreichend aktiv sind oder den semantischen *default*-Fall darstellen. Auf jeden Fall müssen die Informationen im expliziten Fokus hinreichend eindeutig sein. Die direkte Verarbeitung komplexer Szenario-Informationen stellt jedoch erhebliche Ansprüche an den Rezipienten, der über genügend Verarbeitungsressourcen verfügen muss, um sie angemessen verwenden zu

können. Wenn Leser/innen nicht über die notwendige verbale Arbeitsgedächtniskapazität verfügen, dann können sie das Szenario nicht direkt verarbeiten, sondern werden eine andere Verarbeitungsstrategie wählen (müssen). So werden sie z.B. komplexere Verarbeitungsschritte an das Satzende verlagern, wo im Vergleich zur online-Konstruktion im Binnensatz genügend Verarbeitungsressourcen allokiert werden können. Da bestimmte Texte und verschiedene Aufgabentypen unterschiedliche Anforderungen an die Verarbeitungsressourcen der Rezipienten stellen (z.B. einfaches Lesen vs. Bewerten eines Satzes), ist hier ggf. von einer komplexen Interaktion von Rezipienten-, Text- und Aufgabenmerkmalen auszugehen.

Long und de Ley überprüften diese Hypothese ebenfalls mit einem *probe-recognition*-Experiment und zwar sehr ähnlich wie MacDonald und MacWhinney. Allerdings berücksichtigten sie zusätzlich noch die Arbeitsgedächtnis-Kapazität ihrer Vpn. Entsprechend ihrer Annahme fanden sie in der Tat einen frühen IC-Fokus-Effekt, allerdings nur, wenn die Probanden über genügend Verarbeitungsressourcen verfügten. Darüber hinaus konnten sie nachweisen, dass der IC-Fokus-Effekt nicht zu einer Aktivierung des ic-kompatiblen Referenten führt, sondern vielmehr zu einer Deaktivierung des Nicht-Referenten (vgl. auch Gernsbacher, 1989). In einem weiteren Experiment replizierten sie ein Experiment von Garnham et al. (1996), welches bisher als Evidenz dafür herangezogen wurde, dass IC-Informationen nur spät wirksam sind. Auch Long und de Ley fanden lediglich einen späten IC-Effekt, allerdings galt dies nur für die Gruppe der Vpn mit niedriger Arbeitsgedächtnis-Kapazität. Wenn genügend Verarbeitungsressourcen vorhanden waren, dann wurden die IC-Informationen bereits im Satzinneren verarbeitet.

An dieser Stelle sieht man noch einmal, dass sich der IC-Effekt nicht unbedingt starr in frühe automatische und späte strategische aufteilen lässt. Vielmehr ist davon auszugehen, dass das menschliche Sprachverarbeitungssystem hinreichend flexibel ist, unterschiedliche Aufgaben und Prozesse unter Berücksichtigung der aktuell gebildeten Struktur, des eintreffenden Inputs und der verfügbaren Ressourcen variabel auszuführen. Die Rekonstruktion des IC-Fokus-Effekts durch Long und de Ley macht jedoch ebenfalls deutlich, dass es wenig Sinn macht, den Sprachverarbeitungsprozess ausschließlich auf automatische, gedächtnis-basierte Prozesse oder nur auf kohärenz-basierte Integrationsprozesse zurückführen zu wollen. Nicht nur, dass die Grenze zwischen den beiden Prozessklassen offensichtlich unklar bzw. verschiebbar ist. Denn die Ergebnisse machen ja deutlich, dass die schnellen, automatischen Fokus-Effekte den späteren Erklärungsprozessen gerade *zuarbeiten*. Allerdings ist auch bezüglich der Studie von Long und de Ley noch einmal darauf hinzuweisen, dass sie nicht den IC-

Kongruenz-Effekt untersuchen, sondern vielmehr den IC-Fokus-Effekt, der als Indikator für die frühe Wirksamkeit von Szenarien-Informationen interpretiert wird.

Es ist auffällig, dass alle Ansätze, die sich für eine frühe Wirkung stark machten, nicht mit Lesezeit- oder Frage-Maßen arbeiteten, sondern ausschließlich mit dem *probe-recognition*-Paradigma. Eine solche Korrelation von theoretischer Fundierung und methodischem Inventar ist in der Psychologie keine Seltenheit, sondern ist sogar recht oft anzutreffen (vgl. z.B. Groeben, 1997). In Bezug auf die Erfassung pronominaler Auflösungsprozesse werden *probe-recognition*-Experimente insbesondere deshalb eingesetzt, weil argumentiert wird, dass sich mit ihnen automatische Prozesse erheben lassen (vgl. McKoon & Ratcliff, 1992). Von daher ist es nicht verwunderlich, dass Ansätze, die von einer frühen Lokalisation des IC-Effekts ausgehen, zu dieser Methode greifen, Ansätze, die jedoch eine späte Wirkung von IC-Informationen annehmen, andere Methoden präferieren, z.B. Lesezeit-Methoden. Ich möchte mich an dieser Stelle nicht in die Diskussion um die Angemessenheit der einen oder der anderen Methode einmischen (vgl. hierzu z.B. Gordon, Hendrick & Ledoux Foster, 2000; im Überblick: Mitchell, 2004). Für die Verallgemeinerbarkeit des IC-Fokus-Effekts ist diese Konzentration auf lediglich eine einzige Erhebungsmethode jedoch kritisch. Vor diesem Hintergrund ist eine neuere Studie von Koorneef und van Berkum (2006) bedeutsam, weil sie den frühen IC-Fokus-Effekt auch mit zwei Verfahren nachweisen konnten, die sonst nur von den Integrations-Ansätzen verwendet werden, nämlich einer *self-paced-reading*- und einer *eye-tracking*-Studie.

In dieser untersuchten sie (wie McKoon et al., 1993) kleine 3-Satz-Diskurse wie in (17):

- (17)
1. David and Linda were both driving pretty fast.
 2. At a busy intersection they crashed hard into each other.
 - 3a. David apologized to Linda because he according to the witnesses was the one to blame.
 - 3b. Linda apologized to David because he according to the witnesses was not the one to blame.

Der grundsätzliche Aufbau solcher Stimuli ist inzwischen wohlbekannt und bedarf wohl keiner Erläuterung mehr. Auf zwei Besonderheiten des Materials sei dennoch hingewiesen. Zum einen verwendeten Koorneef und van Berkum immer gender-disambiguierte Stimuli, d.h. der Referent des Pronomens war bereits durch morpho-syntaktische Merkmale eindeutig identifizierbar. Zum anderen verwendeten sie sowohl in der ic-kompatiblen als auch in der ic-inkompatiblen Fortsetzung eine gemeinsame Konstituente (hier: *according to the witnesses*), welche für eine gewisse Strecke in beiden Sätzen eine Gemeinsamkeit herstellt. Damit reagieren sie auf einen Vorwurf von Garnham (2001), dass die bisher berichteten IC-Effekte auf sprachliche Unterschiede zwischen den beiden Fortsetzungen zurückführt werden können.

Denn wenn sich die beiden Fortsetzungen in andern Eigenschaften als nur der IC-Kompatibilität voneinander unterscheiden (z.B. in ihrer Komplexität oder Plausibilität), dann liegt u.U. eine Konfundierung vor. Eine weitere Besonderheit ist, dass sie die Form des Pronomen (*he*) konstant hielten, da berechtigterweise eingewendet worden war (vgl. z.B. Garnham et al., 1992), dass im Niederländischen (wie auch im Deutschen) das Pronomen *zij* (*sie*) ambig ist (da es sowohl die 3. Sg. fem. als auch die 3. Pl. fem./mask. bezeichnen kann).

Koornneef und van Berkum fanden nun in einer wortweisen *self-paced-reading*-Studie, dass Vpn auf den ersten beiden Wörtern nach dem Pronomen ihre Lesegeschwindigkeit abbremsen. Auch in *eye-tracking*-Daten wurden die gleichen Stellen länger fixiert. Sie interpretieren dies als einen Indikator für frühe Integrationsschwierigkeiten für den erwartungskonträren Referenten. Komplementär hierzu fanden van Berkum et al. (2007; vgl. auch Nieuwland & van Berkum, 2006) in einer begleitenden EKP-Studie auf dem Pronomen eine P600-ähnliche Komponente, die gewöhnlich als Index eines syntaktischen Reanalyse-Prozesses in Reaktion auf eine syntaktische Anomalie interpretiert wird. Da sich der Effekt auf einen Satzabschnitt bezieht, für den die Kongruenz von expliziter Begründung und implizitem Verb-Bias nicht relevant ist, handelt es sich um einen IC-Fokus-Effekt. Für die satzfinale Integration berichten Koornneef und van Berkum keine Daten. Es ist jedoch zu vermuten, dass sie hier ebenfalls einen IC-Effekt gefunden haben.

Interessant ist jedoch v.a., wie Koornneef und van Berkum den gefundenen Effekt theoretisch einbetten. Sie begnügen sich nämlich nicht einfach damit, ihn als einen frühen Fokus-Effekt zu interpretieren, der sich daraus ergibt, dass mit dem ic-kompatiblen Pronomen eine Refokussierung notwendig wird (vgl. z.B. McKoon et al., 1993). Vielmehr vertreten sie die Auffassung, dass die Satzverarbeitung stark inkrementell verläuft, d.h. integrative Prozesse finden nicht nur satzfinal, sondern vielmehr für jedes Wort einzeln statt (vgl. auch Bornkessel & Schleswesky, 2006). IC-Kausalitätsinformationen können also sofort dazu benutzt werden, eine möglichst weitgehende Diskursrepräsentation aufzubauen. Und Abweichungen, die in Bezug auf diese frühe Diskursrepräsentation festgestellt werden (z.B. dadurch, dass man beim ic-inkompatiblen Pronomen merkt, dass die bisherige Interpretation inadäquat ist), werden sofort dazu benutzt, diese zu berichtigen. Die Fortsetzung des Diskurses wird also direkt auf seine Plausibilität geprüft und in Abhängigkeit von dieser Entscheidung dann in die vorliegende Ereignisstruktur eingebaut. Syntaktischer Strukturaufbau und Diskursinterpretation werden gleichzeitig und in wechselseitiger Abhängigkeit voneinander vorgenommen. Diese Auffassung unterscheidet sich wesentlich von einem rein fokus-basierten Mechanismus. Bei diesem führt die IC eines Verbs lediglich dazu, dass ein bestimmter Charakter attentional

zentriert wird. IC ist vielmehr ein sprachliches Mittel, das vorwärts-gerichtet („proaktiv“ sensu Garnham, 2001) dazu dient, Erwartungen über die kausale Struktur des Diskurses aufzubauen. Von daher verhält sich die IC eines Verbs auch nicht wesentlich anders als andere verbale Informationen. Sie sind halt nur kausal mit der Ereignisstruktur des zugrunde liegenden Szenarios verknüpft. Entsprechend sollten übergreifende Diskurs-Informationen auch in der Lage sein, sie ggf. überschreiben zu können.

2.1.4 Diskussion

In der Besprechung der theoretischen Konzeptionen und empirischen Befunde zum IC-Effekt ist deutlich geworden, dass es – zumindest anfänglich – zwei deutlich verschiedene Auffassungen darüber gab, wie dieser Effekt zustande kommt. Der sog. integrations-basierte Ansatz geht davon aus, dass der Effekt dadurch entsteht, dass ic-inkompatible Informationen schwerer in ein kohärentes Diskursmodell zu integrieren sind. Ebenso lassen sich nur schwer ic-inkompatible Inferenzen zur Erklärung des im Hauptsatz genannten Ereignisses generieren. Dieser Auffassung zufolge sollten sich IC-Effekte primär, wenn nicht sogar ausschließlich am Ende des Satzes finden lassen. Die komplementäre fokus-basierte Sichtweise des Effekts geht hingegen in ihrer ursprünglichen Formulierung davon aus, dass die IC zu einer salienteren Repräsentation des durch den Kausalitäts-Bias ausgezeichneten Aktanten führt. Dieser ist deshalb bei der späteren pronominalen Auflösung bereits voraktiviert. Der ic-inkompatible Referent ist hingegen weniger zugänglich und lässt sich deswegen wesentlicher schwerer auflösen. Stattdessen ist eine Refokussierung notwendig, die sich u.U. in einem frühen Inkongruenzeffekt auf dem Pronomen selbst niederschlägt.

Nach der Charakterisierung des gedächtnis- und des erklärungs-basierten Ansatzes in Kap. 1 fällt es relativ leicht, die Verbindungspunkte zwischen diesen allgemeinen Modellen der Sprachverarbeitung und den beiden speziellen Erklärungsansätze des IC-Effekts zu sehen. Das Erklärungsmodell des IC-Effekts als Fokus-Effekt ist offensichtlich mit dem gedächtnis-basierten Ansatz assoziiert, während das Erklärungsmodell des IC-Effekts als Integrations-schwierigkeit offensichtliche Parallelen zum erklärungs-basierten Ansatz aufweist. Im ersten Fall sind hierarchie-niedrige, allgemeine Gedächtnisprozesse für den Effekt verantwortlich, während sich im zweiten Fall die Verarbeitungsschwierigkeit daraus ergibt, dass Rezipienten/innen ein möglichst kohärentes Diskursmodell aufbauen bzw. erhalten wollen. Aber auch personell gibt es Schnittpunkte zwischen den Vertretern/innen der beiden übergreifenden Positionen und den Vertretern/innen der unterschiedlichen Erklärungsansätze des IC-Effekts: So ist z.B. die Gruppe um McKoon entscheidend sowohl an der Formulierung des *memory-based*

view als auch der fokus-basierten Erklärung des IC-Effekts beteiligt. Und die Gruppe um Garnham hat nicht nur eine vehemente Verteidigung des erklärungs-basierten Ansatzes vorgelegt, sondern vertritt auch für den IC-Effekt ein Erklärungsmodell, das entscheidend auf das Konzept eines kohärenten Mentalen Modells zurückgreift.

Aufgrund der theoretischen, methodologischen und personellen Überlappungen ist es auch nicht verwunderlich, dass sich das Grundmuster der allgemeinen Debatte beim IC-Effekt im Miniaturformat redupliziert: Beide Erklärungsansätze des IC-Effekt werden als sich wechselseitig ausschließend betrachtet und offensichtliche Parallelen ausgeblendet. Auch wird übersehen, dass sich die beiden Erklärungsansätze eigentlich komplettieren. Dabei ist im Fall des IC-Effekts noch wesentlich offensichtlicher, dass sie sich nicht gegenseitig ausschließen: Sie beziehen sich ja noch nicht einmal auf den gleichen empirischen Sachverhalt! Der IC-Fokus-Effekt geht von einem Inkongruenz-Effekt aus, der sich aus einem Mismatch von Verb-Bias mit dem nachherigen Pronomen ergibt. In der Tat ist anzunehmen, dass für einen solchen frühen Effekt vergleichsweise hierarchie-niedrige, gedächtnis-basierte Prozesse ausschlaggebend sind. Der IC-Kongruenz-Effekt bezieht sich hingegen auf die Tatsache, dass es schwieriger ist, eine kausale Verbindungsrelation zwischen der Hauptsatz-Proposition und der Nebensatz-Proposition zu inferieren. Es ist plausibel anzunehmen, dass solche Prozesse v.a. hierarchie-hohe, kontrollierte Prozesse involvieren. Es besteht also durchaus die Möglichkeit, dass beide Effekte gleichzeitig und unabhängig voneinander existieren.

Allerdings erscheint es sogar sinnvoll, sie theoretisch aufeinander zu beziehen: Frühe Fokussierung und späte Integration sind bei der Verarbeitung von (in)kongruenten IC-Informationen nicht unabhängig voneinander, sondern bedingen sich gegenseitig bzw. sind aufeinander angewiesen. So strukturieren die hierarchie-niedrigen Prozesse die Ereignisstruktur des Satzes vor und erleichtern damit den späteren Auflösungsakt. Deswegen ist auch davon auszugehen, dass ein größerer Integrationsbedarf am Satzende besteht, wenn keine antizipatorische Fokussierung vorgenommen werden konnte. In der Terminologie des *bonding-resolution*-Ansatzes würde das heißen, dass der frühe IC-Fokus-Effekt ein früher *bonding*-Prozess ist, während der späte IC-Kongruenz-Effekt eher einen späten Auflösungsprozess darstellt, der die finale Struktur des Diskursmodells herstellt und einer Kohärenz- bzw. Plausibilitätsprüfung unterzieht. Und je nachdem, wie stark das *bonding* ausfällt, kann die integrative Auflösung des Pronomens bzw. die Konstruktion einer gemeinsamen Gesamtinterpretation schwerer oder leichter vorgenommen werden.

Diese Sichtweise der Relation von IC-Fokus- und IC-Integrations-Effekt ergibt sich auch direkt aus der ereignisstrukturellen Rekonstruktion des IC-Effekts, wie er am Ende der

linguistischen Analyse der PSYCH-Verben in Abschnitt 2.1.2 entwickelt wurde. Dort wurde in Anschluss an Croft (1993) dafür argumentiert, dass IC-Verben als komplexe Szenarien betrachtet werden können, die eine inhärente Ereignisstruktur haben. Diese spezifiziert, welche Ereignisse am Szenario beteiligt sind und in welcher Relation sie zueinander stehen. Ein Verb wie *anklagen* beschreibt z.B. eine Situation, in der eine Person *x* eine andere Person *y* beschuldigt. Es handelt sich also um ein Subereignis, bei dem *x* auf *y* einwirkt ($x \rightarrow y$). Zusätzlich zu dem vom Verb beschriebenen Zentralereignis werden jedoch noch weitere Ereignisse vom Verb mit präsupponiert, insbesondere ein vorhergehendes Ereignis, das kausal für das fokale Subereignis verantwortlich ist. Im obigen Beispiel wird z.B. der Aktant *y* zuvor irgendwie auf Person *x* eingewirkt haben ($y \rightarrow x$), weshalb diese ihn nun anklagt. Sehr ähnlich hatte auch Grimshaw (1990) in Bezug auf die kausative Lesart der PSYCH-Verben argumentiert. Allerdings musste sie in ihrer Begründung noch auf Unterschiede in der Aktionsart zurückgreifen, was durch den Rekurs auf eine implizite verbale Ereignisstruktur nicht mehr nötig ist.

Inwieweit ist die Annahme einer impliziten Ereignisstruktur zur Erklärung des IC-Effekts nützlich? Meines Erachtens hilft sie v.a. dabei, die Relation der beiden IC-Effekte zueinander zu klären. Wenn man einen Satz wie *Peter beschuldigt Hans* verarbeitet, dann wird mit dem Verb *beschuldigen* ein Szenario-Skript aktiviert, welches explizit zunächst einmal nur spezifiziert, dass es eine Person *Peter* gibt, welche eine andere Person *Hans* BESCHULDIGT. Implizit sind jedoch mit dem Verb *beschuldigen* noch eine ganze Reihe von weiteren Informationen verbunden. Dazu gehört auch das Wissen darüber, dass man jemanden nicht grundlos beschuldigen kann (es gibt also ein implizites Vorläufer-Ereignis) und dass die Beschuldigung Konsequenzen für den Beschuldigten hat (es gibt also implizite, nachfolgende Ereignisse etc.). Diese implizite Ereigniskette, bei der der explizite BESCHULDIGEN-Akt nur der lexikalisch sichtbare Teil ist, muss nicht unbedingt vollständig in die diskursive Satzrepräsentation eingefügt werden. Sie ist vielmehr nur latent mit dem Verb verknüpft und wird bedarfsweise abgerufen, sofern es die Interpretation des Satzes verlangt. In der Terminologie des SMF-Ansatzes bedeutet das, dass sich das BESCHULDIGEN-Szenario im impliziten Fokus befindet und prinzipiell sofort reaktivierbar ist.

Entscheidend ist nun, was passiert, wenn der Kausalkonnektor *weil* verarbeitet wird. Dieser führt zu einer Aktivierung des impliziten, im BESCHULDIGEN-Skript bereits latent angelegten Vorläufer-Akts der Beschuldigung, deren Urheber *Hans* ist. Entsprechend wird *Hans* an dieser Stelle voraktiviert. Eine pronominale Bezugnahme auf *Hans* ist danach vergleichsweise einfach. Dies ist der IC-Fokus-Effekt. Wenn die nachfolgende, explizite Begründung

sich nicht auf das implizite und bereits voraktivierte Teilereignis bezieht, dann gibt es allerdings ein Problem. Man erwartet skript-konform Informationen über *Hans*, erhält aber welche über *Peter*. An dieser Stelle wird nicht nur klar, dass die kanonische Struktur des Skripts verletzt ist, sondern es müssen auch verschiedene Reparaturprozesse gestartet werden, um die Kohärenz der Diskursrepräsentation (und das heißt hier zunächst einmal des BESCHULDIGEN-Skripts) zu erhalten. Eine solche repair-Maßnahme ist z.B., dass man ein zusätzliches Ereignis in die Ereignisstruktur einbaut und diese noch validiert. Dies ist der IC-Integrations-Effekt. Wichtig ist v.a. die Tatsache, dass IC-Fokus- und IC-Integrations-Effekt im Rahmen dieses Ansatzes beide notwendig sind, um die komplexe Wirkung der IC auf die pronominale Auflösung zu erklären. Ohne die frühe Voraktivierung des kausal salienten Szenario-Ereignisses wäre die nachfolgende Integration solcher skript-kompatibler Informationen nicht erleichtert.

Diese ereignisstrukturell fundierte Sichtweise ist m.E. besonders geeignet, die drei beschriebenen Perspektiven auf die IC, Sozialpsychologie, Linguistik und Sprachpsychologie, miteinander zu verbinden. Deswegen ist es m.E. auch besonders erfreulich, dass in der allerjüngsten Diskussion zum IC-Effekt auch von anderen Autoren/innen sehr ähnliche Vorstellungen vertreten werden. So stellen z.B. Pickering und Majid (in press) ebenfalls fest, dass die IC keine Eigenschaft des Verbs selbst ist. Vielmehr handelt es sich bei ihr um eine Eigenschaft der von einem Verb beschriebenen Ereignisse. Der IC-Effekt beruht ihrer Auffassung nach auf den in der Beschreibung dieses Ereignis angelegten Gründen: „it is this event representation that is used to *infer* the cause or consequence“ (Pickering & Majid, in press, p. 7). Die Konstruktion dieser Ereignisstruktur wird dabei von verschiedenen Faktoren beeinflusst. Einerseits zählt hierzu selbstverständlich die Verbsemantik selbst, andererseits allerdings auch andere Eigenschaften der Ereignis-Aktanten (wie Animiertheit, Status, Geschlechtszugehörigkeit etc.). Alle Faktoren zusammen determinieren, welche vorherigen und nachfolgenden Ereignisse mit dem beschriebenen Ereignis verbunden sind und welche sich wie leicht konstruieren lassen. Ein bedeutender Prädiktor für die Abrufbarkeit eines kausal vorgehenden Subereignisses scheint dabei v.a. seine Typikalität zu sein (vgl. Corrigan, 1992). Das heißt, je typischer ein Ereignis ist bzw. je obligatorischer die Abfolge seiner inhärenten Subereignisse ist, desto stärker wird der IC-Bias für das vom Verb beschriebene Ereignisses ausfallen (vgl. auch Majid, Sanford & Pickering, 2006; Simner & Pickering, 2005). Diese Erkenntnis ist v.a. deshalb von Bedeutung, weil sie noch einmal verdeutlicht, dass der diskursive Kontext von entscheidender Wichtigkeit für die ereignisstrukturelle Rekonstruktion ist und bis auf die Ebene der syntaktischen Verarbeitung „durchschlagen“ kann (vgl. hierzu auch Härtl, 2001, und Wunderlich, 1997).

2.2 Gender-Cue

Es ist eine traditionelle und sicher auch intuitiv zunächst einmal plausible Annahme, dass ein „harter“ morpho-syntaktischer Gender-Cue die pronominale Auflösung schnell beeinflusst und zu einer sichereren Auflösung des Pronomens führt. Hingegen sollten „weiche“ semantische oder diskursive Informationen (wie z.B. der IC-Cue) eher unsicher sein und erst spät im Verarbeitungsprozess verfügbar werden. Exakt diese Logik lag ja auch den Überlegungen von Caramazza et al. (1977) zugrunde, als sie gender-disambiguierte Sätze als *Kontroll*-Bedingung einsetzten: Der morpho-syntaktische Gender-Cue sollte eine weitergehende semantische Analyse überflüssig machen und diese gewissermaßen blockieren. Wie dargestellt, hat sich dieses Muster nicht gezeigt: Der IC-Effekt wurde vom Gender-Cue nicht blockiert, obwohl die Gender-Disambiguierung selbst einen starken Einfluss auf die Auflösungsgeschwindigkeit hatte.

Von daher erstaunt es nicht, dass die nachfolgende Forschung zum Einfluss von Gender-Informationen auf die pronominale Auflösung sehr gemischte Ergebnisse hervorgebracht hat. Auch hier ist es wieder so, dass einige Studien eher für eine späte Verarbeitung von gender-bezogenen Informationen sprechen, andere hingegen auch frühe Einflüsse feststellen konnten. Diese Debatte ist nicht zu verwechseln mit der in Kap. 1 besprochenen Kontroverse, ob Gender-Merkmale dazu in der Lage sind, rein syntaktische Verarbeitungsmechanismen (wie das c-Kommando) zu beeinflussen, für die normalerweise eine sehr frühe Verarbeitung angenommen wird. Das war ganz offensichtlich der Fall, da sich gezeigt hat, dass Gender-Informationen bei hinreichender Fokussierung im Zweifelsfall syntaktische constraints überschreiben können. Gender-Informationen und c-Kommando interagieren also im Laufe des pronominalen Auflösungsprozesses. Da syntaktische Informationen ausgesprochen schnell verarbeitet werden, könnte man deswegen annehmen, dass auch Gender-Informationen schnell verarbeitet werden müssen. Allerdings wissen wir leider nicht, *wo* im Verarbeitungsprozess diese Interaktion zu lokalisieren ist. Es könnte durchaus auch sein, dass syntaktische Informationen zwar früh, Gender-Informationen jedoch spät verarbeitet werden und es erst bei der finalen Satzinterpretation zur Interaktion der beiden Faktoren kommt.

Im Folgenden geht es um die weitergehende Fragestellung, ob Gender-Informationen früh in den Prozess der pronominalen Auflösung eingreifen oder nicht. Klarer in Bezug auf die üblichen Zwei-Phasen-Modelle der pronominalen Auflösung (frühe syntaktische vs. späte semantisch-pragmatische Verarbeitung) formuliert, entspricht dies der Frage, ob Gender-Informationen eher als ein primär syntaktisches oder eher als ein semantisches Merkmal an-

gesehen werden sollten (für eine vergleichende Diskussion vgl. Comrie, 1999). Und in Bezug zum *bonding-resolution*-Ansatz heißt dies weiterhin, ob Gender-Informationen bereits das initiale *bonding* von Anapher und Antezedens beeinflussen oder aber lediglich zur finalen Auflösung beitragen. Dabei werde ich wieder dafür argumentieren, dass beides der Fall ist: Gender-Informationen haben sowohl syntaktische als auch semantische Funktionen und beeinflussen deswegen auch das frühe *bonding* und die späte *resolution*-Phase gleichzeitig.

2.2.1 Gender-Informationen als später *resolution*-Faktor

Es erstaunt, dass es nur vergleichsweise wenig systematische Forschung zum Einfluss von Gender-Informationen auf die pronominale Auflösung gibt. Das liegt vielleicht daran, dass die intuitive Annahme, dass der Gender-Cue den Verstehensprozess früh beeinflusst und das Antezedens eindeutig determiniert, in der Literatur ausgesprochen dominant vertreten wurde. Ehrlich (1980, p. 254) formuliert dies am klarsten (ähnlich: Vonk, 1984; 1985a; Crawley, Stevenson & Kleinman, 1990; Stewart et al., 2000):

Initially, a reader examines antecedents for features such as gender. If there is only one antecedent that agrees with the pronoun, a reader can assign it to the pronoun before completing a more detailed analysis of the sentence.

Die Vorstellung ist also, dass Gender-Informationen als eine Art initialer Filter fungieren: Wenn ein Pronomen bezüglich seines Gender-Merkmals mit dem Referenten übereinstimmt, dann kommt es a) zu einer direkten Integration dieser Informationen und es findet b) keine weiterführende semantische Analyse statt.

Allerdings haben sich beide Annahmen in den nachfolgenden empirischen Studien zunächst einmal als nicht haltbar erwiesen. Gegen die Annahme, dass Gender-Informationen eine weiterführende semantische Analyse blockieren spricht schon der Befund von Caramazza et al. (1977), dass IC-Informationen auch dann verarbeitet werden, wenn ein Gender-Cue vorhanden ist. Aber auch die Annahme, dass Gender-Informationen früh detektiert und integriert werden, ist durchaus problematisch. Die vielleicht stärkste Evidenz gegen eine (frühe) Nutzung von Gender-Informationen bei der pronominalen Auflösung stammt aus diversen *probe-recognition*-Experimenten. Tyler und Marslen-Wilson (1980) untersuchten z.B. Sätze wie (18) und fanden, dass nach einem gender-markierten Pronomen kein differenzieller Aktivierungsvorteil für einen gender-kongruenten Referenten im Vergleich zu einem gender-inkongruenten besteht (vgl. auch Marslen-Wilson, Tyler & Koster, 1993):

(18) The sailor tried to save the cat, but he/it fell overboard instead.

Gernsbacher (1989) konnte diesen Befund replizieren: Es gab keine frühe Aktivierung des gender-kongruenten Referenten, vielmehr war er erst am Satzende differenziell voraktiviert (bzw. der Nicht-Referent unterdrückt: vgl. Gernsbacher, 1990). Ähnlich Ergebnisse liegen von MacDonald und MacWhinney (1990), McDonald und MacWhinney (1995) und de Vincenzi (1999) für *probe-recognition*-Daten vor. Es gibt aber auch *eye-tracking*-Daten (z.B. Ehrlich & Rayner, 1983) und Lesezeit-Daten (z.B. Garnham et al., 1992; Stevenson & Vitkovitch, 1986), die ebenfalls gegen eine frühe Aktivierung von Gender-Informationen sprechen. Van Gompel und Liversedge (2003) argumentieren ähnlich in Bezug auf die kataphorische Verwendung von Pronomen.

Zusammenfassend kann festgehalten werden, dass es ziemlich umfassende und starke empirische Evidenz dafür gibt, dass Gender-Informationen entgegen der intuitiven Überzeugung *nicht* früh im pronominalen Auflösungsprozess verwendet werden. Einflüsse des Gender-Cues sind lediglich am Satzende nachweisbar, die initiale Satzverarbeitung wird von ihm hingegen nicht beeinflusst. Die offensichtliche Nichtbeachtung von Gender-Informationen ist sehr erstaunlich, weil der Gender-Cue doch ein morpho-syntaktisch vergleichsweise eindeutig kodiertes und reliables Merkmal ist, das sich zudem leicht extrahieren lässt. Noch verwirrender ist, dass sich gleichzeitig nachweisen lässt, dass morpho-syntaktisch realisierte Numerus-Informationen bereits sehr früh im pronominalen Auflösungsprozess genutzt werden (de Vincenzi, 1999).

Greene et al. (1992) gehen sogar noch einen Schritt weiter: Im Rahmen der Minimalismus-Debatte vertreten sie die These, dass der Gender-Cue lediglich strategisch verarbeitet werden kann. Das heißt, gender-disambiguierte Pronomen wurden in ihrer Studie noch nicht einmal satzfinal aufgelöst, wenn „automatische“ Verarbeitungsbedingungen vorlagen. Erst als die zur Verfügung stehende Lesezeit ausreichend hoch angesetzt war und zusätzlich noch Verständnisfragen gestellt wurden, fanden sie (wie Gernsbacher, 1989) eine Voraktivierung des gender-kongruenten Referenten. Gender-Informationen sind ihrer Auffassung zufolge nicht hinreichend, um eine automatische Auflösung zu garantieren, sondern werden erst bei einer evtl. strategischen Auflösung genutzt. In dieser Hinsicht kontrastieren sie den Gender-Cue mit dem IC-Cue. Denn wie in Abschnitt 2.1.3.2 dargestellt, geht die Gruppe um McKoon et al. gleichzeitig davon aus, dass IC-Informationen automatisch verarbeitet werden. Denn wenn Gender- und IC-Informationen zusammenwirken, dann kann der Referent eines Pronomens auch durch einen rein gedächtnis-basierten Mechanismus eindeutig identifiziert werden.

Aparterweise stimmt die Gruppe um Garnham dem Minimalismus an dieser Stelle ausnahmsweise einmal zu (vgl. Garnham et al., 1992; Garham et al. 1996): Auch sie nehmen an,

dass Gender-Informationen strategisch verarbeitet werden, IC-Informationen jedoch automatisch. Sie stützen diese Behauptung damit, dass sie in ihren Lesezeit-Experimenten nur dann einen Einfluss des Gender-Cues nachweisen konnten, wenn die Gender-Informationen hinreichend salient und reliabel waren. Konkret fanden sie nur dann Gender-Effekte, wenn keine Filler-Items präsentiert wurden, die von den Gender-Informationen ablenkten, und wenn die Vpn Verständnisfragen bearbeiteten, die eine Auflösung des Pronomens verlangten. Wenn das Stimulusmaterial und die Untersuchungsprozedur die Bedeutung des Gender-Cues nicht unterstrichen, dann ließen sich auch keine Gender-Effekte finden. Garnham et al. interpretieren dies dahingehend, dass der Gender-Cue nur in einem satzfinalen Checking-Mechanismus genutzt wird, mit dem die bisher aufgebaute Diskursrepräsentation noch einmal überprüft wird. Gender-Informationen bieten sich ihrer Meinung nach für eine solche konfirmatorische Checking-Prozedur ganz besonders an, weil sie dem Text einfach und leicht entnommen werden können.

McDonald und MacWhinney (1995) fanden ebenfalls, dass Gender-Informationen im Laufe der pronominalen Auflösung erst spät genutzt werden. Sie begründen dies damit, dass Gender-Informationen für sich allein genommen selten hinreichend sind, um einen Referenten eindeutig zu identifizieren. Denn im Gegensatz zu den textuell verarmten Stimuli psycholinguistischer Experimente sind in natürlichen Diskurssituationen meist mehrere Aktanten (auch des gleichen Geschlechts) gleichzeitig aktiv. Von daher sind Gender-Informationen meist nur in *Kombination* mit anderen Faktoren (z.B. pragmatischer Art) nützlich. Deshalb ist es durchaus auch sinnvoll, dass der Auflösungs-Mechanismus mit ihrer Nutzung wartet, bis zusätzliche Informationen verfügbar werden.

Zusammenfassend lässt sich feststellen, dass zumindest die Evidenz für eine späte Nutzung von Gender-Informationen im Laufe der pronominalen Auflösung vergleichsweise gut ist. Das heißt, zumindest zur Etablierung bzw. Bestätigung des Verstehensproduktes am Ende eines Satzes ist der Gender-Cue für die pronominalen Auflösung relevant. Darüber hinaus gibt es Evidenz dafür, dass die Nutzung von Gender-Informationen von strategischen Prozessen beeinflusst werden kann oder zumindest mit diesen interagiert. Die meisten Autoren/innen gehen deswegen auch davon aus, dass Gender-Informationen v.a. bei der strategischen Analyse der Anapher während der *resolution*-Phase verwendet werden. Deswegen konnten in der minimalistischen Studie von Greene et al. (1992) auch kein satzfinalen Auflösungs-Effekte des Gender-Cues gefunden werden konnte: Es stand schlichtweg nicht genügend Zeit zur Verfügung, um solche ressourcen-intensiven *resolution*-Prozesse vornehmen zu können.

2.2.2 Gender-Information als *bonding*-Faktor

Wie sieht es aber mit dem frühen Einfluss von Gender-Merkmalen im Auflösungsprozess aus? War hier die anfängliche Intuition, dass sich das lexikalisch fest kodierte Genus-Merkmal sofort auf die Auflösung eines Pronomens auswirkt, wirklich so falsch? Gegen diese Annahme spricht zunächst einmal, dass sie vor dem Hintergrund der allgemeinen Architektur des Sprachverarbeitungssystems eher unplausibel wäre. So gibt es genügend Evidenz dafür, dass Gender-Informationen auch in Bezug auf andere Aspekte des Satzparsings bereits sehr früh genutzt werden können, z.B. bei der Verarbeitung syntaktisch ambiger Sätze (zumindest laut der *Construal Theory* von Frazier und Clifton, 1996, vgl. jedoch z.B. Mitchell, Cuetos, Corley & Brysbaert, 1995). Warum sollte die pronominale Auflösung hier eine Ausnahme darstellen? Darüber hinaus ist der gesamte Prozess der syntaktischen, semantischen und pragmatischen Verarbeitung eines Satzes ein hochkomplexer Vorgang, der zudem unter zeitlich extrem beschränkten Bedingungen operiert. Dies spricht dafür, dass die Satzverarbeitung hochinkrementell ist und alle verfügbaren Informationen sofort verwendet (vgl. hierzu auch Just & Carpenters, 1980, *immediacy hypothesis*). Gender-Informationen als lexikalisch kodiertes Merkmal sind sogar extrem früh verfügbar (wofür u.a. lexikalische Priming-Daten sprechen, vgl. im Überblick Friederici & Jacobsen, 1999), warum also sollten sie nicht auch bei der pronominalen Auflösung früh genutzt werden?

Es gibt auch empirische Evidenz dafür, dass Gender-Informationen bereits früh in den pronominalen Auflösungsprozess eingreifen. So fand z.B. Shillcock (1982) in einem cross-modalen Priming-Paradigma sehr wohl eine differenzielle Voraktivierung des Referenten direkt auf dem Pronomen. Dieser Befund widerspricht direkt den Ergebnissen von Marslen-Wilson und Tyler (1980). Allerdings war der Referent in der Studie von Shillcock immer auch im aktuellen Aufmerksamkeits-Zentrum, wodurch die Aussagekraft dieser Studie eingeschränkt ist. Vonk (1984) fand in einer *eye-tracking*-Studie ebenfalls frühe Gender-Effekte direkt auf dem Pronomen. Allerdings gab es den Effekt nur, wenn die Vpn eine Verifikationsaufgabe bearbeiteten, bei einer pronominalen Auflösungsaufgabe jedoch nicht. Vonk führt dies darauf zurück, dass die Nutzung unterschiedlicher Auflösungs-Cues offensichtlich von den unterschiedlichen Aufgabencharakteristika moderiert wird.

Die Ergebnisse dieser frühen Studien sind aufgrund der genannten Konfundierungen also noch mit einer gewissen Vorsicht zu betrachten. Studien, die in jüngerer Zeit vorgelegt wurden, sprechen hier eine deutlichere Sprache. So fanden z.B. Arnold, Eisenband, Brown-Schmidt & Trueswell (2000) in einer *visual-world*-Studie sehr frühe Gender-Effekte. In die-

sem Paradigma bekommen die Vpn auditiv einen Stimulussatz vorgespielt und blicken währenddessen auf eine visuelle Szene (meist auf einem Computer-Bildschirm). Gleichzeitig werden ihre Blickbewegungen gemessen. Stimuli waren Satzpaare wie in (19), und die entsprechenden Bilder zeigten entweder die Aktanten in der beschriebenen Konstellation oder nicht:

- (19) Donald is bringing some mail to Mickey/Minnie while a violent storm is beginning.
He's/She's carrying an umbrella, and it looks like they're both going to need it.

Die Vpn fixierten jeweils sehr früh (d.h. 200 ms nach dem Pronomen) die durch den Gender-Cue indizierte Person und konnten dadurch entscheiden, ob der Satz der dargestellten Szene entspricht oder nicht.

Für eine schnelle Nutzung von Gender-Informationen bei der pronominalen Auflösung sprechen aber auch verschiedene neuropsychologische Studien. So konnten z.B. van Berkum, Brown und Hagoort (1999b) in einer EKP-Studie nachweisen, dass (Relativ-)Pronomen, die gender-inkongruent sind, einen sofortigen syntaktischen Reanalyse-Prozess hervorrufen (vgl. auch van Berkum et al., 2007). In der Tat wären auch alle neuropsychologischen Studien, die für einen frühen IC-Effekt sprechen, ohne die gleichzeitige Annahme eine frühe Verarbeitung von Gender-Informationen uninterpretierbar: Denn in diesen Studien ist es ja gerade das Gender-Merkmal, welches IC-(In-)Kongruenz indiziert (vgl. z.B. Koornneef & van Berkum, 2006; Nieuwland & van Berkum, 2006). Eine weitere EKP-Studie von Osterhout, Bersick und McLaughlin (1997) macht darüber hinaus deutlich, dass sich frühe Gender-Mismatch-Effekte nicht nur auf das lexikalisch-grammatische Gender-Agreement beschränken. Sie fanden auch dann eine frühe syntaktische Reanalyse-Komponente, wenn sich ein weibliches Reflexiv-Pronomen auf eine stereotyp männliche Diskursrolle (Arzt) bezog. Aufgrund solcher und ähnlicher Befunde enthalten aktuelle neuropsychologische Sprachverarbeitungsmodelle wie z.B. das *extended Argument Dependency*-Modell von Bornkessel und Schlewsky (2006) auch eine frühe Gender-Checking- bzw. *agreement*-Zuweisungs-Funktion.

Inzwischen liegen aber auch sprachpsychologische Daten im engeren Sinne vor, die frühe Gender-Effekt bei der pronominalen Auflösung mit behavioralen Daten belegen: Rigalleau und Caplan (2000; Exp. 1) untersuchten z.B. Sätze, in denen lediglich ein Referent vorkam und ein nachfolgendes Pronomen entweder mit diesem gender-kongruent war oder nicht (*Brian cried in front of the grave; he/she had a tissue.*). Wenn Referent und Pronomen nicht gender-kongruent waren, dann brauchten die Vpn in einem crossmodalen *naming*-Paradigma länger, das Pronomen korrekt auszusprechen. In zwei weiteren Experimenten konnte sie zeigen, dass dies auch dann der Fall ist, wenn ein gender-kongruenter Referent pragmatisch unplausi-

bel ist (da die pragmatische Plausibilität durch IC-Kongruenz manipuliert wurde, stimmt dieser Befund mit den Ergebnissen von Koornneef & van Berkum, 2006, überein). In einer weiteren Studie (Rigalleau, Caplan & Baudiffier, 2004) replizierten sie diesen Befund noch einmal mit längeren Texten und einer *probe-recognition*-Methode.

Für die Fragestellung, ob gedächtnis- und erklärungs-basierte Verarbeitungsprozesse bei der pronominalen Auflösung miteinander interagieren, ist jedoch v.a. relevant, dass auch das gedächtnis-basierte Resonanz-Modell einen frühen Gender-Effekt vorhersagt (vgl. hierzu auch Kap. 1.2.1). Denn eine wichtige Eigenschaft des Resonanz-Prozesses ist, dass er passiv und untrestringiert ist. Alle Gedächtniselemente, die Merkmale mit dem gerade verarbeiteten Textelement teilen, werden aktiviert und fangen an zu resonieren. Sowohl Myers und O'Brien (1998) als auch O'Brien und Myers (1999) haben darauf hingewiesen, dass es generell bei Anaphern zu einem sog. Distraktor- oder Fan-Effekt kommt: Alle Gedächtniselemente, die ein Merkmal mit dem gerade verarbeiteten Textelement teilen, werden gleichzeitig und gleichmäßig aktiviert. Dies führt zu einer langsameren Konvergenzrate des Resonanzprozesses, weil sich die Signalstärke gewissermaßen auf alle möglichen Antezedenten aufteilt. Wie McKoon et al. (1996; vgl. auch Gerrig & McKoon, 2001; Greene, Gerrig, McKoon & Ratcliff, 1994) nachweisen konnten, gilt der Distraktor-Effekt auch für Pronomen: Das Gender-Merkmal eines Pronomens fungiert wie jedes andere Wortmerkmal auch als Abruf-Cue im Resonanz-Prozess und alle potenziellen Referenten resonieren mit ihm. Wenn zwei Referenten das gleiche Merkmal tragen, dann teilt sich das Signal auf beide gleichmäßig auf. Wenn nicht noch weitere Merkmale zum Gender-Cue hinzutreten, dann werden beide Referenten gleich stark aktiviert und es kann gedächtnis-basiert keine Auflösung vorgenommen werden. Allerdings scheint ein eindeutiger Gender-Cue allein auch nicht für eine differenzielle Voraktivierung des Antezedens auszureichen (vgl. Greene et al., 1992). Erst wenn noch weitere Faktoren zu diesem hinzutreten, ist eine automatische Auflösung möglich (IC-Informationen bei McKoon et al., 1993; Fokus-Informationen bei Shillcock, 1982).

Der Befund, dass der Gender-Cue allein zur Identifikation eines pronominalen Referenten nicht ausreicht, ist jedoch vielleicht auch der Tatsache geschuldet, dass die meisten sprachpsychologischen Erkenntnisse lediglich mit Material in der englischen Sprache durchgeführt werden: Das Englische ist eine Sprache mit einem extrem verarmten Genus-System, d.h. Gender-Informationen werden kaum morphologisch realisiert (vgl. Comrie, 1999). Von daher ist es gut möglich, dass ein Teil der berichteten Null-Effekte darauf zurückzuführen ist, dass Genus-Merkmale einfach aufgrund ihrer fehlenden Oberflächenrealisierung nicht eindeutig angesprochen werden konnten (vgl. hierzu die Ergebnisse von Garnham, Oakhill, Ehrlich

& Carreiras, 1995, bzw. Carreiras, Garnham, Oakhill & Cain, 1996, die frühe Gender-Effekte im Französischen und Spanischen nachweisen konnten; Cacciari, Carreiras & Cionini, 1997, zeigen ähnliches für das Italienische).

Die empirische Befundlage ist also uneinheitlich. Ein Teil der berichteten Studien fand keine frühen Gender-Effekte, andere sprechen jedoch dafür, dass Gender-Informationen sogar sehr schnell genutzt werden können. Diese Diskrepanz lässt sich durch die Beobachtung auflösen, dass sich die verschiedenen Studien auf unterschiedliche empirische Sachverhalte beziehen. Studien wie die von Greene et al. (1992), die keine frühen Effekte des Gender-Cues feststellen konnten, untersuchten die Fragestellung, ob der Gender-Cue zu einer (relativen) Voraktivierung des gender-kongruenten Referenten führt. Dies wird in einer experimentellen Situation überprüft, in der entweder nur einer oder aber zwei gender-kompatible Referenten zur Verfügung stehen. Wie dargestellt, ist die Evidenz für einen solchen Gender-Fokus-Effekt eher schwach. In Studien wie der von Rigalleau et al. (2004), in der ein früher Effekt von Gender-Informationen nachgewiesen werden konnte, wird hingegen die Fragestellung untersucht, ob die *Inkongruenz* von Pronomen und Antezedent zu frühen Verarbeitungsschwierigkeiten führt. Wie dargestellt, gibt es vergleichsweise starke Evidenz dafür, dass solche Sätze als problematisch erlebt werden und die Anomalie sofort entdeckt wird (vgl. Garnham, 2001).

Mit McKoon et al. kann deshalb festgestellt werden, dass Gender-Informationen v.a. die Funktion haben, die zentrale Stellung eines vom Diskurs sowieso schon bereitgestellten Aktanten noch einmal zu bestätigen. Gender-Informationen alleine reichen jedoch nicht dazu aus, diesen Diskursreferenten zu etablieren. Friederici und Jacobsen (1999, p. 482) vertreten eine sehr ähnliche Position (vgl. auch van Berkum et al., 1999):

From the available data on gender priming, we conclude that during normal language comprehension gender information does not preselect particular (gender agreeing) lexical candidates, but that a postlexical checking mechanism evaluates the gender congruency of the incoming element.

Genau diese frühe Prüfung auf Gender-Kongruenz ist das, was im Rahmen des *bonding-resolution*-Ansatzes von Garrod und Sanford (1989) als *bonding* bezeichnet wird. Denn in der *bonding*-Phase der pronominalen Auflösung wird ja gerade evaluiert, ob ein bestimmtes Antezedens (z.B. aufgrund seiner morpho-syntaktischen Merkmale) als Bindungsstelle für das Pronomen in Frage kommt. Wenn dies nicht der Fall ist, dann wird das *bond* direkt aus der möglichen Kandidatenmenge eliminiert. Die Etablierung eines *bonds* allein ist jedoch nicht hinreichend, um eine nachherige Auflösung des Pronomens zu gewährleisten. Hierzu ist der ressourcenintensive *resolution*-Prozess notwendig.

2.2.3 Diskussion

Insgesamt liegt damit für den Gender-Cue eine recht ähnliche Situation vor wie für den IC-Effekt. In der relevanten Literatur werden zwei Erklärungsmodelle des Gender-Effekts gegeneinander gestellt: Eine Position geht davon aus, dass Gender-Informationen lediglich spät in den pronominalen Auflösungsprozess eingreifen und der strategischen Verarbeitung bedürfen, welche v.a. die die satzfinale Integration betrifft. Die Gegenposition geht jedoch davon aus, dass Gender-Informationen bereits sehr früh in der pronominalen Auflösung verwendet werden. Dies macht sich u.a. in frühen Mismatch-Effekten bemerkbar. Wie ausführlich dargestellt, gibt es Evidenz für beide Positionen. Dies liegt v.a. daran, dass sie sich auf unterschiedliche empirische Sachverhalte beziehen: Für den Integrations-Effekt ist v.a. die Passung der Gender-Informationen in das bislang etablierte Diskursmodell entscheidend. Dieses lässt sich anscheinend besser und schneller aufbauen bzw. evaluieren, wenn lediglich ein gender-kompatibler Referent im kohärenz-stiftenden Auflösungsprozess bearbeitet werden muss. Für den Mismatch-Effekt ist hingegen relevant, ob das Pronomen überhaupt mit dem gerade aktiven Aktanten im Diskursmodell kompatibel ist. Wenn kein oder nur ein pragmatisch unplausibler Referent vorhanden ist, dann ergeben sich sehr früh Verarbeitungsschwierigkeiten und es werden entsprechende repair-Prozesse notwendig. In Anschluss an Rigalleau et al. (2000; 2004) muss deshalb festgestellt werden, dass es durchaus möglich, wenn nicht sogar wahrscheinlich ist, dass beide Effekte zusammen und gleichzeitig existieren. Der frühe Mismatch-Effekt (den Rigalleau et al. *confirmation effect* nennen) operiert im Rahmen der frühen *bonding*-Phase und evaluiert mögliche Bindungsstellen für eine Pronomen. Der Integrations-Effekt (den Rigalleau et al. *disengagement effect* nennen) gehört jedoch zur späten *resolution*-Phase, welche die etablierten *bonds* in Hinblick auf ihre Integrierbarkeit beurteilt und die finale Auflösung vornimmt.

Allerdings werden in der relevanten Literatur nicht nur ein, sondern zwei frühe Effekte diskutiert. Bei dem einen handelt es sich um einen frühen Fokus-Effekt, der v.a. für die Vertreter/innen einer gedächtnis-basierten Position relevant ist. Hier geht es um die Frage, ob ein disambiguierender Gender-Cue zu einer differenziellen Voraktivierung eines der beiden Aktanten führt, weshalb dieser allein durch automatische Prozesse als Referent des Pronomens ausgewählt werden kann. Wie gesehen gibt es wenig Evidenz dafür, dass es zu einer solchen differenziellen Voraktivierung kommt. Der andere frühe Gender-Effekt ist der Mismatch-Effekt, der prüft, ob das Pronomen in Bezug auf seine Gender-Informationen mit dem derzeit salientesten oder plausibelsten Diskursreferenten kongruent ist.

Handelt es sich hierbei eigentlich um zwei oder vielleicht nur um einen Effekt? Diese Frage lässt sich nur schwer beantworten, da beide ja nur in unterschiedlichen Situationen und teilweise auch nur mit unterschiedlichen Methoden erfasst werden können. Es macht wenig Sinn, den Aktivierungsgrad verschiedener Referenten mittels *probe-recognition* zu testen, wenn sich das Pronomen aus Gründen mangelnden Gender-Agreements auf keinen von ihnen bezieht. Und umgekehrt kann man lange nach einer P600-Komponente suchen, wenn keine Agreement-Verletzung vorliegt. Auch von ihrer Charakteristik her unterscheiden sich die beiden Prozesse erheblich voneinander: Bei dem einen handelt es sich um einen allgemeinen Resonanz-Prozess, der zu einem bestimmten Aktivierungsmuster in einem assoziativ-semanticen Netzwerk führt. Bei dem anderen handelt es sich um eine spezifische *Prüfprozedur*, die gezielt nur die für diesen Prozess relevanten Informationen sucht und evaluiert. Allerdings ließe sich durchaus darüber spekulieren, dass sie identisch sind: So könnte sich der Mismatch-Effekt gerade daraus ergeben, dass das Pronomen sein Gender-Merkmal in die episodische Textspur „funkt“ und keine (hinreichend starke) Antwort zurückkommt. Wenn jedoch mindestens ein Antezedens ein kongruentes Gender-Merkmal „zurücksendet“, dann kann ein *bond* etabliert (und in der *resolution*-Phase evaluiert) werden. Das Signal selbst muss aber nicht stark genug sein, um zu einer differenziellen Voraktivierung des Referenten zu führen oder sogar seine Auflösung zu gewährleisten. Wie gesagt, ist dies jedoch alles reine Spekulation. Um solche Vermutungen auch nur ansatzweise prüfen zu können, müsste zumindest die Gender-Kongruenz des Pronomens manipuliert werden, was jedoch in Studien zum IC-Effekt eigentlich fast nie getan wird.

Wichtiger ist es, auch hier wieder darauf hinzuweisen, dass die beiden Effekte nicht nur gleichzeitig, aber unabhängig voneinander existieren, sondern vielmehr aufeinander bezogen sind und sich gegenseitig voraussetzen: Ohne die (frühe) Etablierung einer *bonding*-Landestelle kann später keine Auflösung vorgenommen werden. Umgekehrt reicht aber die Etablierung eines *bonds* auch nicht aus, um eine erfolgreiche Auflösung zu gewährleisten. Hierfür muss das *bond* vielmehr später noch in die aktuelle Diskursrepräsentation integriert werden, evtl. mittels kohärenz-stiftender Auflösungsinferenzen. Dabei ist durchaus möglich, dass – wie z.B. Rigalleau et al. (2004) mutmaßen – die Leichtigkeit der späteren Auflösung von der Effizienz des frühen *bondings* abhängig ist. Dies ist v.a. dann wahrscheinlich, wenn das *bonding* nicht nur zu einer differenziellen Voraktivierung der potentiellen Referenten führt, sondern gleichzeitig Nicht-Referenten unterdrückt.

2.3 Interaktion der beiden Cues und Hypothesen

In Kap. 1 wurde ein integratives Rahmenmodell der pronominalen Auflösung vorgestellt, das davon ausging, dass für die pronominale Auflösung sowohl gedächtnis-basierte als auch erklärungs-basierte Prozesse relevant sind. Die beiden Prozessklassen setzen sich gegenseitig voraus und interagieren dynamisch miteinander: Frühe, hierarchie-niedrige Gedächtnisprozesse dienen dazu, die finale, gedächtnis-basierte Integration vorzubereiten. Vor dem Hintergrund dieses allgemeinen theoretischen Rahmens scheint die theoretische Rekonstruktion sowohl des IC- als auch des Gender-Effekts sehr viel versprechend zu sein. Denn für beide Cues werden ja in der Diskussion zwei unterschiedliche Erklärungsmodelle vorgeschlagen, welche ihre Effekte entweder auf hierarchie-niedrige, gedächtnis-basierte oder aber hierarchie-hohe, erklärungs-basierte Prozesse zurückführen wollen. Beim IC-Effekt wird dabei eine Fokus-Erklärung einer Integrations-Erklärung gegenübergestellt; beim Gender-Effekt ein früher *confirmation*-Effekt einem späten *disengagement*-Effekt. Es konnte jedoch gezeigt werden, dass es empirisch ziemlich wahrscheinlich ist, dass für die Effekte der beiden Auflösungs-Cues nicht nur einer der beiden Prozesse verantwortlich ist, sondern beide gleichzeitig. Darüber hinaus wurde dafür argumentiert, dass die verschiedenen Teilprozesse nicht nur unabhängig voneinander auftreten, sondern eng aufeinander bezogen sind, wofür insbesondere Überlegungen des *bonding-resolution*-Ansatzes von Garrod und Sanford (1989) herangezogen wurden. Zusammengefasst bedeutet das, dass es in der Literatur Anhaltspunkte dafür gibt, dass die eingangs skizzierte integrative Rahmenperspektive – zumindest in Bezug auf die beiden diskutierten Auflösungs-Cues – theoretisch und empirisch plausibel ist.

Auf der anderen Seite wurde am Ende des Einleitungskapitels aber auch darauf hingewiesen, dass zur empirischen Überprüfungen der postulierten dynamischen Interaktion von gedächtnis- und erklärungs-basierten Verarbeitungsprozessen bei der pronominalen Auflösung eine experimentelle Ausgangssituation benötigt wird, in der spezifische Textmerkmale dazu in der Lage sind, beide Prozessarten auch differenziell und selektiv hervorzurufen. Und am Anfang des vorliegenden Kapitels wurde dafür argumentiert, dass mit dem IC- und dem Gender-Effekt zwei Textmerkmale vorliegen, welche dies evtl. gewährleisten. Denn zumindest anfänglich wurde ja der IC-Cue eher als ein hierarchie-hoher, pragmatischer und primär die Satzintegration betreffender Auflösungsfaktor betrachtet, während für den lexikalischen Gender-Cue eine vergleichsweise hierarchie-niedrige, morpho-syntaktische und primär die initiale pronominale Analyse betreffende Verarbeitung angesetzt wurde. Wie skizziert lässt sich dieses vereinfachte Bild nicht halten: Beide Cues beeinflussen sowohl hierarchie-hohe als

auch hierarchie-niedrige Prozesse und beeinflussen sich dabei gegenseitig. Was theoretisch schön und wünschenswert ist, scheint nun empirisch problematisch zu sein, da von einer differenziellen und selektiven Beeinflussung nicht mehr gesprochen werden kann. Wenn beide Cues beide Prozesse hervorrufen, wie kann man diese dann noch auseinander halten?

Die Antwort auf diese Frage erhält man, indem man abschließend die Interaktion der beiden Cues betrachtet. Für den Gender-Effekt (bzw. die Gender-Effekte) bringt das zunächst einmal wenig. Denn hier lassen sich keine Vorhersagen generieren, warum sich die beiden Teileffekte (*confirmation* vs. *disengagement*) in Abhängigkeit vom IC-Cue unterschiedlich verhalten sollten: Egal ob mit oder ohne IC-Cue, der Gender-Effekt sollte immer eine frühe, hierarchie-niedrige Checking- und eine späte, hierarchie-hohe Integrations-Komponente umfassen. Wollte man die Prozesse noch weiter voneinander trennen, dann wären weitere Manipulationen notwendig, z.B. die Einführung von pronominalen Konstellationen, in denen es zu einem Gender-Mismatch von Pronomen und Referent kommt (vgl. z.B. Rigalleau & Caplan, 2000; Rigalleau et al., 2004).

Interessant ist jedoch, was die Interaktion der beiden Cues für die beiden IC-Effekte bedeutet. Nicht umsonst war ja in der gesamten Diskussion um den IC-Effekt (seit Caramazza et al., 1977) die Frage bezüglich der Interaktion der beiden Faktoren absolut zentral. Angesichts der Tatsache, dass Caramazza et al. (1977) nachgewiesen haben, dass die beiden Cues nicht miteinander interagieren, scheint es zunächst einmal auch hier nicht möglich zu sein, differenzielle Vorhersagen für die beiden IC-Effekte zu machen. Wenn man sich allerdings die Literatur zum IC-Effekt etwas genauer ansieht, dann offenbaren sich ein paar interessante Muster. Eine Betrachtung der Interaktion der beiden Cue-Effekte verkompliziert sich allerdings dadurch, dass nicht in allen Untersuchungen alle möglichen Kombinationen des IC- und des Gender-Cues erhoben worden sind. So wird z.B. bei McKoon et al. (1993) der IC-Effekt nur untersucht, wenn das Pronomen gleichzeitig gender-disambiguiert ist, während z.B. Ehrlich (1980) nur den Gender-Cue manipuliert, den IC-Cue jedoch konstant hält. Wenn man die Studien, welche die Bestimmung der Interaktion der beiden Faktoren überhaupt zulassen, einmal durchgeht, dann sieht man, dass die Mehrheit keine Interaktion der beiden Faktoren findet (z.B. Springstone, 1976; Caramazza et al., 1977; Vonk, 1984; Vonk, 1985a; Stewart et al., 2000). Ein paar berichten hingegen (nur) dann einen IC-Effekt, wenn das Pronomen gleichzeitig gender-disambiguiert ist (z.B. Garnham & Oakhill, 1985, Garnham et al., 1992; McDonald & MacWhinney, 1995), andere wiederum berichten jedoch (nur) dann einen IC-Effekt, wenn das Pronomen nicht gender-disambiguiert ist (z.B. Garnham et al., 1996). Eine solche empirische Befundlage, die sich größtenteils aus Null-Befunden, zum anderen aber aus

heterogenen Einzelbefunden in unterschiedliche Richtungen zusammensetzt, ist zwar in der Psychologie nicht ungewöhnlich, in diesem Fall aber doch sehr bemerkenswert: Es kann vermutet werden, dass das Interaktionsmuster der beiden Cues von weiteren Variablen moderiert wird.

Weiterhin ist auffällig, dass alle Studien, die einen frühen IC-Fokus-Effekt nachgewiesen haben, entweder immer nur mit einer gender-disambiguierten Bedingung gearbeitet haben (McKoon et al., 1993; Koornneef & van Berkum, 2006) oder aber den Effekt lediglich beobachteten, wenn gleichzeitig ein eindeutiger Gender-Cue vorlag (McDonald & MacWhinney, 1995; Garnham et al., 1996). Eine weitere Studie von Garrod et al. (1994), die zwar nicht den IC-Effekt direkt, aber eine pragmatische Unplausibilität durch Verbinformationen des Nebensatzes untersuchte, kommt zu dem gleichen Ergebnis. Unter der Voraussetzung, dass es sich beim IC-Fokus-Effekt entweder um einen Checking-Mechanismus beim *bonding* handelt oder aber um einen schnellen Refokussierungsprozess, macht das ja auch Sinn: Man kann ja nur dann die Inkongruenz von Verb-Bias und nachfolgender Auflösung feststellen, wenn gleichzeitig eine andere, unabhängige und verlässliche Informationsquelle vorhanden ist, die einem sagt, dass es sich um eine Inkongruenz handelt. Im Rahmen eines eher diskursmodellorientierten Ansatzes entspricht dies der Feststellung, dass der Gender-Cue dabei hilft, die einzelnen Aktanten hinreichend distinkt zu machen, sodass sie leichter zu identifizieren sind und schneller mit dem relevanten Weltwissen in Verbindung gebracht werden können (Garnham & Oakhill, 1985; Garnham et al. 1992). Auch im Rahmen des SMF-Ansatzes von Sanford und Garrod ist davon auszugehen, dass gender-disambiguierte Aktanten schneller vom expliziten in den impliziten Fokus gemappt werden können. Die spätere Integration ist damit gewissermaßen schon vorbereitet und kann schneller und leichter vorgenommen werden.

Für die Bedingung ohne Gender-Cue ist hingegen von einer anderen Konstellation auszugehen. Hier ging ja schon die anfängliche Auffassung eher dahin, dass Integrationsschwierigkeiten aufgrund von IC-Inkompatibilität v.a. (bzw. nur) dann auftreten sollten, wenn keine disambiguierenden Gender-Informationen zur Verfügung stehen. Wie zahlreich gezeigt, war die Annahme, dass der IC-Effekt in diesem Fall vollkommen verschwindet, zu stark formuliert. Wir wissen inzwischen, dass morpho-syntaktische Merkmale nicht ausreichend sind, um die durch die IC-Inkompatibilität hervorgerufenen Integrationsschwierigkeiten zu eliminieren. Im Rahmen des hier vertretenen integrativen Rahmenmodells der pronominalen Auflösung ist das plausibel: Das finale, durch hierarchie-niedrige Verstehensprozesse aufgebaute Verstehensprodukt muss abschließend einer Evaluation unterzogen werden, welche die Kohärenz des Texts gewährleistet und in Abhängigkeit von verschiedenen Text- und Lesermerkmalen

unterschiedlich „gründlich“ ausfallen kann. Die Intuition, die Caramazza et al. (1977) dazu gebracht hat, den Gender-Cue als Kontroll-Bedingung zu verwenden, war jedoch schon sehr richtig: Wenn ein Gender-Cue den Satz nicht bereits früh disambiguiert und dadurch entsprechende Reanalyse-Prozesse in Gang bringt, dann trägt die satzfinale Integration die Hauptlast der schwierigeren Integration. Ohne Gender-Cue muss der gesamte Verstehensprozess in der strategischen Phase gewissermaßen „nachgeholt“ werden. Früher Fokus- und später Integrationseffekt stehen also in der Tat in einem wechselseitigen Abhängigkeitsverhältnis und ergänzen sich wechselseitig: Wenn – bei Vorliegen eines eindeutigen Gender-Cues – bereits früh mit der Verarbeitung begonnen (bzw. ein stärkeres *bond* etabliert) werden kann, dann ist der Integrationsprozess weniger aufwändig. Wenn hingegen keine frühe Verarbeitung vorgenommen werden kann (oder sogar ein falsches *bond* vertieft verarbeitet wird), dann ist die finale Integration umso aufwändiger. Eine solche trade-off-Beziehung verschiedener Prozesse ist jetzt aber für globale Maße, die entweder nicht den Zeitpunkt eines Prozesses oder aber nicht seine Zusammensetzung aus unterschiedlichen Teilprozessen erfassen, geradezu tödlich: Frühe, hierarchie-niedrige Prozesse und späte, hierarchie-hohe Prozesse mitteln sich gegenseitig aus und man kann in beiden Gender-Bedingungen nur noch feststellen, dass ein Effekt vorlag. Dass dieser auf sehr unterschiedlichen Prozessen basiert, die zu unterschiedlichen Zeitpunkten unterschiedlich stark ausgeprägt sind, wird hingegen „übersehen“.

Zusammenfassend lässt sich damit feststellen, dass sich durch die Interaktion der beiden Faktoren unterschiedliche Vorhersagen bezüglich der beiden IC-Effekte machen lassen: Der IC-Fokus-Effekt sollte nur dann auftreten, wenn gleichzeitig ein disambiguierender Gender-Cue vorhanden ist, der entweder eine Refokussierung ermöglicht und damit die Integration vorbereitet. Der IC-Integrations-Effekt sollte zwar in beiden Gender-Bedingungen auftreten (Integration ist obligatorisch), allerdings in der Bedingung ohne Gender-Cue stärker ausfallen, weil zuvor keine oder die falschen Elaborationsprozesse stattgefunden haben. Damit ist abschließend das übergeordnete Ziel dieses Kapitels erreicht. Das in Kap. 1 noch vergleichsweise grob skizzierte Rahmenmodell der pronominalen Auflösung konnte in Bezug auf Gender- und IC-Informationen hinreichend konkretisiert werden, um vorhersagen zu können, wie, wann und warum hierarchie-niedrige und hierarchie-hohe Prozesse in den pronominalen Auflösungsprozess eingreifen. Tab. 2.4 stellt die verschiedenen postulierten Teilprozesse und ihre Relation noch einmal schematisch dar:

<i>Effekt</i>	<i>Teileffekt</i>	<i>Zeitpunkt</i>	<i>Prozesstyp</i>	<i>Interaktion Gender-Cue</i>
Gender	Confirmation	Früh	Hierarchie-niedrig	---
	Disengagement	Spät	Hierarchie-hoch	---
IC	Fokus	Früh	Hierarchie-niedrig	Stärker, wenn disambiguiert
	Integration	Spät	Hierarchie-hoch	Stärker, wenn nicht disambiguiert

Tab. 2.4: Vorhersagen für die verschiedenen Teileffekte des Gender- und des IC-Manipulation.

Beim Gender-Effekt wird zwischen einer hierarchie-niedrigen, frühen *confirmation*-Komponente und einer hierarchie-hohen, späten *disengagement*-Komponente unterschieden. Hinsichtlich des IC-Cues gilt ähnliches: Hier steht ein hierarchie-niedriger, früher Fokus-Effekt einem späten, hierarchie-hohen Integrationseffekt gegenüber. Die vorangegangene Diskussion hat ebenfalls deutlich gemacht, dass zumindest für den IC-Effekt mit Interaktionen der beiden Faktoren zu rechnen ist. Der frühe IC-Fokus-Effekt kann sich prinzipiell nur dann auswirken, wenn überhaupt eine Inkongruenz-Relation durch einen gleichzeitig anwesenden Gender-Cue signalisiert wird. Für den IC-Integrations-Effekt ist hingegen anzunehmen, dass er besonders stark ausfällt, wenn keine vorherige Verarbeitung stattfinden konnte. IC-Fokus- und IC-Kongruenz-Effekt stehen damit zumindest partiell in einer trade-off-Relation.

Es ist allerdings nicht einfach, dass in Tab. 2.4 zusammengefasste theoretische Modell empirisch zu überprüfen. Das liegt v.a. an der postulierten trade-off-Relation zwischen dem IC-Fokus- und dem IC-Integrations-Effekt, die in Abhängigkeit von der Ausprägung des Gender-Cues jeweils unterschiedlich ausfällt. Wie bereits festgestellt, können globale Erhebungsmaße eine solche Konstellation nicht erfassen. Vielmehr sind noch weitere Spezifikationen notwendig, entweder zum zeitlichen Verlauf der Effekte oder bezüglich der Effektzusammensetzung. Entsprechend finden sich in Tab. 2.4 auch noch zwei weitere Spalten, welche die aufgelisteten Teileffekte noch weiter spezifizieren und voneinander dissoziieren.

Die erste Dissoziations-Dimension betrifft den *Zeitpunkt*, an dem sich die verschiedenen Teileffekte nachweisen lassen sollten. So kann theoretisch z.B. aus der Zwei-Phasen-Struktur des *bonding-resolution*-Ansatzes (aber nicht nur diesem) abgeleitet werden, dass hierarchie-niedrige *bonding*-Prozesse früh im Satzverlauf auftreten sollten, während hierarchie-hohe *resolution*-Prozesse vornehmlich satzfinal zum tragen kommen. Und entsprechend ist davon auszugehen, dass sich der Gender-*confirmation*-Effekt und der IC-Fokus-Effekte, als frühe *bonding*-Prozesse direkt auf dem Pronomen oder zumindest kurz danach nachweisen lassen

sollten. Der *Gender-disengagement*- und der IC-Integrations-Effekt sind hingegen *resolution*-Prozesse, die lediglich satzfinal auftreten.

Diese grundsätzliche Gleichsetzung von Verarbeitungsprozess und Verarbeitungszeitpunkt (*bonding* = hierarchie-niedrig = früh bzw. *resolution* = hierarchie-hoch = spät) liegt den meisten psycholinguistischen Methoden zugrunde, die keine unabhängige Erhebung verschiedener Prozesskomponenten vornehmen (können) und diese indirekt erschließen müssen. So wird z.B. in einer *self-paced-reading*-Studie lediglich eine AV erhoben, nämlich die zum Lesen eines bestimmten Textsegmentes benötigte Zeit. Diese ist jedoch zunächst einmal nur ein unspezifischer online-Indikator, der wenig über die zugrunde liegenden Prozesse sagt (vgl. auch Richter, 2003). In den meisten Studien wird deshalb über die Position des Effekts auf den zugrunde liegenden Prozess geschlossen: Finden sich z.B. erhöhte Lesezeiten am Satzende, dann handelt es sich um einen Integrations-Effekt (vgl. z.B. Stewart et al., 2000), finden sich hingegen erhöhte Lesezeiten bereits auf dem Pronomen, dann handelt es sich um einen Fokus-Effekt (z.B. Koornneef & van Berkum, 2006). Bei *probe-recognition*-Studien ist die Lage prinzipiell ähnlich. Auch hier wird nur eine AV erhoben (differenzielle Voraktivierung des Referenten) und die Interpretation ergibt sich durch die Position: Findet sich bereits ein früher Effekt, dann spricht dies für einen hierarchie-niedrigen Prozess, ist er hingegen spät lokalisiert, dann könnte es auch ein hierarchie-hoher Prozess sein.

Auch wenn die genannte Gleichsetzung von Verarbeitungsprozess und Verarbeitungszeitpunkt in der Sprachpsychologie durchaus üblich ist, so ist sie auch nicht unproblematisch. Dadurch, dass sie zwei unabhängige Dimensionen miteinander konfundiert, sind evtl. Unterschiede hinsichtlich des zugrunde liegenden Prozesstyps nicht mehr erfassbar, wenn sie zum gleichen Zeitpunkt ausgeführt werden. Zwar ist kaum zu bestreiten, dass es eine (sachlogische) Korrelation zwischen Prozesstyp und Verarbeitungszeitpunkt gibt, die sich aus der Architektur des Sprachverarbeitungssystem ergibt. Allerdings ist nicht davon auszugehen, dass sie generell gilt: So ist es durchaus möglich, dass auch hierarchie-niedrige Prozesse am finalen *resolution*-Prozess beteiligt sind (z.B. dadurch, dass automatisch Wissensbestände aktiviert werden, welche zur Evaluation der gebildeten Diskursrepräsentation verwendet werden, vgl. Long & Lea, 2005). Interessant ist jedoch auch der umgekehrte Fall, nämlich dass hierarchie-hohe Integrationsprozesse bereits früh wirksam werden und der Aufbau des Diskursmodells streng inkrementell erfolgt. Eine solche Position wird für den IC-Effekt u.a. von Koornneef und van Berkum (2006) vertreten. Aber auch in Bezug auf den Gender-Effekt wurden entsprechende Vorschläge gemacht (vgl. z.B. Friederici & Jacobsen, 1999). An dieser Stelle wird deutlich, dass die Gleichsetzung von Verarbeitungsprozess und Verarbeitungszeitpunkt

eine *Voraussetzung* ist, die alle genannten Verfahren implizit machen müssen. Wie jede Voraussetzung kann sie im Einzelfall gerechtfertigt sein oder auch nicht. Ob sie angemessen ist, lässt sich zwar theoretisch begründen, muss jedoch empirisch jeweils überprüft werden (können). Dafür braucht noch eine weitere, zum Verarbeitungszeitpunkt orthogonale Dissoziations-Dimension. Dies ist insbesondere für den Fall wünschenswert, wenn eine Methode verwendet werden soll, die keine Unterscheidung zwischen unterschiedlichen Verarbeitungszeitpunkten zulässt.

Die zweite in Tab. 2.4 genannte Dissoziations-Dimension betrifft deswegen auch den *Prozesstyp* direkt, welcher den verschiedenen Teileffekten zugrunde liegt. Sowohl der *Gender-confirmation*-Effekt als auch der IC-Fokus-Effekt basieren auf vergleichsweise hierarchie-niedrigen, primär gedächtnis-basierten Teilprozessen und sollten sich entsprechend auch nur in solchen Maßen niederschlagen, die solche Prozesse erfassen. Der *Gender-disengagement*-Effekt und der IC-Integrations-Effekt sind jedoch inhärent erklärungs-basierte Prozesse. Sie sollten sich Prozessmaßen abbilden, die selektiv hierarchie-hohe Verarbeitungskomponenten erfassen. Das zentrale methodische Problem, was sich an dieser Stelle ergibt, besteht darin, zwei Prozessmaße zu finden, mit denen sich gedächtnis- und erklärungs-basierte Verarbeitungsprozesse gleichzeitig, aber auch trennscharf erfassen lassen.

Dies ist allerdings eine sehr starke Anforderung, die von den meisten sprachpsychologischen Methoden nicht erfüllt wird. Diese erlauben entweder keine Differenzierung zwischen den verschiedenen Prozesstypen oder aber sie verwenden lediglich eine AV, die ihre gedächtnis- vs. erklärungs-basierte Interpretation durch andere Kontextfaktoren erhält (z.B. dem Erhebungszeitpunkt s.o.). Die einzige Methode, die zumindest ansatzweise die geschilderten Anforderungen erfüllt, sind m.E. *eye-tracking*-Daten. Dies allerdings auch nur unter der Voraussetzung, dass die traditionelle Unterscheidung zwischen Erstlese-Zeiten und Reanalyse-Prozessen (Regressionen etc.) die beiden Prozessklassen auch adäquat und verzerrungsfrei abbildet (was nicht unbedingt gegeben sein muss, vgl. z.B. Mitchell, 2004). Neuropsychologische Verfahren wie z.B. EKP-Studien konzentrieren sich hingegen von vorneherein nur auf vergleichsweise frühe Prozesse und können eher stichpunktartig bestimmte Teilaspekte des Verarbeitungsprozesses beleuchten (Führt ein ic-inkompatibles Pronomen zu einer direkten Reanalyse?). Sie besitzen aber eine zu hohe Auflösung, um den Prozess der Satzverarbeitung in seiner Gesamtheit abbilden zu können. Was stattdessen benötigt wird, ist ein Verfahren, welches a) einen Einblick in die Satzverarbeitung als Ganzes gibt, b) hierarchie-niedrige und -hohe Prozesse *gleichzeitig* erfassen kann, und c) die einzelnen Prozesse nicht mit ihrem Auftreten innerhalb des Verarbeitungsprozesses konfundiert bzw. sogar eine positionsunabhängig

Erfassung zulässt. Darüber hinaus sollte sich natürlich d) plausibel herleiten lassen, dass die einzelnen AVn in der Tat als differenzielle Indikatoren gedächtnis- und erklärungs-basierter Prozesse rekonstruiert werden können.

Da eine Methode, die alle vier Voraussetzungen gleichzeitig erfüllt, nach meinem Wissen bisher nicht existiert, bleibt einem wenig anderes übrig, als ein bereits bestehendes Verfahren gewissermaßen zu „kapern“ und durch geschickte Auswertungstechniken das Beste aus ihnen zu machen, was in dieser Hinsicht möglich ist. Konkret sollen dabei in der vorliegenden Arbeit zwei sehr unterschiedliche Erhebungsmethoden zum Einsatz kommen: Einmal eine pronominale Auflösungsaufgabe wie bei Caramazza et al. (1977), die lediglich eine globale Analyse der Prozesse erlaubt (Experiment I). Zum anderen soll aber auch ein *self-paced-reading*-Verfahren eingesetzt werden, welches zusätzlich die Unterscheidung verschiedener Verarbeitungszeitpunkte ermöglicht (Experiment II). Beide Verfahren generieren zunächst einmal nur eine einzige AV, nämlich Reaktions- bzw. Lesezeiten. Im nächsten Kapitel soll deswegen ein Auswertungsverfahren vorgestellt werden, mit dem aus dieser „monistischen“ AV zwei unabhängige Indikatoren extrahiert werden können, die selektiv den Anteil an gedächtnis- und erklärungs-basierter Verarbeitung erfassen.

3. Erfassung gedächtnis- und erklärungs-basierter Verarbeitungsprozesse mittels Reaktionszeitverteilungsanalysen

3.1 Reaktionszeitmethoden in der Psychologie: Möglichkeiten und Grenzen

In der Psychologie werden häufig Reaktionszeiten (engl. *reaction time* oder auch *response time*, deswegen im Folgenden mit RT abgekürzt) gemessen, um bestimmte (latente) psychische Zustände, Merkmale, Prozesse, Fähigkeiten etc. zu erfassen. Dabei versteht man unter dem Begriff *RT* zumeist ziemlich pauschal die Zeit, die eine Person braucht, um in einem psychologischen Experiment eine bestimmte Aufgabe zu bearbeiten. Obwohl RT-Methoden eine lange Tradition in der Experimentellen Psychologie haben, die weit ins 19. Jh. zurückreicht (vgl. z.B. die Arbeiten von Donders und Wundt), werden sie v.a. seit der sog. Kognitiven Wende der 1960er Jahre häufig verwendet (vgl. Gardner, 1985). Entsprechend sind RTn ein sehr populäres Maß v.a. in den besonders stark kognitionspsychologisch ausgerichteten Domänen der Psychologie, wie z.B. der Gedächtnis- (Ratcliff, 1978; Sternberg, 1966), Aufmerksamkeits- (Stroop, 1935; Posner, 1978) und Lernforschung (Logan, 1988). Aber auch in den verschiedenen Teilgebieten der Sprachpsychologie sind sie extrem dominant.

Die Liste der experimentellen Paradigmen, in deren Rahmen RTn erhoben werden, reicht dabei von der sog. einfachen (*simple*) Reaktionszeit, die man braucht, um auf einen bestimmten Stimulus zu reagieren (wie dies z.B. für viele Untersuchungen im psychophysikalischen Bereich charakteristisch ist), über Wahlreaktionszeiten bei Entscheidungsaufgaben (die in der Kognitionspsychologie vielleicht am häufigsten besetzte Fallgruppe) bis hin zu den einfachen Lesezeiten für einzelne Wörter, Sätze oder Texte (wie sie bei vielen sprachpsychologischen Fragestellungen verwendet werden). Dabei spielt es zunächst einmal keine Rolle, wie und wo die Reaktion gemessen wurde: Eine RT kann die Zeit sein, die eine Person braucht, um eine bestimmte Taste oder einen Hebel zu drücken (das ist der prototypische Fall). Aber sie kann auch jede andere zeitgebundene Messung einer geistigen oder körperlichen Aktivitäten sein. Auch ist der Begriff „Reaktion“ sehr offen gehalten: Er kann sich sowohl auf bewusste, langsame und ressourcenintensive Entscheidung beziehen als auch auf den simplen Akt der Reizwahrnehmung.

Bei der Durchsicht der möglichen Anwendungsgebiete wird zweierlei direkt deutlich: Erstens handelt es sich bei der Reaktionszeitmessung (Chronometrie) nicht um eine einheitliche Methode, sondern eher um eine sehr große und breite Klasse von ganz unterschiedlichen Verfahren, die jeweils ihre besonderen Eigenschaften haben. Und zweitens kann man mit ei-

niger Berechtigung behaupten, dass die Reaktionszeitmessungen die zentrale und charakteristischste Methode der experimentell-kognitiven Psychologie ist. Und wenn man bestimmte physiologische und neuropsychologische Methoden (wie z.B. EEG-Messungen) nicht sogar direkt als eine bestimmte Unterform der Reaktionszeitmessung klassifiziert, dann ist klar, dass sie sich mehr oder weniger direkt aus ihr entwickelt haben und eine enge konzeptuelle Verbindung zu ihr aufweisen.

Der Grund, warum sich RT-Methoden gerade im Rahmen der kognitionspsychologischer Fragestellungen einer so ausgesprochenen Beliebtheit erfreuen, ist recht simpel: Mit ihnen verbindet sich die Hoffnung, einen Blick in die *black box* des menschlichen Geistes werfen zu können. Genauer verspricht man sich von ihnen, dass die Beschreibung der temporalen Eigenschaften des Antwortprozesses eine Erfassung der Architektur des „psychischen Apparates“ (Freud) erlaubt. Auch erhofft man sich, so die elementaren mentalen Prozesse, die zwischen Stimulus und Antwort angesiedelt sind, identifizieren zu können (woran sich noch einmal zeigt, dass die Kognitionspsychologie konzeptuell das S-R-Schema des Behaviorismus übernommen hat, vgl. z.B. Erb, 1997).

Seit den Pionierarbeiten von Donders (1868/1969) ist die Grundidee der RT-Messung, dass sich unterschiedliche psychische Merkmale, Zustände, Prozesse etc. in unterschiedlichen zeitlichen Verläufen niederschlagen (sollten): Eine Person, die weniger intelligent ist, sollte nicht nur seltener, sondern auch langsamer zur Lösung eines Problems gelangen (Reaktionszeitmessung zur Erfassung von latenten Merkmalen/Fähigkeiten). Und wenn zur Lösung einer bestimmten Aufgabe zwei mögliche Prozesse A und B zur Verfügung stehen, die unterschiedlich viel Zeit beanspruchen, dann kann man aufgrund der Lösungszeit u.U. darauf schließen, welchen Lösungsweg eine Person eingeschlagen hat (Reaktionszeitmessung zur Erfassung von mentalen Prozessen).

Der „philosophische Kern“ der Methodik besteht darin, dass mentale Zustände und Prozesse nach Descartes als „res cogitans“ im Gegensatz zur „res extensa“ zunächst einmal keine Ausdehnung haben und damit auch nicht in einem strengen messtheoretischen Sinne erfassbar bzw. messbar sind. Mit einer Ausnahme: Die einzige Eigenschaft mentaler Ereignisse, die direkt gemessen werden kann, ist ihre Dauer, d.h. ihre zeitliche Erstreckung. Vor diesem Hintergrund lässt sich nun der Anspruch reaktionszeitbasierter Methoden präziser formulieren (vgl. Luce, 1986): Unter der Voraussetzung, dass die menschliche Psyche nicht vollkommen unstrukturiert ist, sondern vielmehr eine gewisse Systematik besitzt, die u.a. spezifiziert, auf welche Art und Weise Informationen verarbeitet werden können, stehen mitunter mehrere theoretisch mögliche Wege zur Informationsverarbeitung zur Verfügung. Jeder dieser „Pfade“

durchs kognitive System weist unterschiedliche zeitliche Eigenschaften auf. Und die zentrale Zielsetzung der Reaktionszeitmessung besteht darin, wie beim *reverse engineering* allein durch eine Erfassung der temporalen Eigenschaften des Systems Aussagen über seine Struktur und die beteiligten Prozesse treffen zu können.

Der indirekte Schluss von der Antwortdynamik eines Systems auf seine interne Struktur ist besonders dann attraktiv, wenn davon auszugehen ist, dass sich der infrage stehende Prozess nicht anders erfassen lässt. Das ist v.a. bei solchen Prozessen der Fall, die so elementar und feingliedrig sind, dass sie sich der vergleichsweise „grobkörnigen“ Auflösung unseres Bewusstseins entziehen. Die Überlegung ist hierbei, dass automatische Prozesse, selbst wenn sie schnell und latent ablaufen, immer noch eine bestimmte zeitliche Erstreckung haben oder die Dynamik des Reaktionsprozesses verändern. Damit sind sie dann aber auch weiterhin mittels chronometrischer Verfahren erfassbar; hinreichende Auflösung des eingesetzten Verfahrens vorausgesetzt.

Man kann (mit einiger Berechtigung) beklagen, dass chronometrische Verfahren eine unrühmliche Rolle bei der „Entmündigung“ des Subjekts durch die moderne Psychologie spielen (vgl. Erb, 1997). Denn zumindest wenn sie unvernünftig bzw. übergeneralisierend eingesetzt werden, verleiten sie zu der Annahme, dass *alle* psychischen Prozesse zum automatischen und unbewussten Typ gehören. (Auch hier trifft der berühmte Ausspruch von Maslow, 1966, zu, dass es sehr verführerisch ist, alles als Nagel zu betrachten, wenn man nur einen Hammer hat.) Diese Annahme ist m.E. sowohl aus theoretischen als auch empirischen Gründen falsch und nicht ernsthaft zu vertreten. Da auf der anderen Seite jedoch auch klar ist, dass es in der Tat solche elementar-automatischen Prozesse gibt, ist die entgegen gesetzte Position, die davon ausgeht, dass sich alle mentalen Operationen bewusst erfassen lassen, ebenso unhaltbar (wobei sich über Umfang und Inhalt der beiden Prozessklassen trefflich streiten lässt). Wichtig ist an dieser Stelle nur, dass chronometrische Verfahren zumindest potenziell einen wertvollen methodischen Zugang zu einer ganzen von Klasse von nicht-bewusstseinfähigen, latenten Phänomenen eröffnen, die uns sonst verschlossen bliebe.

In Bezug auf den Gegenstandsbereich der vorliegenden Arbeit ist nun festzuhalten, dass sie (ausschließlich oder zumindest größtenteils) dem latenten Prozesstypus angehören. Für die gedächtnis-basierten Sprachverarbeitungsprozesse ist dies ganz offensichtlich: Wie in Kap. 1.2 dargestellt, werden diese ganz dezidiert als elementar-basale Gedächtnisprozesse expliziert, da der sog. *memory-based view* davon ausgeht, dass sich (alle) Prozesse der Sprachverarbeitung auf solche automatischen Such- und Abrufprozesse im Langzeitgedächtnis reduzieren lassen. Bei der Teilklasse der erklärungs-basierten Verarbeitungsprozesse sind hingegen

zumindest teilweise auch höhere kognitive Mechanismen involviert. Es wird explizit davon ausgegangen, dass diese Verarbeitungsprozesse zumindest teilweise strategisch und willentlich steuerbar sind. Man kann deswegen davon ausgehen, dass diese Prozessklasse in der Tat bewusstseinsfähig ist oder zumindest Auswirkungen auf bewusstseinsfähige Zustände hat. Allerdings gilt dies sicherlich nicht für alle Teilprozesse: Einige, wie z.B. die schnelle Kohärenzprüfung sprachlicher Strukturen und viele Prozesse der Inferenzgenerierung, sind ebenfalls habitualisiert und laufen unterschwellig ab. Solche Prozesse lassen sich kaum anders als mit einer reaktionszeitbasierten Methode erfassen, zumal die temporale Verlaufsdynamik für ihre Interaktion eine Rolle spielt.

Auch wenn die vorliegende Untersuchung durch ihren spezifischen Gegenstandsbereich notwendigerweise auf ein reaktionszeitbasiertes Verfahren angewiesen ist, so muss sie dieses gleichzeitig auch sehr vorsichtig einsetzen. Denn RT-Methoden haben wie alle psychologischen Erhebungs- und Auswertungsverfahren ihre spezifischen Begrenzungen. Das ist an sich noch nicht schlimm, weil keine Methode zu einer umfassenden und verzerrungsfreien Abbildung eines Gegenstandes führt und immer spezifische Vor- und Nachteile hat. Allerdings ist wichtig, möglichst viel über ihre Grenzen zu wissen, wenn man sie anwenden und die mit ihr erzielten Ergebnisse angemessen interpretieren will.

Vor diesem Hintergrund ist ein großes Problem aller RT-Verfahren zunächst einmal ihre mangelhafte Reliabilität. Denn obwohl RTn (heutzutage) vergleichsweise einfach und billig zu erheben sind und in jeder Untersuchung gleich in sehr großer Zahl anfallen, heißt dies noch lange nicht, dass sie eine gute Messung im messtheoretischen Sinne konstituieren und eine hohe Datenqualität besitzen. Im Gegenteil: RTn weisen typischerweise eine hohe intraindividuelle Variabilität auf und werden extrem durch Lern- und/oder Ermüdungseffekte beeinflusst. Desgleichen ist leicht vorstellbar, dass bei jedem einzelnen Durchgang sehr viele unterschiedliche Prozesse ablaufen und damit auch viele innere und äußere Störvariablen auf den Antwortprozess einwirken. Dies führt dazu, dass RT-Daten insgesamt sehr „verrauscht“ sind und einen großen Anteil an Fehlervarianz enthalten. Nur im Extremfall führen solche aufgabenfremden Prozesse zu sog. Ausreißern, also stark divergenten Datenpunkten, die durch Ablenkungen oder Fehlverhalten bei der Aufgabenbearbeitung entstehen (zu frühes Drücken der Antworttaste etc.). Meist sind mit Fremdprozessen kontaminierte Datenpunkte weniger extrem und so mit den „korrekten“ RTn vermischt, dass sie sich im Nachhinein nicht von ihnen trennen lassen (vgl. z.B. Ratcliff, 1993).

Allerdings ist bei RTn nicht nur die intra-, sondern auch die interindividuelle Variabilität ausgesprochen hoch. Das heißt, es gibt meist sehr große und nicht immer unbedingt kon-

stante Unterschiede zwischen den einzelnen Vpn. Einige Vpn antworten extrem schnell, während andere sehr lange brauchen. Dies stellt v.a. ein eminentes Problem für *between*-Designs dar (die deswegen bei RT-Methoden mit Vorsicht behandelt werden müssen, vgl. Salthouse & Hedden, 2002): Meist ist die Variabilität zwischen den einzelnen Personen sehr viel größer als die zwischen den infrage stehenden Untersuchungsbedingungen. Ein Aspekt, der wesentlich zur hohen interindividuellen Variabilität beiträgt, ist dabei die zusätzliche Problematik des sog. *speed-accuracy-trade-offs* (vgl. z.B. Meyer et al., 1988). Dieser bezieht sich darauf, dass bei allen RT-Methoden die Reaktionsgeschwindigkeit zwangsläufig mit der Reaktionsqualität konfundiert ist, weshalb es für Vpn möglich ist, eine schnellere Aufgabenbearbeitung gegen eine höhere Fehlerrate „einzutauschen“ oder umgekehrt eine fast fehlerfreie Bearbeitung mit einer längeren Antwortzeit zu „bezahlen“. Verschiedene Personen können sich nun darin unterscheiden, welches der beiden Kriterien von ihnen höher gewertet wird. Unterschiede in ihren Reaktionszeiten müssen dann nicht zwangsläufig damit zu tun haben, dass sie sich z.B. in einer unterschiedlichen Untersuchungsbedingung befinden. Aber auch intra-individuell kann das Antwortkriterium im Laufe der Untersuchung wechseln. Schlimmstenfalls ist ein solcher Wechsel mit den experimentellen Bedingungen konfundiert. Typischerweise wird versucht, dieses Problem in kognitionspsychologischen Untersuchungen dadurch zu umgehen, dass eines der beiden Kriterien (meist die Korrektheitsrate) auf einen sehr hohen Wert fixiert wird. Das heißt, es werden lediglich Aufgaben verwendet, bei denen Korrektheitsraten über 90% garantiert sind (Lexikalische Entscheidungsaufgaben, einfache Signalreaktion etc.). Denn wenn sowieso alle Antworten richtig sind, dann kann es nicht mehr zu einem Trade-off der Kriterien kommen, und alle Unterschiede bilden sich allein in der Bearbeitungsgeschwindigkeit ab.

Damit sind wir auch schon beim zweiten und sogar noch größeren Problem von RT-Methoden angelangt, nämlich ihrer Validität. Denn auch wenn sich mit ihnen sehr große Hoffnungen verbinden, so ist fraglich, ob es überhaupt so einfach möglich ist, durch RTn bestimmte kognitive Prozesse identifizieren zu können. Ein Grund, der zunächst einmal gegen diese Annahme spricht, ist, dass die RT eine sehr komplexe Variable ist, die (je nach Perspektive) entweder viel zu viel oder aber viel zu wenig abbildet als das, was man eigentlich erfassen will. Auf der einen Seite sind RTn gewissermaßen eine Aggregat-Variable, da alle Aufgaben (auch wenn sie noch so einfach sind) immer sehr viele mentale Prozesse gleichzeitig involvieren, die alle zur Gesamt-RT beitragen. Von daher weiß man nachher auch nicht genau, was man denn nun eigentlich erfasst hat. Denn wenn man nachher nur einen einzigen Wert hat und man weiß, dass mindestens zwei Prozesse zu ihm beitragen, dann kann von dem einen

Wert nicht auf den Anteil der beiden Teilprozesse zurückgeschlossen werden. Diese Problematik wird noch komplizierter, wenn sich das „Mischungsverhältnis“ der beteiligten Teilprozesse zwischen den einzelnen Untersuchungsbedingungen ändert. In der RT-Literatur (vgl. z.B. Luce, 1986) wird dies als *Dekompositionsproblem* bezeichnet. Dieses ist vergleichbar mit einem unterbestimmtem Gleichungssystem: Wenn man drei Variablen hat und nur zwei restringierende Gleichungen, dann hat dieses Gleichungssystem keine eindeutige Lösung. Das heißt, es gibt unendlich viele Lösungen und wir können die Lösungsmenge nicht weiter eingrenzen. Ähnlich verhält es sich nun auch mit der Gesamt-RT: Wir wissen nur, dass viele Prozesse gleichzeitig zu ihr beitragen, nicht jedoch, welche dies genau sind und wie sie sich zueinander verhalten. Der Punkt ist nun, dass wir es auch nie rausbekommen werden, da sich ohne weitere Annahmen alle Prozesskonstellationen begründen lassen. Die grundsätzliche Underdeterminiertheit der RT-Dekomposition ist auch der Grund dafür, warum eigentlich alle psychologischen Fragestellungen, zu deren Lösung RT-Methoden einmal antraten (serielle vs. parallele Architektur, selbst-terminierende vs. exhaustive Gedächtnissuche etc., vgl. z.B. Sternberg, 1966, 1969) immer noch ungelöst sind: Entsprechende Parametrisierung vorausgesetzt, kann jedes kognitive Modell jede mögliche RT-Verteilung generieren (vgl. Townsend, 1990; zum speziellen Problem des sog. *model mimicking* vgl. Abschnitt 3.6).

Auf der anderen Seite wird gegen RT-Verfahren eingewendet, dass der Informationsgehalt von RTn einfach zu gering ist, um valide Schlüsse auf kognitive Prozesse rechtfertigen zu können. Denn das Problem bei RT-Verfahren ist ja, dass sie uns zwar sehr genau sagen, wann etwas passiert (bzw. wohl eher: wie lange es gedauert hat), jedoch zunächst überhaupt gar nichts darüber, *was* denn passiert ist. Vielmehr sind alle Aussagen über die relevanten Prozesse, die von RTn abgebildet werden, entweder ungeprüfte Annahmen oder aber es sind lediglich indirekte Schlüsse, die nur dann gezogen werden können, wenn man bereits sehr spezifische Annahmen über die mentale Architektur gemacht hat (die sich nicht immer alle prüfen, manchmal noch nicht einmal rechtfertigen lassen). Im Extremfall tragen RTn also nur noch weiter zu dem für die psychologische Forschung sowieso schon typischen leeren und zirkulären Theoriebildungen bei, da sie immer bereits voraussetzen (müssen), was sie nachher beweisen.

Alle diese Kritiklinien konvergieren nun dahingehend, dass sie darin übereinstimmen, dass die physikalische Zeit zur Ausführung eines Prozesses allein noch nicht hinreichend ist, um mentale Prozesse und Strukturen zu identifizieren. Dies ist auch einer der Hauptgründe dafür, warum eine starke Tendenz besteht, RT-Verfahren durch andere methodische Zugangsweisen und Daten zu ergänzen. Zwei Spielarten sind hier zu unterscheiden:

Für die eher subliminal-elementare Prozessebene werden heutzutage meist neuropsychologische bzw. psychophysiologische Verfahren verwendet (vgl. im Überblick Posner, 2005), die (wieder einmal) versprechen, die mentale *black box* mit „Inhalt“ füllen zu können. So erhält man z.B. bei EKP-Verfahren nicht nur Informationen darüber, wie lange die Aufgabenbeantwortung gedauert hat, sondern auch Informationen darüber, *wo* (im Gehirn, nicht im mentalen System!) der Prozess zu lokalisieren ist. Gleichzeitig lassen sich über die Identifikation typischer EKP-Komponenten auch Aussagen darüber treffen, *welcher Art* der Prozess ist (im sprachpsychologischen Bereich z.B. N400 = Semantik, ELAN = frühe syntaktische Verarbeitung, P600 = späte syntaktische Repair-Prozesse etc.). Die Validität solcher Zuordnungen soll hier nicht diskutiert werden (vgl. hierzu z.B. Kutas, van Petten & Kluender, 2006, oder van Berkum, 2004), wichtig ist nur, dass diese Verfahren genau deshalb attraktiv sind, weil sie offensichtlich eine Lücke schließen (wollen), die RT-Verfahren alleine nicht zu schließen in der Lage sind.

Im Bereich der prinzipiell bewusstseinsfähigen Prozesse (wie z.B. beim Problemlösen) werden hingegen auch häufig verbale Daten ergänzend zur reinen RT erhoben, z.B. mittels der Technik des Lauten Denken (vgl. Ericsson & Simon, 1980; 1993). Auch hier liegt der Grund auf der Hand: Verbaldaten liefern genau das, was RT-Daten nicht haben, nämlich Semantik. Dadurch, dass sich Personen dazu äußern, was sie in einem bestimmten Zeitraum gemacht haben, sind die Prozesse in dieser Zeitspanne viel besser interpretierbar. Allerdings muss im sprachpsychologischen Bereich wegen der offensichtlichen Interferenz-Gefahr von Aufgabenbearbeitung und Erhebungsverfahren auf Formen des vor- bzw. nachzeitlichen Lauten Denkens zurückgegriffen werden (z.B. im Rahmen der sog. Drei-Strang-Methode, vgl. Magliano & Graesser, 1992).

Als Fazit lässt sich vielleicht festhalten, dass RTn so etwas wie das Fast-food der Kognitionspsychologie sind: Sie sind schnell und billig zu haben, aber ernährungstechnisch fast wertlos bzw. bei monokultureller und einseitiger Nutzung sogar langfristig gesundheitsschädlich. In Bezug auf die Fragstellung dieser Arbeit im Speziellen fällt das Fazit nicht weniger ernüchternd aus: Auf der einen Seite kann zur Erfassung der dynamischen Wechselwirkung gedächtnis- und erklärungs-basierter Prozesse bei der Sprachverarbeitung eigentlich nur eine reaktionszeitbasierte Methodik ernsthaft in Frage kommen; deshalb ist es auf der anderen Seite enttäuschend zu erfahren, dass gerade diese eher schlecht dazu in der Lage ist, diese Prozesse trennscharf und valide zu erfassen.

Allerdings lässt sich etwas tun. Denn auch wenn RT-Methoden aufgrund der ihnen inhärenten Grenzen mit Vorsicht zu verwenden sind, so sind sie bei vernünftigem Gebrauch

doch aussagekräftiger, als man anfänglich denkt. Erschreckend ist allerdings die Tatsache, wie selten man in psychologischen Studien eine solche sachgemäße und vorsichtige Verwendung antrifft: RTn sind halt etwas, „was man so erhebt“. Über die Interpretation macht man sich meist keine genauen Gedanken, und wenn, dann erst im Nachhinein. Im Allgemeinen herrscht eine Interpretation vor, die RT mit so etwas wie dem „mental“en Verarbeitungsaufwand gleichsetzt. Entsprechend lässt sich ein RT-Unterschied zwischen zwei Untersuchungsbedingungen meist recht flexibel (und damit potenziell empirisch leer) interpretieren.

Unter einem „verantwortungsvollen“ Umgang mit der Variable RT verstehe ich v.a. einen Umgang, der sich im Vorfeld der Untersuchung um eine möglichst valide Interpretation potenzieller Effekte bemüht. Dies ist jedoch nur möglich, wenn man schon im Vorhinein weiß, welche Effektmuster von welchen Prozessen verursacht werden (sollten). Allgemeiner formuliert ist eine möglichst klare semantische Eingrenzung der zu beobachtenden Effekte durch entsprechende versuchsplanerische Maßnahmen notwendig: Es sollte möglichst so sein, dass die experimentellen Manipulationen im Rahmen des angenommenen Verarbeitungsmodells nur noch eine einzige Interpretation offenlassen. Das setzt allerdings wiederum eine möglichst starke theoretische Einbettung voraus, da eine maximal gehaltvolle und damit potenziell valide Interpretation von RT-Ergebnissen nur dann vorgenommen werden kann, wenn das zugrunde liegende Prozessmodell möglichst spezifisch und differenziert expliziert ist. Grundsätzlich können die Ergebnisse aller RT-basierten Studien immer nur vor dem Hintergrund eines bestimmten Verarbeitungsmodells interpretiert werden, und der Gehalt der Interpretation bestimmt sich wesentlich aus dem Gehalt der theoretischen Modellierung. Ein gutes und möglichst vollständiges theoretisches Modell allein reicht jedoch noch nicht. Man muss auch wissen, wie sich das theoretisch spezifizierte Modell empirisch auswirkt, und auch das möglichst präzise. Zumindest soweit das zugrunde liegende theoretische Modell eine Quantifizierung erlaubt, ist es deshalb sinnvoll, RT-Methoden gleichzeitig in Kombination mit Simulationsstudien einzusetzen, da diese eine möglichst genaue Vorhersage der temporalen Eigenschaften eines bestimmten theoretischen Modells erlauben. Nicht nur, dass unter diesen Umständen ein besserer Vergleich zwischen theoretischem Modell und empirischer Beobachtung möglich ist (und damit auch eine stärkere und gehaltvollere Hypothesenprüfung), auch trägt ein solches Vorgehen zur *Theoriebildung* bei, weil man unter dem Zwang einer konkreten Implementierung eines theoretischen Ansatzes gezwungen ist, sich umfassender und tiefer mit einem Problem zu beschäftigen als es heutzutage bei den meisten psychologischen Theorien der Fall ist.

Aufgrund der skizzierten Probleme der RT-Messung wird in der vorliegenden Arbeit auf das im psycholinguistischen Kontext übliche, aber problematische Vorgehen einer rein mittelwerts-basierten Auswertung verzichtet bzw. diese um weitere reliabilitäts- und validitätssteigernde Auswertungsschritte ergänzt. Da für den konkreten Gegenstandsbereich bis jetzt noch keine präzisen Verarbeitungsmodelle entwickelt worden sind, bietet sich als idealer Kompromiss v.a. eine Analyse ganzer *RT-Verteilungen* an. Im Gegensatz zum herkömmlichen Vorgehen, bei dem die in einer Untersuchungsbedingungen angefallenen RTn auf einen einzigen Wert reduziert werden, wird bei der RT-Verteilungsanalyse das ganze Spektrum an Informationen ausgeschöpft, die in der Verteilung der RTn enthalten sind. Das ermöglicht zum einen eine adäquatere und effizientere statistische Auswertung als dies bei einer traditionellen Analyse der Fall ist. Der entscheidende Punkt ist aber, dass die RT-Verteilungsanalyse validere Aussagen über die der RT zugrunde liegenden Prozesse erlaubt, ohne gleichzeitig zu weit gehende Prozessannahmen formulieren zu müssen. Denn während beim aggregierten Mittelwert nicht mehr unbedingt klar ist, was er eigentlich genau abbildet, werden bei der RT-Verteilungsanalyse mehrere Charakteristika des zugrunde liegenden Antwortprozesses gleichzeitig erfasst, die wiederum mit verschiedenen kognitiven Prozessstypen korreliert sind. Insbesondere ist es deshalb möglich, reliable und valide Prozessindikatoren für die beiden theoretischen relevanten Prozessklassen der gedächtnis- und erklärungs-basierten Verarbeitungsprozesse zu spezifizieren und Hypothesen über ihr Zusammenwirken zu testen.

3.2 Jenseits des Mittelwertes: Modellierung von Reaktionszeitverteilungen

Wie im letzten Abschnitt dargestellt, besteht eines der fundamentalen Probleme der RT-Forschung darin, dass RTn zunächst einmal theoretisch unterspezifiziert sind und für sich selbst nur sehr wenig sagen. Diese Problematik der RT-Methodik wird noch dadurch verstärkt, dass RTn in den meisten Fällen auf eine Art und Weise ausgewertet werden, die sie noch informationsärmer macht, als sie eigentlich sind. Im folgenden Abschnitt wird dafür argumentiert, dass eine Beschränkung der RT-Datenbeschreibung auf einen Wert der zentralen Tendenz allein (meistens den Arithmetischen Mittelwert) die in RT-Daten vorhandenen Information nicht optimal nutzt und darüber hinaus auch zu artifiziellen Befunden führen kann. Danach wird dann mit der RT-Verteilungsanalyse eine Methode vorgestellt, die nicht nur zu einer deskriptiv adäquateren Darstellung der empirischen Effekte führt, sondern zumindest ansatzweise auch in der Lage ist, die theoretische Leere der RT-Methodologie mit Inhalt zu füllen.

3.2.1 Probleme der mittelwerts-basierter Auswertung

Damit klar ist, was kritisiert wird, sei einleitend kurz ein prototypisches RT-Experiment skizziert: Normalerweise ist in einem repräsentativen RT-Experiment in der Kognitionspsychologie eine vergleichsweise einfache Aufgabe zu bearbeiten, wie z.B. eine Lexikalische Entscheidungsaufgabe. Dies gewährleistet einerseits, dass die RTn möglichst kurz und damit vergleichsweise „rein“ sind, da aufgrund der nur kurzen Bearbeitungszeit nur wenig andere kognitive Prozesse intervenieren können. Andererseits wird durch die Einfachheit der Aufgabe (die manchmal die Grenze zur Stupidität überschreitet) sichergestellt, dass fast nur korrekt geantwortet wird, wodurch das Problem des *speed-accuracy-trade-offs* (künstlich) umgangen wird (s.o.). Abgesehen davon, dass die ganze Zeit die gleiche Aufgabe zu bearbeiten ist, werden Stimuli in unterschiedlichen Untersuchungsbedingungen präsentiert, wobei von jeder Person meist mehrere Stimuli pro Untersuchungsbedingung bearbeitet werden. Die einzelnen RT-Werte pro Person und Untersuchungsbedingung werden danach aufgrund der notorischen Unreliabilität einer einzelnen RT-Messung (s.o.) aggregiert, wobei als Aggregatwert meist das Arithmetische Mittel als Maß der zentralen Tendenz verwendet wird (seltener der Median, vgl. Miller, 1988). Die RT-Mittelwerte pro Person (im Folgenden „mittlere RT“ genannt) werden dann nachfolgend mittels eines inferenzstatistischen Verfahrens (meist der Varianzanalyse) auf Unterschiede zwischen den Untersuchungsbedingungen geprüft. Wenn sich die Mittelwerte der verschiedenen Untersuchungsbedingungen (gemittelt über alle Personen) vor dem Hintergrund der Grundvariabilität der Daten hinreichend stark voneinander unterscheiden, dann wird das inferenzstatistische Testverfahren ein signifikantes Ergebnis erbringen und man nimmt an, dass das manipulierte Merkmal in der Tat einen (mehr oder weniger starken bzw. mehr oder weniger theoretisch relevanten) Effekt auf die Bearbeitungszeit der Aufgabe hat.

Der weitaus größte Teil aller kognitionspsychologischen Untersuchungen, welche auf RTn basieren, laufen nach diesem oder einem ähnlichen Schema ab. In der sprachpsychologischen Forschung wird (bis auf ganz wenige Ausnahmen) eigentlich immer so gearbeitet. Die Problematik, die mit diesem Vorgehen verbunden ist, wird dann deutlich, wenn man sich anschaut, was bei einem solchen Experiment zunächst einmal konkret entsteht: Jede Person bearbeitet eine bestimmte Anzahl von Items (sagen wir 20) in einer Untersuchungsbedingung. Damit haben wir für diese Person auch zunächst einmal nicht einen, sondern 20 Werte pro Untersuchungsbedingung. Und diese 20 Werte sind nicht alle gleich, sondern werden auf eine bestimmte Art und Weise variieren. Statistisch sind sie deshalb als Realisierung einer zugrun-

de liegenden Zufallsvariablen anzusehen, die eine bestimmte *Verteilung* hat (vgl. z.B. Papoulis & Pillai, 2002). Eine erste grafische Approximation dieser Verteilung erhält man, indem man sich ein Histogramm dieser RTn betrachtet. Abb. 3.1 zeigt das Histogramm einer hypothetischen Versuchsperson, die 100 Items bearbeitet hat:

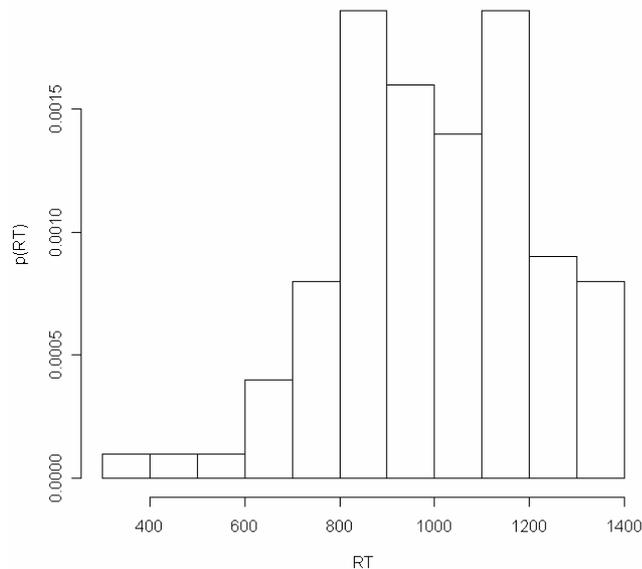


Abb. 3.1: RT-Verteilung einer hypothetischen Vpn ($N = 100$; $RT \sim N(1000,200)$).

Auf der x-Achse sind die Werte der 100 RTn abgetragen, welche die Person abgegeben hat. Auf der y-Achse ist hingegen die *Wahrscheinlichkeitsdichte* ($p(RT)$) abgebildet; d.h. hier finden sich Informationen darüber, *wie häufig* ein bestimmter Wert auftaucht (allerdings relativiert und nicht wie bei einem üblichen Histogramm in absoluten Häufigkeiten). Auch wenn man es der Abbildungen nicht unbedingt ansieht, ist die dargestellte empirische Häufigkeitsverteilung eine Zufallsstichprobe aus einer Normalverteilung mit einem Mittelwert von 1000 ms und einer Standardabweichung von 200 ms. Es gibt einige wenige Werte um 400 ms und einen deutlichen Masseschwerpunkt zwischen 800-1200 ms (der allerdings leicht bimodal erscheint). Eigentlich würde man (vergleichbar zum linken Rand) auch noch einen „Ausläufer“ der Verteilung auf der rechten Seite erwarten, wo sich wenige Beobachtungen im Bereich um 1600 ms befinden. Dieser ist jedoch aufgrund der zufälligen Stichprobenvariabilität offensichtlich empirisch nicht realisiert.

In diesem Beispiel ist die Welt noch schön und einfach: Da die Verteilung (zumindest die zugrunde liegende) normalverteilt ist, sollten sich die beiden „Schwänze“ der Verteilung gegenseitig aufheben, weshalb sich bei einer Aggregation der 100 Werte zu einem einzigen in der Tat ein Mittelwert von 1000 ms ergibt. (Auch hier geht allerdings schon Information verloren, nämlich wie weit die Verteilung streut, s.u.). Komplizierter wird es schon, wenn

man berücksichtigt, dass RTn (aus Gründen, die in Abschnitt 3.6 besprochen werden) eigentlich immer rechtsschief verteilt sind (vgl. z.B. Ratcliff, 1993). Abb. 3.2 zeigt eine solche Verteilung.

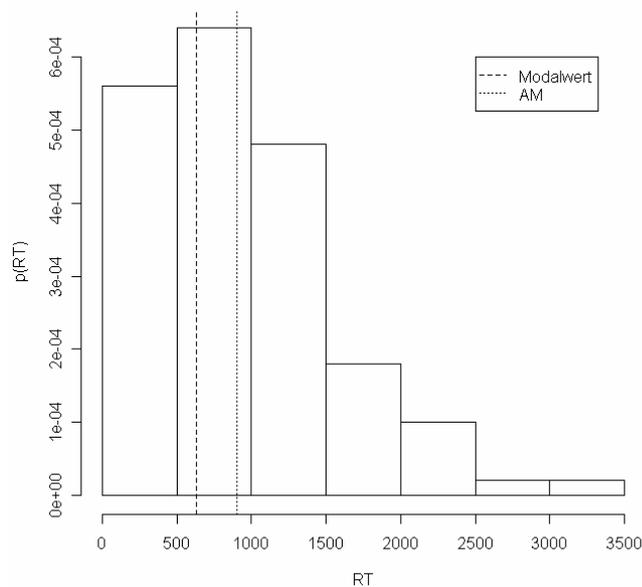


Abb. 3.2: Noch eine RT-Verteilung einer hypothetischen Vpn ($N = 100$; $RT \sim Weibull(0,1000,1.5)$).

Im Gegensatz zur Normalverteilung in Abb. 3.1 hat die Verteilung ein deutliches „Übergewicht“ auf der linken Seite und einen längeren Schwanz auf der rechten. Dies führt dazu, dass nicht mehr so richtig klar ist, was denn nur der beste Beschreibungspunkt für die Verteilung ist. Im Gegensatz zur symmetrischen Normalverteilung, bei der Modalwert, Median und Arithmetisches Mittel identisch sind, divergieren diese Werte bei schiefen Verteilungen: Im vorliegenden Beispiel liegt der (theoretische) Modalwert z.B. bei 629 ms, der (theoretische) Mittelwert jedoch bei 903 ms (vgl. Abb. 3.2). Auch ist klar, dass der Mittelwert ein immer schlechterer Beschreibungswert für die Verteilung ist, je länger der rechte Schwanz wird: Es gibt dann immer mehr hohe Werte in der Verteilung, die das Arithmetische Mittel nach oben ziehen, selbst wenn sich der Masseschwerpunkt am linken Rand nicht verändert.

An diesem Beispiel lassen sich (mindestens) zwei Dinge klar machen: a) Das Arithmetische Mittel ist nicht das einzige (und manchmal auch nicht das beste) statistische Maß, mit dem sich die Verteilung einer Zufallsvariablen beschreiben lässt. Insbesondere bildet es nicht die ganze in einer Zufallsverteilung vorhandene Information ab, sondern selektiv nur einen bestimmten Teilaspekt. b) Neben der zentralen Tendenz haben Zufallsverteilungen noch andere Eigenschaften (z.B. eine bestimmte Schiefe), die u.U. ebenfalls die Ausprägung des mittleren RT-Werts beeinflussen. Beide Punkte zusammen sollen noch einmal an einem Beispiel veranschaulicht werden.

Hierfür empfiehlt es sich, provisorisch eine bestimmte theoretische Verteilungsform, nämlich die sog. Ex-Gauss-Verteilung (vgl. Ratcliff, 1978), für die empirisch beobachtete RT-Verteilung anzunehmen, weil sich dann die Darstellung wesentlich vereinfacht (vgl. Balota & Spieler, 1999; Spieler et al., 2000). Die Ex-Gauss-Verteilung (die ausführlicher in Abschnitt 3.3.2 beschrieben wird) ist ein Konvolut aus der Normal- und der Exponential-Verteilung, die sich zumindest empirisch sehr gut zur Beschreibung von RT-Verteilungen eignet (vgl. Luce, 1986). Der Begriff „Konvolut“ bezieht sich (im Gegensatz zu dem der „Mixtur“) auf die Tatsache, dass sich jeder Datenpunkt aus einem normal- und aus einem exponentialverteilten Anteil gleichzeitig zusammensetzt. In Abb. 3.3 a) –c) ist dies dargestellt:

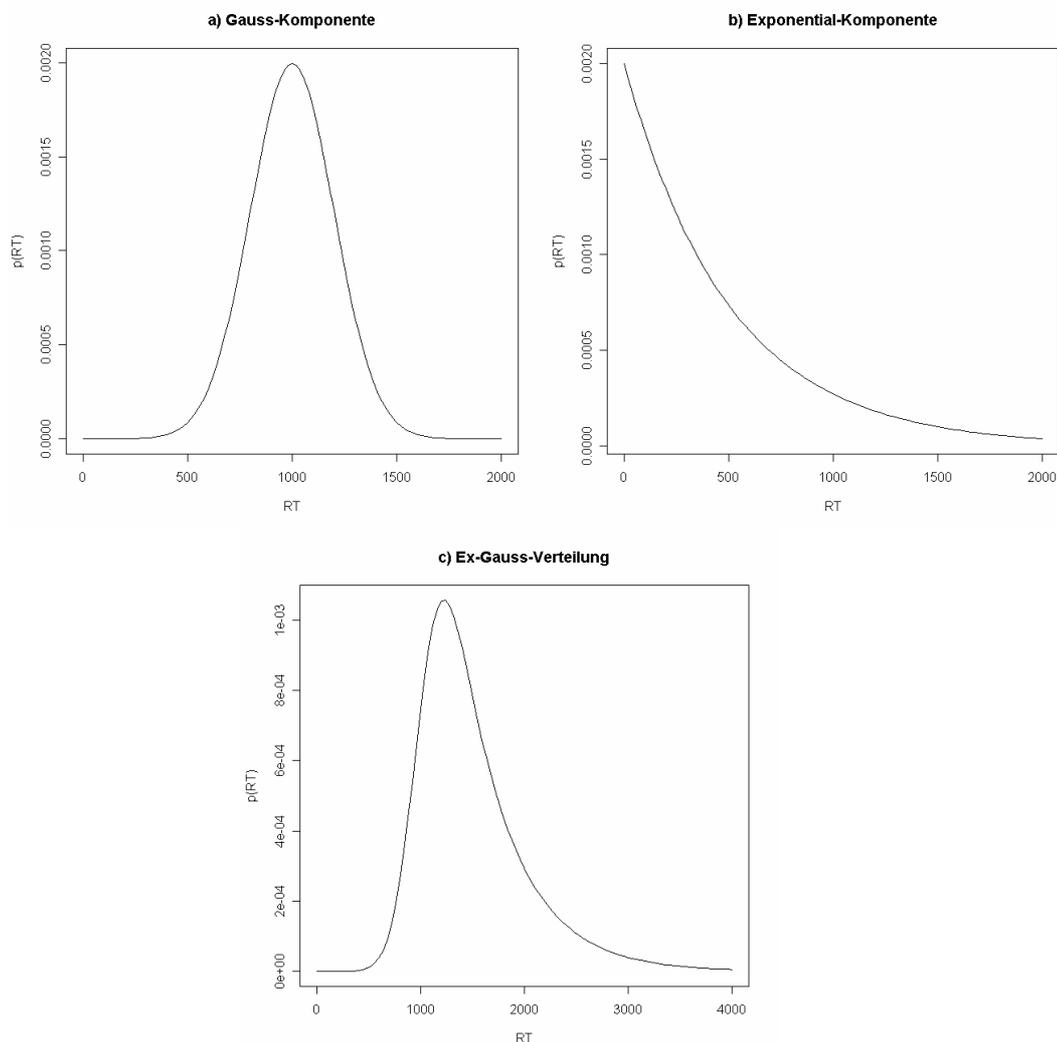


Abb. 3.3: a) Normal- ($N(1000,200)$), b) Exponential-Komponente ($E(500)$) und c) Ex-Gauss-Konvolut.

Teilabb. a) zeigt eine Normalverteilung mit einem Mittelwert (μ) von 1000 und einer Standardabweichung (σ) von 200, Teilabb. b) stellt eine Exponential-Verteilung mit einem Skalierungsparameter (τ) von 500 dar und Teilabb. c) das Konvolut beider Komponenten. Offensichtlich wird bei der Ex-Gauss-Verteilung die Gesamtverteilung immer schief, je größer der Exponential-Anteil an der Verteilung ist, also der τ -Parameter. Der Mittelwert der

Ex-Gauss-Verteilung ist leicht zu berechnen, weil er sich direkt aus der Summe des μ -Parameters der Normalverteilungs- und des τ -Parameters der Exponential-Komponente ergibt (im vorliegenden Beispiel also 1500 ms beträgt).

Der entscheidende Punkt ist nun, dass der Mittelwert der Verteilung (der sich aus dem Gleichungssystem $AM = \mu + \tau$ ergibt) unterdeterminiert ist. Das heißt dann zunächst einmal konkret, dass sich ein und derselbe Mittelwert aus unterschiedlichen „Mischungsverhältnissen“ der Normal- und der Exponential-Komponente ergeben kann und eine Betrachtung des Mittelwerts allein wenig über die zugrunde liegende Verteilung aussagt. In Abb. 3.4 sind z.B. drei Ex-Gauss-Verteilungen abgetragen, die alle zur gleichen mittleren RT führen (1500 ms). Allerdings ist auf den ersten Blick ersichtlich, dass sich die zugrunde liegenden Verteilungen erheblich voneinander unterscheiden. Während die Kombination ($\mu = 1000$, $\tau = 500$) fast normalverteilt ist, ist die Kombination ($\mu = 500$, $\tau = 1000$) extrem schief verteilt.

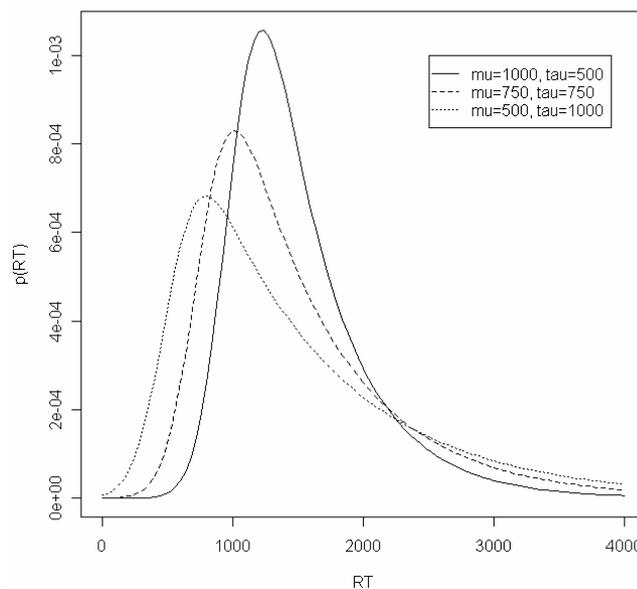


Abb. 3.4: Ex-Gauss-Verteilungen mit unterschiedlichen Parameter-Konstellationen

Die entscheidende Frage ist nun, ob es überhaupt einen Unterschied macht, wie die zugrunde liegende Verteilung aussieht. Und die emphatische Antwort, die in der vorliegenden Arbeit auf diese Frage gegeben wird, lautet: Ja! Es wird behauptet, dass verschiedene Aspekte einer Reaktionszeitverteilung (wie ihre zentrale Tendenz, ihre Skalierung und ihre Schiefe) viele Informationen über die reaktionsgenerierenden kognitiven Prozesse vermitteln, die bei einer Betrachtung lediglich der mittleren RT vollkommen ungenutzt bleiben. So wird das zentrale Problem der RT-Messung, nämlich ihre Informationsarmut, nur noch künstlich und m.E. vollkommen unnötig verschärft. (Allerdings wird es durch die Berücksichtigung differenzierterer Verteilungscharakteristika auch nicht vollkommen behoben).

Speziell wird in der vorliegenden Arbeit mit drei Beschreibungsmerkmalen einer RT-Verteilung gearbeitet, nämlich ihrer „Lokation“, „Skalierung“ und „Form“. Eine präzise Definition dieser drei Konzepte findet sich in Appendix A. Hier reicht vorläufig der Hinweis, dass mit dem Begriff „Lokation“ die Platzierung einer Verteilung auf der x-Achse gemeint ist und mit dem Begriff „Skalierung“ die Variabilität der Verteilung (also ihre Dispersion). Am komplexesten ist jedoch das Konzept der „Form“ einer Verteilung. Grob gesagt, bezieht es sich darauf, wie die Verteilung „aussieht“, also ihre konkrete Gestalt. Eine genauere Definition lässt sich eigentlich erst dann geben, wenn man eine konkrete Verteilungsfamilie betrachtet. Im obigen Beispiel der Ex-Gauss-Verteilung ist mit der Form der Verteilung im Wesentlichen das Mischungsverhältnis von Normal- und Exponentialkomponente gemeint, das dafür verantwortlich ist, dass die drei dargestellten Verteilungen so unterschiedlich aussehen.

Darüber hinaus wird in der vorliegenden Arbeit davon ausgegangen, dass sich verschiedene kognitive Prozessklassen in unterschiedlichen Verteilungscharakteristika niederschlagen. Eine detaillierte Interpretation der verschiedenen Aspekte von RT-Verteilungen soll erst entwickelt werden, wenn das dieser Arbeit zugrunde liegende Verteilungsmodell eingeführt wurde (vgl. Abschnitt 3.3). Die Grundidee ist jedoch die, dass sich vergleichsweise periphere und hierarchie-niedrige kognitive Prozesse (speziell: gedächtnis-basierte Sprachverarbeitungsprozesse) primär in der Lokation einer Verteilung abbilden, während hierarchie-hohe, zentrale Prozesse (speziell: erklärungs-basierte Sprachverarbeitungsprozesse) die Skalierung und Form der Verteilung beeinflussen. Obwohl diese generelle Zuordnung durchaus nicht unumstritten ist (vgl. z.B. Luce, 1986; eine ausführliche Diskussion der Thematik findet sich in Abschnitt 3.6), so hat sie doch eine lange Tradition in der Literatur zur RT-Verteilungsanalyse (vgl. im Überblick z.B. Balota & Spieler, 1999; Schwarz, 2001).

Zusammenfassend kann festgehalten werden, dass die RT-Verteilungsanalyse die zugrunde liegende RT-Verteilung präziser beschreibt als ein rein mittelwerts-basiertes Vorgehen. Darüber hinaus führt sie u.U. zu theoretisch gehaltvolleren Prozessindikatoren, welche die dem Antwortprozess zugrunde liegenden kognitiven Prozesse besser spezifizieren als der Mittelwert, bei dem verschiedene Antwortprozesse miteinander konfundiert sind. Speziell für den Gegenstandsbereich der vorliegenden Arbeit verbindet sich mit ihr die Hoffnung, dass sich durch eine differenzierte Betrachtung der verschiedenen Verteilungscharakteristika unabhängige Indikatoren für die beiden theoretisch relevanten Prozessklassen der gedächtnis- und erklärungs-basierten Sprachverarbeitungsprozesse entwickeln lassen, die dynamisch im Laufe der Satzverarbeitung erhoben werden können.

3.2.2 Methoden der Reaktionszeitmodellierung

Im letzten Abschnitt wurde dargestellt, dass die RT-Verteilungsanalyse zu einer präziseren und informativeren Beschreibung der zugrunde liegenden RT-Verteilung führt. Allerdings ist bekanntlich ja immer, wo Sonne ist, auch Schatten. Nicht nur, dass solche Analysen wesentlich schwieriger und aufwändiger sind als eine herkömmliche Auswertung. Darüber hinaus gilt auch hier der Energieerhaltungssatz: Wenn man mehr Informationen haben will, dann müssen diese irgendwoher kommen. Und entsprechend ist für die reliable Schätzung von RT-Verteilungskomponenten eine sehr viel größere Datenbasis notwendig als für die Schätzung allein der mittleren RT.

Denn der entscheidende Unterschied zwischen einer mittelwerts-basierten Auswertung und der RT-Verteilungsanalyse besteht darin, dass bei letzterer die einzelnen RT_n als Realisierungen einer zugrunde liegenden Zufallsvariable aufgefasst werden (vgl. einführend van Zandt, 2000, 2002). Man macht also eigentlich keine Aussagen mehr über die in einem Experiment beobachteten, realen RT_n , sondern vielmehr über die nicht direkt beobachtbare, theoretische Zufallsvariable, die jedoch aus den vorliegenden Daten geschätzt wird. Und da alle relevanten Informationen über eine Zufallsvariablen in ihrer Wahrscheinlichkeitsdichtefunktion (engl. *probability density function*, im Folgenden *pdf*) bzw. ihrer Verteilungsfunktion (engl. *cumulative distribution function*, im Folgenden *cdf*) enthalten sind, besteht das eigentliche Problem darin, wie sich diese reliabel schätzen lassen (auf die sog. *hazard*-Funktion einer Zufallsvariablen wird hier nicht eingegangen, vgl. hierzu Luce, 1986).

Es gibt zwei grundsätzlich unterschiedliche Ansätze zur Lösung dieses Problems. Zum einen kann man versuchen, möglichst keine a priori-Annahmen über die Form der relevanten Verteilungsfunktion zu machen und diese vielmehr vollkommen frei aus den vorliegenden Daten zu schätzen. Hierfür steht eine Reihe von *verteilungsfreien (non-parametrischen)* Verfahren zur Verfügung, die eben keine bestimmte Verteilungsfunktion (mit ihren Parametern) vorgeben. Zu dieser Verfahrensklasse gehören u.a. sog. *Moment-Methoden*, welche die Momente und Kumulanten der Zufallsvariablen direkt berechnen (vgl. Appendix A.3 und A.4).

Eine weitere Verfahrensklasse sind sog. non-parametrische Dichte-Schätzer (wie *Kernel Density Estimation*, *Histogram Estimation*, *B-spline-Verfahren*, vgl. Altman, 1992; Silverman, 1986). So unterschiedlich die einzelnen Verfahren dieser Klasse auch funktionieren, ist ihnen allen gemeinsam, dass sie eine möglichst optimale Kurve zur Beschreibung einer empirischen Häufigkeitsverteilung finden (Abb. 3.5 zeigt dies am Beispiel eines Kernel-Schätzers, vgl. Venebles & Ripley, 2002):

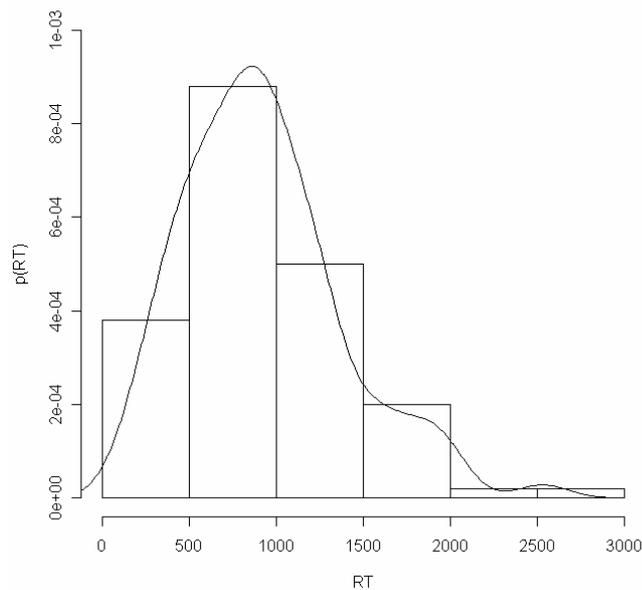


Abb. 3.5: Non-parametrischer Kernel-Schätzer für eine RT-Verteilung.

Obwohl solche non-parametrische Verfahren eine sehr attraktive Form der Datenmodellierung sind und insbesondere Kernel-Schätzer und B-spline-Verfahren im Rahmen der sog. Funktionalen Datenanalyse in den letzten Jahren enorme Fortschritte gemacht haben (s. z.B. Ramsay & Silverman, 2002, 2005), bieten sie sich für die vorliegende Problemlage weniger an. Zum einen tendieren sie dazu, die Daten zu gut anzupassen (d.h. sie tendieren beim sog. *Bias-Variance-Dilemma* eher zu einer Minimierung der Bias-Komponente, vgl. Alisch, 2001). Damit können sie zwar vergleichsweise feinkörnige Aspekte einer empirischen Verteilung abbilden, die sich dann jedoch nicht mehr generalisieren lassen. Diese Problematik äußert sich v.a. darin, dass diese Verfahren äußerst anfällig für Ausreißer sind (besonders wenn sie auf höheren Momenten einer Verteilung basieren, bei denen sich die Abweichung „auf-exponentieren“; vgl. Ratcliff, 1979). Die Schätzungen sind dann hochgradig instabil und man braucht wesentlich mehr Datenpunkte, um eine stabile Verteilung zu erhalten. (Dies ist auch genau der Grund, warum sich non-parametrische Schätzmethoden in der Simulationsstudie von van Zandt, 2000, als vergleichsweise ineffizient erwiesen haben). Ein weiterer Nachteil verteilungsfreier Verfahren ist, dass es bei ihnen keine einheitlichen Parameter gibt, die in allen Untersuchungsbedingungen gleich sind. Sie eignen sich deswegen eher für eine möglichst detaillierte Darstellung der in einem Datensatz vorhandenen Muster. Für den inferenzstatistisch-konfirmatorischen Vergleich verschiedener Aspekte mehrerer RT-Verteilungen eignen sie sich zunächst einmal weniger, weil eine numerische Kontrastierung häufig nur indirekt möglich ist (vgl. aber Ramsay & Silverman, 2005).

Ein parametrisches Verfahren erscheint deswegen im vorliegenden Fall attraktiver. Solche Verfahren basieren darauf, dass sie die Verteilungsfunktion einer Zufallsvariable nicht „von Null“ (also mit maximaler Freiheit, aber auch mit maximalen Kosten) schätzen, sondern vielmehr ein bestimmte Verteilungsform a priori vorgeben und nur noch die Parameter dieser Verteilungsfunktion so anpassen, dass sie die empirische Verteilung möglichst gut abbildet. Das Problem der Schätzung der pdf- und cdf-Funktionen wird dann ein Problem der Modellpassung, da ein (statistisches) Modell vorgegeben wird und geprüft wird, wie gut es der Realität entspricht. Der Vorteil ist, dass weniger Informationen zur Spezifizierung der Verteilungsfunktion benötigt werden (da ein Großteil der hierfür benötigten Informationen bereits vorgegeben wird); der Nachteil ist, dass die Schätzungen komplett an der Realität vorbeilaufen können, wenn die zugrunde gelegte Verteilungsfunktion nicht zu den Daten passt. Der „Fit“ der Verteilung (mit optimal geschätzten Parametern) ist deswegen jeweils zu evaluieren.

Wenn man sich dazu entschlossen hat, eine parametrische Verteilung zu „fitten“, dann stellt sich direkt die anschließende Frage, welche Verteilung man nehmen soll. Die statistische Literatur zur RT-Modellierung stellt hier eine große Auswahl von Funktionen zur Auswahl (für einen Überblick vgl. Luce, 1986; van Zandt, 2000). Dabei ist die bereits ansatzweise beschriebene Ex-Gauss-Verteilung sicherlich die Verteilung, die bis jetzt in der kognitionspsychologischen Forschung am häufigsten verwendet wurde. Sowohl aus theoretisch-inhaltlichen als auch empirisch-statistischen Gründen wird jedoch in der vorliegenden Arbeit nicht die Ex-Gauss-Verteilung verwendet, sondern die sog. Weibull-Verteilung. Der hauptsächliche Grund hierfür ist, dass für diese Verteilung ein statistisches (Hierarchical Bayes) Schätzverfahren zur Verfügung steht, welches eine besonders effiziente Schätzung von RT-Verteilungskomponenten ermöglicht (vgl. Rouder, Sun, Speckman, Lu & Zhou, 2003).

Dies ist notwendig, weil man selbst mit einem parametrischen Verfahren immer noch sehr viel mehr Datenpunkte braucht, als wenn man lediglich die zentrale Tendenz einer Zufallsvariable schätzen will. So wird z.B. für die Ex-Gauss-Verteilung häufig die Daumenregel angegeben (vgl. z.B. Ratcliff, 1979; Heathcote et al., 1991), dass man mindestens 100 Datenpunkte pro Person und Untersuchungsbedingung haben sollte, um eine hinreichend reliable Schätzung ihrer Parameter mittels Maximum Likelihood-Verfahren zu erhalten (Studien, welche die Formen verschiedener RT-Verteilungen miteinander vergleichen, erheben mitunter sogar Tausende von Datenpunkten pro Person, vgl. z.B. Ratcliff & Rouder, 1998).

Die Tatsache, dass die RT-Verteilungsanalyse sehr datenintensiv ist, ist einer der Hauptgründe dafür, dass sie selbst in der kognitionspsychologischen Forschung nur relativ selten verwendet wird (für ähnliche Probleme bei neuropsychologischen Sprachexperimenten

vgl. jedoch van Berkum, 2004). Aus den gleichen Gründen scheint eine RT-Verteilungsanalyse auch in der vorliegenden Untersuchung eigentlich unmöglich: Selbst wenn in den geplanten Experimenten lediglich vier Untersuchungsbedingungen miteinander verglichen werden, können selbst hierfür nicht genügend Datenpunkte erhoben werden, weil es schlichtweg nicht genügend Verben mit einem ausreichend starken IC-Bias gibt (die meisten IC-Studien untersuchen lediglich 20-40 Verben!).

Ein Ausweg aus diesem Dilemma besteht darin, zwar weniger Datenpunkte pro Person zu erheben, aber anschließend die Daten verschiedener Personen mit einer quantil-basierten Methode so zu „poolen“, dass eine hinreichend stabile Schätzung gewährleistet ist (sog. *Vincentizing*, vgl. Ratcliff, 1979). Es lässt sich jedoch zeigen, dass eine solche Form der Aggregation bei formvariablen Verteilungen zu artifiziellen Aggregatverteilungen führen kann, d.h. die Gesamtverteilung bildet keine der Einzelverteilungen mehr angemessen ab. Denn Vincentizing-Verfahren sind nur für Funktionen, die ein sog. *location-scale*-Format haben, verzerrungsfrei (vgl. Speckman & Rouder, 2004; Rouder & Speckman, 2004). Für Funktionen, die zur Familie der sog. *location-scale-shape*-Funktionen gehören (wie z.B. die Weibull-Verteilung) oder sich nicht in ein *location-scale*-Format bringen lassen (wie z.B. die Ex-Gauss-Verteilung), ist diese Form der Aggregation jedoch nicht statthaft (für eine formale Definition der verschiedenen Funktions-Familien vgl. Appendix A). Stattdessen soll in der vorliegenden Arbeit eine andere Lösung für das Problem realisiert werden, dass in sprachpsychologischen Untersuchungen auf der Satzebene nicht genügend Datenpunkte für eine RT-Verteilungsanalyse zur Verfügung stehen. Diese basiert darauf, dass ein hierarchisches Schätzverfahren eingesetzt wird, welches auch mit vergleichsweise wenigen Datenpunkten zu reliablen Schätzern führt.

Im Folgenden wird zunächst die Weibull-Verteilung als ein mögliches statistisches Modell zur Modellierung von RT-Verteilungen eingeführt und mit der Ex-Gauss-Verteilung kontrastiert, weil diese bekannter ist und es relevante Überschneidungen bei der Interpretation der Parameter beider Verteilungen gibt (3.3). Anschließend wird dargelegt, dass sich die Parameter der Weibull-Verteilung – trotz zahlreicher mathematischer Widrigkeiten – auch unter Bedingungen, wie sie für die vorliegende Studie realistischerweise anzunehmen sind, hinreichend reliabel schätzen lassen (3.4) und sich ihr Fit evaluieren lässt (3.5). Der Rest des Kapitels widmet sich dann dem vergleichsweise wichtigsten Punkt, nämlich wie sich die Parameter der Weibull-Verteilung theoretisch gehaltvoll interpretieren lassen (3.6). Auch hier wird wieder die Ex-Gauss-Verteilung als bekannter Vergleichspunkt herangezogen.

3.3 Die Weibull-Verteilung als RT-Verteilungsmodell

3.3.1 Eigenschaften der Weibull-Verteilung

Die sog. Weibull-Verteilung ist nach dem schwedischen Mathematiker und Physiker Waloddi Weibull (1887-1979) benannt, der diese Funktion 1939 zum ersten Mal systematisch beschrieb (Weibull, 1939; 1951) und zur Modellierung der Lebensdauer und Bruchfestigkeit verschiedener Materialien benutzte. Obwohl Weibull also bereits sehr früh die breite Anwendbarkeit der Funktion demonstriert hatte, hatte sie ihren eigentlichen „Durchbruch“ erst in den 1960er Jahren. Dann allerdings explosionsartig, was sich sowohl in der Anzahl der theoretischen Artikel als auch in ihrer zunehmenden Verwendung bei den verschiedensten Fragestellungen niederschlägt (manche sprechen sogar von einer „Weibull-Euphorie“). Ihr hauptsächlichliches Verwendungsgebiet ist – nach wie vor – die Überlebens- und Reliabilitätsanalyse von Materialien, und zwar aus Gründen, die nachfolgend noch erläutert werden.

Drei Gründe sind es, welche die Weibull-Verteilung zu einer für viele Forschungsbereiche sehr attraktiven Funktion machen (vgl. Johnson et al., 1994, Logan, 1992):

1. Sie ist eng verwandt mit der bekannten und häufig zur Modellierung von Wachstums- und Verfallsprozessen genutzten Exponentialverteilung (vgl. Johnson et al., 1994, ch. 19). Genauer gesagt: Die Weibull-Verteilung ist eine Generalisierung der Exponentialverteilung und enthält diese als Spezialfall. Denn eine exponentialverteilte Zufallsvariable wird durch den in (1) dargestellten funktionellen Zusammenhang beschrieben (vgl. Hallinan, 1993):

$$(1) \quad p(y) = e^{-y} \quad \text{wobei} \quad y = p(x) = \frac{x - \theta}{\sigma} \quad (\theta < x; \sigma > 0).$$

Dabei ist θ der Lokations-Parameter und σ der scale-Parameter der Verteilung. Die Exponentialverteilung gehört damit zur Familie der sog. *location-scale*-Funktionen (vgl. Appendix A.1). Der Lokationsparameter θ ist in diesem Fall ein shift-Parameter, der angibt, um wie viele Einheiten der Ansatzpunkt der Funktion auf der x-Achse in Relation zum Nullpunkt „verschoben“ ist. Der scale-Parameter σ indiziert hingegen den „Krümmungsgrad“ der Verteilung: Je kleiner er ausfällt, desto steiler ist ihr Verlauf (vgl. Abb. 3.6).

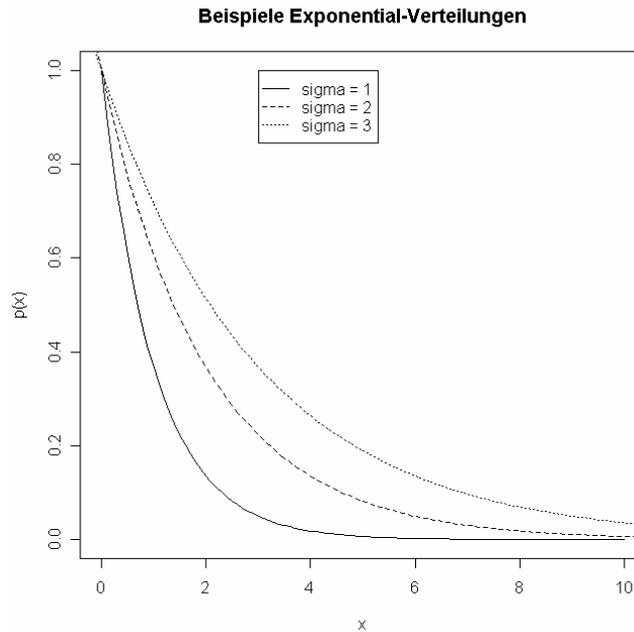


Abb. 3.6: Exponentialverteilung mit drei unterschiedlichen scale-Parametern. Der Lokations-Parameter ist auf 0 fixiert.

Die Weibull-Verteilung ist nun eine direkte Erweiterung der Exponentialverteilung, weil bei ihr der Exponent der Exponentialverteilung selbst wieder exponiert wird. Durch diese Ergänzung ergibt sich dann die in (2) dargestellte Funktion. Die Exponential- und die Weibull-Verteilung unterscheiden sich also nur durch den neu hinzugekommenen Parameter c .

$$(2) \quad p(y) = e^{-y} \quad \text{wobei} \quad y = p(x) = \left(\frac{x - \theta}{\sigma} \right)^c \quad (\theta < x; \sigma > 0, c > 0).$$

Die korrespondierende pdf-Funktion ist in (3) angegeben und ihre cdf-Funktion in (4):

$$(3) \quad pdf = p_X(x) = \frac{c}{\sigma} \left(\frac{x - \theta}{\sigma} \right)^{c-1} e^{-\left(\frac{x - \theta}{\sigma} \right)^c},$$

$$(4) \quad cdf = P_X(x) = 1 - e^{-\left(\frac{x - \theta}{\sigma} \right)^c}.$$

Die Weibull-Verteilung gehört damit zur Familie der *location-scale-shape*-Funktionen (vgl. Appendix A.2). Der neu hinzugekommene Parameter c ist der shape-Parameter der Verteilung. Wenn er den Wert 1 annimmt, ist die Weibull-Verteilung offensichtlich mit der ursprünglichen Exponentialverteilung identisch. Wenn c den Wert 2 annimmt, dann wird die Weibull-Funktion zur sog. Rayleigh-Funktion (vgl. Johnson et al., 1994, ch. 18). Durch die mathematische Nähe der Weibull-Verteilung zu diesen beiden anderen Verteilungen ist ihre Berechnung wesentlich vereinfacht, was ein weiterer Grund dafür ist, warum in vielen Modellen auf sie zurückgegriffen wird.

Der Erwartungswert der Weibull-Verteilung ergibt sich nach (5)

$$(5) \quad E(X) = \Gamma\left(1 + \frac{1}{c}\right) \times \sigma + \theta;$$

wobei σ der scale-Parameter der Verteilung ist, c der shape-Parameter und θ der shift-Parameter. Das Zeichen Γ steht hingegen nicht für einen Parameter der Verteilung, sondern ist eine Abkürzung für die Gamma-Funktion (*nicht*: Gamma-Verteilung!), also das 2. Eulersche Integral ($\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt$). Die Varianz der Weibull-Verteilung wird dann analog durch die

Funktion (6) beschrieben:

$$(6) \quad \text{Var}(X) = \left(\Gamma\left(1 + \frac{2}{c}\right) - \Gamma\left(1 + \frac{1}{c}\right)^2 \right) \times \sigma^2.$$

2. Die Existenz eines zusätzlichen shape-Parameters ist dann auch schon direkt der zweite Grund für die ausgesprochen weite Verbreitung der Weibull-Funktion. Denn dadurch kann sie im Gegensatz z.B. zur Exponentialverteilung sehr viele unterschiedliche Formen annehmen. Sie ist deshalb ausgesprochen flexibel und kann viele empirische Verteilungen besser (d.h. fehlerfreier) modellieren als die „starre“ Exponentialverteilung.

Für den Bereich $c \leq 1$ ist die Wahrscheinlichkeitsdichtefunktion einer Weibull-verteilten Zufallsvariable (im Folgenden der Einfachheit halber schlicht Weibull-Verteilung genannt) umgekehrt J-förmig mit einem Maximum an ihrer unteren Grenze. Bei $c = 1$ entspricht ihre Krümmung der der Exponentialverteilung, bei Werten unter 1 ist sie sogar noch gekrümmter. Für den Wert $c > 1$ sieht der Verlauf der Weibull-Funktion hingegen glockenförmig aus, mit einem eindeutigen Gipfel bei dem Wert $x = \sigma \left(\frac{c-1}{\sigma} \right)^{\frac{1}{c}} + \theta$. Im Intervall $c = [1, 3.6)$ ist die Verteilung rechtsschief ($\beta_1 > 0$, vgl. Appendix A.3). Die Bei einem Wert von $x \approx 3.6$ ist die Weibull-Verteilung approximativ normal verteilt und im Intervall $(3.6, \infty]$ linksschief ($\beta_1 < 0$). Die Kurtosis der Verteilung (β_2 , vgl. Appendix A.3) hat ein Minimum an der Stelle $x = 3.35$ und steigt zu beiden Grenzen monoton an. Abb. 3.7 zeigt verschiedene Kurven der sog. Standardnormal-Weibull-Verteilung (mit $\theta = 0$ und $\sigma = 1$), deren Verlauf alleine vom shape-Parameter abhängt.

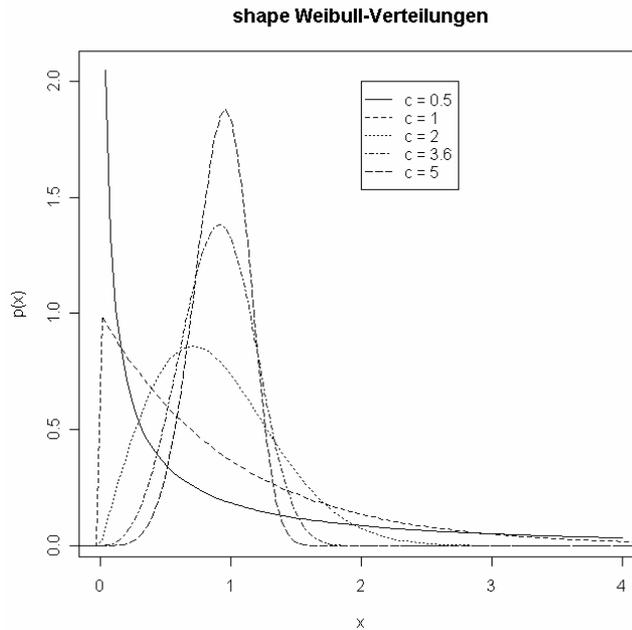


Abb. 3.7: Abhängigkeit der Form der Weibull-Verteilung vom shape-Parameter c ($\theta = 0$, $\sigma = 1$).

Für den Bereich der Reaktionszeit-Modellierung ist hingegen von besonderer Wichtigkeit, dass die Weibull-Verteilung gerade in ihrem rechtssteilen Bereich (c zwischen 1 und 3.6) eine sehr große Ähnlichkeit zu vielen typischen, empirisch vorfindbaren Reaktionszeitverteilungen aufweist. Von großem Vorteil ist auch, dass durch eine einfache Erhöhung des shape-Parameters ein stufenloser Übergang von einem extrem exponentiell verteilten Verlauf bis hin zu einer approximativen Normalverteilung erreicht werden kann. Ähnlich wie bei der Ex-Gauss-Verteilung (vgl. Abschnitt 3.3.2) gibt der shape-Parameter so etwas wie das „Mischungsverhältnis“ zwischen der exponentiellen und normalen Reaktionszeitkomponente an, allerdings auf eine etwas andere Weise. Wenn der shape-Parameter gegen den Wert 1 geht, dann wird die Verteilung von der Exponentialkomponente dominiert, bei einem shape-Wert von 3.6 hingegen von der Normalkomponente. Zwischenwerte sprechen hingegen für eine Mischung beider Komponenten. Menschliche Reaktionszeitverteilungen weisen shape-Werte zwischen $c = 1.5$ und 3.5 auf, wobei der Modalwert bei ca. 2 liegt (vgl. Logan, 1992; Rouder et al. 2003). Linksschief sind menschliche Reaktionszeitverteilungen eigentlich nie, obwohl sie – insbesondere bei kleinen Stichproben – aufgrund der Stichprobenvariabilität manchmal so erscheinen.

Beim shift-Parameter θ handelt es sich um einen Lokationsparameter, d.h. seine Funktion besteht darin, die Verteilung auf der x -Achse zu verorten oder zu justieren. Im Gegensatz jedoch zum Mittelwertparameter der Normalverteilung ist dieser „Ort“ nicht der Mittelpunkt der Verteilung, (graphisch: ihre Spitze), sondern vielmehr ihr Anfangspunkt, d.h. die Mindestausprägung der Verteilung. Die Normalverteilung kennt einen solchen Minimalwert gera-

de nicht, da sich ihre beiden Ränder zwar asymptotisch dem Wert 0 annähern, aber kein „kleinster“ oder „größter“ möglicher Datenpunkt existiert. Im Gegensatz zur Normal-, aber auch zur Ex-Gauss-Verteilung verfügt die Weibull-Verteilung also über eine klar definierte untere Grenze. Obwohl sie dadurch als RT-Verteilungsmodell theoretisch wesentlich plausibler ist (s.u.), erschwert sich ihre Schätzung dadurch erheblich (vgl. Appendix A.3). Rein graphisch führt eine Veränderung dieses shift-Parameters dazu, dass sich die gesamte Verteilung gewissermaßen nach links bzw. rechts auf der x-Achse verschiebt (vgl. Abb. 3.8).

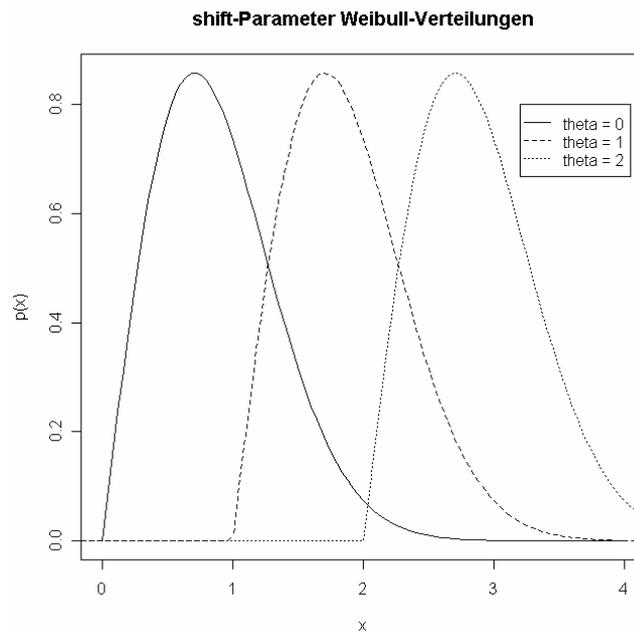


Abb. 3.8: Einfluss des shift-Parameters auf Weibull-Verteilung ($\sigma = 1$, $c = 2$).

Für den scale-Parameter σ gilt hingegen, dass er die Skalierung der Verteilung festlegt, allerdings ohne die grundsätzliche Form der Verteilung zu verändern, denn dafür ist ja bereits der shape-Parameter zuständig. Ähnlich wie bei der Normalverteilung wird die Verteilung durch eine Erhöhung dieses Parameters breiter, d.h. ihre Datenpunkte verteilen sich über einen größeren Abschnitt der x-Achse. Gleichzeitig – und das ist bei der Weibull-Verteilung das auffälligere Merkmal – verschiebt sich ein nicht unwesentlicher Anteil der Verteilungsmasse zu den Rändern, wodurch der Gipfel immer niedriger ausfällt (vgl. Abb. 3.9). Der Ansatzpunkt der Verteilung bleibt hingegen identisch – im Gegensatz z.B. zur Normalverteilung.

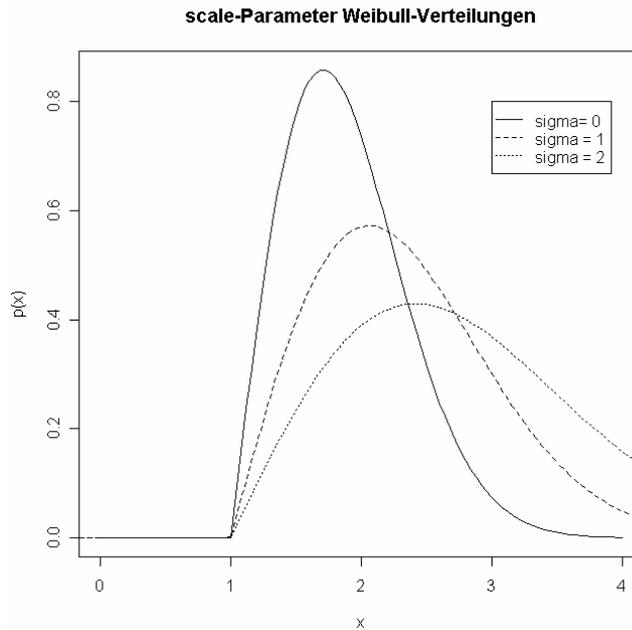


Abb. 3.9: Einfluss des scale-Parameters auf Weibull-Verteilung ($\theta = 1$, $c = 2$).

3. Der dritte und vielleicht wichtigste Grund für die große Beliebtheit der Weibull-Verteilung ist ihre stringente theoretische Verbindung zur Mathematik der Extremwerte, also der Minima und Maxima von Verteilungen. Warum sollte sich jemand für die Extremwerte eine Verteilung interessieren? Wie bereits erwähnt, spielt die Weibull-Verteilung ja eine prominente Rolle in der Überlebensanalyse (z.B. der Materialermüdung). So will man vielleicht manchmal wissen, wie lang eine bestimmte Kette (z.B. die des Fahrstuhls des Psychologischen Instituts!) noch halten wird. Und da ja bekannt ist, dass eine Kette immer so stark ist wie ihr schwächstes Glied, ist die Lebensdauer der Kette letztendlich eine Funktion der minimalen Lebensdauer ihrer einzelnen Glieder. Um herauszubekommen, wann die Kette als Ganzes versagen wird, muss man also die Lebensdauer der einzelnen Glieder bestimmen und das Minimum identifizieren. Und wenn man nun nicht nur eine einzige Kette untersucht, sondern ganz viele, dann kann man sich leicht vorstellen, dass nicht bei jeder Kette der gleiche Minimalwert auftreten wird. Vielmehr wird die Ausprägung dieses Minimalwerts von Kette zu Kette variieren und selbst eine Zufallsvariable mit einer bestimmten Verteilung bilden. Konkret: Wenn man von allen Ketten den Minimalwert aufschreibt und aus zählt, wie häufig welcher Wert vorgekommen ist, dann lässt sich die Verteilung der erzielten Häufigkeiten durch eine Kurve beschreiben, die *asymptotische Extremwertverteilung* genannt wird.

Insgesamt gibt es drei dieser asymptotischen Extremwertverteilungen (Typ I-III), und welche im einzelnen Fall die richtige ist, bestimmt sich dadurch, welche Charakteristika die jeweilige Ausgangsfunktion hat, für die das Minimum definiert wurde (vgl. Fisher & Tippett, 1928; Gnedenko, 1943; Gumbel, 1958). Wenn die Ausgangsfunktion z.B. eine Normalverteilung

lung war, dann wird die asymptotische Extremwertverteilung des Minimums vom Typ I sein (auch Gumbel-Verteilung genannt). Wenn die Ausgangsfunktion hingegen von unten begrenzt ist (also einen bestimmten, mindestens notwendigen Wert annehmen wird) und nach oben hin unbegrenzt ist, dann ist die Verteilung des Minimums vom Typ III. Diese ist jedoch mit der Weibull-Verteilung identisch. Oder anders ausgedrückt: Wenn eine bestimmte Variable eine untere Schranke hat (z.B. weil sie nur positive Werte annehmen kann, was sowohl für Lebensdauern als auch Reaktionszeiten per definitionem der Fall ist), dann sind die Minimalwerte dieser Variablen mathematisch zwingend weibull-verteilt. Diese Tatsache ist v.a. für psychologische Modelle von Bedeutung, welche die Weibull-Verteilung zur Modellierung von Abrufprozessen aus dem Langzeitspeicher heranziehen (vgl. z.B. Logan, 1992).

3.3.2 Die Ex-Gauss-Verteilung und ihre Beziehung zur Weibull-Verteilung

In der Psychologie ist die Modellierung von Reaktionszeiten mittels der Weibull-Verteilung vergleichsweise unüblich. Wenn überhaupt eine RT-Verteilungsanalyse vorgenommen wird, dann wird hierfür fast ausschließlich die Ex-Gauss-Verteilung verwendet. Von daher ist es nützlich, kurz darzustellen, wie sich die Parameter der Weibull-Verteilung zu denen der Ex-Gauss-Verteilung verhalten.

Die Ex-Gauss-Verteilung war bereits in den 1960er Jahren als mögliches statistisches Modell für Reaktionszeitverteilungen vorgeschlagen worden (vgl. McGill, 1963; Hohle, 1965), verschwand dann allerdings zunächst wieder für ein Jahrzehnt von der Bildfläche. Ihr „Comeback“ wurde dann ganz wesentlich von Roger Ratcliff initiiert, welcher gegen Ende der 1970er Jahre der Ex-Gauss-Verteilung sowohl durch ihre empirische Verwendung (Ratcliff, 1978) als auch durch die Aufarbeitung ihrer mathematischen und schätztheoretischen Grundlagen (Ratcliff, 1979) wieder Aufwind gab. Aber auch danach war ihr Einsatz zunächst nur ein paar Spezialisten vorbehalten (Luce, 1986; Hockley, 1984; Dawson, 1988). Sehr einflussreich für ihre „Popularisierung“ war ein Tutorial-Artikel zum Stroop-Effekt (Heathcote et al. 1991) und die Tatsache, dass zunehmend Programme zur Schätzung ihrer Parameter geschrieben wurden, die vergleichsweise einfach zu benutzen waren (Heathcote, 1996; Cousineau & LaRoche, 1997, Dolan et al., 2002; im Überblick: Cousineau et al., 2004). Dies machte es auch für „normal sterbliche“ Psychologen/innen möglich, die Ex-Gauss-Verteilung in ihrer Forschung zu verwenden (vgl. u.a. z.B. Leth-Steensen, King Elbaz & Douglas, 2000; Spieler, Balota & Faust, 1996; Wixted & Rohrer, 1993; Rohrer & Wixted, 1994).

Obwohl sich die Ex-Gauss- und die Weibull-Verteilung oberflächlich, also allein von ihrer Form her, ziemlich ähnlich sehen und auch ein vergleichbar flexibles Verhalten aufwei-

sen, besteht zwischen den beiden Verteilungen mathematisch gesehen ein großer Unterschied: Bei der Weibull-Verteilung handelt es sich um eine eigenständige Verteilung, bei der Ex-Gauss-Verteilung hingegen um ein Konvolut, also um eine Mischung zweier unabhängiger Verteilungskomponenten. Entsprechend ergeben sich Erwartungswert und Varianz der Verteilung auch ganz einfach als Summen der Mittelwerte und Varianzen der beiden beteiligten Komponenten (vgl. (7) und (8)):

$$(7) \quad E(X) = \mu + \tau,$$

$$(8) \quad \text{Var}(X) = \sigma^2 + \tau^2.$$

Die scheinbare Einfachheit der Ex-Gauss-Verteilung bricht allerdings zusammen, wenn man sich ihre Wahrscheinlichkeitsdichte- und Verteilungsfunktion ansieht (vgl. (9) und (10)), wobei Φ für das Integral der Standardnormalverteilung steht):

$$(9) \quad \text{pdf} = p_x(x) = \frac{1}{\tau} e^{\frac{\mu-x}{\tau} + \frac{\sigma^2}{2\tau^2}} \Phi\left(\frac{x-\mu}{\sigma} - \frac{\sigma}{\tau}\right),$$

$$(10) \quad \text{cdf} = P_x(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) - e^{\frac{\sigma^2}{2\tau^2} \frac{\mu-x}{\tau}} \Phi\left(\frac{\mu-x}{\sigma} - \frac{\sigma}{\tau}\right).$$

Hinzu kommt, dass sich die Ex-Gauss-Verteilung nicht in location-scale-(shape)-Form bringen lässt (Thomas & Ross, 1980; vgl. auch Appendix A.2). Dies hat Konsequenzen: U.a. lassen sich bestimmte Formen der Datenaggregation (wie z.B. das verbreitete Vincentizing) trotz des Vorschlags von Ratcliff (1979) nicht sinnvoll bei der Ex-Gauss-Verteilung anwenden, weil die gemittelte Gesamtverteilung keine forminvariante Abbildung der Einzelverteilungen ist (Rouder & Speckman, 2004). Es gibt allerdings noch eine alternative Parametrisierung, unter der die Ex-Gauss-Verteilung sehr wohl location-scale-Format annimmt (vgl. Heathcote et al., 2002, Rouder et al., 2005). Diese entsteht dadurch, dass man die Verteilung mit einem eigenen shape-Parameter versieht, der sich aus dem Verhältnis des τ -Parameters zum σ -Parameter ergibt. Dieser shape-Parameter bestimmt nun allein die höheren Eigenschaften der Funktion (wie Schiefe und Kurtosis), während der μ - und σ -Parameter location- und scale-Funktion übernehmen.

Ein theoretischer Nachteil (aber praktischer Vorteil bei der Parameterschätzung, vgl. Appendix A.4) der Ex-Gauss-Verteilung ist, dass sie (im Gegensatz zur Weibull- und Exponentialverteilung) keine untere Grenze besitzt. Das heißt, nichts an der funktionalen Beschreibung der Verteilung hindert sie daran, auch negative Werte anzunehmen. Dies ist nun für den Reaktionszeitbereich klar inakzeptabel, weil keine Versuchsperson antworten kann, bevor eine Untersuchung begonnen hat. Auch ist es für Reaktionszeitverteilungen aufgrund physio-

logischer und psychologischer Überlegungen sinnvoll, einen bestimmten Minimalwert anzunehmen, der mindestens notwendig ist, um eine (ernsthafte) Antwort zu generieren. Praktisch ist diese Fehlspezifikation der Ex-Gauss-Verteilung allerdings in den allermeisten Fällen kein wirkliches Problem, weil für den Fall, dass das Verhältnis von Mittelwert zu Standardabweichung der Normalkomponente 3:1 beträgt, die Wahrscheinlichkeit für negative Werte lediglich 0.14% beträgt. Das heißt, dass empirisch sowohl die meisten Datenpunkte als auch die sich ergebenden Parameterschätzungen positiv ausfallen; allerdings ragen die generierten Kurven manchmal in den negativen Bereich hinein. Auf theoretischer Ebene lässt sich das Problem z.B. dadurch beheben, dass man entweder die Ex-Gauss-Verteilung trunziert oder aber entsprechende Parameterrestriktionen formuliert (vgl. Dolan et al., 2002).

Zusammenfassend lässt sich feststellen, dass sich die Ex-Gauss-Verteilung zwar theoretisch nur schlecht als plausibles Reaktionszeitmodell begründen lässt (zu weiteren Argumenten, die gegen die Ex-Gauss-Funktion sprechen, vgl. Schwarz, 2001, und Luce, 1986), andererseits aber ausgezeichnet in der Lage ist, empirische Reaktionszeitverteilungen adäquat zu beschreiben (vgl. z.B. Ratcliff & Murdock, 1976, zum Vorteil der Ex-Gauss-Verteilung gegenüber der Gamma- oder Lognormal-Verteilung). Heathcote (1996) führt als einen möglichen Grund für den guten Fit der Ex-Gauss-Verteilung die Tatsache an, dass sie – im Gegensatz zu anderen RT-Verteilungen – über eine eigenständige Gauss-Komponente verfügt und deshalb einfach einen Teil des additiven und normalverteilten Messfehlers mitfitten kann. In der Terminologie der mathematischen und kognitiven Modellierung nennt man dieses Phänomen „overfitting“ (Myung & Pitt, 2002; Pitt & Myung, 2002, Roberts & Pashler, 2000).

Logan (1992, Appendix B) untersuchte erstmals in einer Simulationsstudie, in welcher Beziehung die Parameter der Ex-Gauss- zur Weibull-Verteilung stehen. Dafür generierte er mehrere unterschiedliche Ex-Gauss-Verteilungen jeweils mit einer distinkten Parameterkonstellation und fittete diese dann mittels einer Weibull-Verteilung. Die Parameter der beiden Verteilungen standen in sehr eindeutigen Beziehungen, die sich wie folgt beschreiben lassen: Eine Erhöhung des μ -Parameters bildete sich in einer korrespondierenden Erhöhung des shift-Parameters θ ab. Eine Erhöhung des σ -Parameters führte hingegen tendenziell zu einer Absenkung des shift-Parameters, aber zu einer Erhöhung des scale- und shape-Parameters. Und schließlich führte eine Erhöhung des τ -Parameters zu einem erniedrigten shape-Parameter. Angesichts der Tatsache, dass eine Weibull-Verteilung mit einem shape-Parameter von 1 exponential- und mit einem shape-Parameter von 3.6 normalverteilt ist, ist diese Beziehung sehr plausibel. Allgemein gab es in der Studie von Logan (1992) eine hohe Korrespondenz zwischen den beiden Verteilungen: Die Reliabilität des Weibull-Fits (definiert als quadrierte Kor-

relation zwischen den Kurven der Ex-Gauss- und der Weibull-Verteilung) war mit $r^2 = .992$ ausgesprochen hoch.

Dieser Befund steht jedoch in Widerspruch zu einer Simulationsstudie von Dolan et al. (2002, Studie 3), die mittels eines χ^2 -Verfahrens eine vergleichsweise schlechte Passung der Weibull-Verteilung ermittelte, wenn die Daten in Wirklichkeit ex-gauss-verteilt waren. Ansonsten zeigte die Studie von Dolan et al. das gleiche Beziehungsmuster zwischen den Parametern der beiden Verteilungen, wie es von Logan (1992) beschrieben worden war. Interessanter ist hingegen, dass Dolan et al. (2002) auch die Gegenrichtung realisierten, nämlich wahre Weibull-Daten mittels der Ex-Gauss-Verteilung zu fitten. Leider gingen sie dabei nicht besonders systematisch vor, sodass es schwierig ist, die Beziehungen zwischen den Parametern genau zu beschreiben. Allgemein scheint für die Weibull-Verteilung die *Kombination* der Parameter entscheidender zu sein, da es kaum eindeutige Beziehungen zwischen den einzelnen Parametern gab. Allerdings etablierte sich eine klare Asymmetrie in der Schätzgüte: Während Weibull-Parameter für Ex-Gauss-Daten nur schlecht schätzbar waren, war umgekehrt der Fit der Ex-Gauss-Verteilung für Weibull-Daten sehr viel besser.

Rouder et al. (2005) überprüften ebenfalls in einer Simulationsstudie die funktionellen Beziehungen zwischen den Weibull- und den Ex-Gauss-Parametern. Allerdings verwendeten sie zur besseren Vergleichbarkeit die Ex-Gauss-Verteilung in ihrer location-scale-Parametrisierung (s.o.). Auch hier ergab sich eine sehr einfache Beziehung zwischen den beiden Funktionen: Der μ -Parameter der Ex-Gauss-Verteilung bildet sich allein im shift-Parameter ab und der σ -Parameter in der scale-Komponente. Die Veränderung des Ex-Gauss-shape-Parameters hatte hingegen komplexere Auswirkungen: Seine Erhöhung führte einerseits zu einer Erhöhung des scale-Parameters und zu einer Absenkung des shape-Parameters. Dies ist ein Hinweis darauf, dass sich der τ -Parameter der Ex-Gauss-Verteilung primär im scale-Parameter der Weibull-Verteilung niederschlägt.

Zusammenfassend kann man feststellen, dass beide Verteilungen essenziell die gleiche Aufgabe erfüllen: Beide zerlegen eine empirische Verteilung (und damit auch ihren Erwartungswert) in zwei Teilkomponenten, allerdings jeweils unterschiedlich. Bei der Ex-Gauss-Verteilung wird die Verteilung in eine Normalkomponente und eine Exponentialkomponente aufgesplittet, wobei sich die Normalkomponente aus zwei Parametern zusammensetzt, die sich prinzipiell gleichförmig verhalten (je höher der Mittelwert einer Normalverteilung, desto höher fällt auch seine Varianz aus). Bei der Weibull-Verteilung findet hingegen eine Zerlegung in eine shift- und eine scale/shape-Komponente statt, wobei diese zweite Komponente durch die Verknüpfung des scale- und des shape-Parameters gebildet wird, die sich tenden-

ziell gegenläufig zueinander verhalten (je höher der scale- Parameter ausfällt, desto niedriger wird der shape-Parameter ausfallen). An dieser Stelle ist es alleine wichtig festzuhalten, dass es einerseits eine enge Korrespondenz zwischen der Normal- und der shift-Komponente gibt, andererseits aber auch eine funktionale Äquivalenz zwischen der Exponential- und der scale/shape-Komponente.

3.4 Schätzung der Weibull-Verteilung mittels Hierarchischer Bayes-Verfahren

Hat man sich einmal dafür entscheiden, die Weibull-Verteilung als parametrisches Modell zur Modellierung der RT-Verteilung zu verwenden, stellt sich die Frage, wie sich ihre Verteilungsparameter am besten schätzen lassen. Dieser Aspekt, der im Wesentlichen die Reliabilität der verschiedenen Schätzverfahren betrifft, ist leider sehr entscheidend und vielleicht noch wichtiger, als die Wahl der zu schätzenden Verteilungsfunktion selbst. Was nützt einem eine theoretisch überlegene Verteilungsfunktion, wenn man sie nicht vernünftig schätzen kann? Denn wie bereits dargelegt sind die Unterschiede zwischen den verschiedenen für RT-Verteilungen vorgeschlagenen Verteilungsfunktionen (vgl. z.B. Luce, 1986) ziemlich gering und oftmals kann empirisch gar nicht entschieden werden, welche im Einzelfall angemessener ist. Von daher kann es im Einzelfall durchaus rational sein, eine leichte Fehlspezifikation des Modells in Kauf zu nehmen, wenn sich dadurch eine noch gröbere Verzerrung durch stark fehlerbehaftete Parameterschätzungen vermeiden lässt. Dies gilt insbesondere, wenn die Funktion des Modells primär in der deskriptiven Datenbeschreibung liegt und sich die Interpretation der Parameter der konkurrierenden Modelle stark ähneln. Konkret besteht der Konflikt im vorliegenden Fall in der Frage, ob eher die Weibull- oder vielmehr die Ex-Gauss-Verteilung als parametrisches Modell angesetzt werden sollte: Einerseits ist die Weibull-Verteilung eine ausgesprochen plausible Verteilungsfunktion für RT-Daten (s.o.), andererseits ergeben sich praktische Vorteile bei der Schätzung der Ex-Gauss-Parameter.

Allgemein sind die Ausgangsbedingungen zur Schätzung von RT-Verteilung in der vorliegenden Untersuchung (wie bei fast allen kognitionspsychologischen Experimenten) nicht besonders gut (vgl. Rouder et al., 2005): Meist steht nur eine begrenzte Menge an Versuchspersonen zur Verfügung und der Umfang einer Untersuchung (und damit die Anzahl der Items pro Person) ist aus den verschiedensten Gründen (wie z.B. Aufmerksamkeits- oder Motivationseffekte) ebenfalls stark restringiert. Im sprachpsychologischen Kontext und speziell bei Untersuchungen zur Satz- bzw. Diskursverarbeitung verstärkt sich diese prinzipielle Problematik noch einmal erheblich: Allein schon aufgrund der Stimuluslänge können nur vergleichsweise wenige Items pro Bedingung bearbeitet werden. Im vorliegenden Fall kommt

noch hinzu, dass Verben ausreichend starker Impliziter Kausalität eher selten im Sprachkorpus auftauchen.

Die Tatsache, dass sich die Effizienz der Parameterschätzung eines RT-Modells hauptsächlich durch die Anzahl der zur Verfügung stehenden Datenpunkte bestimmt, ist der Grund dafür, warum für RT-Verteilungsanalysen typischerweise von einer Vpn sehr viele Items bearbeitet werden. Hinzukommt, dass die Variabilität von RT-Daten in den meisten psychologischen Domänen ausgesprochen hoch ist. Sowohl zwischen einzelnen Items als auch zwischen verschiedenen Vpn bestehen sehr große Unterschiede. Deshalb lassen sich kleinere Itemmen- gen auch nicht unbedingt durch größere Stichprobenumfänge ausgleichen: Zum einen wird die Schätzung durch diese zusätzliche Variabilität sogar noch ineffizienter, zum anderen sind die einzelnen Datenpunkte nicht unabhängig voneinander.

Während die bisher genannten Problempunkte für alle RT-Modelle gleichermaßen zutreffen, kommen noch einige spezielle Probleme bei der Schätzung der Weibull-Verteilung hinzu. Denn leider verhält sie sich ausgesprochen exzentrisch und lässt sich nur sehr schwer und nur unter speziellen Bedingungen schätzen. Für eine Darstellung gängiger Schätzverfahren für die Weibull-Verteilung und ihre Probleme sei hier auf Appendix A.4 verwiesen. In Appendix A.5 findet sich eine Beurteilung über die relative Effizienz der verschiedenen Verfahren. Hier reicht die Feststellung aus, dass in der vorliegenden Arbeit eine relativ neue Schätzmethode verwendet werden soll, nämlich das Hierarchical Bayes-Verfahren von Rouder et al. (2003), welches Ergebnisse liefert, die nicht nur besser ausfallen als für andere Verfahren zur Schätzung der Weibull-Verteilung, sondern auch eine ähnliche Reliabilität wie die Ex-Gauss-Verteilung aufweisen. Dieses Schätzverfahren soll kurz skizziert werden.

Einer der Gründe, warum Bayes-Verfahren bei der Schätzung der Weibull-Funktion eine höhere Effizienz besitzen, ist, dass sie zusätzlich zur Likelihood-Funktion auch Prior-Informationen berücksichtigen (vgl. Appendix A.4). Ein Problem besteht allerdings dann, wenn nicht genügend Vorwissen vorhanden ist, um die Priorverteilung sinnvoll spezifizieren zu können. In solchen Fällen bietet sich der Einsatz eines sog. Hierarchical Bayes-Verfahren an (vgl. Lee, 2004, ch. 8). Ähnlich wie bei der verwandten Methodenklasse der Empirical-Bayes-Verfahren (zur Relation von Hierarchical- und Empirical-Bayes-Verfahren, vgl. Malinverno & Briggs, 2004) wird an der grundsätzlichen Idee einer Priorverteilung festgehalten, gleichzeitig aber berücksichtigt, dass nicht genügend Informationen vorhanden sind, um sie explizit formulieren zu können. Die fehlende Information wird vielmehr direkt den Daten entnommen: Während der Parameterschätzung wird ein Prior „konstruiert“ wird, der einer-

seits der Datenlage gerecht wird, sich andererseits aber auch nicht übermäßig den Stichprobencharakteristika anpasst (und damit nicht mehr generalisierbar wäre).

Damit bieten hierarchische Bayes-Modelle eine Möglichkeit, beide Grundprobleme der RT-Verteilungsanalyse zusammen zu lösen (für ähnliche Vorteile im Kontext Hierarchischer Linearer Modelle, vgl. Raudenbush & Bryk, 2002): Erstens sind die meisten psychologischen Merkmale sehr variabel und weisen eine hohe interindividuelle Varianz auf. Für die Schätzung der Weibull-Verteilung bedeutet das, dass die shift-, scale- und shape-Parameter von Person zu Person stark variieren. Ein hierarchisches Modell verfügt jedoch über zwei separate Modellkomponenten, mit denen die *within*- und *between*-Varianz der Daten gleichzeitig modelliert werden können: Es wird angenommen, dass die Parameter des Modells zwar für jede Person unterschiedlich ausfallen, aber dass sie alle aus einer gemeinsamen „Eltern“-Verteilung stammen. Durch die Einbeziehung von geteilten Stichproben-Informationen können dann zweitens die Parameter auf der Individualebene auch dann reliabel geschätzt werden, wenn nur wenige Datenpunkte zur Verfügung stehen. Denn durch die hierarchische Prior-Verteilung gewinnen die einzelnen Schätzer an Stärke und abweichende Datenpunkte werden entsprechend der allgemeinen Verteilungscharakteristik „korrigiert“.

In Anlehnung an die Arbeiten von Berger und Sun (1993; vgl. auch Davison & Louzada-Neto, 2000) haben Rouder et al. (2003) ein hierarchisches Bayes-Modell zur Schätzung der Weibull-Verteilung formuliert, welches speziell für den Bereich der RT-Modellierung optimiert ist. Bei diesem wird für jede Person ein eigener Satz an Weibull-Parametern bestimmt, aber angenommen, dass die Individual-Parameter aus einer übergeordneten Prior-Verteilung stammen, die wiederum aus den Daten geschätzt wird. Der shift-Parameter wird nicht hierarchisch modelliert, sondern besitzt nur einen einfachen (uninformativen) Prior, der einfach alle Werte eines bestimmten Intervalls als gleichwahrscheinlich ansetzt. Die Prior-Verteilung des scale- und shape-Parameters sind hingegen hierarchisch und bestehen aus zwei Stufen. Die erste Stufe ist im Prinzip ein ganz normaler Prior, der beschreibt, wie die einzelnen Individualparameter innerhalb der Population variieren (*within*-Varianz). Allerdings wird im Gegensatz zum klassischen Bayes-Verfahren angenommen, dass die Prior-Verteilung *unbekannt* ist. Lediglich die generelle Form des Priors wird vorausgesetzt, in diesem Fall, dass er gamma-verteilt ist. Die genaue Form des Priors wird allerdings aus den Daten geschätzt.

Die Gamma-Verteilung ist eine Verteilungsfamilie, die sehr viele unterschiedliche Formen annehmen kann (von fast normal bis zur Gleichverteilung, vgl. Johnson et al., 1994). Sie hat zwei Parameter, einen scale- und einen shape-Parameter. Die beiden Parameter für den scale-Prior werden η_1 und η_2 , die für den shape-Prior ξ_1 und ξ_2 genannt (vgl. (11)). Da sie

bestimmen, wie der Prior der ersten Stufe aussehen wird, nennt man sie die *Hyperparameter* des Priors.

$$(11) \quad \begin{aligned} \text{scale}_i &\sim \text{Gamma}(\eta_1, \eta_2) \\ \text{shape}_i &\sim \text{Gamma}(\xi_1, \xi_2) \end{aligned}$$

Hier liegt nun der eigentliche Unterschied zwischen dem klassischen und dem hierarchischen Bayes-Modell: Bei ersterem werden diese Parameter als bekannt vorausgesetzt (es gibt nur eine Prior-Verteilung, und die gilt für alle). Mathematisch gesehen sind sie also Konstanten, die vorgegeben werden und nicht geschätzt werden müssen. Hierarchische Modelle gehen jedoch davon aus, dass die Form der Priorverteilung in Abhängigkeit von Charakteristika der zweiten Ebene variieren kann (*between-Varianz*). Die Parameter der Priorverteilung sind also Zufallsvariablen, die verschiedene Ausprägungen annehmen können.

Da die genaue Ausprägung der Hyperparameter nicht vorgegeben ist, sondern vielmehr aus den Daten geschätzt wird, muss auch für die Hyperparameter wieder eine eigene Prior-Verteilung angenommen werden, die man entsprechend *Hyperprior* nennt. Im Modell von Rouder et al. wird angenommen, dass auch die Hyperparameter gamma-verteilt sind, und zwar der scale-Hyperprior mit den Parametern a_1 und b_1 bzw. a_2 und b_2 und der shape-Hyperprior mit den Parametern c_1 und d_1 bzw. c_2 und d_2 (vgl. (12)).

$$(12) \quad \begin{aligned} \eta_k &\sim \text{Gamma}(a_k, b_k); k = \{1,2\} \\ \xi_k &\sim \text{Gamma}(c_k, d_k); k = \{1,2\} \end{aligned}$$

Die Parameter für den Hyperprior der zweiten Ebene (a, b, c, d) werden wie gewohnt als fixe Parameter vorgegeben. Rouder et al. nehmen spezifische Hyperprior-Verteilungen an, die von ihnen als „vaguely informative“ bezeichnet werden, d.h. sie sind nicht total uninformativ, sondern berücksichtigen einige generelle Informationen über RT-Verteilungen (wie z.B. die Tatsache, dass sie typischerweise rechtsschief sind etc.). Rouder et al. berichten jedoch, dass die Wahl der Hyperpriors keinen großen Einfluss auf die Schätzung hat. Wenn der Hyperprior einmal spezifiziert ist, dann werden die Werte der Hyperparameter auf der Populationsebene aus den Daten geschätzt. Diese Hyperparameter spezifizieren dann die konkrete Form des Priors, aus dem wiederum die Parameter für die einzelnen Individuen geschätzt werden. Zur Schätzung aller Populations- und Personen-Parameter wird in dem Modell von Rouder et al. der sog. Gibbs-Sampler verwendet (vgl. hierzu Kap. 4).

Hier sollte nur die Grundstruktur des Verfahrens skizziert werden sowie die Fakten, die für das Verständnis des nachfolgenden Vorgehens absolut unverzichtbar sind. Für alle Details

der Berechnung bitte ich, im Zweifelsfall auf die Publikationen von Rouder et al. selbst zurückzugreifen. Konkret wurde in der vorliegenden Arbeit das sog. *single-population-model* von Rouder et al. benutzt. Dieses geht lediglich von einer einzigen Untersuchungsbedingung aus, weitere Faktorstufen auf der Ebene der Hyperparameter wurden nicht implementiert (vgl. hierzu Rouder et al., 2005). Das heißt, der Vergleich der einzelnen Untersuchungsbedingungen erfolgte nicht „innerhalb“ des Modells selbst, sondern wurde nachträglich vorgenommen: Für jede Untersuchungsbedingung wurde einzeln ein eigenes *single-population*-Modell geschätzt und die resultierenden Weibull-Parameter pro V_p dann abgespeichert. Die weitergehende inferenzstatistische Auswertung erfolgte dann im üblichen Rahmen des Allgemeinen Linearen Modells. Alle Schätzungen wurden mit dem Programm R 2.01 vorgenommen (R Development Core Team, 2004), wobei eine von Jeff Rouder zur Verfügung gestellte Syntax-Routine adaptiert und modifiziert wurde (Rouder, 2005b).

3.5 Evaluation der Modellpassung

Hat man einmal – egal mit welcher Methode – möglichst optimale Parameter für die Weibull-Verteilung gefunden, dann stellt sich die anschließende Frage, wie gut die so spezifizierte Verteilung in der Lage ist, die empirisch vorgefundenen Daten abzubilden. Genereller gesprochen ist das die Frage danach, wie gut das mathematische Modell mit der Realität übereinstimmt. Diese Relation wird in der Literatur meist als „Fit“ bezeichnet (Myung & Pitt, 2002). Nur wenn der Fit des Modells halbwegs zufrieden stellend ausfällt, ist man auch hinreichend berechtigt, das Modell als (vorläufige) Approximation der Realität beizubehalten und mit ihm z.B. weitere, spezifischere Hypothesen zu prüfen (vgl. Pitt & Myung, 2003).

Es gibt sehr viele unterschiedliche Methoden, den Fit eines Modells zu bestimmen (vgl. hierzu im Überblick Myung & Pitt, 2002). Zu den bekanntesten und meistverwendeten gehört jedoch sicherlich die große Familie der χ^2 -Methoden (Read & Cressie, 1988). Diese sind nicht nur bei Modellierung von Strukturgleichungsmodellen Standard, sondern werden vielfach auch zur Evaluation des Fits univariater Verteilungsfunktionen herangezogen. Die Logik des Verfahrens ist ungefähr so: Wenn wir (nach Parameterschätzung) die Wahrscheinlichkeitsdichtefunktion des theoretischen Modells kennen und wissen, wie viele Beobachtungen vorgenommen wurden, dann lässt sich berechnen, wie viele Beobachtungen in einen bestimmten Datenbereich fallen sollten. Wenn wir z.B. einen IQ-standardisierten Intelligenztest an 100 Personen durchführen, dann sollten bei Gültigkeit des Normalverteilungsmodells mehr Werte im Bereich zwischen 85 und 115 liegen als in den beiden Randgebieten. Wenn das nicht der Fall ist, dann ist die Normalverteilungsannahme für diesen Datensatz problematisch. Genauso

wird nun auch für RT-Verteilungen der mögliche Wertebereich der Verteilung in verschiedene Abschnitte (sog. *bins*) eingeteilt und berechnet, wie viele Beobachtungen bei Gültigkeit einer bestimmten Verteilungsfunktion mit einer spezifischen Parameterkombination in diesen Abschnitt fallen sollten. Das Gleiche wird dann auch mit der empirisch beobachteten RT-Verteilung gemacht, und die beiden sich ergebenden Häufigkeiten pro *bin* miteinander verglichen. Über alle *bins* hinweg lässt sich ein χ^2 -Wert nach der Formel

$$(13) \quad \chi^2 = \sum_{i=1}^m \frac{(y_i - nf(y_i | \hat{\theta}))^2}{nf(y_i | \hat{\theta})}$$

berechnen, der bei Gültigkeit des Modells mit df (N bins – N Modellparameter – 1) asymptotisch χ^2 -verteilt ist. Das heißt, bei voller Gültigkeit des Modells nimmt die Statistik den Wert 0 an und zunehmend höhere Werte deuten auf eine größere Diskrepanz zwischen Modell und Realität hin. Dieser Missfit kann jedoch – wie das bei jeder Statistik so ist – auch zufällig zustande gekommen sein, weswegen er entsprechend inferenzstatistisch geprüft werden sollte. Wenn ein gegebener χ^2 -Wert also so groß ausfällt, dass er über den kritischen Wert eines beliebig angesetzten Signifikanzniveaus fällt und damit als „signifikant“ bezeichnet wird, dann ist das Modell abzulehnen.

Soweit die Theorie; praktisch ergeben sich im vorliegenden Fall eine Reihe von Problemen, die sich sowohl auf die Weibull-Verteilung als relevante Verteilungsfunktion beziehen, als auch auf die spezifische Datensituation der Studie. So haben sowohl Dolan et al. (2002) als auch Rouder et al. (2005) darauf hingewiesen, dass die Verwendung von χ^2 -Verfahren bei der Weibull-Verteilung wegen ihrer Irregularität problematisch ist (s. Appendix A.4). Die χ^2 -Statistik der Weibull-Verteilung konvergiert bei steigendem N nicht zur nominalen χ^2 -Verteilung, ist also schlichtweg nicht χ^2 -verteilt. Stattdessen wird sowohl der Erwartungswert als auch die Standardabweichung dieses χ^2 -Werts systematisch größer ausfallen, wobei sich jedoch leider das Ausmaß dieses Bias nicht vorhersagen (und damit auch nicht korrigieren) lässt. Man muss also mit der Tatsache leben, dass die χ^2 -Methode im Fall der Weibull-Verteilung zu einer zu progressiven Signifikanzentscheidung kommt und damit Modelle u.U. auch trotz Gültigkeit abgelehnt werden.

Hinzu kommt das Problem, dass in der vorliegenden Arbeit ja mit eher kleinen Itemstichproben gearbeitet wird. Das führt dazu, dass sich für jede Verteilung auch nur vergleichsweise wenige *bins* extrahieren lassen, insbesondere wenn man beachtet, dass χ^2 -Verfahren eigentlich nur bei erwarteten Zellfrequenzen größer 5 anwendbar sind. Denn dann müssen *bins* mit Zellfrequenzen kleiner 5 sogar noch gepoolt werden, wodurch sich ihre An-

zahl evtl. noch weiter reduziert. Bei 40 Datenpunkten und der (sinnvollen) Restriktion offener Randkategorien führt dies zu maximal 7 bins (3 df). Für 20 Items resultieren sogar nur 3 bins, d.h. in diesem Fall stehen keine Freiheitsgrade mehr zur Modellprüfung zur Verfügung. Darüber hinaus kann man sich leicht vorstellen, dass so wenige Vergleichskategorien nur zu einer sehr groben Auflösung führen, weswegen sich schon bei normaler Stichprobenvariabilität relativ schnell Abweichungen zwischen Modell und beobachteter Verteilung ergeben.

Beide Gründe zusammen führen dazu, dass die Aussagekraft des beschriebenen χ^2 -Verfahrens im vorliegenden Fall stark eingeschränkt ist. Allerdings heißt das nun auch wieder nicht, dass es zu vollkommen sinnlosen Aussagen führt. Man sollte es lediglich mit Vorsicht handhaben und nur als ein Kriterium unter vielen bewerten. Konkret wird in der Arbeit wie folgt vorgegangen: In den Fällen, wo dies möglich ist, d.h. bei allen Analysen, die auf mindestens 40 Datenpunkten beruhen, soll das beschriebene χ^2 -Verfahren zur Evaluation der Modellpassung herangezogen werden. Dabei werden – wie bei der Passung univariater Verteilungen üblich – zwei Kennwerte berichtet. Zum einen ein über alle Personen- und Bedingungskombinationen gemittelter χ^2 -Wert, zusammen mit einem ebenfalls über alle Einzelmodelle gemittelten p -Wert zur Signifikanzbeurteilung. Die resultierenden Werte können als durchschnittlicher Fit des Modells auf der Ebene der Einzelmodelle interpretiert werden. Zum anderen wird der Prozentsatz signifikanter Einzelmodelle als grober Indikator der globalen Modellpassung auf Ebene des gesamten Experimentes berichtet. Um die Progressivität der resultierenden χ^2 -Statistik zumindest ansatzweise zu berücksichtigen, wird hierbei ein Signifikanzniveau von $\alpha = .01$ angesetzt. Die konkrete Implementierung des Verfahrens lehnt sich eng an die von Heathcote (1996, 2004) beschriebene Version an, wobei als Mindestwertebereich für ein bin ein Wert von 25 ms (Experiment I) bzw. 5 ms (Experiment II) angesetzt wurde.

3.6 Kognitive Interpretation der Parameter der Weibull- und Ex-Gauss-Verteilung

Es stellt sich nun abschließend die Frage, ob sich die Parameter der Weibull-Verteilung psychologisch sinnvoll interpretieren lassen. Was bilden die Parameter der Weibull-Verteilung eigentlich ab? Das sind Fragen, welche die *Validität* der geschätzten Reaktionszeit-Parameter betreffen. Die vorliegende Arbeit schließt sich einer Interpretation der Weibull-Parameter an, bei der einerseits auf eine zwar reiche, aber im vorliegenden Fall unrealistische Substanzinterpretation verzichtet wird (vgl. hierzu z.B. Logan, 1992), die aber andererseits auch nicht auf ein rein deskriptives Niveau zurückfällt. Vielmehr wird versucht, die positiven Aspekte beider Extrempole der deskriptiven und der substanzialen statistischen Mo-

dellbildung konstruktiv miteinander zu verbinden, indem zwar prinzipiell eine eher deskriptive Perspektive eingenommen, gleichzeitig aber eine vorsichtige Interpretation der einzelnen Parameter im Rahmen einer (seriellen) Zwei-Prozess-Theorie vorgenommen wird (vgl. v.a. Rouder et al., 2003, 2005; für eine ähnliche Argumentation bezüglich der ex-Wald-Verteilung s. Schwarz, 2001). Genauer gesagt wird postuliert, dass sich in den Parametern der Weibull-Verteilung selektiv zwei basale kognitive Prozessklassen abbilden, die grob als „peripher“ vs. „zentral“ bezeichnet werden können (vgl. Schneider & Shiffrin, 1977; Hasher & Zacks, 1979).

Während eine solche Sichtweise für die Weibull-Verteilung noch vergleichsweise neu ist (vgl. Rouder et al., 2003), wird sie für die Ex-Gauss-Verteilung schon seit jeher vertreten. Das ist nicht weiter erstaunlich, weil eine Zwei-Prozess-Interpretation schon gewissermaßen in ihre mathematische Grundstruktur fest eingebaut ist. Denn wie in Abschnitt 3.3.2 erläutert, besteht die Ex-Gauss-Verteilung aus der additiven Verknüpfung einer Normal- und einer Exponential-Komponente. Von daher ist es eine relativ nahe liegende Idee, beide Verteilungskomponenten als Resultat zweier funktional und zeitlich selbstständiger Verarbeitungsschritte aufzufassen, die sich dann zu einer Gesamtverteilung zusammenmischen. Allerdings unterscheiden sich die verschiedenen Ansätze darin, welche kognitiven Teilprozesse von den beiden Teilkomponenten abgebildet werden. Das liegt auch daran, dass die Ansätze aus sehr unterschiedlichen psychologischen Teilgebieten stammen. So wird z.B. in der eher entscheidungstheoretisch orientierten RT-Literatur (im Überblick Luce, 1986) meist zwischen einer „Entscheidungs“- und einer „Residual“-Komponente unterschieden, wobei letztere z.B. alle Prozesse der Stimuluswahrnehmung und -verarbeitung umfasst und erstere den eigentlichen Entscheidungsprozess. Die eher psychomotorisch orientierte RT-Literatur setzt hingegen die „Organisation“ und die „Ausführung“ einer Bewegung als die beiden relevanten Komponenten der Reaktionszeit an (vgl. Hohle, 1965). In der Gedächtnisforschung wird wiederum zwischen einer Initiationsphase der Gedächtnissuche und einem anschließenden kontinuierlichen Suchprozess unterschieden (vgl. Rohrer & Wixted, 1994). Auch wenn diese Interpretationen zunächst sehr unterschiedlich wirken, so haben Balota und Spieler (1999) festgestellt, dass ihnen allen eine Unterscheidung zwischen eher stimulus-bezogenen, peripheren und automatischen Prozessen einerseits und stimulus-fernen, zentralen und kontrollierten Prozessen andererseits zugrunde liegt.

Interessanter Weise bestand allerdings zumindest in der frühen Literatur zur Ex-Gauss-Verteilung Uneinigkeit darüber, welche der beiden Verteilungs-Komponenten welche Prozesse abbildet. Während McGill (1963) die peripheren Prozesse der Exponential- und die zentra-

len der Normalverteilungs-Komponente zugeordnete, vertritt Hohle (1965) genau die gegenteilige Position. Inzwischen besteht jedoch weitgehend Einigkeit darüber, dass Antwortlatenzen für vergleichsweise hierarchie-niedrige Prozesse schnell, wenig variabel und approximativ normal verteilt sind, während Antwortlatenzen für vergleichsweise hierarchie-hohe Teilprozesse langsam, varianzreich und meist (rechts)schief sind. Das heißt, dass es zumindest von der empirischen Befundlage her wahrscheinlicher ist, dass die periphere Komponente normal- und die zentrale exponentialverteilt ist.

Dabei ist klar, dass sich sowohl über die Unterscheidungen der verschiedenen Prozessarten als auch über ihre Zuordnung zu den beiden RT-Komponenten trefflich streiten lässt. Man denke hier nur an die Debatte, ob eine kategorische Trennung zwischen automatischen und kontrollierten Prozessen überhaupt sinnvoll ist oder ob hier nicht eher ein fließender Übergang angesetzt werden sollte (vgl. z.B. Logan, 1988; Goschke, 1996). Genauso erscheint uns heute die etwas starre Zuordnung bestimmter Prozessarten zu verschiedenen RT-Komponenten angesichts unserer besseren Kenntnis der ausgesprochen komplexen kognitiven und physiologischen Teilprozesse etwas vorschnell (vgl. z.B. Heathcote et al., 1991). Gleiches ließe sich von der Plausibilität einer ausschließlich seriellen Architektur oder einer allzu schematisch gehandhabten Unterscheidung zwischen „frühen“ und „späten“ Teilprozessen sagen (Townsend, 1990). Im Vorgriff auf die spätere Diskussion sei hier nur angemerkt, dass es im Folgenden zunächst einmal nur um den heuristischen Wert einer nur sehr groben Unterscheidung zwischen zwei unterschiedlichen Prozessarten geht, die sich auf höchstem Abstraktionsniveau als „hierarchie-niedrig“ vs. „hierarchie-hoch“ bezeichnen lassen. Das heißt, die gewählte Interpretation ist zwar in der Tat eine Zwei-Prozess-Interpretation, darüber hinaus sollten jedoch keine überdifferenzierenden Annahmen über Art und Zusammenwirken der Prozesse gemacht werden. Ihren eigentlichen Wert erhält die Interpretation von RT-Verteilungs-Parametern m.E. sowieso erst, wenn man sie im Kontext einer konkreten Untersuchungssituation betrachtet (s.u.).

In jüngerer Zeit wurde auch für die Parameter der Weibull-Verteilung eine ähnliche Zwei-Prozess-Interpretation vorgeschlagen wie für die Ex-Gauss-Verteilung (vgl. Rouder et al., 2003, 2005). Andere Ansätze vertreten hingegen meist „substanzielle“ Interpretationen unter Rückgriff auf ihre Verbindung zur Mathematik der Extremwerte (vgl. z.B. Ida, 1980; Logan, 1992). Genauso wie bei der Ex-Gauss-Verteilung lässt sich die Zwei-Prozess-Interpretation der Weibull-Parameter an der mathematischen Struktur der Verteilung festmachen: Denn wie die Formel ihres Erwartungswertes zeigt (vgl. Formel (5) in Abschnitt 3.3.1), findet auch hier eine additive Zerlegung des Erwartungswertes in zwei Teilkomponenten statt.

Die erste Teilkomponente wird vom shift-Parameter gebildet, die andere besteht hingegen aus dem mit einer Funktion des shape-Parameters gewichteten scale-Parameter. Von daher scheint hier eine Entsprechung vorzuliegen: Genauso wie die Normalkomponente der Ex-Gauss-Verteilung bildet der shift-Parameter hierarchie-niedrige, periphere Prozesse ab, während sich der scale- und der shape-Parameter wie die Exponentialkomponente der Ex-Gauss-Verteilung auf die zentralen, hierarchie-hohen Prozesse beziehen. Wie in Abschnitt 3.3.2 dargestellt, zeigt sich dieses Muster auch empirisch.

Mitunter wird dem shape-Parameter allerdings noch eine andere Funktion zugesprochen (vgl. Logan, 1992; Rouder et al. 2003, 2005): So geht z.B. Logan (1992) in seiner *instance theory of automaticity* davon aus, dass sich der zunehmende Automatisierungsgrad eines Prozesses in steigenden shape-Werten abbildet, weil bei der Automatisierung eines Prozesses zunehmend mehr Gedächtnis-Instanzen für einen schnellen automatischen Abruf zur Verfügung stehen. Ähnlich argumentieren auch Rouder et al. (2003) dafür, dass Veränderungen des kognitiven Prozessablaufes (Veränderung der Lösungsstrategie, Einfügung zusätzlicher Teilprozesse etc.) zu Veränderungen des shape-Parameter führen. Allerdings gibt es durchaus Gründe, an dieser Annahme zu zweifeln. So bilden sich z.B. Unterschiede in der Driftrate des Diffusionsmodells sowohl im scale- als auch im shape-Parameter ab, obwohl die Prozessarchitektur identisch geblieben ist (vgl. Rouder et al., 2005).

Konkret wird in der vorliegenden Arbeit angenommen, dass sich die in Kap. 1 ausführlich diskutierte Unterscheidung von gedächtnis- und erklärungs-basierten Verarbeitungsprozessen selektiv im shift- bzw. scale-/shape-Parameter der Weibull-Verteilung niederschlägt. Denn gedächtnis-basierte Prozesse sind im Rahmen des *memory-based view* per definitionem Prozesse, die auf einem schnellen, passiven Prozess der Resonanz basieren. Ein bestimmtes sprachliches Element löst bei seiner Verarbeitung eine schnelle parallele Suche im Speichersystem aus, wobei zuvor verarbeitete Elemente genauso aktiviert werden wie evtl. relevantes Weltwissen (vgl. Kap. 1.2). Es handelt sich also um keine Suche im Sinne eines bewussten oder sogar steuerbaren Prüfprozesses, sondern es wird bottom-up einfach alles aktiviert, was irgendwie in einer assoziativen Verbindung mit dem aktuellen Input steht. In ähnlicher Weise wurde auch der *bonding*-Prozess des *bonding-resolution*-Ansatzes von Garrod und Sanford (1989) als ein automatischer *pattern-matching*-Prozess beschrieben, der vollkommen ohne Zutun des/der Rezipienten/in aktiviert wird und ein *bond* zwischen Anapher und potenziellem Antezedens etabliert. Aufgrund dieser Charakteristika ist es eine relativ nahe liegende Idee, den passiv-automatischen Resonanz-Prozess mit der hierarchie-niedrigen, primär stimulus-getriebenen Prozessklasse zu identifizieren, die sich im v.a. im shift-Parameter abbildet.

Im Gegensatz hierzu ist der charakteristische Prozesstypus der erklärungs-basierten Sprachverarbeitung ein aktiver Prozess der Inferenzgenerierung. Dabei kommt schon im Schlagwort des „search after meaning“ der konstruktivistischen Inferenztheorie zum Ausdruck, dass es sich hierbei zumindest partiell um eine gerichtete top-down-Suche handelt, bei der gezielt und ressourcenintensiv nach relevanten Informationen „recherchiert“ wird. Und wenn diese erfolgreich identifiziert wurden, dann kommen noch weitere Prozesse hinzu, wie z.B. Produktionsprozesse zur Konstruktion der Inferenz oder eine finale Kohärenzprüfung (vgl. Kap. 1.2). Wichtig ist v.a., dass all diese Prozesse vergleichsweise hierarchie-hoch sind und der zentralen Verarbeitung bedürfen: „It is difficult for us to envision how a reader might actively search memory without invoking some sort of executive processor“ (Gerrig & O’Brien, 2005, p. 234). Von daher ist es auch hier vergleichsweise plausibel anzunehmen, dass sowohl die erklärungs-basierten Sprachverarbeitungsprozesse im Allgemeinen als auch die *resolution*-Prozesse der pronominalen Auflösung zur Klasse der hierarchie-hohen, analytischen Prozesse gehören. Und diese wird im Rahmen der Zwei-Prozess-Interpretation der Weibull-Verteilung vom scale- bzw. shape-Parameter erfasst.

Diese Zwei-Komponenten-Interpretation der Weibull-Parameter hört sich sehr einfach und plausibel an. Von daher stellt sich die Frage, ob sie nur ein schöner Traum der Forschung ist oder ob es auch empirische Evidenz für sie gibt. Die übliche Art, die Interpretation eines kognitiven Modells zu überprüfen, besteht darin, sie mittels sog. *selectiv-influence*-Studien zu validieren. Grundidee solcher Studien ist es, dass bestimmte kognitive Prozesse durch externe Variablen selektiv beeinflusst werden können. Will man nun wissen, ob ein Modell-Parameter wirklich einen bestimmten kognitiven Prozess abbildet, dann muss man eine solche externe Variable identifizieren, sie als Faktor experimentell manipulieren und dann beobachten, ob sie in der Tat selektiv nur diesen Parameter beeinflusst, andere hingegen nicht. Und da die meisten kognitiven Modelle mehrere Parameter besitzen, lassen sich verschiedene Manipulationen miteinander kombinieren von so die konvergente und diskriminante Validität der Parameter zu bestimmen.

Für die Weibull-Verteilung liegen bisher lediglich zwei Studien vor, die jedoch in die gleiche Richtung weisen (vgl. Rouder et al., 2005): So wird sowohl der symbolische Distanzeffekt (die Unterscheidung zwischen zwei Ziffern ist schwieriger, wenn sich ihre Werte ähneln; vgl. Moyer & Landauer, 1967) als auch die Reaktionszeit, um unterschiedlich helle Quadrate von einem weißen Hintergrund zu unterscheiden (vgl. Ratcliff & Rouder, 1998), allein vom scale-Parameter getragen. Der shift-Parameter von diesen Manipulationen hingegen nicht beeinflusst. Da angenommen wird, dass beide Manipulationen lediglich Entschei-

dungsprozesses beeinflussen (weil sie sich die Diskriminierbarkeit der Stimuli nicht verändert), kann in beiden Fällen der Zwei-Prozess-Ansatz (vorläufig) beibehalten werden.

Da somit für die Weibull-Verteilung bis jetzt noch kaum *selective-influence*-Studien vorliegen, ist es sinnvoll, zur Beurteilung der empirischen Gültigkeit der Zwei-Prozess-Interpretation der Weibull-Parameter approximativ auch auf vergleichbare Studien für die Ex-Gauss-Verteilung zurückzugreifen. Denn für die Ex-Gauss-Verteilung liegen wesentlich mehr *selective-influence*-Studien vor und noch mehr, die nicht speziell zur Validierung der Parameter der Verteilung geplant wurden, aber dafür genutzt werden können. Die Befundlage – die zu umfangreich ist, um sie hier umfassend darstellen zu können – ist sehr gemischt, zumindest soweit sie sich überhaupt sinnvoll beurteilen lässt. Ein Grundproblem von *selectiv-influence*-Studien ist nämlich, dass ein Urteil darüber, ob eine solche Studie erfolgreich war, natürlich von der Richtigkeit der gewählten Manipulation und unserem Wissen über die Struktur der beteiligten Prozesse abhängig ist. Denn zum einen kann es ja durchaus sein, dass ein und dieselbe Manipulation mehrere Parameter gleichzeitig beeinflusst (so wie z.B. bekannt ist, dass im Rahmen des Signaldetektions-Paradigmas viele Manipulationen sowohl die Sensitivität als auch den Bias verändern). Zum anderen werden viele kognitive Prozesse und Aufgaben heute anders konzeptualisiert als früher, meist komplexer. So werden z.B. viele kognitive Prozesse, die man in den 1960er Jahren (noch) für seriell hielt, heute parallel modelliert, wobei die Gedächtnissuche hier nur das prominenteste Beispiel ist. Entsprechend hat sich auch die Sichtweise auf die beteiligten Teilprozesse seit der ersten *selective-influence*-Studie von Hohle (1965) teilweise dramatisch verändert. Man denke z.B. an die vielfältigen Uminterpretationen, die allein der Stroop-Effekt erdulden musste (vgl. MacLeod, 1991): Was heute noch automatisch ist, kann morgen schon kontrolliert sein! Und wer weiß schon genau, wie selbst ein so simpler Aufgabentypus wie die Lexikalische Entscheidungsaufgabe genau funktioniert (vgl. z.B. Ratcliff et al., 2004, und Yap et al., 2006, für einen Vergleich verschiedener Ansätze). Trotz solcher ziemlich prinzipieller Probleme sind die bisher vorgelegten Befunde noch lange nicht wertlos; auch sprechen sie nicht gegen die Logik des *selectiv-influence*-Designs generell. Allerdings machen sie deutlich, dass die Validierung von Parametern, die latente kognitive Prozesse abbilden sollen, ein schwieriges und auch langwieriges Unterfangen ist, das manchmal komplizierter ausfällt als ursprünglich gedacht (vgl. hierzu auch Schwarz, 2001).

Wenn man allerdings die Ergebnisse der bislang vorliegenden Studien zur Ex-Gauss-Verteilung zusammennimmt, dann gibt es durchaus Evidenz dafür, dass der τ -Parameter für höhere kognitive Funktionen und speziell Entscheidungsprozesse indikativ ist: Er korreliert nicht nur substanziell ($r = -.4$) mit der allgemeinen Intelligenz (g ; vgl. Blanco & Alvarez,

1994; Juhel, 1993), sondern hat auch diagnostischen Wert zur Erkennung von Störungsbildern wie ADHS (Leth-Steensen, King Elbaz & Douglas, 2000) oder Alzheimer-Demenz (Spieler et al., 1996), die gemeinhin mit einer Störung kognitiver Kontrollprozesse in Verbindung gebracht werden. Bei der Lexikalischen Entscheidungsaufgabe verursachen Frequenzeffekte fast ausschließlich Effekte im τ -Parameter (Plourde & Besner, 1997). Dies sowie die Tatsache, dass sich solche τ -vermittelten Frequenzeffekte nur bei der Lexikalischen Entscheidungs-, nicht aber bei einer parallelen Benennungsaufgabe ergeben (Balota et al., 1999), spricht für die Zwei-Prozess-Interpretation, weil angenommen wird, dass nur bei der Lexikalischen Entscheidungsaufgabe eine Entscheidungskomponente involviert ist (vgl. Andrews & Heathcote, 2001; Balota, Cortese, Sergent-Marshall, Spieler & Yap, 2004). Darüber hinaus sind auch die meisten bekannten Rekognitionseffekte (wie z.B. Listenlänge, Abstand zwischen Lern- und Testphase etc.) rein τ -vermittelt (vgl. Ratcliff, 1978; Ratcliff & Murdock, 1976, Hockley, 1982; Rohrer & Wixted, 1994).

Den μ -Parameter der Ex-Gauss-Verteilung beeinflussen hingegen tendenziell eher Manipulationen, die perzeptuelle, zeitliche oder motorische Teilkomponenten betreffen. Insbesondere verursachen v.a. Such- und Mustererkennungsprozesse selektive Veränderungen im μ -Parameter. Bei der Lexikalischen Entscheidungsaufgabe werden Erkennbarkeits-, Wiederholungs- und Lexikalitätseffekte v.a. vom μ -Parameter getragen (Plourde & Besner, 1997, Yap & Balota, 2007). Weiterhin beeinflussen den μ -Parameter: a) bei Rekognitions- und *free recall*-Aufgaben die Präsentationsdauer der zu lernenden Wörter und ihre Wiederholungsfrequenz (Wixted & Rohrer, 1993; Rohrer & Wixted, 1994); b) bei recency-Aufgaben (Angabe, welches Wort zuletzt genannt wurde) die serielle Position (Hacker, 1980); c) bei wahrnehmungsbezogenen Aufgaben die Vorperiode (z.B. Hohle, 1965); d) bei Suchaufgaben im Neisser-Paradigma eigentlich fast alle Manipulationen (Hockley, 1984); e) im Stroop-Paradigma wird der Erleichterungs-Effekt v.a. vom μ -Parameter getragen, während der Interferenz-Effekt sowohl im μ - als auch im τ -Parameter zu verorten ist (Heathcote et al., 1991, Mewhort et al., 1992).

Während die bisher berichteten Befunde *cum grano salis* für die Zwei-Prozess-Interpretation sprachen, gibt es mitunter auch Ergebnisse, die für diesen Interpretations-Ansatz problematisch sind oder zumindest nicht direkt von ihm erklärt werden können: So ist zunächst einmal nicht direkt ersichtlich, warum die Listenlänge bei Rekognitionsaufgaben allein zu einer Erhöhung des τ -Parameters führt, bei visuellen Suchaufgabe hingegen zu einer Erhöhung des μ -Parameters. Genauso gibt es (zunächst) keine plausible Erklärung dafür, warum sich bei der Lexikalischen Entscheidungsaufgabe Wiederholungseffekte bei hochfrequen-

ten Wörtern allein im μ -Parameter zeigen, bei niedrigfrequenten Wörtern hingegen auch im τ -Parameter (Balota et al., 1999). Und warum gibt es Frequenzeffekt gerade bei Kategorisierungsaufgaben nur in der Normalkomponente? Von solchen unklaren oder sogar divergierenden Befunden ließe sich noch eine ganze Reihe aufzählen.

Aufgrund solcher und ähnlicher Befunde stehen manche Autoren (z.B. Luce, 1986), dem ganzen Projekt, eine konsistente Zuordnung zwischen verschiedenen Reaktionszeitparametern und distinkten kognitiven Teilprozessen zu finden, eher skeptisch gegenüber. Andere, wie z.B. Heathcote et al. (1991, Andrews & Heathcote, 2001), nehmen von einer substanziellen Interpretation von RT-Parametern ganz Abstand und propagieren stattdessen eine rein deskriptive Verwendung der Ex-Gauss-Verteilung. Angesichts der Differenziertheit und Heterogenität des dargestellten Befundmusters sollte man jedoch solch kategorische Urteile nicht vorschnell fällen. Denn schon bei cursorischer Betrachtung der dargestellten Befunde wird deutlich, dass man den Erfolg der Weibull- und Ex-Gauss-Verteilung, zwischen analytischen und nicht-analytischen kognitiven Prozessen zu differenzieren, nicht pauschal beurteilen kann. Vielmehr scheint der Erklärungswert der Parameter je nach untersuchter Domäne bzw. experimentellem Paradigma zu variieren. Da mir keine einzige sprachpsychologische Arbeit bekannt ist, in welcher entweder die Weibull- oder die Ex-Gauss-Verteilung auf Satzebene verwendet wurde, ist schwer abzuschätzen, wie tragfähig eine Zwei-Komponenten-Interpretation der Parameter im vorliegenden Fall schlussendlich sein wird. Es wäre deshalb ausgesprochen wünschenswert, wenn es noch andere sprachpsychologische Studien gäbe, die ebenfalls mit RT-Verteilungsanalysen arbeiten würden, weil sich dann die verschiedenen Effekte miteinander vergleichen ließen. Solange dieses jedoch nicht der Fall ist, wird sich der eigentliche Wert der hier vertretenen Interpretation erst bei ihrer Anwendung wirklich offenbaren. Und mangels „externer“ Vergleichspunkte zur Absicherung der gewählten Interpretation müssen zu ihrer Sicherung verstärkt „interne“ Validierungsaspekte berücksichtigt werden, welche die Konstruktvalidität der Parameter betreffen. In Kap. 2 wurden zwei Dissoziationsdimensionen vorgestellt, mit denen sich im Rahmen der vorliegenden Untersuchung hierarchie-niedrige *bonding*- und hierarchie-hohe *resolution*-Prozesse voneinander trennen lassen. Die erste bezog sich auf den unterschiedlichen zeitlichen Verlauf der beiden Prozesse, die andere auf differenzielle Interaktionen der IC-bezogenen Verarbeitungsprozesse mit dem Gender-Cue.

Als Fazit lässt sich festhalten, dass bei Durchsicht der vorliegenden Literatur durchaus Ansatzpunkte für die Validität der Zwei-Prozess-Interpretation gefunden werden konnten. Weitere Aspekte lassen sich im Laufe der Untersuchung noch empirisch prüfen. Unterm

Strich bleibt jedoch eine gewisse Unsicherheit bezüglich der exakten Interpretation der Weibull-Parameter, die sich aufgrund der mangelhaften empirischen Befundlage auch nicht ganz ausräumen lässt. An dieser Stelle bleibt dann nur der Hinweis auf die altbekannte Tatsache, dass Daten zunächst einmal Daten sind und lediglich ihre Interpretation Gegenstand der wissenschaftlichen Kontroverse sein kann. Von daher ist es möglich und steht jedem frei, die berichteten Parameter im Zweifelsfall rein deskriptiv als Beschreibung der angefallenen RT-Verteilung zu interpretieren – eine Beschreibung immerhin, die um ein Vielfaches differenzierter ist als die Betrachtung allein der mittleren RT.

4. Simulationsstudien

4.1 Einleitung

In Kap. 3 wurde dargestellt, dass es zur Untersuchung der Interaktion gedächtnis- und erklärungs-basierter Verarbeitungsprozesse bei der Satzverarbeitung sinnvoll ist, nicht nur die mittlere RT zu analysieren. Vielmehr bietet es sich an, die Verteilung der RTn in ihrer Gesamtheit zu betrachten, weil dies eine detaillierte und theoretisch reichhaltigere Hypothesenprüfung erlaubt. Es wurde dargestellt, dass sich die Weibull-Verteilung als ein adäquates RT-Modell erweist und sich ihre Parameter heuristisch als Indikatoren der beiden relevanten Prozessklassen interpretieren lassen. Zudem steht mit dem von Rouder et al. (2003) entwickelten Hierarchical Bayes-Verfahren eine statistische Methode zur Verfügung, die eine effiziente Parameterschätzung auch bei kleiner Datenbasis verspricht. Ein solche ist aber für eine sprachpsychologische Untersuchung auf der Satzebene als realistisch anzusetzen. Da die Durchführbarkeit des gesamten Vorhabens kritisch auf der Effizienz des skizzierten Verfahrens von Rouder et al. beruht, ist es sinnvoll, einige methodische Vorüberlegungen anzustellen, um eine sinnvolle und möglichst effiziente Datenanalyse zu gewährleisten.

Hierzu wurden im Vorfeld der Untersuchung zwei Simulationsstudien durchgeführt, die zum einen sicherstellen sollten, dass sich das Verfahren überhaupt für die erwartbare Datensituation eignet, und um Randbedingungen zu identifizieren, die ihren Einsatz begünstigen und deshalb versuchsplanerisch zu berücksichtigen sind. In diesem Sinne verfolgten die Simulationsstudien gleich mehrere Ziele auf einmal:

1. Zunächst ging es darum, das von Rouder et al. (2003) entwickelte Hierarchical Bayes-Verfahren auch für Reaktionszeitbereiche zu erproben, wie sie in der vorliegenden Studie wahrscheinlich anfallen würden. In der vorliegenden Untersuchung werden zwei verschiedene Erhebungsverfahren eingesetzt (vgl. Kap. 2): In Experiment I kommt eine pronominale Auflösungsaufgabe zum Einsatz, bei der Vpn Sätze lesen, die ein ambiges Pronomen enthalten. Ihre Aufgabe ist es, so schnell wie möglich den intendierten Referenten des Pronomens zu benennen. In Experiment II sollen hingegen wortweise Lesezeiten erhoben werden, d.h. die Vpn lesen einen Satz Wort für Wort, die hierfür benötigte Zeit wird jeweils einzeln gemessen.

Wichtig ist, dass sowohl bei einer pronominalen Auflösungsaufgabe als auch bei wortweisen Lesezeiten ganz andere Daten anfallen als in „normalen“ Reaktionszeitexperimenten. Während z.B. die Antwortlatenzen bei einer Lexikalischen Entscheidungsaufgabe vergleichsweise kurz sind (mit Mittelwerten um 600 ms), lesen und bearbeiten die Vpn bei der pronominalen Auflösungsaufgabe ganze Sätze auf einmal, was mehrere Sekunden in An-

spruch nimmt. Hinzukommt, dass „typische“ Reaktionszeitenverteilungen weniger variabel sind, da die zu bearbeitenden Aufgaben sehr einfach und klar strukturiert sind (Wort aussprechen; Entscheidung über Reiz vs. kein Reiz, Entscheidung Wort vs. Nicht-Wort; kategoriale Entscheidung etc., s.o.). Bei solchen Aufgaben werden kaum Fehler gemacht, und sie beanspruchen nur sehr wenige kognitive Operationen gleichzeitig. Im Vergleich hierzu erfasst die pronominale Auflösungsaufgabe als unspezifisches Globalmaß des Auflösungsaufwandes alle Prozesse der Satzverarbeitung gleichzeitig: von den primär perzeptuellen Teilprozessen über den lexikalischen Zugriff und die morpho-syntaktische und semantische Verarbeitung bis hin zum Aufbau der Diskursrepräsentation. Insgesamt ist dies ein vergleichsweise „langer“ kognitiver Verarbeitungsweg, der entsprechend mehr Zeit in Anspruch nimmt und durch die Einbeziehung vieler Teilprozesse variabler und instabiler ausfällt.

In Experiment II werden hingegen wortweise Lesezeiten als lokale Online-Indikatoren erhoben. Dadurch lassen sich die dynamischen Veränderungen der beteiligten Verarbeitungsprozesse im Laufe der Satzverarbeitung erfassen. Wortweise Lesezeiten haben ebenfalls bestimmte Charakteristika, die für „normale“ Reaktionszeitdaten eher unüblich sind. Zum einen fallen Leszeiten typischerweise sehr variabel aus, da sie stark von den verschiedensten linguistischen Faktoren auf der Wort-, Satz- und Diskursebene beeinflusst werden. Sie sind jedoch auch von vielen allgemeinen Variablen der Untersuchungsperson und -situation abhängig (Arbeitsgedächtniskapazität, Aufmerksamkeit, Lautstärkepegel während der Aufgabenbearbeitung etc.) und ausgesprochen fehleranfällig. Sie können leicht Werte zwischen 200 ms (Verarbeitung kurzer phraseninterner Wörter mit gerigem semantischen Gehalt und klarer syntaktischer Funktion, z.B. Artikel) und 1 s annehmen (z.B. Verarbeitung satzfinaler Wörter oder syntaktisch ambiger Strukturen; vgl. hierzu Just, Carpenter & Wooley, 1982, Graesser et al., 1980, 1986). Aber selbst Leszeiten für funktional äquivalente Wörter sind meist sehr vari- anreich, stark ausreißerbehaftet und häufig bimodal verteilt. Diese Ausgangssituation (niedrige Parameterausprägung bei gleichzeitig hoher Variabilität) setzt die zugehörige Verteilung unter starken Stress und ist prinzipiell für alle Schätzverfahren zunächst problematisch.

Zusammengefasst bedeutet das, dass man nicht einfach davon ausgehen kann, dass sich alle Ergebnisse der bisherigen RT-Forschung reibungslos auf die vorliegende Datenlage übertragen lassen. Das gilt auch für primär methodische Aspekte, wie z.B. die Beurteilung der Güte der Parameterschätzung. In Simulationsstudie I soll deswegen die Effizienz des Verfahrens von Rouder et al. (2003) zunächst für RT-Verteilungen überprüft werden, wie sie bei der pronominalen Auflösungsaufgabe erwartbar sind. In Simulationsstudie II wird hingegen die Verteilung von wortweisen Lesezeiten simuliert.

2. Weiterhin ist im Vorfeld der Untersuchung die Frage relevant, wie viele Stimuli überhaupt pro Untersuchungsbedingung benötigt werden, um eine halbwegs reliable Schätzung der Verteilungsparameter auf der Individualebene zu erhalten. Dies hat u.U. große Konsequenzen, und zwar sowohl für das Design der Untersuchung als auch für die eingesetzten Aufgabenarten. Wenn z.B. die vorliegende Itemmenge nicht ausreichen sollte, um den Gender- und den IC-Cue gleichzeitig zu manipulieren, dann muss evtl. auf ein *between*-Design zurückgegriffen werden etc.

Als ein erster Faktor wurde in den Simulationsstudien deswegen die Anzahl der zur Verfügung stehenden Datenpunkte pro Person und Untersuchungsbedingung variiert. So lässt sich bestimmen, wie viele Items von jeder Vpn in jeder Versuchsbedingung bearbeitet werden sollten, um eine akzeptable Schätzung der Verteilungsparameter zu erhalten. Zusätzlich sollte überprüft werden, ob die Größe der verwendeten Stichprobe einen Einfluss auf die Parameterschätzung hat. Deswegen wurde als zweiter Faktor gleichzeitig die Stichprobengröße auf Personenebene variiert. Die Ergebnisse sollen zur Festlegung des optimalen Stichprobenumfangs herangezogen werden und damit eine maximal effiziente Versuchsplanung gewährleisten.

Ein weiterer Punkt, der speziell für Bayes-Verfahren wichtig ist, betrifft die Festlegung der Prior-Verteilung. Diesen sollte man möglichst theoriegeleitet und unabhängig von den eigentlichen Daten im Vorfeld der Untersuchung festlegen, damit diese Wahl nicht die Generalisierbarkeit der Ergebnisse gefährdet. Um einen „intelligenten“ Prior spezifizieren zu können, ist es sinnvoll, möglichst genau zu beobachten, welche Auswirkungen ein bestimmter Prior auf die Parameterschätzung hat und wie sich das Verhalten der Parameterschätzung bei Austausch des Priors verändert.

3. Deswegen ist es auch notwendig, differenziertere Informationen über die Güte der Parameterschätzungen einzuholen als die, die in den entsprechenden Simulationsstudien von Rouder et al. (2003, 2005) berichtet wurden. Dort wird lediglich auf den *Root Mean Square Error* (RMSE) als Maß der Effizienz der Parameterschätzung eingegangen. In den folgenden Simulationsstudien sollen hingegen alle drei klassischen Gütekriterien der Parameterschätzung (Erwartungstreue, Effizienz und Konsistenz) beurteilt werden. Hinzu kommt, dass das Verfahren von Rouder et al. als ein hierarchisches Schätzverfahren explizit mit Informationen zweier Ebenen gleichzeitig operiert, nämlich der Individual- und der Stichprobenebene. Entsprechend soll die Effizienz des Verfahrens getrennt für die beiden Ebenen dargestellt werden. Denn eine Schätzmethode, die zwar für jedes Individuum einzeln ineffizient ist, kann sich auf der Ebene der gesamten Stichprobe durchaus als zufrieden stellend erweisen.

4.2 Simulationsstudie I

4.2.1 Methode

Populationsverteilung und Generierung der wahren Individualwerte. Die „wahre“ Populationsverteilung in Simulationsstudie I war absichtlich so gestaltet, dass sich eine möglichst große Ähnlichkeit zu den erwartbaren Reaktionszeitverteilungen der pronominalen Lösungsaufgabe ergibt. Informelle Voruntersuchungen hatten bereits ergeben, dass die Durchschnittszeit zur Bearbeitung dieser Aufgabe (mit kurzen Sätzen) ca. 3 bis 4 Sekunden beträgt. Darüber hinaus ist bekannt (vgl. z.B. Rouder, 2005), dass bei komplexeren Aufgaben (wie z.B. dem Lesen von Wörtern) der shift- Parameter ungefähr die Hälfte des Mittelwertes umfasst. Entsprechend wurden der wahre Populationswert sowohl des shift- als auch des scale-Parameters auf einen Wert von 2 Sekunden festgesetzt. Diese Wahl ist auch kohärent mit der Annahme, dass allein schon die Enkodierung eines Wortes (Wahrnehmung, lexikalischer Zugriff etc.) ungefähr 200 ms beansprucht (vgl. Balota, 1994). Bei einer Satzlänge von ca. 10 Wörtern ergibt sich ebenfalls ein ungefährender Betrag von 2 Sekunden allein für die basalen Wortverarbeitungsprozesse. Für den shape-Parameter wurde ein wahrer Wert von 2 angenommen, weil dies ein ziemlich typischer Wert für Reaktionszeitverteilungen ist, deren shape-Werte meist zwischen 1.5 und 2.5 liegen (vgl. Logan, 1992; Rouder et al., 2003).

Im Gegensatz zu diesen vergleichsweise nahe liegenden Entscheidungen war die Festlegung der Stichproben-Variabilität wesentlich schwieriger, weil hierzu kaum Befunde vorliegen. Allgemein ist jedoch bekannt, dass die interindividuelle Variabilität für Reaktionszeitdaten sehr hoch und ungefähr normal verteilt ist (vgl. z.B. Ratcliff, 1993). Das heißt, ähnlich wie bei der Intelligenz gibt es habituelle Unterschiede darin, ob eine Person eher langsam oder eher schnell antwortet. Abgesehen von solchen grundsätzlichen Niveauunterschieden weisen Personen dann allerdings meist sehr ähnliche, nicht normal-verteilte Antwortmuster auf (d.h. *interindividuell* sind Reaktionszeiten normalverteilt, *intraindividuell* aber gerade nicht). Entsprechend wurde für die vorliegende Simulationsstudie eine Normalverteilung der Individualwerte angenommen und mangels präziserer Informationen als „worst case“-Szenario eine vergleichsweise hohe Standardabweichung von 25% des Parameterausgangswertes vorgegeben. In (1) ist der Prozess der Generierung der wahren Individualwerte noch einmal zusammengefasst (wobei das Subskript i für die Person $i = 1 \dots i \dots I$ steht).

$$(1) \quad \begin{aligned} \text{shift}_i &\sim N(2,0.5) \\ \text{scale}_i &\sim N(2,0.5) \\ \text{shape}_i &\sim N(2,0.5) \end{aligned}$$

Variiert wurde als ein erster Faktor, wie viele (virtuelle) Individuen aus dieser Grundverteilung gezogen wurden (N Personen). Das heißt, der Stichprobenumfang auf Personenebene wurden systematisch manipuliert und umfasst entweder $I = 20, 40, 60, 80$ oder 100 Personen. Solche Stichprobenumfänge sind für die meisten kognitionspsychologischen Experimente typisch.

Generierung der Individualverteilungen. Nachdem so für jedes Individuum ein eigener wahrer Wert für den shift-, den scale- und den shape-Parameter bestimmt worden war, wurde in einem zweiten Schritt auf der Grundlage dieser Parameterkombinationen eine individuelle Reaktionszeitverteilung generiert. An dieser Stelle kam nun der zweite manipulierte Faktor der Studie zum tragen, da systematisch variiert wurde, auf wie vielen Datenpunkten diese Verteilung pro Individuum beruhte (N Items). Auch hier standen fünf Stufen zur Verfügung, d.h. die Reaktionszeit-Verteilung eines Individuums basierte entweder auf $J = 20, 40, 60, 80$ oder 100 Datenpunkten.

Als Verteilungsfunktion wurde die Weibull-Verteilung mit dem jeweils individualspezifischen Parameter-Set angesetzt, aus dem dann die entsprechende Anzahl an Items zufällig gezogen wurde. In (2) ist das Vorgehen der Generierung der einzelnen Datenpunkte zusammengefasst (wobei „ RT_{ij} “ für das Item $1 \dots j \dots J$ der Person $1 \dots i \dots I$ steht).

$$(2) \quad RT_{ij} \sim weibull(scale_i, shape_i) + shift_i$$

Es muss explizit darauf hingewiesen werden, dass damit die Grundvoraussetzungen für eine erfolgreiche Passung der Weibull-Verteilung unrealistisch gut sind: Nicht nur, dass die Weibull-Verteilung selbst zur Generierung der Datenpunkte verwendet wurde. Darüber hinaus wird auch noch angenommen, dass die Daten nicht irgendwie fehlerbehaftet oder veräuscht sind (z.B. durch verteilungsfremde Prozesse oder Übungseffekte) und die Parameterschätzung nicht durch Ausreißer verzerrt wird. Dies ist natürlich einerseits eine extreme Simplifizierung, die jedoch andererseits die Möglichkeit eröffnet, die Wirkung der manipulierten Faktoren möglichst „rein“ zu erfassen, also ohne interferierende Störeffekte.

Zusammengefasst heißt das, dass (durch eine Approximation des Antwortniveaus und der erwartbaren Effektzusammensetzung sowie durch die Berücksichtigung interindividueller Unterschiede) ein grundsätzlich realistisches Setting angenommen wird, aber eine untypisch gute Datenqualität. Diese Vereinfachung scheint aufgrund der bereits beschriebenen Unterschiede zwischen der aktuellen Datensituation und der sonstigen Reaktionszeitforschung vorläufig gerechtfertigt. Allerdings sollte man im Hinterkopf behalten, dass deswegen die folgenden Ergebnisse lediglich als eine maximal erreichbare Obergrenze interpretiert werden können, die für Planungszwecke evtl. nach unten korrigiert werden sollten.

Parameterschätzung. Das Vorgehen der Parameterschätzung lehnte sich so weit wie möglich an das originale Verfahren von Rouder et al. (2003) an. Es wurde der dort spezifizierte „vague prior“ verwendet (mit der Parametrisierung: $a_1 = 2.0$, $b_1 = 0.1$, $a_2 = 2.0$, $b_2 = 2.85$, $c_1 = 2.0$, $d_1 = 0.02$, $c_2 = 2.0$, $d_2 = 0.04$), der allerdings auf die unterschiedliche Skalierung des scale-Parameters angeglichen wurde (durch Setzung des Parameters $b_2 = 0.25$). Im Gegensatz dazu war der Prior des shape-Parameters bereits optimal auf den angesetzten Wertebereich (um 2) abgestimmt, weil dieser Wert auch von Rouder et al. als reaktionszeittypisch angesehen wird. Der Startwert des scale-Parameters wurde auf einen Wert von $\lambda = \theta^{-\beta} = 0.25$ adjustiert, damit es keine Unterschiede in der Konvergenz-Strecken der beiden Parameter gibt. Der Startwert des shape-Parameters blieb bei einem Wert von 2 (was dem wahren Populationswert entspricht), und der des shift-Parameters lag beim beobachteten Minimum der Verteilung minus 50 ms.

Stufenkombinationen, Stichprobenziehung und abhängige Variablen. Für jede der $5(N \text{ Personen}) \times 5(N \text{ Items})$ Zellkombinationen des Designs wurden 10 Stichproben realisiert, jeweils mit der entsprechenden Anzahl an Individuen und Items. Das heißt, auf der Ebene der Stichproben lagen nachher für jede Zelle 10 Werte vor, die jedoch (je nach Personen- und Item-Anzahl) auf jeweils unterschiedlich vielen Datenpunkten auf der Individual-Ebene beruhten. Mit lediglich 10 Beobachtungen pro Bedingungskombination hat die Simulationsstudie nicht das Auflösungsvermögen, um den „wahren“ asymptotischen Bias des Verfahrens zu ermitteln. Allerdings ist dies auch nicht das Ziel der Simulation, die lediglich überprüfen sollte, wie sich die Methode im angezielten Datenbereich verhält bzw. ob sie überhaupt auf diesen generalisierbar ist. Dafür ist eine Stichprobe von 10 Replikationen pro Zelle ausreichend.

Für die Individual-Ebene ist relevant, wie „gut“ das Verfahren aufgrund der empirisch gegebenen Verteilung die *individuellen* wahren Verteilungsparameter zu schätzen vermag. Zur Beurteilung dieser „Güte“, also der *Effizienz* des Verfahrens, soll der (über alle Stichproben gemittelte) Root Mean Square Error als ein klassisches Maß der Schätzgüte berichtet

werden ($RMSE(\theta) = \sqrt{\frac{1}{I} \sum_{i=1}^I (\hat{\theta}_i - \theta_i)^2}$). Hinzu kommt die (einfache) Abweichung der Schätzer zum wahren Wert als Indikator für die *Erwartungstreue* des Verfahrens

($BIAS(\theta) = \frac{1}{I} \sum_{i=1}^I (\hat{\theta}_i - \theta_i)$): Wenn es im Zuge der Aggregation zu systematischen Positiv-

oder Negativ-Abweichungen kommt, dann sind die Schätzer „biased“, d.h. das Verfahren führt zu einer systematischen Verzerrung der Parameter. Schließlich soll das Verfahren als *konsistent* bezeichnet werden, wenn Effizienz und Bias der Parameter mit zunehmendem

Stichprobenumfang (wie vom zentralen Grenzwert-Theorem verlangt) abnehmen. Für die Individual-Ebene ist hierbei v.a. die Anzahl der Items relevant, während die Anzahl der Personen einen eher geringen Einfluss haben sollte.

Auf der Stichproben-Ebene geht es hingegen nicht um die Übereinstimmung zwischen der individuellen Parameterschätzung und dem individuellem wahren Wert, sondern vielmehr darum, wie gut das Verfahren auf der Gruppenebene in der Lage ist, den relevanten *Populationsparameter* zu schätzen. Entsprechend werden nicht die jeweils individuellen wahren Wert als Kriterium herangezogen, sondern es geht um die Relation des Stichproben-Mittelwertes zum wahren Populationswert. Die *Erwartungstreue* wird dann bestimmt als die Differenz der gemittelten Individualparameter zum Populationswert ($BIAS(\theta) = \frac{1}{N} \sum_{n=1}^N \left(\frac{1}{I} \sum_{i=1}^I \hat{\theta}_i - \theta_{Pop} \right)$) und

die Effizienz ist durch die Standardabweichung dieser Mittelwerte gegeben

($RMSE(\theta) = \sqrt{\frac{1}{N} \sum_{n=1}^N \left(\frac{1}{I} \sum_{i=1}^I \hat{\theta}_i - \theta_{Pop} \right)^2}$). *Konsistent* ist das Verfahren wiederum, wenn Bias

und Effizienz mit steigendem Stichprobenumfang abnehmen, wobei allerdings für die Stichproben-Ebene v.a. die Anzahl der Personen relevant ist. Das heißt, unter der Voraussetzung, dass die Schätzungen auf der gleichen Anzahl von Items basieren, sollten größere (Personen)Stichproben zu effizienteren Ergebnissen führen. Abb. 4.1 stellt das gesamte Vorgehen noch einmal zusammenfassend dar.

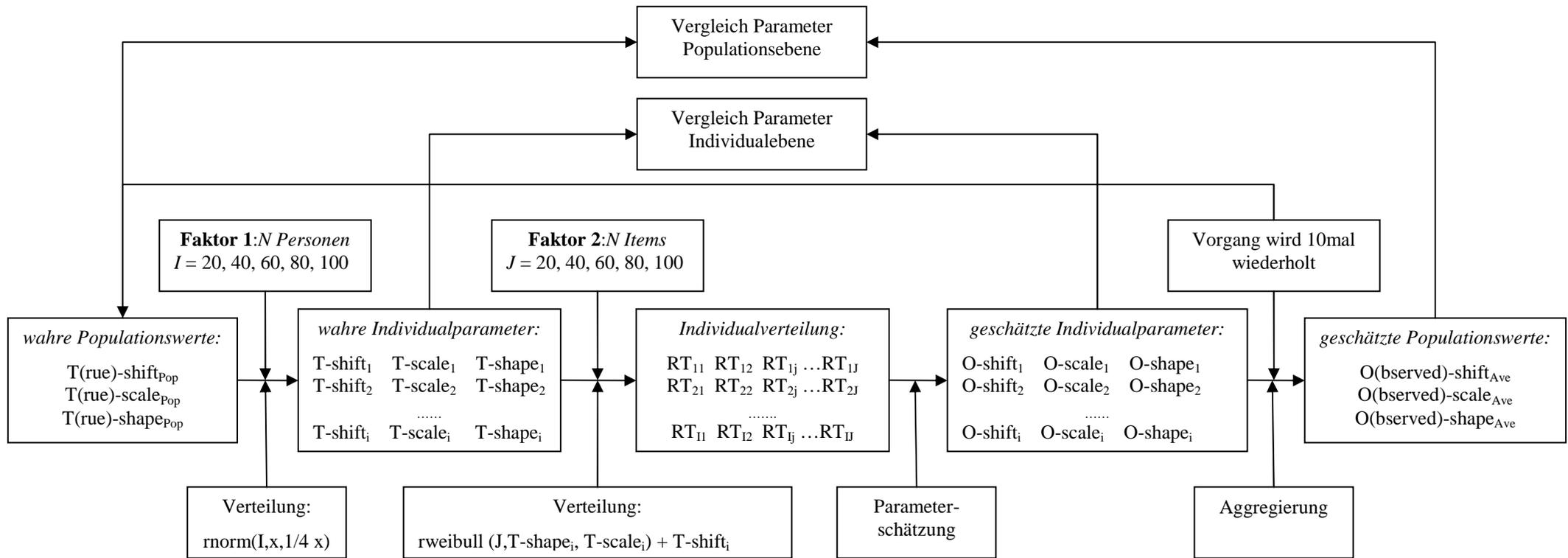


Abb. 4.1: Schematischer Ablauf der Simulationsstudie I.

4.2.2 Ergebnisse

4.2.2.1 Konvergenz

Wie in Appendix A.4 dargestellt, werden bei Bayes-Verfahren die einzelnen Modellparameter dadurch geschätzt, dass man ihre marginale Posterior-Verteilung bestimmt und einen adäquaten Kennwert dieser Verteilung (meist den Mittelwert) als Punktschätzer verwendet. Die Posterior-Verteilung selbst wird heutzutage meist durch *Monte Carlo Markov Chain*-Verfahren wie den Gibbs-Sampler approximiert, indem sukzessiv und iterativ die einzelnen Datenpunkte aus einer theoretischen Population gezogen werden. Da die Stichprobenziehungen zusammen eine Markov-Kette bilden, sind sie stochastisch voneinander abhängig. Das heißt, die aktuelle Stichprobenziehung wird partiell von den zuvor gezogenen Werten beeinflusst. Die Kette beginnt bei einem zuvor spezifizierten Startwert und „bewegt“ sich dann langsam in Richtung der zentralen Tendenz der optimalen Posterior-Verteilung. Nachdem der Konvergenzpunkt erreicht wurde, ist die Kette stabil und oszilliert entsprechend der Charakteristika der Posterior-Verteilung. Wenn Prior und Likelihood-Funktion bestimmte Voraussetzungen erfüllen (die im vorliegend Fall gegeben sind), dann ist mathematisch garantiert, dass die Kette irgendwann konvergieren wird (vgl. Gelman et al., 2004). Allerdings ist unklar, *wann* sie es tut. Es ist deswegen wichtig, in einem ersten Schritt zu untersuchen, ob die MCMC-Kette in der angesetzten Anzahl von Iterationen überhaupt konvergiert und wenn ja, wann. Denn ohne Konvergenz ist die Posterior-Verteilung kein adäquates Abbild der (gesuchten) Parameter-Verteilung, sondern sagt eher etwas über den Suchweg aus. Alle abgeleiteten Parameterschätzer sind dann entsprechend verzerrt. Zur Prüfung der MCMC-Konvergenz wurden sehr viele formale Verfahren und Teststatistiken entwickelt. Da der vorliegende Fall recht eindeutig ist, wird von einer formellen Prüfung (für die es auch keine klaren Konventionen gibt) abgesehen und lediglich eine graphische Inspektion der Konvergenzdynamik vorgenommen (vgl. hierzu z.B. Gill, 2002). Die Konvergenzdynamik der drei Parameter ist separat in den Abb. 4.2-4 dargestellt, wobei für den scale-Parameter β die von Rouder et al. empfohlene *rate*-Parametrisierung verwendet wurde ($\lambda = \theta^{-\beta}$). Die graphische Inspektion umfasst sowohl die Bestimmung des Konvergenzpunktes der MCMC-Kette selbst als auch eine Prüfung des Autokorrelationsmusters und der Posterior-Verteilung. Da sich die Muster der drei Parameter grundsätzlich ähneln, werden sie im Folgenden zusammen besprochen. Desgleichen soll der generelle Befund hier nur an einer einzigen Bedingungskombination demonstriert werden. Hierfür wurde eine möglichst „durchschnittliche“ Bedingungskombination mit 40 Personen und 40 Items gewählt.

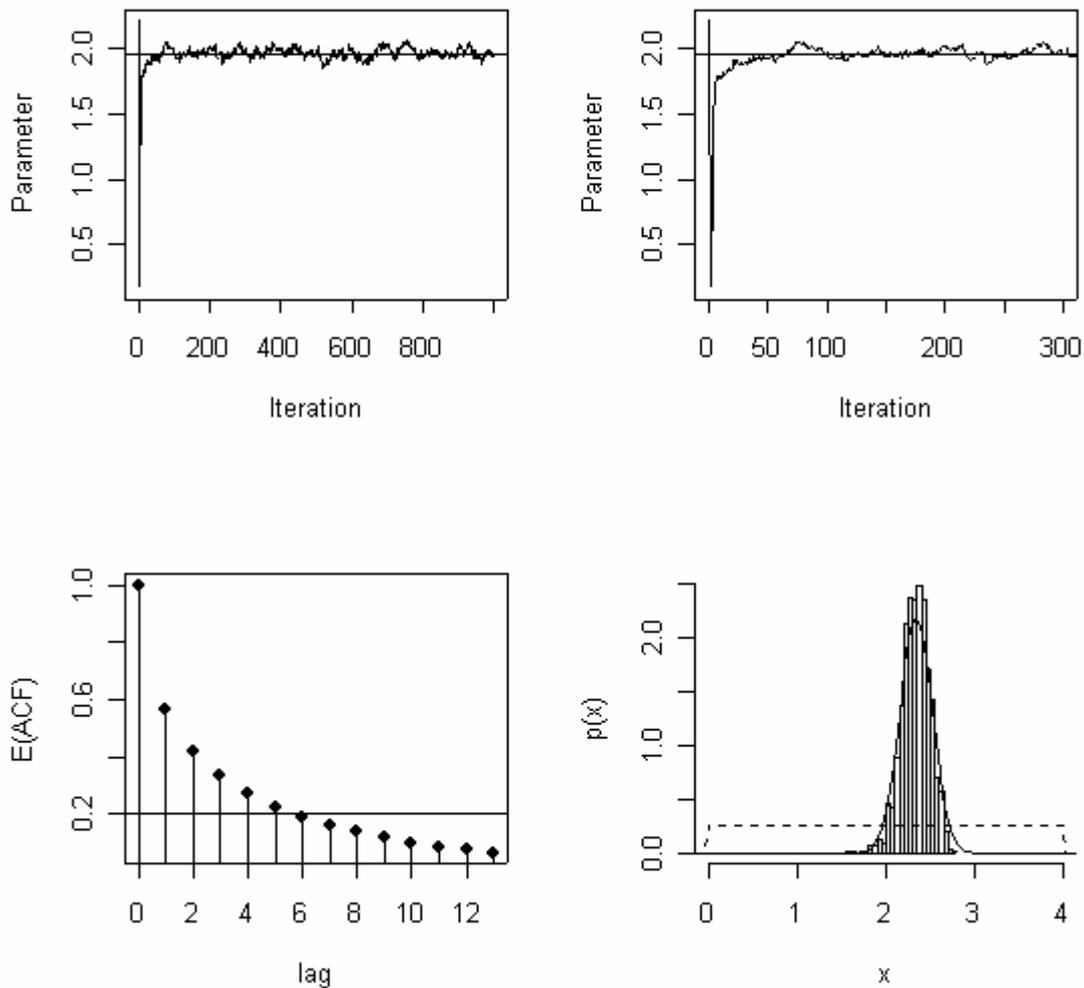


Abb. 4.2: MCMC-Konvergenz des shift-Parameters. a) Verlauf der MCMC-Kette für die gesamte Simulationsstrecke (oben links) bzw. b) nur für die ersten 300 Iterationen (oben rechts); c) Autokorrelationsmuster der MCMC-Kette (unten links), d) Posterior-Verteilung der ersten V_p (durchgezogene Linie) bzw. Priorverteilung (gestrichelt).

1. Teilabb. a) aller drei Abbildungen zeigt jeweils den typischen Verlauf der MCMC-Kette für alle 1000 Iterationen der Simulation, gemittelt über alle Personen (die individuellen Verläufe sehen sehr ähnlich aus, haben jedoch selbstverständlich einen wesentlich höheren Rauschanteil). Nach einer ca. 50-100 Iterationen umfassenden initialen *burn-in*-Phase (vgl. Teilabb. b)) oszillieren alle Parameterschätzungen stabil um den Erwartungswert der Posterior-Verteilung. 2. Die schnelle Konvergenz ergibt sich u.a. daraus, dass die Markov-Ketten ein nur vergleichsweise moderates Ausmaß an Autokorrelation aufweisen: Teilabb. c) zeigt das Autokorrelationsmuster der Kette (über alle Personen gemittelt). Während die Autokorrelation für die ersten beiden *lags* im mittleren Bereich liegt, nimmt sie für die nachfolgenden *lags* schnell (exponentiell) ab und fällt spätestens ab dem achten *lag* unter die .2-Schwelle. 3. Auch eine Inspektion der Posterior-Verteilungen bestätigt den Eindruck einer guten Konvergenz.

Pars pro toto ist in Teilabb. d) jeweils die Posterior-Verteilung der ersten (virtuellen) V_p dargestellt. Nicht-konvergierende oder instabile Ketten würden zu deutlich bimodalen oder gleichverteilten Posterior-Verteilungen führen (vgl. z.B. Gill, 2002). Im vorliegenden Fall sind sie jedoch stets unimodal und approximativ normal verteilt. (Zum Vergleich ist übrigens – mit der gestrichelten Linie – jeweils auch der Prior eines Parameters dargestellt. Schon an diesen Einzelbeispielen erkennt man, dass sich Posterior-Verteilungen ergeben können, die klar vom Prior divergieren).

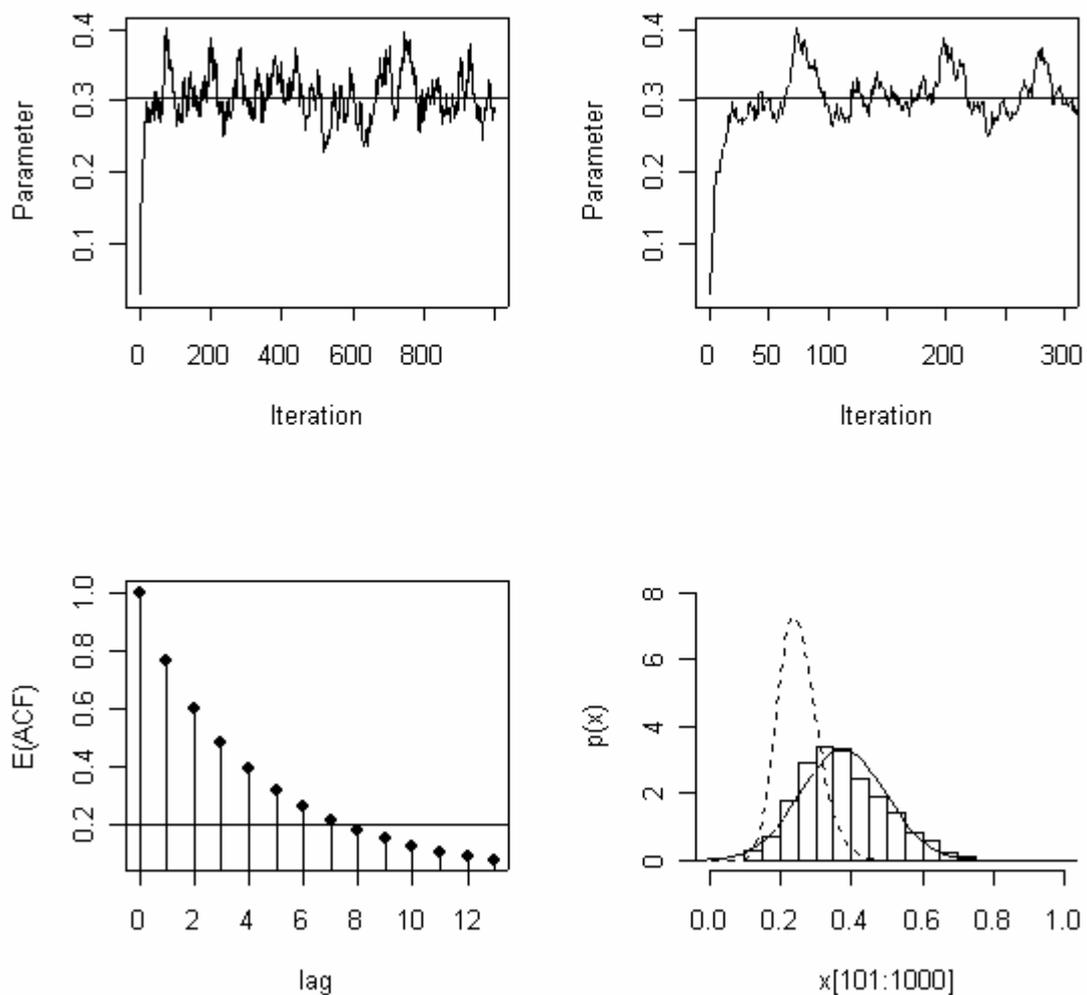


Abb. 4.3: MCMC-Konvergenz des scale-Parameters. a) Verlauf der MCMC-Kette für die gesamte Simulationsstrecke (oben links) bzw. b) nur für die ersten 300 Iterationen (oben rechts); c) Autokorrelationsmuster der MCMC-Kette (unten links), d) Posterior-Verteilung der ersten V_p (durchgezogene Linie) bzw. Priorverteilung (gestrichelt).

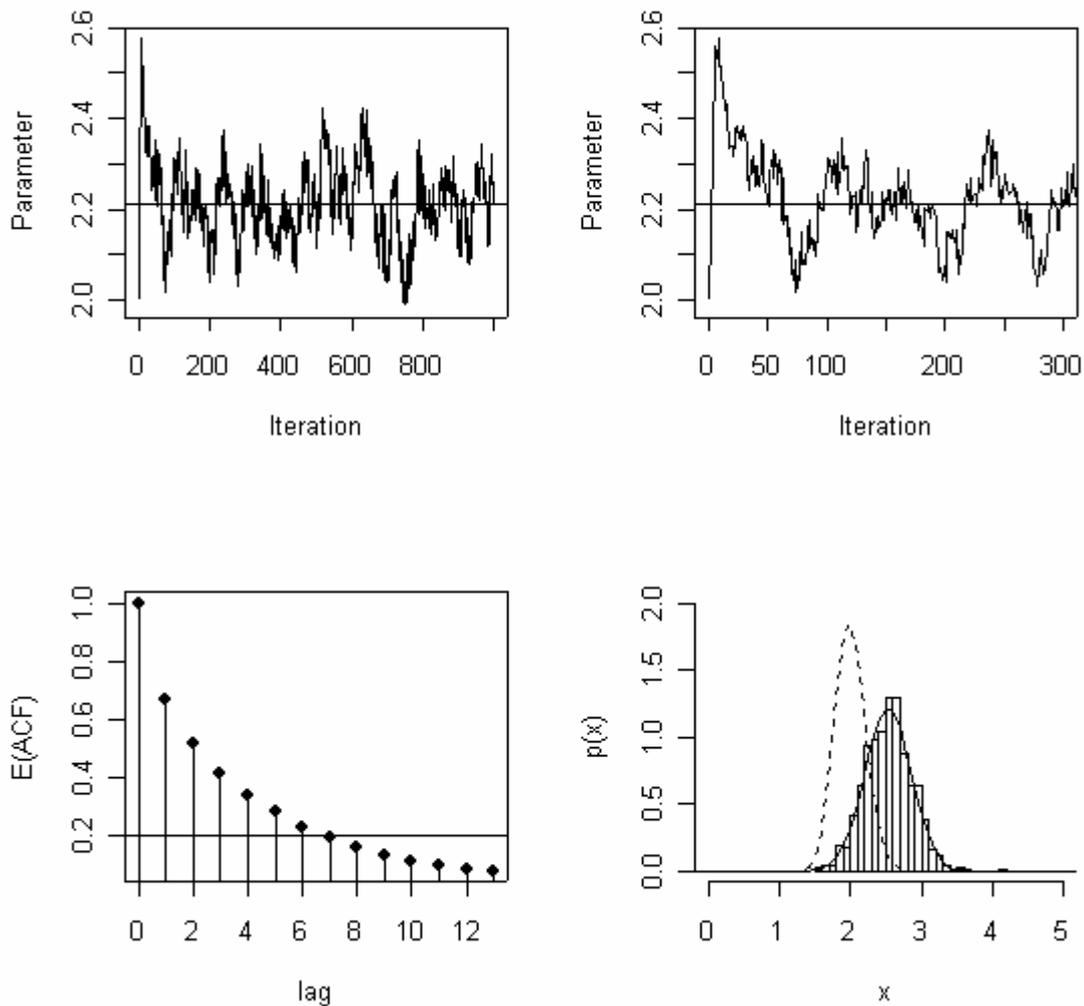


Abb. 4.4: MCMC-Konvergenz des shape-Parameters. a) Verlauf der MCMC-Kette für die gesamte Simulationsstrecke (oben links) bzw. b) nur für die ersten 300 Iterationen (oben rechts); c) Autokorrelationsmuster der MCMC-Kette (unten links), d) Posterior-Verteilung der ersten V_p (durchgezogene Linie) bzw. Priorverteilung (gestrichelt).

Zusammenfassend bedeutet das, dass alle Ketten erwartungsgemäß sehr schnell konvergieren und nur eine geringe Autokorrelation aufweisen. Dieses exzellente Konvergenz-Verhalten gilt prinzipiell für alle drei Parameter, obwohl der scale-Parameter tendenziell etwas schlechter konvergiert. Selbst in diesem Fall ist es jedoch so, dass eine Simulationsstrecke von 1000 Iterationen mit einer *burn-in*-Periode von 100 Iterationen mehr als hinreichend ist bzw. sogar als konservativ bezeichnet werden kann (Rouder & Lu, 2005).

4.2.2.2 Individualparameter

Nachdem die Frage geklärt ist, ob die MCMC-Ketten in der Tat auf einen stabilen Wert konvergieren, stellt sich direkt die nächste Frage, nämlich ob sie auch zu einer guten Schätzung führen. Um diese Frage zu beantworten, wurde zunächst verglichen, wie gut das Verfah-

ren in der Lage ist, die (wahren) Parameter auf der Individualebene zu approximieren. Hierfür soll zunächst ihr Bias evaluiert werden und dann ihre Effizienz. Abschließend wird geprüft, ob die Parameterschätzung konsistent ist.

4.2.2.2.1 Bias

Durchschnittlicher Bias. Abb. 4.5 zeigt den durchschnittlichen Bias der Parameterschätzungen auf der Individualebene, über alle Faktorstufen gemittelt. Er ergibt sich aus dem Abzug des wahren Parameters von dem des geschätzten, weshalb positive Wert für eine *Über-*, negative hingegen für eine *Unterschätzung* stehen.

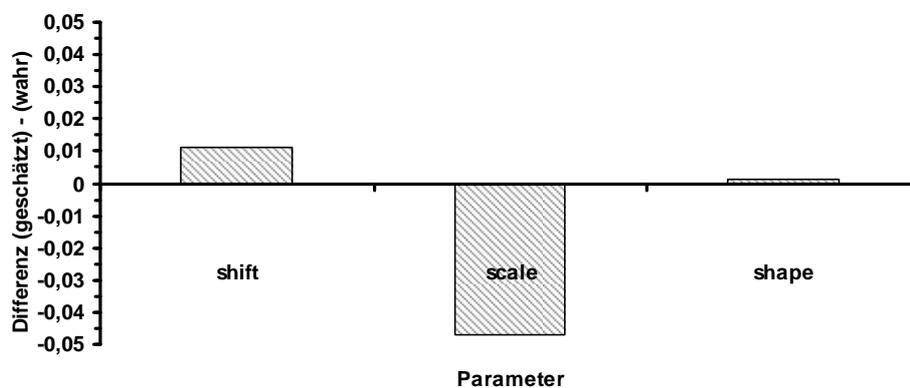


Abb. 4.5: Durchschnittlicher Bias der Parameter auf der Individualebene.

Es ist un schwer zu erkennen, dass es durchschnittlich zu einer leichten Überschätzung des shift-Parameters kommt (ca. 10 ms), während der scale-Parameter um ca. 50 ms unterschätzt wird. Der (relative) Bias für den shift-Parameter ist vergleichsweise klein (ca. 0.5 %), der des scale-Parameters hingegen mit 2.5 % zwar substantiell, aber auch nicht übermäßig groß. Der shape-Parameter scheint hingegen fast gar nicht verschätzt zu werden (bei einem relativen Bias von ca. 0.05 %).

Es ist schwierig zu beurteilen, ob dieses Bias-Muster ungewöhnlich für die Weibull-Verteilung ist. Denn wie bereits einleitend festgestellt, berichten Rouder et al. (2003, 2005) keine Bias-Werte und leider ist mir auch keine andere Simulationsstudie bekannt, die ein äquivalentes Bayes-Verfahren untersucht hätte (Smith & Naylor, 1987, besprechen lediglich Fallbeispiele und Sinha & Sloan, 1988, beziehen sich auf ein modifiziertes Bayes-Verfahren, das auf ML-Schätzern basiert). Die restlichen Simulationsstudien kontrastieren nur (modifizierte) Moment-Schätzer mit (modifizierten) ML-Schätzern. Aber selbst dann ergibt sich kein einheitliches und klares Ergebnismuster, weil die verschiedenen Verfahren jeweils unterschiedliche Verzerrungstendenzen haben, die darüber hinaus auch noch in Abhängigkeit von verschiedenen Randbedingungen (insbesondere der Stichprobengröße und der Schiefe der

Verteilung) unterschiedlich ausfallen. Trotz der Gefahr einer Übergeneralisierung lässt sich der Forschungsstand vielleicht so zusammenfassen: In der Literatur zur Weibull-Schätzung sind zwei Verzerrungsmuster besonders dominant. Das eine besteht darin, dass der shift-Parameter über-, und der scale- und shape-Parameter unterschätzt werden. Es tritt v.a. bei den (klassischen) ML-Schätzern auf und nimmt mit steigender Schiefe der Verteilung (bzw. abnehmenden shape-Parameter) zu. Das zweite prominente Verzerrungsmuster ist dem ersten genau entgegengesetzt, da hier der shift-Parameter unterschätzt wird, scale- und shape-Parameter hingegen überschätzt werden. Es wird insbesondere für verschiedene Moment-Schätzer berichtet, wobei der Bias mit steigender Schiefe der Verteilung meist abnimmt. Das modifizierte Bayes-Verfahren von Sinha und Sloan (1988) bewegt sich irgendwo in der Mitte zwischen diesen beiden Mustern, da es je nach Stichprobenumfang (und wahrscheinlich anderen, bisher unidentifizierten Faktoren) mal wie das erste und mal wie das zweite ausfällt.

Vor diesem Hintergrund lässt sich feststellen, dass das Bias-Muster der vorliegenden Simulation eher dem der ML-Verfahren entspricht. Auch die Stärke der Verzerrung ist ungefähr vergleichbar: Cousineau et al. (2004) berichten für eine Verteilung mit gleicher Schiefe (shape = 2, $N = 40$, ML) relative Bias-Werte von 1 % (shift), 3 % (scale) und 0.5 % (shape), Cohen et al. (1984) für ihre modifizierten Moment-Schätzer Werte von 0.3 %, 0.7 % und 2 %. Für alle drei Parameter zusammen ermittelten Cousineau et al. (2004) eine multivariaten Bias von 0.5 %, der damit ähnlich ausfällt wie z.B. für die Ex-Gauss-Verteilung (vgl. auch Heathcote et al., 2002, 2004).

N Personen. Abb. 4.6 zeigt den Effekt des Stichprobenumfangs der Personenebene auf den Bias der drei Weibull-Parameter. Man sieht deutlich, dass er keinen systematischen Einfluss auf den Bias der Parameter hat.

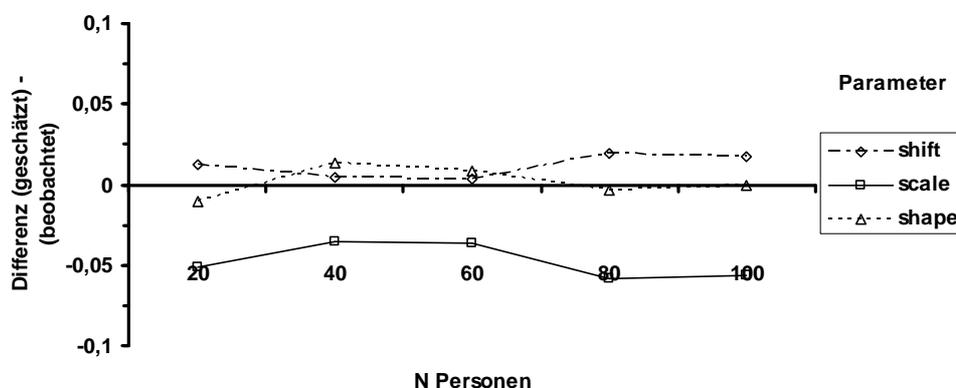


Abb. 4.6: Bias der Weibull-Parameter in Abhängigkeit von der Stichprobengröße auf Personenebene.

N Items. Abb. 4.7 zeigt den Effekt der Itemanzahl auf den Bias der drei Weibull-Parameter.

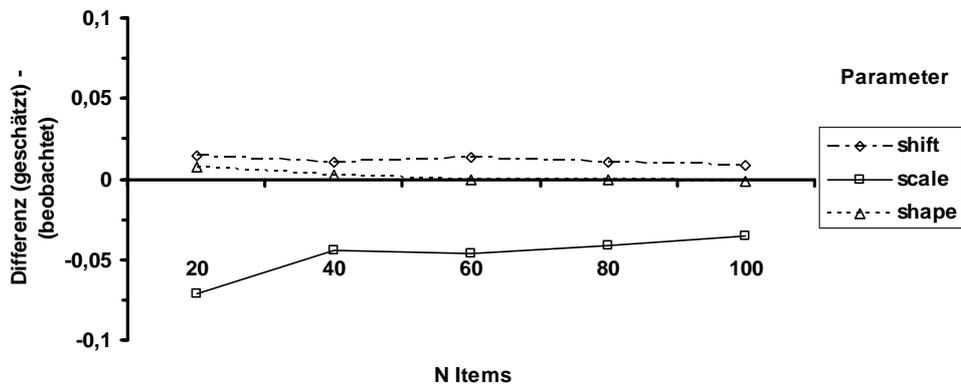


Abb. 4.7: Bias der Weibull-Parameter in Abhängigkeit von der Stichprobengröße auf Personenebene.

Für den Bias des scale-Parameters zeigt sich ein sehr deutlicher Effekt der zur Verfügung stehenden Itemanzahl. Diese Tendenz ist für den Übergang von 20 zu 40 Items am stärksten ausgeprägt (Bias-Reduktion um ca. 1 %). Eine weitere Erhöhung der Itemanzahl führt zu keiner wesentlichen Qualitätssteigerung (Δ 40-100 Items: ca. 0.5 % zwischen 40 und 100 Items). Auf den Bias des shift- und des shape-Parameters hat die Itemanzahl keinen großen Einfluss.

4.2.2.2.2 RMSE

Durchschnittlicher RMSE. Abb. 4.8 zeigt die über alle Faktorstufen gemittelt RMSE-Werte der drei Weibull-Parameter auf der Individualebene.

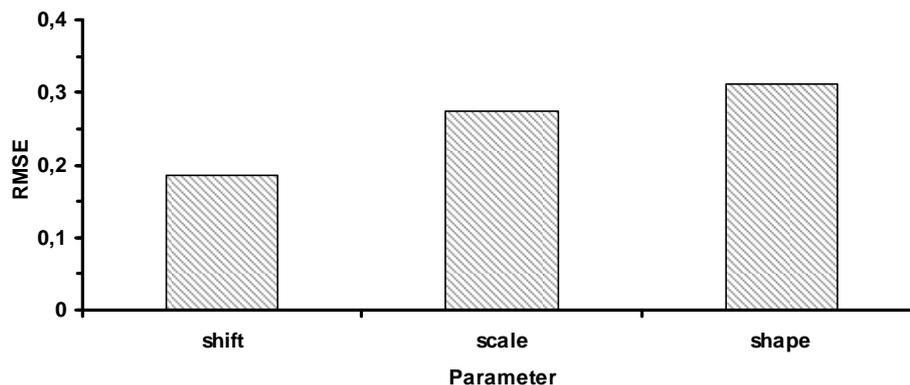


Abb. 4.8: Durchschnittlicher RMSE der Parameter auf der Individualebene.

Beim RMSE handelt es sich um ein Maß der Fehlervarianz, d.h. je höher ein RMSE-Wert ist, desto *ineffizienter* ist die Schätzung. Die Schätzung des shift-Parameters kann deshalb mit einem RMSE-Wert von 0.19 als vergleichsweise reliabel bezeichnet werden, wohingegen die Fehlervarianz für den scale- (0.28) und den shape-Parameter (0.31) schon etwas höher aus-

fällt. Relativ gesehen entspricht dies einem RMSE-Wert von ca. 10 % für den shift-, und jeweils ca. 15 % für den scale- und shape-Parameter.

Ist das jetzt viel oder wenig? Allgemein sagt ein bestimmter RMSE-Wert für sich alleine nur sehr wenig aus (vgl. Myung & Pitt, 2002). Vielmehr ergibt sich der eigentliche Wert dieser Statistik erst im Modellvergleich, also wenn verschiedene Schätzer miteinander kontrastiert werden. Allerdings ist es schwierig, die RMSE-Werte verschiedener Verfahren miteinander zu vergleichen, wenn sie in verschiedenen Studien ermittelt wurden. Meist unterscheiden sich diese auch noch bezüglich anderer Randbedingungen. Ein Vergleich mit der Studie von Rouder et al. (2003, 2005) ist im vorliegenden Fall sicherlich am informativsten. Da dort jedoch nur Stichproben von 20 und 80 Items untersucht wurden, ist es sinnvoll, sich noch ein wenig zurückzuhalten, bis eine Differenzierung der Ergebnisse nach dem Umfang der Itemstichprobe stattgefunden hat.

Schon jetzt lässt sich aber feststellen, dass die RMSE-Werte in einem ähnlichen Bereich liegen wie bei anderen Verfahren auch. Auch ist kaum zu übersehen, dass es einen substantiellen Effizienz-Vorteil für das Hierarchical-Bayes-Verfahren von ca. 5 % gibt. So berichtet z.B. Offinger (1996) für die ML-Schätzung und eine Verteilung mit gleicher Schiefe (shape = 2, $N = 50$) relative RMSE-Werte von 12 % (shift), 16 % (scale) und 22 % (shape). Für die Moment-Methode liegen die RMSE-Werte sogar deutlich höher, nämlich bei 20 %, 25 % und 31 % (vgl. jedoch Saylor, 1977, oder Zanakis, 1979, für effizientere Moment-Methoden).

N Personen. Abb. 4.9 zeigt den Effekt des Stichprobenumfangs der Personenebene auf den RMSE der drei Weibull-Parameter. Man sieht deutlich, dass die Personenanzahl keinen nennenswerten Einfluss auf die Effizienz der Parameterschätzung hat.

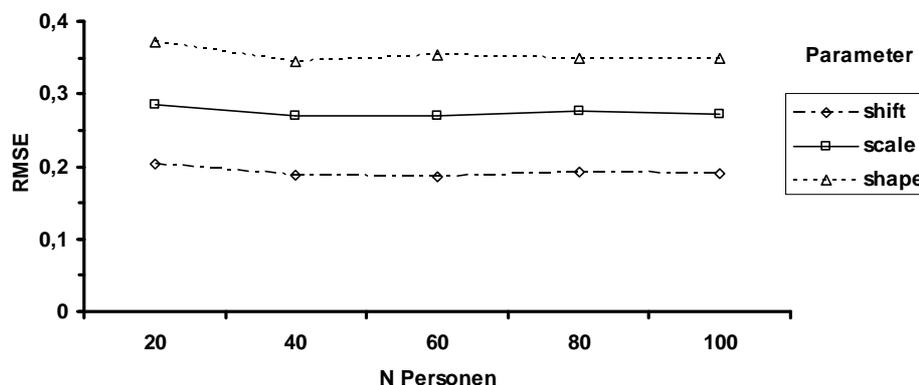


Abb. 4.9: RMSE der Weibull-Parameter in Abhängigkeit von der Stichprobengröße auf Personenebene.

N Items. Abb. 4.10 zeigt den Effekt der Itemanzahl auf den RMSE der drei Weibull-Parameter.

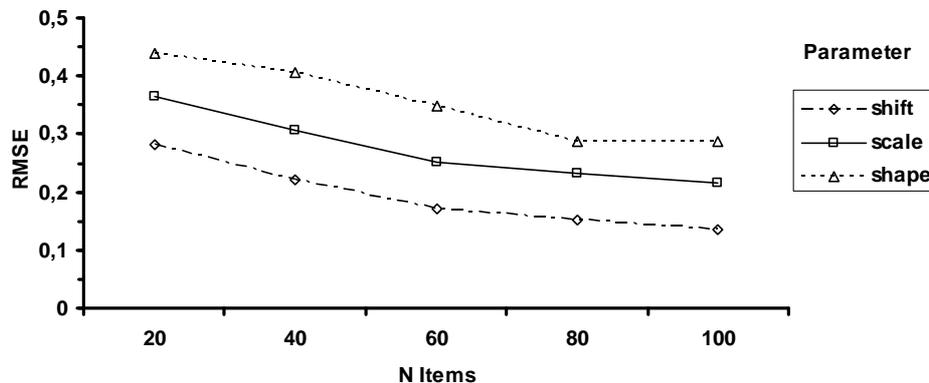


Abb. 4.10: RSME der Weibull-Parameter in Abhängigkeit von der Stichprobengröße auf Personenebene.

Es zeigt sich deutlich, dass eine Erhöhung der Itemanzahl zu abnehmenden RMSE-Werten führt. Dieser Reduktionseffekt (Δ 20-100 Items: ca. 8 %) ist für alle drei Parameter ungefähr gleich stark ausgeprägt.

Nun ist auch ein differenzierter Vergleich mit den Ergebnissen der Simulationsstudien von Rouder et al. (2003, 2005) möglich, bei denen ja lediglich Stichproben von 20 und von 80 Items miteinander verglichen wurden (die Personen-Stichprobengröße war konstant $N = 50$). So berichten z.B. Rouder et al. (2005) für den shape-Parameter, der als einziger in seiner Skalierung zwischen den beiden Untersuchungen vergleichbar ist, einen RMSE-Wert von 0.039 (80 Items) und 0.55 (20 Items). Da die entsprechenden Werte in der vorliegenden Untersuchung bei 0.29 (80 Items) bzw. 0.44 (20 Items) liegen, fällt die Effizienz ziemlich ähnlich aus. Für den Vergleich des shift- und des scale-Parameters ist zu beachten, dass für diese nur der relativierte RMSE aussagekräftig ist, weil die Skalierung der Parameter in den beiden Untersuchungen unterschiedlich war. Das implementierte Version des Verfahrens wies in den beiden relevanten Bedingungen (20 bzw. 80 Items über alle Personen-Stichprobenumfänge gemittelt) exakt den gleichen relativen RMSE für den shift-Parameter auf wie in der Studie von Rouder et al. (2005), nämlich ca. 8 % (80 Items) bzw. 14 % (20 Items). Für den scale-Parameter ergaben sich sogar etwas bessere Ergebnisse. Während Rouder et al. einen relativen RMSE von 14 % (80 Items) bzw. 22 % (20 Items) berichten, liegt er in der vorliegenden Studie bei 12 % und 18 %. Das heißt zusammengefasst, dass sich insgesamt sehr ähnliche, tendenziell sogar etwas bessere Ergebnisse zeigen. Das generelle Niveau und die relative Rangordnung der Effekte sind in beiden Studien ebenfalls vollkommen identisch. Außer Frage steht auch, dass die erzielte Effizienz generell wesentlich höher liegt als für die ML-Schätzung, insbesondere wenn lediglich 20 Items zur Verfügung stehen (vgl. Appendix A.5).

4.2.2.2.3 Zusammenfassung

Insgesamt können die Ergebnisse auf der Individualebene als sehr zufrieden stellend beurteilt werden. Das von Rouder et al. (2003) entwickelte Schätzverfahren scheint sehr robust zu sein und durch seine Übertragung auf langsamere RT-Daten nicht wesentlich an Effizienz zu verlieren: Der RMSE der einzelnen Untersuchungsbedingungen war ähnlich ausgeprägt wie bei den Simulationsstudien von Rouder et al. und besser als bei vergleichbaren Verfahren. Dies gilt insbesondere dann, wenn nur wenige Items zur Verfügung stehen. Eine weitere wichtige Erkenntnis ist, dass die Schätzung des shape- und des scale-Parameters etwas ineffizienter ausfällt als die des shift-Parameters. Im Gegensatz zur Studie von Rouder et al. konnten auch Information zum Bias der Parameter ermittelt werden. Hier zeigte sich, dass der shift-Parameter leicht über-, der scale-Parameter hingegen moderat unterschätzt wird. Die Tatsache, dass der shape-Parameter zwar keinen nennenswerten Bias, aber die höchste Fehlervarianz aufweist, spricht dafür, dass er vergleichsweise instabil ausfällt.

Erwartungsgemäß zeigte sich, dass die Effizienz des Verfahrens mit steigender Itemanzahl zunimmt. Für den Bias gilt dies tendenziell auch, v.a für den scale-Parameter. Das Verfahren kann deshalb als konsistent bezeichnet werden. Etwas erstaunlich ist zunächst, dass die Größe der Personen-Stichprobe keinen wesentlichen Einfluss auf die Parameterschätzung hat. Allerdings spricht dies durchaus für die Qualität der hierarchischen Komponente des Verfahrens, da offensichtlich gerade bei kleinen Stichprobengrößen die Parameterschätzung durch die Einbeziehung des Priors optimiert werden kann. Bei steigendem Stichprobenumfang nimmt der Einfluss des Priors hingegen ab (vgl. Rouder & Lu, 2005). Der Präzisionszugewinn des Verfahrens scheint also v.a auf der hierarchischen Prior-Verteilung zu basieren.

4.2.2.3 Populationsparameter

In einem nächsten Schritt soll verglichen werden, wie gut das Verfahren dazu in der Lage ist, die (wahren) Populationsparameter zu schätzen. Die Darstellung erfolgt wieder für Bias- und Effizienz-Informationen getrennt.

4.2.2.3.1 Bias

Durchschnittlicher Bias. Abb. 4.11 zeigt den durchschnittlichen Bias der Parameterschätzungen auf der Stichprobenebene, über alle Faktorstufen gemittelt. Er ergibt sich aus dem Abzug des wahren Parameters von dem des geschätzten, weshalb positive Wert für eine *Über-*, negative hingegen für eine *Unterschätzung* stehen.

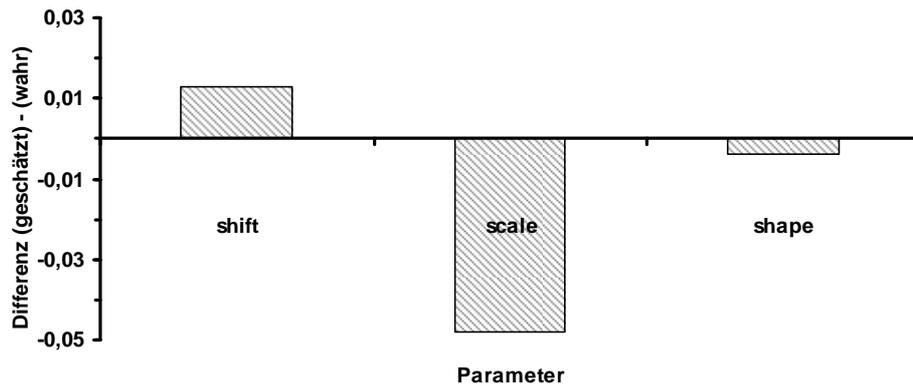


Abb. 4.11: Durchschnittlicher Bias der Parameter auf der Stichprobenebene.

Es ergibt sich ein nahezu identisches Muster wie für die Individualebene: Der shift-Parameter wird (sehr) leicht über-, der scale-Parameter hingegen substantiell unterschätzt (relativer Bias von 0.65 % bzw. 2.5 %). Für den shape-Parameter, der im Gegensatz zur Individualebene nun schwach negativ ausfällt, gibt es hingegen fast keinen Bias (0.01 %).

N Personen. Abb. 4.12 zeigt den Effekt des Stichprobenumfangs auf Stichprobenebene auf den Bias der drei Weibull-Parameter.

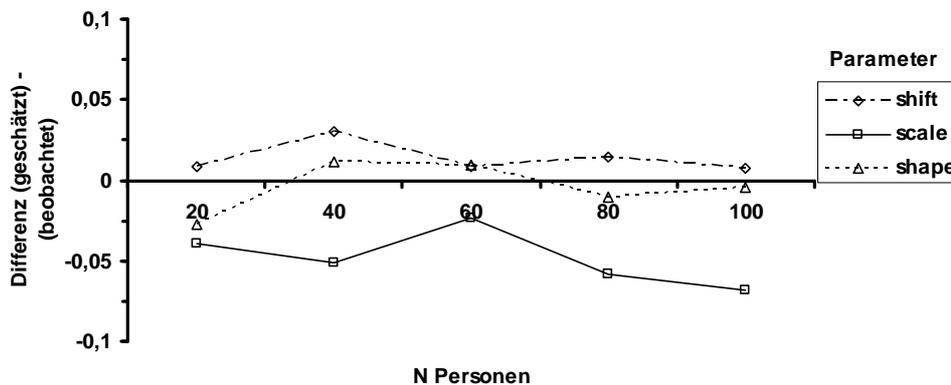


Abb. 4.12: Bias der Weibull-Parameter in Abhängigkeit von der Stichprobengröße auf Personenebene.

Entgegen der eingangs beschriebenen Erwartung hat die Größe der Stichprobe keinen nennenswerten und insbesondere keinen systematischen Einfluss auf den Bias der drei Parameter.

N Items. Abb. 4.13 zeigt den Effekt der Itemanzahl auf den Bias der drei Weibull-Parameter.

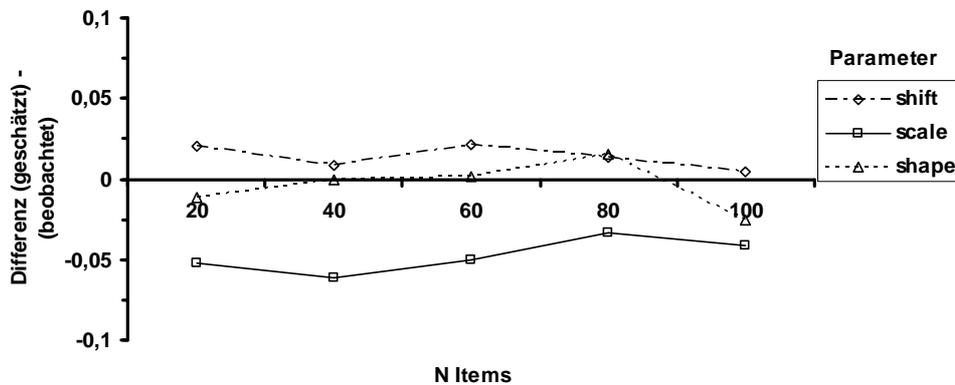


Abb. 4.13: Bias der Weibull-Parameter in Abhängigkeit von der Stichprobengröße auf Personenebene.

Auch die verwendete Itemanzahl hat – im Gegensatz zur Individualebene – keinen großen Einfluss auf den Parameter-Bias der Stichprobenebene. Lediglich für den scale-Parameter gibt es überhaupt einen Reduktionseffekt (um ca. 0.5 %), der jedoch vergleichsweise klein ist.

4.2.2.3.2 RMSE

Durchschnittlicher RMSE. Abb. 4.14 zeigt die durchschnittlichen RMSE-Werte der drei Weibull-Parameter auf der Stichprobenebene, über alle Faktorstufen gemittelt.

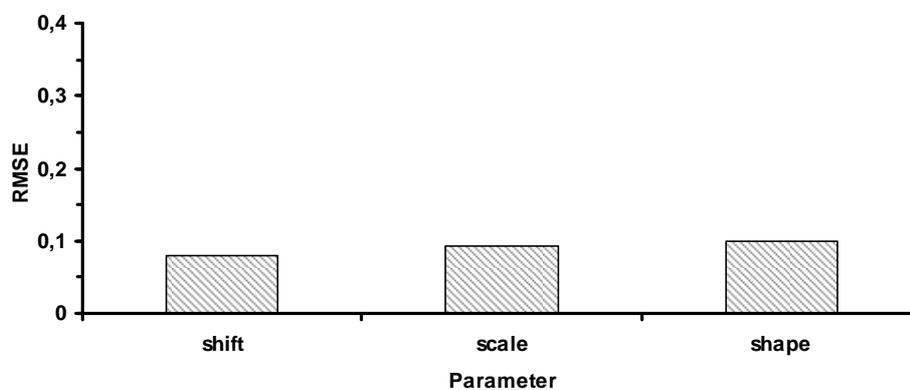


Abb. 4.14: Durchschnittlicher RMSE der Parameter auf der Stichprobenebene.

Wie ein Vergleich mit der korrespondierenden Abb. 4.8 der Individualebene zeigt, kommt es erwartungsgemäß durch die Aggregation der Individualparameter zu einer RMSE-Reduktion um ca. 50-75 %. Die Werte aller drei Parameter liegen nun einheitlich bei ca. 0.1 (relativer Bias von ca. 5 %) und die Asymmetrie der RMSE-Werte ist verschwunden.

N Personen. Abb. 4.15 zeigt den Effekt des Stichprobenumfangs auf den RMSE der drei Weibull-Parameter der Stichprobenebene.

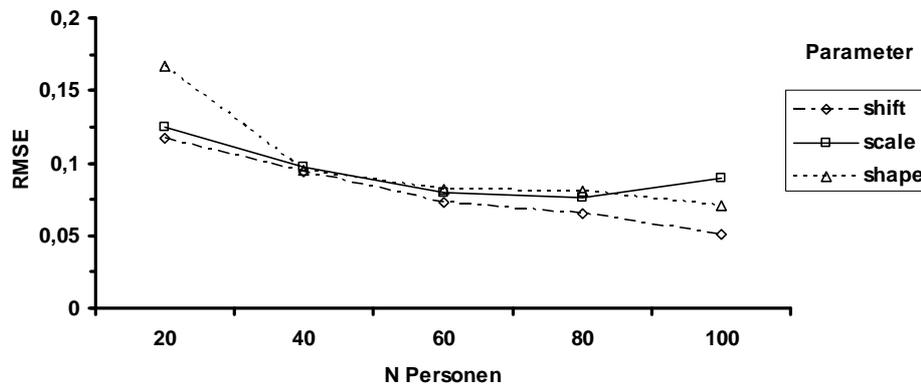


Abb. 4.15: RMSE der Weibull-Parameter in Abhängigkeit von der Stichprobengröße.

Im Gegensatz zum Bias gibt es einen deutlichen Einfluss der Stichprobengröße auf die Effizienz der Parameterschätzung. Dieser Reduktionseffekt fällt für den shift- (2.7 %) und den scale-Parameter (1.9 %) vergleichsweise moderat aus, ist aber für den shape-Parameter mit 4.7 % stark ausgeprägt.

N Items. Abb. 4.16 zeigt den Effekt der Anzahl der Items den RMSE der drei Weibull-Parameter.

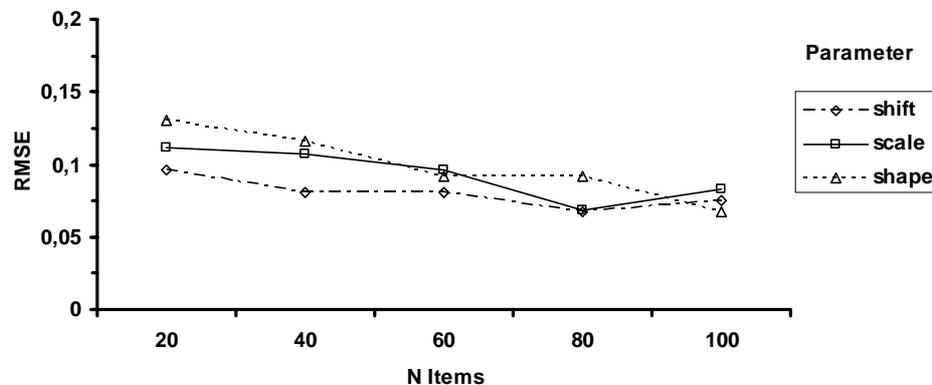


Abb. 4.16: RSME der Weibull-Parameter in Abhängigkeit von der Stichprobengröße auf Personenebene.

Auch die Anzahl der verwendeten Items hat einen Einfluss auf den RSME. Allerdings fallen die Effekte durchweg geringer aus als für die Anzahl der erhobenen Personen: Der Reduktionseffekt (Δ 100-20 Items) beträgt für den shift-Parameter ca. 1 %, der für den scale-Parameter 1.4 % und für den shape-Parameter 3.1 %.

An dieser Stelle ist wieder ein Vergleich mit der Studie von Rouder et al. (2003) möglich, da das dort untersuchte „Gruppenschätzverfahren“ (bei dem die gemittelten Einzelschätzungen zum Mittelwert der Population in Beziehung gesetzt wurden) ungefähr dem aktuellen Vorgehen entspricht. Bei der Stichprobe mit 80 Items ergibt sich für den shift-Parameter ein

relativer RMSE von 1.2 % für den shift-Parameter einer von 1.8 % und für den scale-Parameter einer von 3.8 %. In der vorliegenden Studie liegen die korrespondierenden Werte hingegen bei 3.4 %, 3.5 % und 4.6 %; d.h. sie fallen um ca. 2 % schlechter aus. Für die Stichprobe mit nur 20 Items ergibt sich ein ähnliches Bild, weil auch hier der relative Bias bei Rouder et al. (shift: 2.8 %, scale: 4.2 %, shape: 7.4 %) meist geringer ausfällt als in der vergleichbaren Bedingung der vorliegenden Studie (shift: 4.8 %, scale: 5.6 %, shape: 6.5 %). Nur der shape-Parameter wird etwas effizienter geschätzt.

4.2.2.3.3 Zusammenfassung

Die Analyse der Güte des verwendeten Schätzverfahren auf der Stichprobenebene ergänzt das bisherige Ergebnismuster um einen entscheidenden Aspekt: Bei den meisten kognitionspsychologischen Untersuchungen geht es – im Gegensatz z.B. zur Individualdiagnostik – eben nicht darum, den wahren Parameterwert eines einzelnen Individuums möglichst genau zu ermitteln. Vielmehr können die individuellen Werte ruhig ein wenig verschätzt werden, solange alle Individualschätzer zusammen eine hinreichend genaue Approximation des infrage stehenden Populationswertes gewährleisten. Vor diesem Hintergrund geben die beschriebenen Ergebnisse durchaus Anlass zu Optimismus: Die Effizienz der aggregierten Individualparameter zur Schätzung des wahren Populationsmittelwertes war insgesamt sehr hoch und für alle Parameter ungefähr gleich gut. Dieser Effizienzgewinn war zu erwarten und ergibt sich aus der schlichten Tatsache, dass sich durch die Kombination vieler Einzelinformationen (häufig) ein präziseres Gesamtbild ergibt als für die Summe der Einzelinformation. Im Vergleich zum Originalverfahren von Rouder et al. fällt die Parameterschätzung zwar etwas weniger effizient aus, was aber wahrscheinlich darauf zurückzuführen ist, dass in der vorliegenden Studie bereits für die Populationsverteilung eine überdurchschnittlich hohe Varianz angesetzt wurde.

Bezüglich der Konsistenz des Verfahrens ist zunächst festzustellen, dass es für den Bias keine starken Konsistenzeffekte gibt, sehr wohl aber für die Effizienz des Verfahrens: Erwartungsgemäß steigt die Effizienz mit zunehmender Personen-Stichprobengröße. Überraschend ist hingegen zunächst, dass auch die Erhöhung der Itemanzahl zu einer Reduktion der Fehlervarianz führt. Dies ist jedoch nur eine Konsequenz der Tatsache, dass die beiden Ebenen – Individuum und Stichprobe – nicht unabhängig voneinander sind: Wenn schon auf der Individualenebene Fehlervarianz eingespart werden kann, dann erhöht dies nachfolgend auch die Effizienz der aggregierten Parameterschätzung.

4.3 Simulationsstudie II

4.3.1 Methode

Insgesamt war das methodische Vorgehen vollkommen identisch zu dem in Simulationsstudie I. Die einzigen beiden Abweichungen betrafen zum einen die Ausprägung der (wahren) Populationsparameter und zum anderen die gewählten Priorwerte bei der Parameterschätzung.

Populationsparameter. Simulationsstudie II diente dazu, die Brauchbarkeit des von Rouder et al. entwickelten Verfahrens für wortweise Lesezeiten zu überprüfen. Diese weisen Eigenschaften auf, die für die Modellierung von RT-Verteilung evtl. problematisch sind. So fallen wortweise Lesezeiten kürzer aus und sind wesentlich varianzreicher als typische RT-Daten. Nach Just et al. (1982) liegt der Mittelwert wortweiser Lesezeiten bei ca. 440 ms, mit einer vergleichsweise hohen Standardabweichung von 260 ms. Allerdings muss man sich daran erinnern, dass die von Just et al. berichteten Werte durch Aggregation über *alle* Wörter eines Satzes entstanden sind und nicht durch Aggregation über funktional identische Wörter. Das bedeutet, dass die genannte Standardabweichung zu hoch ausfällt, weil sie zu einem Großteil aus Varianz besteht, die auf Unterschiede zwischen verschiedenen Wörtern zurückgeht (wie z.B. syntaktische und textstrukturelle Effekte). Unter der Prämisse, dass die shift- und die scale-Komponente wiederum ungefähr gleich groß sind (vgl. Rouder, 2005a), wurde bei der Simulation sowohl für den wahren Populations-shift- als auch für den wahren Populations-scale-Parameter ein Wert von 200 ms angesetzt. Als Wert für den shape-Parameter wurde wiederum ein Wert von 2 als plausibel angesehen. Um ein halbwegs realistisches Maß an Parametervariabilität zu simulieren, wurde für alle drei Parameter eine Standardabweichung von einem Viertel ihres Wertes angesetzt. Die zur Generierung der individuellen wahren Werte angesetzten Verteilungen sind zusammenfassend in (3) dargestellt:

$$(3) \quad \begin{aligned} \text{shift}_i &\sim N(0.2, 0.05) \\ \text{scale}_i &\sim N(0.2, 0.05) \\ \text{shape}_i &\sim N(0.2, 0.05) \end{aligned}$$

Parameterschätzung. Wieder wurden Parameter für die Hyperparameter des scale- und des shape-Parameter gewählt, die eine Approximation an die wahren Populationswerte darstellen ($a_1 = 2.0$, $b_1 = 0.1$, $a_2 = 2.0$, $b_2 = 4$, $c_1 = 2.0$, $d_1 = 0.02$, $c_2 = 2.0$, $d_2 = 0.04$). Als Startwerte für das MCMC-Verfahren wurden ein shape-Wert von 2 und ein shape-Wert von $\lambda = \theta^{-\beta} = 40$ angesetzt. Für den shift-Parameter war er das beobachtete Minimum minus 50 ms.

4.3.2 Ergebnisse

4.3.2.1 Konvergenz

Wieder sollen als erstes die Konvergenzeigenschaften der drei Parameter untersucht werden. Das Vorgehen war dabei das gleiche wie in Abschnitt 4.3.1.1. Abb. 4.17-19 zeigen die graphische Inspektion der MCMC-Ketten für die drei Parameter.

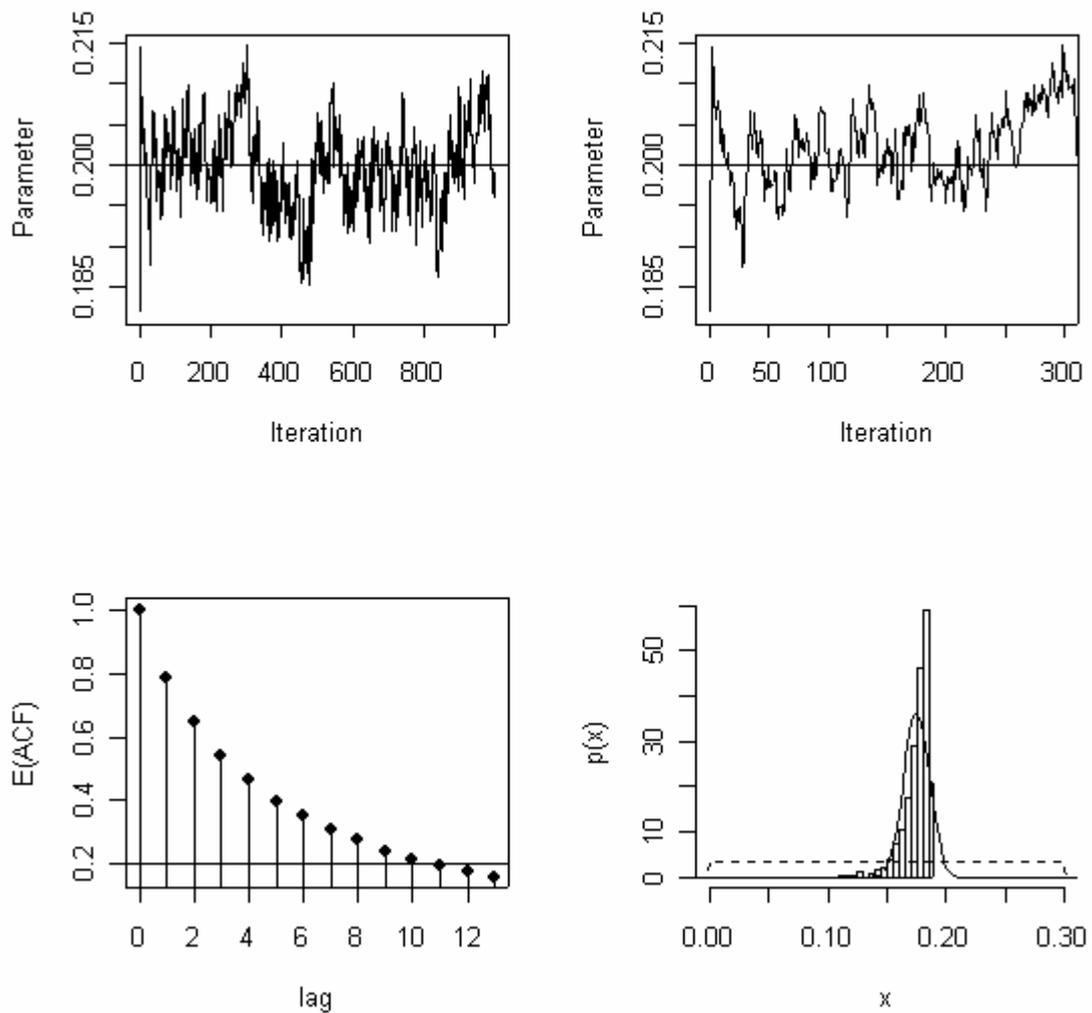


Abb. 4.17: MCMC-Konvergenz des shift-Parameters. a) Verlauf der MCMC-Kette für die gesamte Simulationsstrecke (oben links) bzw. b) nur für die ersten 300 Iterationen (oben rechts); c) Autokorrelationsmuster der MCMC-Kette (unten links), d) Posterior-Verteilung der ersten V_p (durchgezogene Linie) bzw. Priorverteilung (gestrichelt).

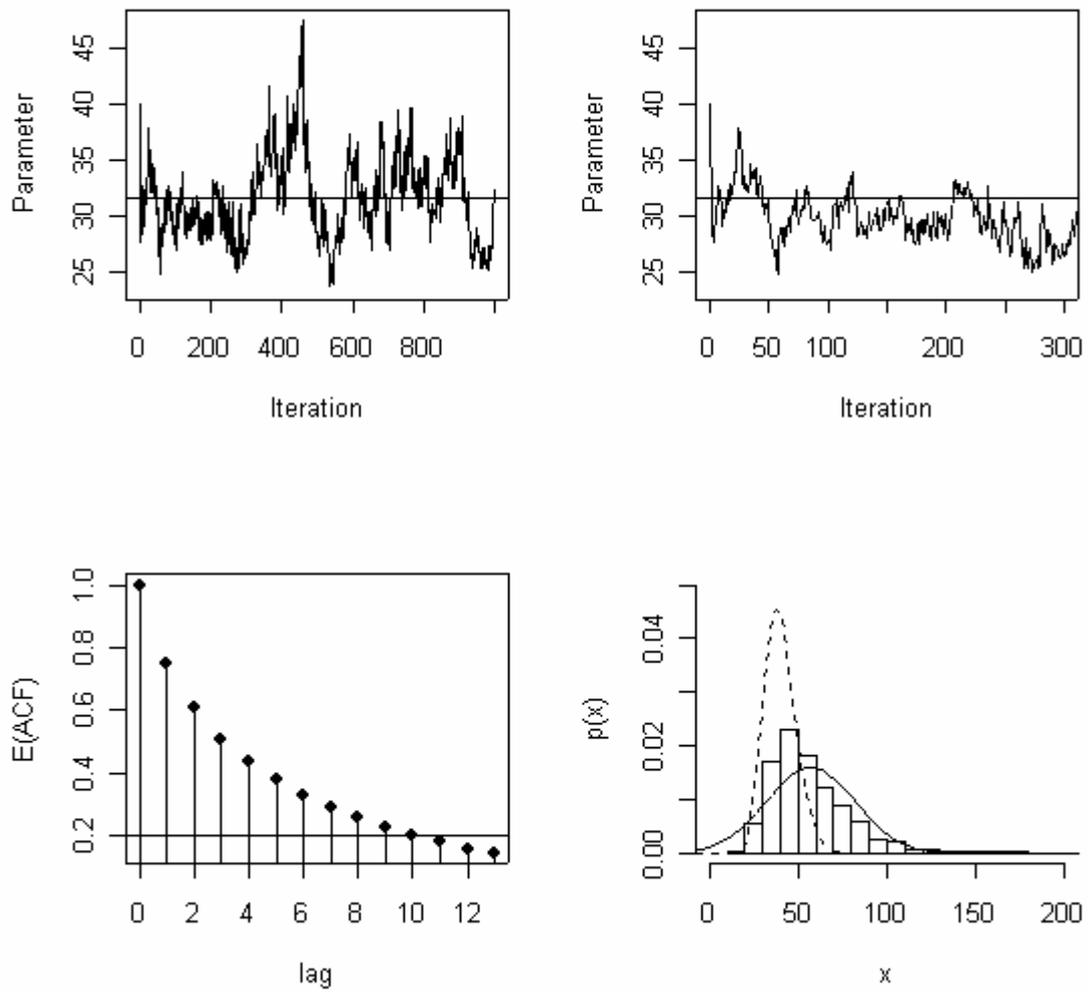


Abb. 4.18: MCMC-Konvergenz des scale-Parameters. a) Verlauf der MCMC-Kette für die gesamte Simulationsstrecke (oben links) bzw. b) nur für die ersten 300 Iterationen (oben rechts); c) Autokorrelationsmuster der MCMC-Kette (unten links), d) Posterior-Verteilung der ersten V_p (durchgezogene Linie) bzw. Priorverteilung (gestrichelt).

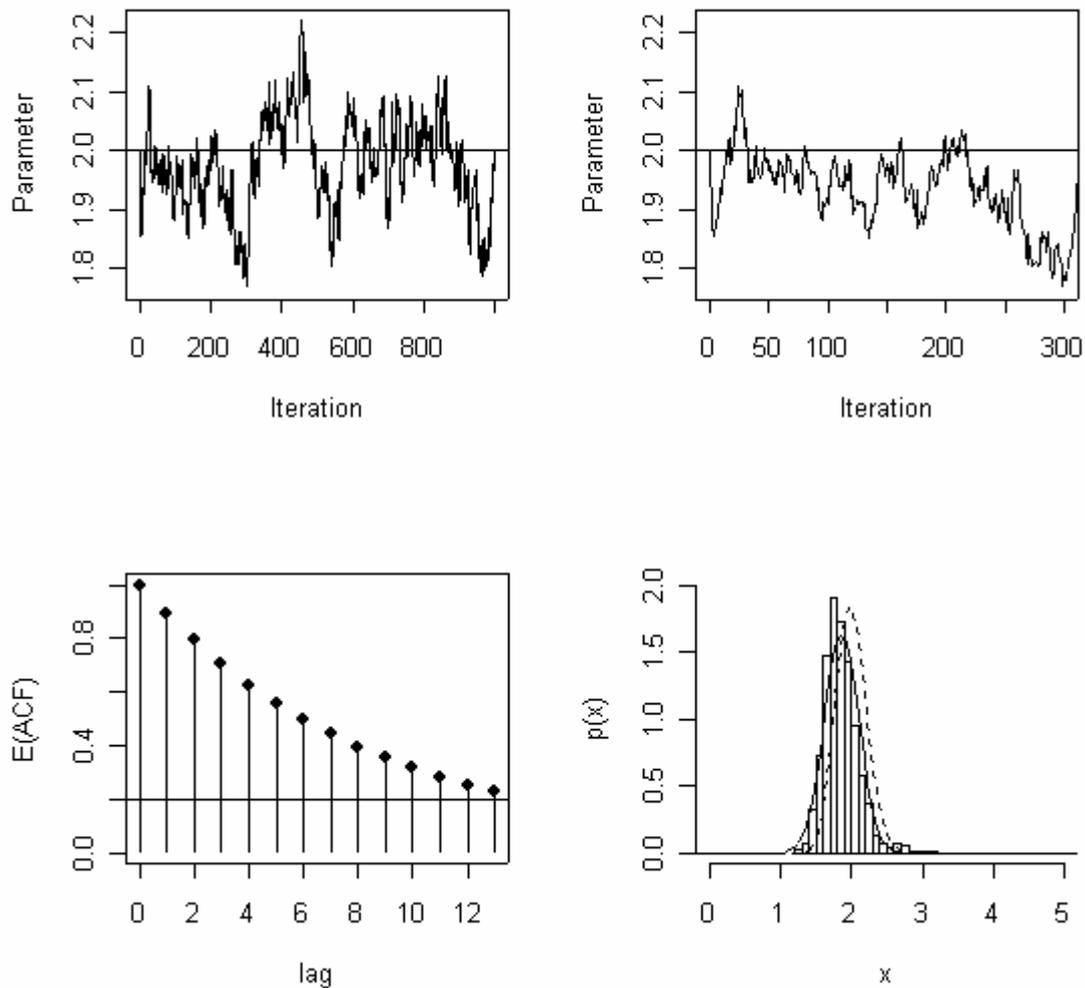


Abb. 4.19: MCMC-Konvergenz des shape-Parameters. a) Verlauf der MCMC-Kette für die gesamte Simulationsstrecke (oben links) bzw. b) nur für die ersten 300 Iterationen (oben rechts); c) Autokorrelationsmuster der MCMC-Kette (unten links), d) Posterior-Verteilung der ersten V_p (durchgezogene Linie) bzw. Priorverteilung (gestrichelt).

Insgesamt ergibt sich ein ähnliches Bild wie in Simulationsstudie I. Die Ketten konvergieren recht schnell und das Ausmaß an Autokorrelation ist überschaubar. Allerdings ist nicht zu übersehen, dass sich für alle drei Parameter etwas schlechtere Ergebnisse zeigen. Nicht nur, dass das generelle Autokorrelations-Niveau höher ist als für die langsamen RTn in Simulationsstudie I. Auch sind die Ketten weniger stabil und insgesamt varianzreicher. Diese Tendenz ist für die Posterior-Verteilung des scale-Parameters am stärksten ausgeprägt. Trotzdem scheint auch in diesem Fall eine Simulationsstrecke von 1000 Iterationen und eine *burn-in*-Periode von 100 Iterationen angemessen zu sein. *Burn-in*-Periode, Startwerte und Priorspezifizierung können deswegen in den nachfolgenden Untersuchungen beibehalten werden.

4.3.2.2 Individualparameter

In einem ersten Schritt sollen wieder die Gütekriterien der Parameterschätzungen auf der Individualebene beurteilt werden. Die Darstellung erfolgt zunächst für den Bias der Parameter. Nachfolgend wird auf ihre Effizienz eingegangen. Im Gegensatz zu Simulationsstudie I ist zu beachten, dass die Parameter unterschiedlich skaliert sind: Während der shift- und der scale-Parameter vergleichsweise kleine wahre Werte von 0.2 haben, liegt der wahre Wert des shape-Parameters bei 2 (also um den Faktor 10 größer). Für den Vergleich der Parameter ist deswegen nur ein Vergleich der relativen und nicht der absoluten Bias- und RMSE-Werte informativ. Diese werden nachfolgend in Prozent direkt angegeben.

4.3.2.2.1 Bias

Durchschnittlicher Bias. Abb. 4.20 zeigt den über alle Faktorstufen gemittelten relativen Bias der Parameterschätzungen auf der Individualebene. Er ergibt sich aus dem Abzug des wahren Parameters vom geschätzten, weshalb positive Wert für eine *Über-*, negative hingegen für eine *Unterschätzung* stehen.

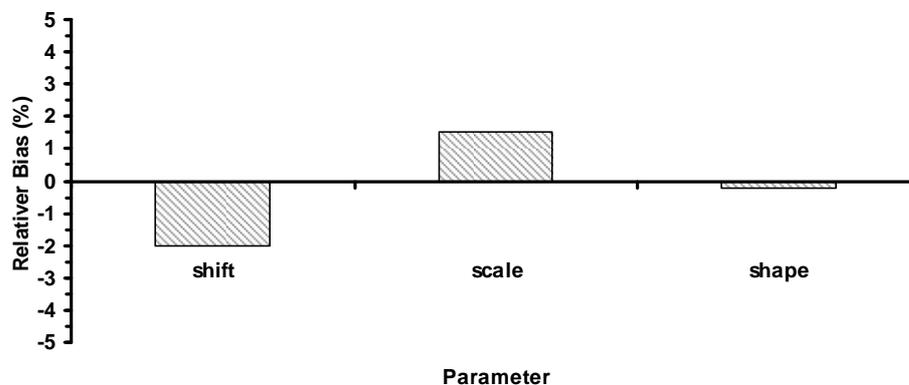


Abb. 4.20: Durchschnittlicher Bias der Parameter auf der Individualebene.

Interessanterweise ergibt sich ein etwas anderes Biasmuster als bei Simulationsstudie I. Während dort der scale-Parameter unterschätzt wurde, wird er nun leicht überschätzt. Gleichzeitig gibt es nun eine substantielle Unterschätzung des shift-Parameters. Relativ gesehen ist die Größe dieser Verschätzung mit Werten von 2 % (shift) bzw. 1.5 % (scale) ungefähr gleich stark ausgeprägt wie in Simulationsstudie I. Der shape-Parameter weist wiederum keine wesentliche Verzerrung auf (0.2 %).

Während also die Stärke des Bias beim Übergang von Simulationsstudie I zu II konstant geblieben ist, hat sich das Verzerrungsmuster umgedreht, welches nun dem typischen der Moment-Verfahren ähnelt. Dass das Bias-Muster sich bei Bayes-Schätzungen mitunter ändert,

wurde in der Literatur bereits mehrfach beschrieben (vgl. z.B. Sinha & Sloan, 1988). Allerdings kann nicht vollkommen ausgeschlossen werden kann, dass das veränderte Bias-Muster durch die Prior-Spezifizierung bedingt ist.

N Personen. Abb. 4.21 zeigt den Effekt des Stichprobenumfangs der Personenebene auf den Bias der drei Weibull-Parameter. Sie zeigt, dass der Bias der Parameter kaum von der Größe der Personen-Stichprobe beeinflusst wird.

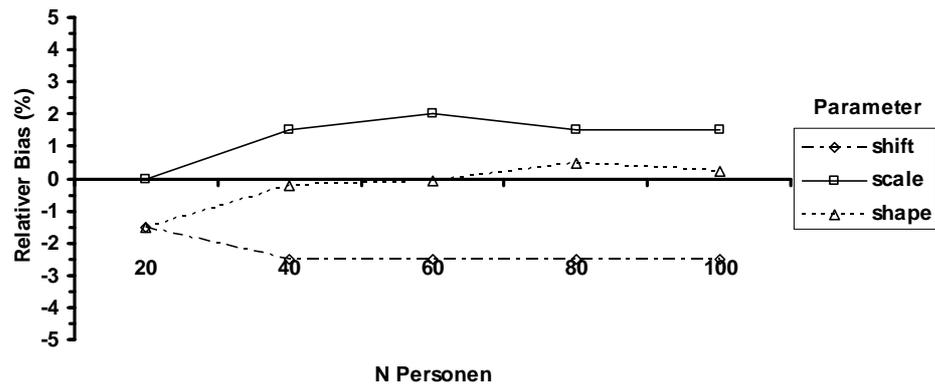


Abb. 4.21: Bias der Weibull-Parameter in Abhängigkeit von der Stichprobengröße auf Personenebene.

N Items. Abb. 4.22 zeigt den Effekt der Anzahl der Items auf den Bias der drei Weibull-Parameter.

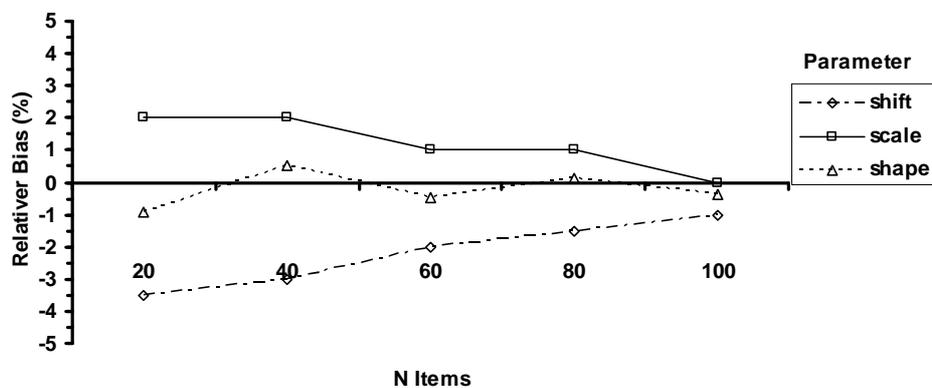


Abb. 4.22: Bias der Weibull-Parameter in Abhängigkeit von der Stichprobengröße auf Personenebene.

Sowohl der Bias des shift- als auch der des scale-Parameters nehmen mit steigender Itemanzahl ab (für beide gilt Δ 20-100 Items: ca. 2.5 %). Der Reduktionseffekt fällt damit sogar etwas stärker aus als in Simulationsstudie I.

4.3.2.2.2 RMSE

Durchschnittlicher RMSE. Abb. 4.23 zeigt die durchschnittlichen RMSE-Werte der drei Weibull-Parameter der Individualebene, über alle Faktorstufen gemittelt.

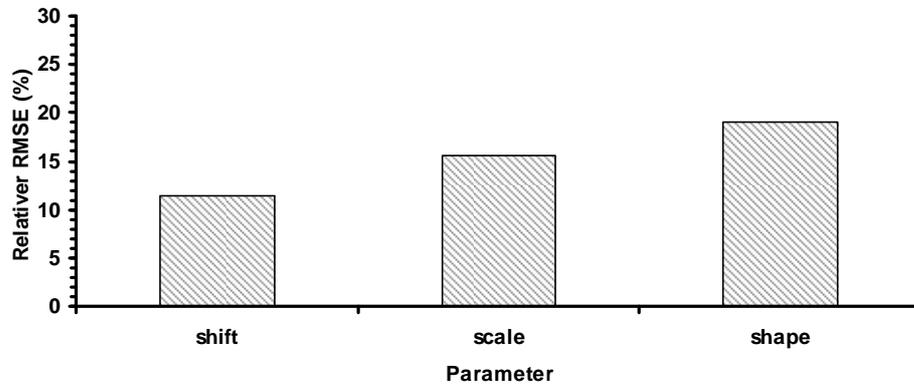


Abb. 4.23: Durchschnittlicher RMSE der Parameter auf der Individualebene.

Der relative RMSE der Parameter (shift: 12 %, scale: 16 %, shape: 19 %) ist vergleichsweise homogen und ähnlich wie in Simulationsstudie I ausgeprägt.

N Personen. Abb. 4.24 zeigt den Effekt des Personen-Stichprobenumfangs auf den RMSE der drei Weibull-Parameter.

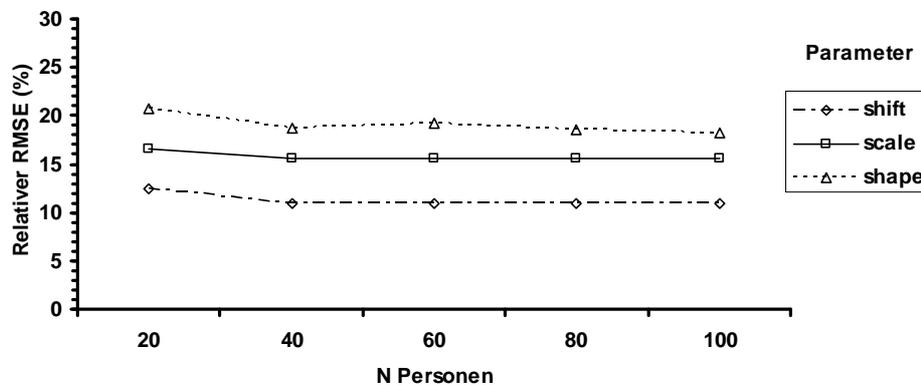


Abb. 4.24: RMSE der Weibull-Parameter in Abhängigkeit von der Stichprobengröße auf Personenebene.

Die Größe der Stichprobe hat keinen nennenswerten Einfluss auf den RMSE der Parameter. Nur der relative RMSE des shape-Parameters sinkt mit steigender Itemanzahl leicht (um ca. 3 %).

N Items. Abb. 4.25 zeigt den Effekt der Itemanzahl auf den RMSE der drei Weibull-Parameter.

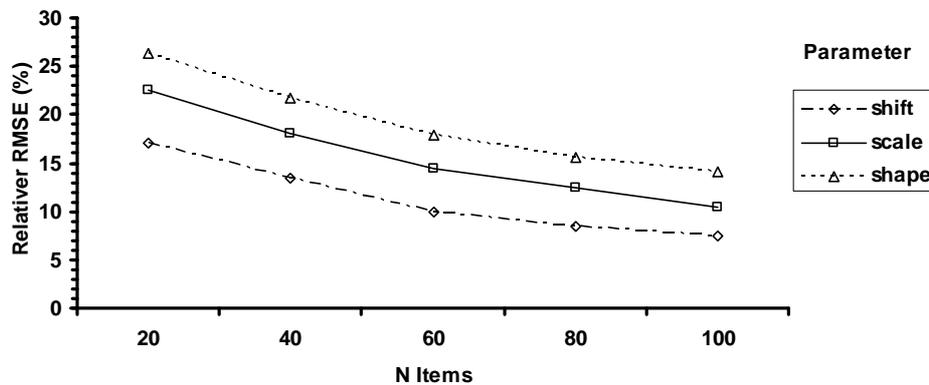


Abb. 4.25: RSME der Weibull-Parameter in Abhängigkeit von der Stichprobengröße auf Personenebene.

Für die Effizienz der Parameterschätzung zeigt sich, dass eine Vergrößerung der Itemanzahl zu niedrigeren RMSE-Werten führt. Dieser Reduktionseffekt (Δ 20-100 Items) beträgt für den shift-Parameter ca. 9 % und für den scale- und shape-Parameter jeweils ca. 12 %. Er ist damit stärker ausgeprägt als in Simulationsstudie I.

4.3.2.2.3 Zusammenfassung

Sowohl die Größe des Bias als auch der RMSE der Parameterschätzer ist sehr ähnlich ausgeprägt wie in Simulationsstudie I. Allerdings hat sich das Bias-Muster verändert: Der scale-Parameter wird über- und der shift-Parameter unterschätzt. Parallel hierzu ist die Effizienz des Verfahrens ein wenig gesunken. Insgesamt ist es jedoch immer noch ähnlich effizient wie die von Rouder et al. (2003) beschriebene Variante. Da sowohl der Bias als auch die Fehlervarianz mit zunehmender Itemzahl abnimmt, kann die Methode durchaus als konsistent bezeichnet werden.

4.3.2.3 Populationsparameter

Nachfolgend geht es um die Eignung des Verfahrens, den wahren Populationswert möglichst originalgetreu zu schätzen. Zunächst wird der Bias der Parameter sowie ihre Effizienz besprochen, nachfolgend soll die Konsistenz des Verfahrens beurteilt werden.

4.3.2.3.1 Bias

Durchschnittlicher Bias. Abb. 4.26 zeigt den relativen Bias der Parameterschätzungen auf der Stichprobenebene, über alle Faktorstufen gemittelt. Er ergibt sich aus dem Abzug des wahren Parameters von dem des geschätzten, weshalb positive Wert für eine *Über-*, negative hingegen für eine *Unterschätzung* stehen.

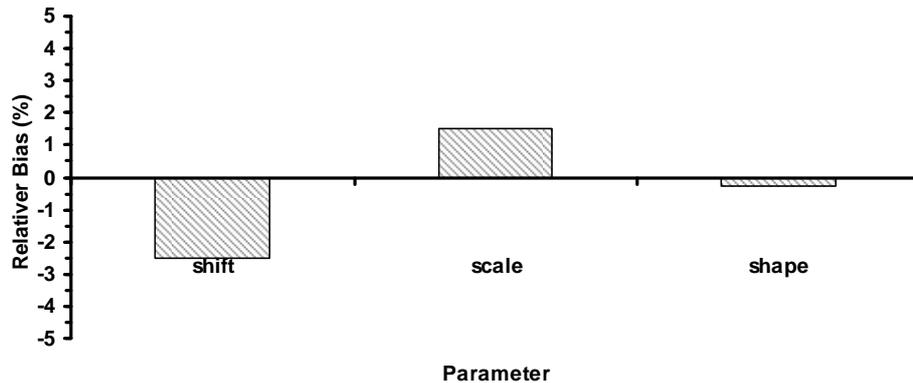


Abb. 4.26: Durchschnittlicher Bias der Parameter auf der Stichprobenebene.

Das Ergebnismuster ist wiederum identisch mit dem der Individualebene: Der shift-Parameter wird unter- und der scale-Parameter überschätzt. Auch die Größe des Bias ist mit 2.5 % für den shift-Parameter, 1.5 % für den scale-Parameter und 0.3 % für den shape-Parameter ungefähr vergleichbar.

N Personen. Abb. 4.27 zeigt den Effekt der Personenanzahl auf die Parameterschätzung auf der Stichprobenebene auf den Bias der drei Weibull-Parameter. Abgesehen vom shape-Parameter, der wieder etwas variabel ist, hat der Personen-Stichprobenumfang keinen nennenswerten Einfluss auf den Bias.

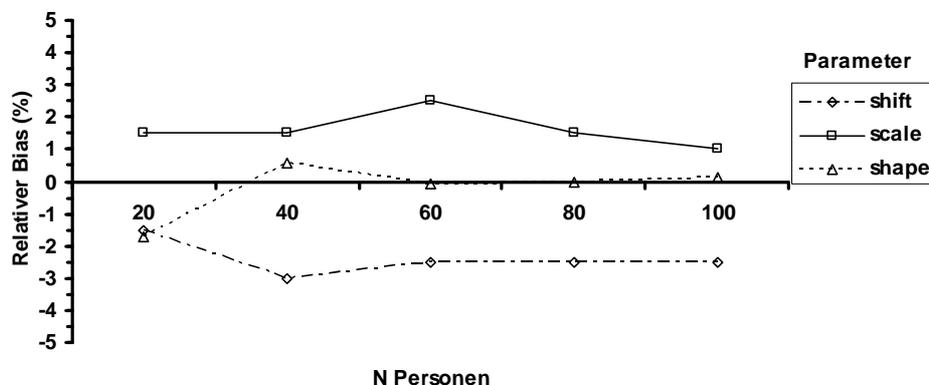


Abb. 4.27: Bias der Weibull-Parameter in Abhängigkeit von der Personen-Stichprobengröße.

N Items. Abb. 4.28 zeigt den Effekt der Itemanzahl auf den Bias der drei Weibull-Parameter.

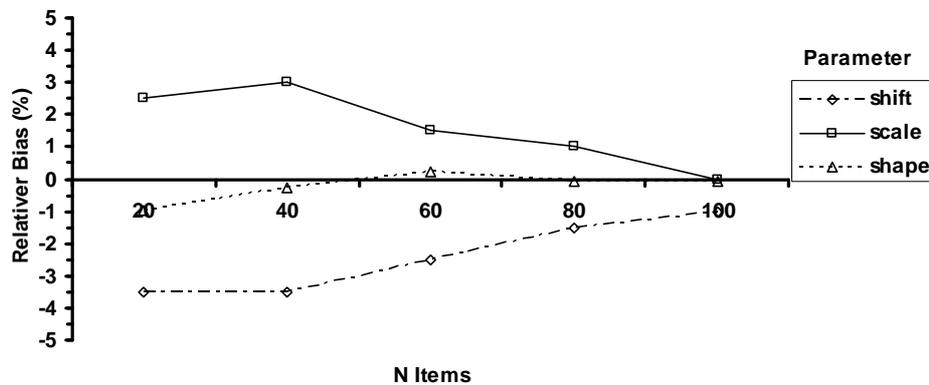


Abb. 4.28: Bias der Weibull-Parameter in Abhängigkeit von der Stichprobengröße auf Personenebene.

Ähnlich wie auf der Individualebene führt die Zahl der zur Verfügung stehenden Items zu einer starken Bias-Reduktion von ca. 2.5 %. Bei einer Itemanzahl von $N = 100$ gibt es keine relevante Verzerrung der Parameter mehr (< 1 %).

4.3.2.3.2 RMSE

Durchschnittlicher RMSE. Abb. 4.29 zeigt die durchschnittlichen RMSE-Werte der drei Weibull-Parameter auf der Stichprobenebene, über alle Faktorstufen gemittelt.

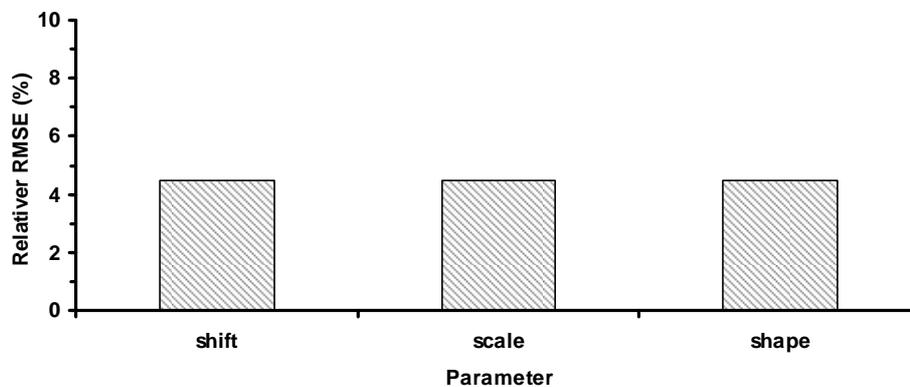


Abb. 4.29: Durchschnittlicher RMSE der Parameter auf der Stichprobenebene.

Der relative RMSE liegt für alle drei Parameter bei ca. 5 %. Das heißt, durch den Übergang zur Stichprobenebene erhöht sich die Effizienz der Schätzung um 5 bzw. 10 %. Auch dieser Befund stimmt mit Simulationsstudie I überein.

N Personen. Abb. 4.30 zeigt den Effekt des Stichprobenumfangs auf den RMSE der drei Weibull-Parameter der Stichprobenebene.

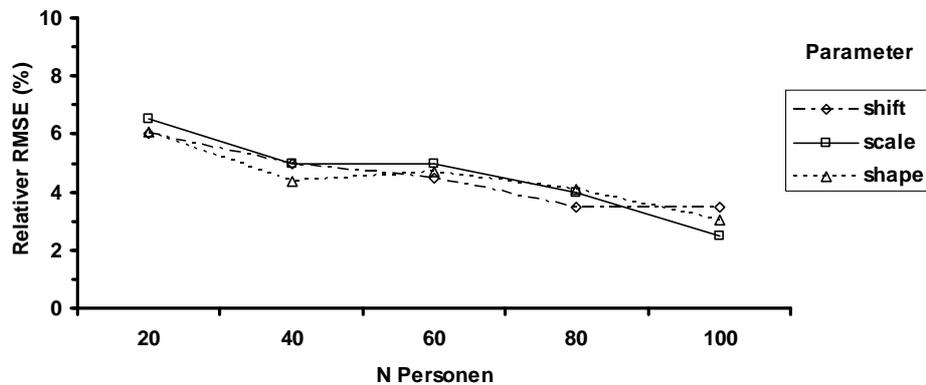


Abb. 4.30: RMSE der Weibull-Parameter in Abhängigkeit von der Stichprobengröße.

Die Effizienz der Parameterschätzung nimmt mit steigender Personenanzahl zu. Diese Steigerung beträgt 2.5 % für den shift-, 4 % für den scale- und 3 % für den shape-Parameter.

N Items. Abb. 4.31 zeigt den Effekt der Anzahl der Items auf den RMSE der drei Weibull-Parameter.

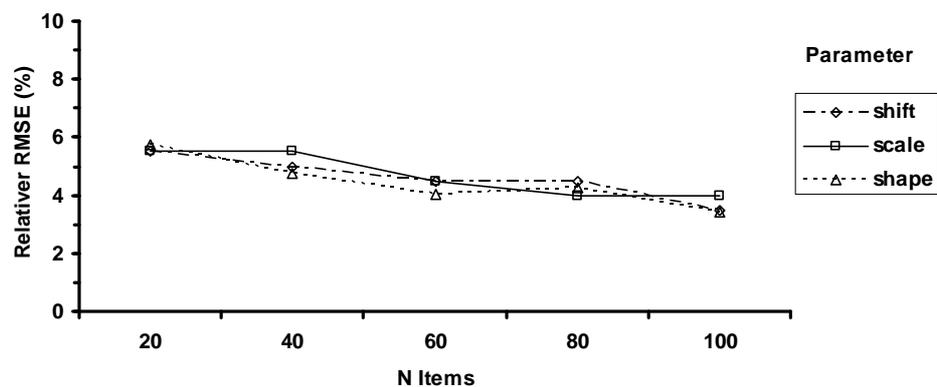


Abb. 4.31: RSME der Weibull-Parameter in Abhängigkeit von der Item-Stichprobengröße.

Auch die Item-Stichprobengröße hat einen Effekt auf die Effizienz der Parameterschätzung auf Stichprobenebene. Allerdings ist der Effekt mit 2 % für den shift-, 1.5 % für den scale- und 2.5 % für den shape-Parameter kleiner ausgeprägt als der für die Personen-Stichprobengröße.

4.3.2.3.3 Zusammenfassung

Auch für die direkte Schätzung der Populationsparameter ergeben sich wieder die gleichen Ergebnisse wie in Simulationsstudie I: Der Bias der Parameter bleibt konstant, während sich die Effizienz der Parameterschätzung durch die Aggregation verbessert. Dabei ist das grundsätzliche Effizienz-Niveau ebenfalls ungefähr vergleichbar mit Simulationsstudie I. Er-

wartungsgemäß zeigt sich die Konsistenz der Parameterschätzung v.a. in einer Effizienzsteigerung durch eine Erhöhung der Personenstichprobe, und zwar sowohl für den Bias als auch die Fehlervarianz der Parameterschätzer. Ebenfalls wie in Simulationsstudie I führt jedoch auch eine Erhöhung der Itemanzahl zu einer Bias-Reduzierung und einer Reliabilitätssteigerung des Verfahrens.

4.4 Diskussion

Insgesamt kann festgestellt werden, dass das von Rouder et al. (2003) beschriebene Verfahren erfolgreich auf zwei neue und dazu sehr heterogene Datenbereiche übertragen werden konnte. Die Effizienz der Parameterschätzung ist insgesamt sehr zufrieden stellend und genauso gut oder sogar noch besser als in der Studie von Rouder et al. Ein Vergleich mit Simulationsstudien zur ML- und Moment-Methode ergibt, dass das Hierarchical-Bayes-Verfahren insgesamt wesentlich effizienter ist als diese. Allerdings ist die Vergleichbarkeit der Ergebnisse dadurch eingeschränkt, dass in den Studien unterschiedliche Parameterkonstellationen und teilweise auch unterschiedliche Effizienzmaße verwendet werden. Die Effizienz des Verfahrens steigert sich noch zusätzlich, wenn die Individualparameter aggregiert werden und zur Schätzung des gemeinsamen Populationsparameters verwendet werden.

Widersprüchlich sind allerdings die Ergebnisse zum Bias der Parameter. Während in Simulationsstudie I der shift-Parameter über- und der scale-Parameter unterschätzt wird, ist es in Simulationsstudie II genau umgekehrt. Der shape-Parameter ist nicht wesentlich verzerrt, obwohl er gleichzeitig der am wenigsten reliable der drei Parameter ist. Im Gegensatz zum shift- und scale-Parameter scheinen sich die Fehler jedoch in diesem Fall gegenseitig aufzuheben. Der Bias für den scale- und den shape-Parameter scheint dabei zumindest in Simulationsstudie I asymptotisch zu sein, weil er weder von der Menge der untersuchten Vpn, noch der Menge an erhobenen Items abhängig ist. Insgesamt ist jedoch das Ausmaß der Verzerrung eher gering, absolut beträgt sie lediglich 1-3 % und dürfte damit eher von theoretischer als praktischer Bedeutung sein. In Simulationsstudie II nimmt der Bias jedoch mit dem Stichprobenumfang auf der Itemebene ab. Ab ca. 100 Items ist er ganz verschwunden. In Bezug auf die Fehlervarianz erwies sich das Verfahren in beiden Simulationsstudien konsistent, d.h. der RMSE nahm mit zunehmendem Stichprobenumfang ab. Dabei war erwartungsgemäß für die Individualebene eher die Itemanzahl entscheidend. Für die Populationsebene erwies sich hingegen sowohl die Größe der Item- als auch der Personen-Stichprobe als bedeutsam.

Unklar ist jedoch, warum es überhaupt zu einer Umkehrung des Bias-Musters zwischen den beiden Simulationsstudien kam. Zumindest scheinen die üblichen in der Literatur zur

Weibull-Verteilung genannten Faktoren wie Stichprobengröße oder Schiefe der Verteilung nicht hierfür verantwortlich zu sein, weil sie in den beiden Studien gleich waren. Damit kommen eigentlich sinnvoller Weise nur noch die Größe der anderen beiden Parametern (bzw. ihre Relation zum konstanten shape-Parameter) oder die veränderte Priorspezifizierung als Ursachen in Frage. Letzteres ist allerdings eher unwahrscheinlich, weil die genaue Priorspezifizierung allgemein einen eher geringen Einfluss auf die Parameterschätzung zu haben scheint. Exemplarisch soll dies für die Parameterschätzung von Simulationsstudie I gezeigt werden. Abb. 4.32 a) und b) zeigen den Bias und den RMSE der Parameterschätzer der Individualebene in Abhängigkeit von verschiedenen Priorspezifizierungen. Prior 1 ist der Prior, der bereits berichtet wurde. Prior 2 ist hingegen der originale „vague prior“, wie er von Rouder et al. (2003) eingesetzt wurde (mit $b_2 = 0.025$). Da der scale-Parameter in dieser Studie um den Faktor 10 niedriger angesetzt worden war, stellt dies eine extreme Fehlspezifikation dar. Entsprechend wird der scale-Parameter in diesem Fall auch noch stärker unterschätzt, allerdings beträgt der relative Bias in diesem Fall immer noch lediglich 4 %. Der Einfluss des Priors macht sich also durchaus bemerkbar, allerdings nur moderat. Auf den RMSE hat der Prior überhaupt keinen Einfluss.

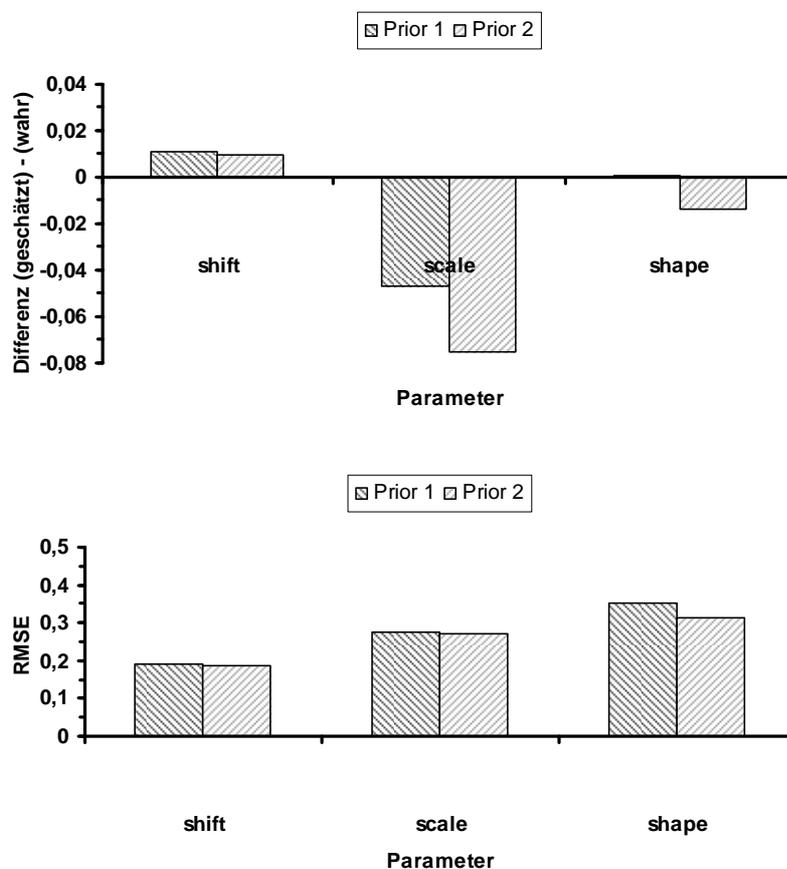


Abb. 4.32: a) Bias und b) RMSE der Parameterschätzung in Simulationsstudie I in Abhängigkeit von der Parameterschätzung.

Unter dem Aspekt der Untersuchungsoptimierung lässt sich abschließend festhalten, dass sich das Verfahren von Rouder auch mit kleineren Personen- und Itemmengen sinnvoll verwenden lässt. Zwar sind die Schätzungen nicht vollkommen fehlerfrei, aber der Anteil an Fehlervarianz steht ist tragbar und mit dem anderer Verfahren vergleichbar, die mit größeren Itemstichproben arbeiten. Problematisch ist hingegen, dass die Parameter leicht überschätzt werden. Dies gilt insbesondere für den scale-Parameter, in Teilen auch für den shift-Parameter. Allerdings ist der Bias vergleichsweise gering und beträgt lediglich 2-3 %.

Aus dem Konsistenzverhalten des Verfahrens kann man weiterhin schließen, dass man zur Effizienzmaximierung im Zweifelsfall eher die Item- als die Personenanzahl steigern sollte. Dafür spricht auch die Tatsache, dass eine Erhöhung der Itemanzahl gleichzeitig die Parameterschätzung auf der Individual- und der Aggregatebene optimiert, die Anzahl der Vpn jedoch lediglich für die Populationsebene relevant ist. Eine Stichprobengröße von $N = 20$ Personen scheint für die geplanten Untersuchungen vollkommen ausreichend zu sein. Eine größere Stichprobe würde zwar einen leichteren Effizienzzuwachs (von ca. 2 %) mit sich bringen. Dieser steht jedoch in keinem angemessenen Verhältnis zum erwartbaren Mehraufwand bei der Erhebung.

Kritischer für die nachfolgenden Untersuchungen ist hingegen die Festlegung der optimalen Itemzahl. Grundsätzlich haben beide Simulationsstudien gezeigt, dass die Anzahl der verfügbaren Datenpunkte pro Person möglichst groß sein sollte. Insbesondere der Übergang von 20 auf 40 Items pro Untersuchungsbedingung scheint einen bedeutsamen Reliabilitätswuchs mit sich zu bringen. Eine darüber hinausgehende Erhöhung auf 60 Items wäre allerdings auch wünschenswert. Dies gilt insbesondere für die Schätzung der Weibull-Parameter bei den wortweisen Lesezeiten, die anscheinend etwas schwieriger zu schätzen sind und einen stärkeren Bias aufweisen. Dem steht allerdings die bereits mehrfach erläuterte Tatsache gegenüber, dass die Anzahl brauchbarer IC-Verben stark begrenzt ist. Zur Untersuchung des in Kap. 2 erläuterten theoretischen Verhältnisses der Effekte des Gender- und des IC-Cues auf die pronominale Auflösung sind mindestens vier Untersuchungsbedingungen notwendig. Zumindest solange ein reines *within*-Design realisiert werden soll, wäre deshalb die vierfache Anzahl von Verben notwendig. Eine realistische Obergrenze für die maximal verfügbare Anzahl von Items pro Untersuchungsbedingung liegt deswegen m.E. bei ca. 40 Items. Denn dann würden insgesamt ca. 160 IC-Verben benötigt. Angesichts der Tatsache, dass die meisten Untersuchungen zum IC-Effekt lediglich 20-40 Verben untersuchen, ist dies eine enorme Menge. Allerdings gibt es im deutschen Sprachkorpus ungefähr 1000 interpersonale Verben, weshalb diese Zahl auch nicht vollkommen unrealistisch erscheint.

5. Material

Das übliche Vorgehen bei Untersuchungen zur IC besteht darin, dass in einem ersten Schritt zunächst verschiedene Verben ausgewählt werden, die vermutlich einen IC-Bias aufweisen. Diese Auswahl wird dann mittels einer Voruntersuchung überprüft bzw. optimiert, indem nur Verben in die Untersuchung aufgenommen werden, die eine bestimmte Mindestausprägung an IC-Bias aufweisen. In einem zweiten Schritt werden dann für diese Verbauswahl die konkreten, in der Untersuchung verwendeten Stimulussätze konstruiert. Diese werden dann meist in einer zweiten Voruntersuchung noch einmal überprüft, v.a. auf ihre Plausibilität (manchmal auch noch einmal auf ihren Bias).

Die Materialerstellung für die vorliegenden Studien folgte dem gleichen Schema. Allerdings musste an einigen Stellen vom etablierten Vorgehen abgewichen werden, und zwar aus Gründen, die mit der Besonderheit der eingesetzten Methodik zu tun haben. Denn wie in Kap. 3 dargestellt, sind für eine reliable Schätzung von RT-Komponenten mehr Datenpunkte notwendig als für die Bestimmung der mittleren RT. Da in den üblichen Studien zur IC nur sehr wenige Items untersucht werden (ca. 20-30 Items), ist hier eine „scharfe“ Verbauswahl möglich, d.h. es werden nur solche Verben untersucht, die einen bestimmten Mindest-Bias aufweisen. In der vorliegenden Untersuchung wird hingegen ein Vielfaches dieser Itemmenge benötigt. Unter diesen Umständen ist klar, dass nicht jedes Item optimal sein kann. Anstatt also durch eine Reihe von Voruntersuchungen einen kleinen, aber möglichst feinen Itemsatz herzustellen, wird im Folgenden aufgrund theoretischer Überlegungen ein bestimmtes Verb- und Itemset vorgegeben. Die Voruntersuchungen dienen dann lediglich der Überprüfung, ob sich die Verben und Stimulussätze der verschiedenen Bedingung hinreichend voneinander unterscheiden.

5.1 IC-Verben

5.1.1 Verbauswahl

Nach Rudolph und Försterling (1997a) gibt es in der deutschen Sprache ca. 1100 interpersonale Verben (Hoffman & Tchir, 1990, und Semin & Marsman, 1994, kommen zu sehr ähnlichen Ergebnissen für das Englische und Niederländische). In der vorliegenden Untersuchung kommt das sog. revidierte *Action-State-Distinction-System* (ASD) von Rudolph und Försterling (1997a) zur Anwendung (vgl. Kap. 2.1.1). Von den 1100 interpersonalen Verben sind ca. 55 % vom Typ Agens-Patiens (AP), 12 % Agens-Evokator (AE), 15 % Stimulus-Experiencer (SE) und 18 % Experiencer-Stimulus (ES). Für jede dieser vier Gruppe existieren

also unselektiert mindesten ca. 120 Verben, sodass es möglich sein sollte, eine Gleichverteilung im verwendeten Itempool zu erreichen.

In einem ersten Schritt alle publizierten Studien wurden zur impliziten Verbkausalität durchgegangen und die dort verwendeten Verben (soweit explizit benannt) in eine Ausgangsliste aufgenommen. Es ergab sich ein Korpus von insgesamt 34 Studien, die den Zeitraum von 1971 (Fillenbaum & Rappaport, 1971) bis 2006 (Koornneef & van Berkum, 2006, Guerry et al., 2006, Crinean & Garnham, 2006) abdeckten. Die extrahierte Liste umfasste insgesamt 313 Verben, die meist dem Englischen, aber auch dem Deutschen, Niederländischen, Französischen und Chinesischen entstammten. Der Umfang der Liste ist durchaus realistisch: Rudolph und Försterling (1997a) untersuchten in ihrer Metaanalyse 246 Verben. Erwartungsgemäß wurden die einzelnen Verben unterschiedlich häufig untersucht: Ungefähr 150 Verben werden nur ein einziges Mal erwähnt (was aber z.T. daran liegt, dass die betrachtete Sprache, wie Französisch oder Deutsch, nicht häufig untersucht wird) und lediglich ca. 30 Verben (alle Englisch) mehr als zehnmal.

In einem zweiten Schritt wurde aus dieser Ausgangsliste eine weitere Liste erstellt, indem deutsche Übersetzungen für die Verben der verschiedenen Sprachen gesucht wurden (nicht alle Verben lassen sich so übersetzen, dass der Kausalitätsbias erhalten bleibt). Dabei konnte die Beobachtung von Rudolph und Försterling (1997a), dass die semantische Rollenverteilung im Englischen und Deutschen weitgehend identisch ist, grundsätzlich bestätigt werden. Des Weiteren erwiesen sich aufgrund der Besonderheiten der vorliegenden Untersuchung eine ganze Reihe von Verben auch in deutscher Übersetzung als unbrauchbar: So können lediglich transitive Verben (mit einer Subjekt- und lediglich einer Objektposition) verwendet werden (was Verben wie *verkaufen* ausschließt), ebenso mussten im Deutschen ambige Verben (z.B. *achten*) und Verben, die nicht gleich frequent in allen Geschlechterkonstellationen verwendet werden (wie z.B. *küssen*), ausgeschlossen werden. Eine weitere große Einschränkung ergab sich durch die spezielle Morphologie des Deutschen (vgl. Bhatt, 1991): Wenn die Anzahl der Wörter pro Satz konstant gehalten werden soll, können alle sog. trennbaren Verbderivate (wie z.B. *anrufen*) nicht verwendet werden, weil bei diesen das Präfix im Hauptsatz satzfinal steht und diese Sätze deshalb um ein Wort länger gewesen wären.

Die Liste mit den deutschen Übersetzungen umfasste schließlich 160 Verben. Es wurde eine Analyse der Kausalitätsrichtung vorgenommen sowie eine Klassifikation im Rahmen des ASD-Systems: Zur Klassifikation wurden die in den vorliegenden Studien berichteten Kennwerte (meist Daten aus Satzergänzungsstudien und Kausalitätsratings) verwendet und die einschlägigen typologischen Tests (vgl. Brown & Fish, 1983a; Au, 1986) semantischer (z.B. füh-

len- bzw. erleben-Paraphrasierung, Beobachtbarkeit) und syntaktischer Art (Imperativ- und Progressiv-Form, aber-Test) durchgeführt (vgl. Kap. 2.1.2). Da alle diese Test nicht wirklich hinreichend sind, um den Verbtyp eindeutig festzulegen (vgl. Hoffman & Tchir, 1990; Edwards & Potter, 1993), hatten sie lediglich eine orientierende Funktion. Auch ist klar, dass die endgültige Festlegung mehr oder weniger „fuzzy“ ist und manchmal substanzielle Graubereiche bestehen. Die letztendliche Einteilung bestand dann aus insgesamt 87 NP1-Verben (AP: 48; SE: 39) und 73 NP2-Verben (AE: 43, ES: 30).

Auch diese Liste konnte noch nicht zur Stimuluskonstruktion verwendet werden. Allerdings erschien damit ein finaler Itempool von 160 Items, in dem alle vier Itemtypen gleich häufig vertreten sind, als durchaus realistisch. Durch erste Versuche der Itemkonstruktion sowie Gespräche mit Kollegen/innen wurde der Verbpool in mehreren Schritten und Rückkopplungsschleifen noch mehrmals modifiziert. Dabei wurden ungeeignete Verben ausgeschlossen und neue durch alternative Übersetzungen der fremdsprachlichen Begriffe sowie die Generierung von Synonymen hinzugefügt. Da gleichzeitig schon erste Versuche der Stimuluskonstruktion stattfanden, waren die Erstellung des Itempools und die Konstruktion der Stimulussätze nicht unabhängig voneinander, sondern vielmehr eng aufeinander abgestimmt. Ein solches Vorgehen ist vor dem Hintergrund einer möglichst zufälligen Auswahl aus dem Universum aller möglichen Items sicherlich zu kritisieren, erwies sich jedoch aufgrund der relativ harten Restriktionen der Verboselektierung als unausweichlich. Insgesamt war der Prozess der Verbauswahl sicherlich nicht willkürlicher als in anderen Studien. Im Gegenteil ist durch die Berücksichtigung ungefähr der sechs- bis achtfachen Itemmenge eine repräsentative Auswahl wahrscheinlicher. Der finale Itempool enthielt, wie intendiert, 160 Verben. Dies ist eine Anzahl, die sich in der in Kap. 4 dargestellten Simulationsstudie als sinnvoll und realistisch zugleich erwiesen hat. Um mögliche Konfundierungen zu vermeiden, war genau die Hälfte der Verben NP1-Verben, die der anderen Hälfte waren vom Typ NP2. Orthogonal hierzu waren 50 % der Verben Handlungs- und die anderen 50 % Zustandsverben.

Ein weiteres Itemmerkmal, welches v.a. im Rahmen attributionstheoretischer Ansätze als ein möglicher Einflussfaktor auf den IC-Effekt diskutiert wird (vgl. Brown & Fisch, 1983a; Au, 1986) ist die Valenz der Verben. Auch wenn sich in der Metaanalyse von Rudolph und Försterling (1997a) keine konsistenten Effekte nachweisen lassen, berichten einzelne Forscherteams (z.B. Franco & Arcuri, 1990; Semin & Marsman, 1994), dass der IC-Effekt bei Verben mit negativer Valenz stärker ausfällt als bei Verben mit positiver oder neutraler Valenz. Entsprechend wurde für jedes Verb des Itempools seine Valenz bestimmt (meist unter Rückgriff auf bisherige Studien, der Rest wurde selbstständig ergänzt). Dabei wurden Verben

wie *lieben* als positiv, Verben wie *verspotten* hingegen als negativ klassifiziert. Verben mit neutraler Valenz (wie z.B. *befragen*) fanden sich vergleichsweise selten, was darauf hindeutet, dass die Darstellung interpersonalen Ereignisse in der Alltagssprache meist eine klare Valenzierung aufweist. Tab. 5.1 zeigt die Anteile der vier Verbtypen getrennt für negative, neutrale und positive Items. Für jede der drei Valenzgruppen wurde ein χ^2 -Tests berechnet, um zu überprüfen, ob alle vier Verbtypen gleichmäßig vertreten sind. Dies war für die negativen und neutralen Verben der Fall (negativ: $\chi^2(1) = 0.53, p > .465$; neutral: $\chi^2(1) = 1.27, p > .285$), nicht aber für die positiven Verben ($\chi^2(1) = 7.83, p < .005$), weil hier Verben vom Typ AP unterrepräsentiert sind. Wie jedoch in der Tabelle deutlich zu sehen, weist dieser Verbtyp ausgleichend vermehrt neutrale Valenzen auf, wahrscheinlich, weil sich diese Verben auf behaviorale Handlungen und nicht valenzierte Zustände beziehen. Wenn man neutrale und positive Verben aggregiert, zeigen sich entsprechend auch keine Unterschiede zwischen den vier Verbtypen ($\chi^2(1) = 0.37, p > .540$). Die Power der Tests war jeweils ausreichend, um Effekte von mittlerer Stärke ($w = .3$, vgl. Cohen, 1988) mit einer Wahrscheinlichkeit von $1-\beta = 5\%$ abzusichern (alle Power-Berechnungen wurden mit dem Programm G-Power 3 berechnet, vgl. Faul, Erdfelder, Lang & Buchner, 2007).

Valenz	IC-Bias		Verbtyp		Total
			Zustand	Handlung	
negativ		NP1	24	24	48
		NP2	22	16	38
	Total		46	40	86
neutral		NP1	4	14	18
		NP2	2	2	4
	Total		6	16	22
positiv		NP1	12	2	14
		NP2	16	22	38
	Total		28	24	52

Tab. 5.1: Beziehung zwischen Valenz eines Verbs und den vier verschiedenen Verbtypen.

Alle anderen Merkmale der Verben (wie Frequenz etc.) wurden bei der Verbauswahl nicht kontrolliert, auch weil das Material eine möglichst repräsentative Auswahl aus dem deutschen Sprachkorpus darstellen sollte. Alle diese Merkmale werden nachfolgend dadurch kontrolliert, dass alle relevanten experimentellen Manipulationen *within*-Items-Faktoren sind. Das heißt, jedes der Verben wird in allen experimentellen Bedingungen gleichermaßen gezeigt, weshalb es zu keiner systematischen Konfundierung von Itemmerkmalen und experimentellen Faktoren kommen kann.

5.1.2 Voruntersuchung I: Kontrolle IC-Bias

Die erste Voruntersuchung sollte sicherstellen, dass die untersuchten Verben überhaupt einen hinreichend starken IC-Bias aufweisen. Dieser Treatment-Check wird üblicherweise mittels einer Satzergänzungsstudie realisiert (vgl. z.B. Garnham et al., 1996; Stewart et al., 2000; Koornneef & van Berkum, 2006), bei der Personen vorgegebene Satzfragmente (wie *Peter mochte Hans, weil er...*) frei ergänzen können und evtl. zusätzlich noch die intendierte Referenz des Pronomens angeben. Wenn die in den Satzfragmenten verwendeten Verben in der Tat einen gewissen IC-Bias haben, dann sollten in den freien Fortsetzungen systematisch entweder NP1- oder NP2-ERgänzungen überwiegen.

Konkret wurde so vorgegangen, dass alle 160 ausgewählten IC-Verben mit jeweils zwei gleichgeschlechtlichen Aktanten (zur Zuweisung vgl. Abschnitt 5.2) mit einem *weil*-Konnektiv und einem geschlechtskongruenten Pronomen zu Satzfragmenten gruppiert wurden (Beispiel: *Ada bemitleidete Nadine, weil sie...*). Die eine Hälfte der Satzfragmente verwendete zwei weibliche Aktanten, die andere Hälfte zwei männliche. Die Sätze wurden in eine pseudo-randomisierte Reihenfolge gebracht, indem sie alphabetisch nach dem Anfangsbuchstaben des Subjekts sortiert wurden. Es wurden insgesamt zwei Abfolgen hergestellt, wobei die Sätze in der einen alphabetisch auf-, in der anderen absteigend angeordnet waren. Die Vpn wurden instruiert, jedes Satzfragment durch eine Begründung der im Hauptsatz genannten Situation zu ergänzen. Die Fortsetzung sollte vergleichsweise kurz sein (der hierfür zur Verfügung stehende Platz war auf eine Zeile begrenzt) und sollte möglichst spontan vorgenommen werden. Die Vpn wurden darauf hingewiesen, dass sie bei der Formulierung des Ergänzungssatzes grundsätzlich vollkommen frei sind. Sie durften jedoch lediglich auf eine der im Hauptsatz genannten Personen Bezug nehmen. Die andere sollte weder direkt noch indirekt im Nebensatz erwähnt werden (wie in *Peter kontrollierte Richard, weil er Richard/ihm misstraute*). Diese Restriktion war deshalb notwendig, weil es sonst möglich ist, die kausal saliente Person „durch die Hintertür“ wieder in den Satz einzuführen (vgl. Guerry et al., 2006). Abschließend wurden die Versuchspersonen noch gebeten, diejenige Person durch Umkreisung zu markieren, auf die sie sich bezogen hatten. Die verschiedenen Fragebogenversionen wurden in einer übergreifenden Lehrveranstaltung verteilt, konnten jedoch eigenständig innerhalb einer Woche bearbeitet werden. Die Versuchspersonen erhielten 1.5 Vp-Stunden als Vergütung für ihre Teilnahme.

Stichprobe. Insgesamt gaben 30 Personen ihren Fragebogen im vereinbarten Zeitraum wieder ab (Rücklaufquote: 75 %). Die Stichprobe bestand ausschließlich aus Studierenden des

2. Fachsemesters des Diplom-Studiengangs Psychologie (29 Frauen, 1 Mann), die durchschnittlich 28.3 Jahre alt waren ($SD = 9.9$). Alle Versuchspersonen gaben an, Deutsch als Erstsprache erworben zu haben.

Referenz-Markierung. Grundlage der Auswertung waren die Umkreisungen, mit denen die Vpn selbst markiert hatten, wen sie mit dem Pronomen des Nebensatzes meinten. Allerdings wurden alle Umkreisungen mit dem Inhalt der Satzergänzung abgeglichen, fehlende Umkreisungen ergänzt und offensichtliche Fehlumkreisungen korrigiert (1.3 %). Darüber hinaus wurden alle Sätze, die fehlerhaft bearbeitet wurden (weil entweder eine Plural-Interpretation des Pronomens *sie* oder der nicht gewählte Aktant zusätzlich im Nebensatz genannt wurde), von der Analyse ausgeschlossen (ca. 7.1 %). Für die verbleibenden Datenpunkte wurde jeweils kodiert, um welches der 160 Verben es sich handelte und ob eine Vpn das Subjekt (Wert: 1) oder das Objekt (Wert: 2) zur Auflösung gewählt hatte. Entsprechend dem Vorgehen von Koornneef und van Berkum (2006) wurden die Werte nachfolgend über alle Vpn pro Verb gemittelt: Ein Wert von 1.0 bedeutet dann einen maximalen Bias in Richtung zum Hauptsatz-Subjekt (NP1) und ein Wert von 2.0 einen maximalen Bias in Richtung zum Hauptsatz-Objekt (NP2). Tab. 5.2 zeigt die entsprechenden Bias-Werte für die vier Verbarten.

IC-Bias	Verbtyp		Total
	Zustand	Handlung	
NP1	1.10 (0.08)	1.27 (0.20)	1.18 (0.17)
NP2	1.76 (0.19)	1.75 (0.17)	1.76 (0.18)
Total	1.42 (0.36)	1.52 (0.30)	1.47 (0.34)

Tab. 5.2: Mittelwerte und Standardabweichungen (in Klammern) des Verb-Bias für die vier Verbarten.

Eine *by-items*-Varianzanalyse mit den beiden Faktoren *IC-Bias* (2: NP1 vs. NP2, between) und *Verbtyp* (2: Zustands vs. Handlungsverb, between) ergab einen ausgesprochen varianzstarken Haupteffekt für den Faktor *IC-Bias* ($F(1,156) = 480.22, p < .001, \eta_p^2 = .755$): NP1-Verben zeigten in der Tat einen sehr starken Bias in Richtung Hauptsatz-Subjekt, NP2-Verben hingegen einen ungefähr ebenso starken in Richtung Hauptsatz-Objekt (vgl. Abb. 5.1). Der Haupteffekt des Faktors *Verbtyp* war ebenfalls signifikant ($F(1,156) = 9.24, p < .003, \eta_p^2 = .056$): Bei Zustandsverben wurde das Subjekt als Referent des Pronomens präferiert. Ein signifikanter, aber schwacher Interaktionsterm ($F(1,156) = 11.42, p < .001, \eta_p^2 = .068$) indizierte, dass der Unterschied zwischen NP1- und NP2-Verben bei den Zustandsverben stärker ausfiel ($F(1,156) = 319.87, p < .001, \eta_p^2 = .672$) als bei den Handlungsverben ($F(1,156) = 171.77, p < .001, \eta_p^2 = .524$).

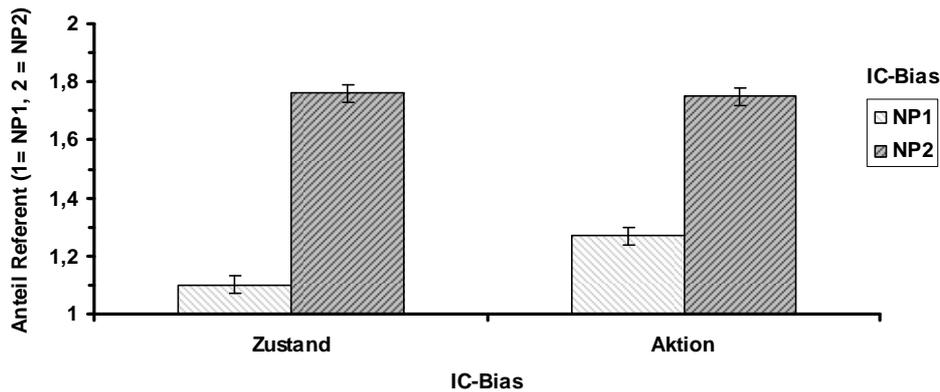


Abb. 5.1: Mittelwerte und Standardabweichungen des Verb-Bias für die vier Verbarten.

Zusammenfassung. Unabhängig vom Bias der einzelnen Verben gibt es eine generelle Tendenz zur Subjekt-Auflösung: Dafür spricht einerseits, dass der Gesamtmittelwert knapp unter dem zu erwartenden Wert von 1.5 liegt, andererseits aber auch die Tatsache, dass der NP1-Bias (mit einer Abweichung von .32 vom zu erwartenden Mittelwert) stärker ausfällt als der NP2-Bias (Abweichung von .26). Da der ersteingeführte Aktant einen gewissen Fokus-Vorteil hat, ist diese (eher leichte) Tendenz gut erklärbar (sog. *first mention advantage* vgl. Gernsbacher, 1989).

Abgesehen von dieser allgemeinen, sehr leichten Auflösungstendenz zum Satzsubjekt weist der überaus starke Unterschied zwischen den NP1- und den NP2-Verben darauf hin, dass die Manipulation des IC-Bias ausgesprochen gut gelungen ist. Einschränkend muss allerdings hinzugefügt werden, dass die Manipulation insgesamt schwächer ausfällt als in vergleichbaren Studien. So berichten z.B. Koornneef und van Berkum (2006) einen mittleren NP1-Bias von 1.14 und einen mittleren NP2-Bias von 1.95. Dieser Umstand ist hauptsächlich der Tatsache geschuldet, dass im Gegensatz zu den bereits existierenden Studien eine sehr viel größere Itemanzahl verwendet wurde (z.B. Koornneef & van Berkum, 2006, untersuchten nur 40 Verben). Von daher ist es verständlich, dass nicht jedes Item einen vergleichbar extremen Bias aufweist. Wenn nur die jeweils 20 stärksten NP1- und NP2-Verben ausgewählt werden, dann ergeben sich ein NP1-Bias von 1.04 und ein NP2-Bias von 1.95. Allerdings ist im vorliegenden Kontext nur wichtig, dass die Bedingungen sich insgesamt voneinander unterscheiden.

Ein Wort noch zum Unterschied im IC-Bias zwischen den Handlungs- und den Zustandsverben: Es ist ein gut bekannter und häufig replizierter Befund, dass der IC-Bias für Zustandsverben stärker ausfällt als für Handlungsverben (vgl. z.B. Rudolph & Försterling, 1997a). Die vorliegenden Ergebnisse können dies bestätigen. Allerdings fällt dieser Unterschied vergleichsweise klein aus und ist (prinzipiell) vernachlässigbar (solange er nicht mit

den einzelnen experimentellen Faktoren konfundiert ist). Allerdings ist im vorliegenden Kontext unklar, ob die Diskrepanz zwischen den Handlungs- und Zustandsverben wirklich darauf zurückzuführen ist, dass der IC-Bias bei Zustandsverben generell stärker ist, oder ob vielmehr selektiv nur die AP-Verben kausal weniger eindeutig interpretiert werden.

5.2 Stimulussätze

5.2.1 Konstruktionsprinzip der Stimulussätze

Nach der zufrieden stellenden Auswahl von Items mit IC-Bias wurden in einem zweiten Schritt die relevanten Stimulussätze zu jedem Verb konstruiert. Allerdings ist noch einmal darauf hinzuweisen, dass Verbauswahl und Itemkonstruktion zwar prinzipiell unabhängig voneinander erfolgten, aber durch antizipatorische Rückkopplungsschleifen aufeinander bezogen waren: Es macht wenig Sinn, ein bestimmtes Verb in einem sehr aufwändigen Selegierungsverfahren auszuwählen und zu validieren, wenn absehbar ist, dass sich keine guten Stimulussätze zu ihm konstruieren lassen.

Aufgrund der benötigten Itemanzahl kamen als konkret zu bearbeitende Stimuli nur Einzelsätze in Frage. Dies ist zwar der in den meisten Studien zur impliziten Verbkausalität der übliche Stimulustyp, allerdings verwenden zumindest einige Studien (z.B. die Gruppe um McKoon, vgl. aber auch Koornneef & van Berkum, 2006) auch „Mini-Texte“ aus drei Sätzen, um ansatzweise einen diskursiven Kontext zur Verfügung zu stellen. Dies wäre m.E. eigentlich die theoretisch zu bevorzugende Variante gewesen: Zum einen ist es nun einmal ein Merkmal von Sprache, dass wir fast nie Einzelsätze, sondern in der Regel in einen Diskurskontext eingebettete Information verarbeiten. Solche Stimuli haben deswegen eine größere Realitätsnähe. Zum anderen ist es in solchen kleinen „Geschichten“ einfacher, eindeutig disambiguierte Auflösungen für ein Pronomen zu generieren. Allerdings wären mit diskursiv eingebetteten Stimuli im vorliegenden Untersuchungskontext auch Nachteile verbunden: Zum einen sind sie zu umfangreich und würden zu einer unangemessen langen Untersuchungszeit führen. Auch zielt die vorliegende Studie auf die Erfassung relativ feinkörniger Prozesse auf der Mikro-Ebene der Satzverarbeitung ab, weshalb konfundierende Kontexteffekte eher zu vermeiden sind. Dennoch ist die Abwesenheit eines natürlichen, auflösungs-lizenzierenden Diskurs-Kontextes als eine der größten Einschränkungen der vorliegenden Studie zu betrachten. Allerdings ist unwahrscheinlich, dass dieses Defizit die Validität der Ergebnisse allzu sehr beeinträchtigt: Studien, die diskursive Stimuli verwenden, kommen prinzipiell zu qualitativ sehr ähnlichen Ergebnissen wie Studien, die lediglich Einzelsätze untersuchen. Allerdings

ergeben sich auch einige wichtige Unterschiede, auf die im Laufe der folgenden Studien noch eingegangen wird.

Praktisch ergibt sich durch die Festlegung auf die Untersuchungseinheit „Satz“ zunächst einmal das Problem, in nur einem einzigen Satz ein Szenario zu etablieren, das sowohl eine mit der IC des Verbs kongruente als auch eine inkongruente Fortführung zulässt. Darüber hinaus ist die zur Verfügung stehende Satzstruktur stark restringiert: Die Logik der Studien zur IC bringt es mit sich, dass ein (kausales) Nebensatzgefüge untersucht werden muss, bei dem im Hauptsatz (HS) zwei Aktanten (NP1 und NP2) genannt werden. Diese sind mit einem IC-Verb miteinander verbunden, welches entweder dem Aktanten NP1 oder dem Aktanten NP2 mehr kausales Gewicht beimisst. An diesem Hauptsatz muss sich ein kausaler Nebensatz anschließen, der mit einer kausalen Konjunktion (meist *weil*) eingeleitet werden muss (da der Effekt bei anderen Konjunktion nicht auftritt, vgl. Abschnitt 2.1.3.1).

Vor diesem Hintergrund ist es wichtig zu betonen, dass alle verwendeten Sätze in der Tat eine *kausale* Interpretation hatten: Traxler et al. (1997a, 1997b) haben darauf hingewiesen, dass „weil“-Konnektive *kausal* (*Die Straße ist nass, weil es regnet.*) oder aber *diagnostisch* (*Es regnet, weil die Straße nass ist.*) verwendet werden können, wobei die Information des Nebensatzes in diesem Fall nur Indikatoren für die Behauptung des Hauptsatzes spezifiziert. Traxler et al. konnten nachweisen, dass diagnostische *weil*-Nebensätze weitaus schwieriger zu verarbeiten sind als kausale (wobei m.E. fraglich ist, ob dies daran liegt, dass es sich um einen diagnostischen Nebensatz handelt, oder schlicht daran, dass solche Sätze pragmatisch defizitär sind). Die Verbindung zur vorliegenden Problemstellung besteht darin, dass eine „diagnostische“ Interpretation des Kausalkonnektivs *weil* u.U. eine adäquate „Hintertür“ ist, um einen Nebensatz entgegen der Richtung der IC zu formulieren. In der Tat haben einige Versuchspersonen in der Voruntersuchung I (vgl. Abschnitt 5.1) genau diese Strategie angewendet. So formulierte Vp 2 z.B. den Satz *Elke beschwindelte Herta, weil sie log*. Hier ist der Akt des Lügens kein *Grund* für den Akt des Schwindelns, sondern eher seine Manifestation. Ähnliches gilt auch, wenn man jemanden *hindert*, indem man ihm *im Weg steht* oder jemanden *überrumpelt*, dadurch dass man *voreilig handelt* etc. Wenn aber in den Nebensätzen überhaupt gar keine Gründe genannt werden, die mit der IC konfliktieren, dann ist die Bildung solcher Sätze viel einfacher. Obwohl die Unterscheidung zwischen kausalen und diagnostischen *weil*-Nebensätzen unscharf ist und fließende Übergänge existieren, wurde in der vorliegenden Arbeit darauf geachtet, dass alle Stimuli eine klare Kausalinterpretation hatten. Dies wurde dadurch gewährleistet, dass solche Sätze ausgeschlossen wurden, bei denen das Konnektiv *weil* durch *indem* oder *dadurch dass* paraphrasiert werden konnte. Dies ist ein geeigne-

ter Test dafür, ob in einem Satz der kausale oder der diagnostisch-instrumentelle Gebrauch überwiegt.

Darüber hinaus ist klar, dass in dem Nebensatz ein (ambiges) Pronomen erwähnt werden muss (ansonsten kann man ja nicht den Auflösungsprozess untersuchen). Darüber hinaus muss eine, die intendierte Auflösung lizenzierende Begründung gegeben werden, die entweder mit der Kausalitätsrichtung des Verbes übereinstimmt (*ic-kompatibel*) oder aber nicht (*ic-inkompatibel*). Das impliziert schon einmal folgendes Satzschema, ohne dass bis jetzt auch nur eine überflüssige Festlegung vorgenommen wurde (vgl. (1)):

- (1) Schema: $[[NP1]_{HS-NP} [IC\text{-Verb } NP2]]_{HS-VP}]_{HS}, [[\text{weil}] [Pro]_{NS-NP} [V]_{NS-VP}]_{NS}$

Erläuterung: Generell markieren eckige Klammern Konstituenten. Der Hauptsatz enthält also minimal die beiden Aktanten NP1 (als Subjekt) und NP2 (als Objekt) sowie das verbindende (transitive) IC-Verb. NP1 stellt eine eigene Nominalphrase dar, Verb und NP2 bilden zusammen eine Verbalphrase. Der Nebensatz enthält nach der initialen Konjunktion notwendigerweise das Pronomen als Subjekt (Nebensatz-NP) und mindestens ein Verb (als minimale Nebensatz-VP). Wenn sich das Pronomen eindeutig lediglich auf einen der beiden Aktanten des Hauptsatzes beziehen soll, dann muss es in der 3. Person Singular stehen (s.u. zur Problematik, dass im Deutschen die 3. Pers. Sing. Fem. und die 3. Pl. morphologisch identisch sind). Abb. 5.2 zeigt ein Beispiel für dieses minimale Satzschema:

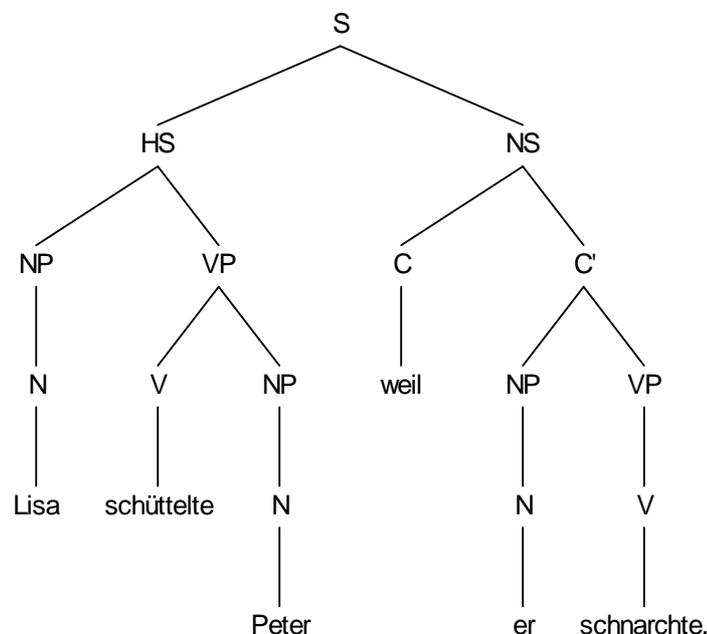


Abb. 5.2: Darstellung der Minimalform des untersuchten Satzgefüges als Baumstruktur (approximativ).

Mit dem skizzierten Minimalschema lassen sich Sätze wie *Lisa schüttelte Peter, weil er schnarchte.* bilden. Freiheitsgrade bestehen eigentlich nur noch bezüglich der genauen Ausgestaltung der Hauptsatz- und Nebensatz-Komponente des Satzgefüges. Aus Gründen der

Ökonomie wurde die Festlegung getroffen, dass sowohl der Hauptsatz als Ganzes als auch die Nebensatz-NP in der Tat minimal realisiert werden sollten, d.h. hier durften keine weiteren Elemente dem Satz hinzugefügt werden. Des Weiteren wurde festgelegt, dass als Aktanten lediglich Eigennamen verwendet werden sollten, da diese einerseits eine möglichst sparsame Art der NP-Realisierung darstellen (sie benötigen z.B. keine Artikel) und andererseits bekannt ist, dass pronominale Auflösungseffekte mit Eigennamen stärker ausfallen als bei definiten NPs (vgl. Badecker & Staub, 2002). Alle Sätze bestanden also bis zur Einführung des aufzulösenden Pronomens bereits aus 4 Wörtern, die in der ic-kompatiblen und der ic-inkompatiblen Fassung jeweils identisch sind. Zur Festlegung einer möglichst eindeutigen und überzeugenden Lesart, welche (in)kongruent mit dem IC-Bias des Hauptsatz-Verbs ist, verblieb unter diesen Umständen lediglich noch der restliche Nebensatz. Unter der Voraussetzung, dass die Sätze nicht allzu lang sein sollten, dafür aber gleich viele Wörter enthalten, ist das eine schwierige Aufgabe.

Das wird sofort klar, wenn man sich das obige Beispiel *Lisa schüttelte Peter, weil er schnarchte*. betrachtet. Dieser Satz ist vielleicht isoliert betrachtet ein wenig seltsam, aber doch irgendwie sinnvoll, sofern sich das Pronomen *er* auf Peter bezieht (was ja nicht zwangsläufig der Fall sein muss!). Zumindest lässt sich an der Tatsache, dass wir den Satz *Lisa schüttelte Peter, weil sie schnarchte*. als noch missratener betrachten, ersehen, dass eine Situation, in der eine Person eine andere schüttelt, weil diese schnarcht, mit unserem Weltwissen über das Schlaf- und Schüttelverhalten von Menschen kompatibel ist, nicht aber der umgekehrte Fall: Nicht nur, dass die Tätigkeit des Schnarchens nicht die des Schüttelns impliziert, darüber hinaus sind die beiden Tätigkeiten in diesem Fall (außer unter ziemlich abstrusen Randbedingungen) sogar pragmatisch inkompatibel. Die Preisfrage ist nun: Was können wir statt dem Verb *schnarchen* in den Satz einfügen, damit sich eine Lesart ergibt, bei der die referenzielle Beziehung zwischen *sie* und *Lisa* ähnlich zwingend ist wie die zwischen dem schnarchenden Peter und *er*? Der/die Leser/in möge einmal ausprobieren, ob er/sie hierfür ein einzelnes Verb findet, ohne gleichzeitig ein grobes pragma-semantisches Foul zu begehen. Es steht zu vermuten, dass er/sie sich nicht nur nach vergleichsweise kurzer Zeit in einem Zustand milder Verzweiflung befindet, sondern auch die Notwendigkeit verspürt, mehr als nur ein Wort als Satzergänzung zuzulassen.

Vor dem Hintergrund dieser antagonistischen Anforderungen wurde letztlich folgendes, liberales Schema für die Nebensatz-VP gewählt: Als Tribut an die standardisierte Satzlänge musste sie in jedem Fall genau vier Wörter umfassen. Ebenfalls musste sie mindestens noch ein weiteres Argument erfordern (meist eine Nominalphrase als Objekt, manchmal auch ein Adverb). Ansonsten war die Binnenstrukturierung jedoch frei, d.h. das Objekt konnte a) mit einem fakultativen Artikel (bzw. Quantor) und/oder einem Adjektiv noch weiter spezifiziert werden (z.B. *viele dumme Scherze machte*), b) das Verb konnte durch ein Adverb und/oder einen Intensifier (*ganz, besonders*) ergänzt werden (z.B. *plötzlich keine Luft bekam*), c) das Verb musste nicht singular realisiert sein, sondern konnte (durch die Hinzufügung von Modal-

und Hilfsverben) auch eine komplexe Verbalgruppe sein (z.B. *entlarvt werden*, *akzeptieren können* etc.). In (2a-e) ist ein (wahrscheinlich unvollständiges) Schema für die ergänzende Nebensatz-VP dargestellt sowie jeweils einen Beispielsatz für alle vier Verbtypen (der intendierte Referent ist jeweils das Hauptsatz-Subjekt):

- (2a) Schema NS-VP: [[(Det) (Adj) N]_{NP} (Adv) V (Mod) (Aux)]
- (2b) AP: Ingeborg belog Irene, weil sie Angst vor Vorwürfen hatte.
- (2c) AE: Peter belohnte Heinz, weil er etwas gut machen wollte.
- (2d) SE: Rabea begeisterte Ines, weil sie gewaltige Fortschritte gemacht hatte.
- (2e) ES: Pamela beneidete Lola, weil sie viel weniger Geld verdiente.

Auf der Grundlage dieses Schemas wurde für jedes der 160 Verben jeweils eine Fortsetzung generiert, bei dem sich das Pronomen pragmatisch auf das Hauptsatz-Subjekt bezog, und eine, bei der das Pronomen pragmatisch auf das Hauptsatz-Objekt referierte. Für Verben mit NP1-Bias sind erstere ic-kompatibel (IC-Cue+) und letztere ic-inkompatibel (IC-Cue-), bei NP2-Verben ist es genau umgekehrt. In (3) und (4) ist der Unterschied zwischen den beiden Itemversionen dargestellt:

- (3a) SE, IC+ (Bias: NP1, Referenz: NP1):
Rabea begeisterte Ines, weil sie gewaltige Fortschritte gemacht hatte.
- (3b) SE, IC- (Bias: NP1, Referenz: NP2):
Rabea begeisterte Ines, weil sie schnell zu beeindrucken war.
- (4a) ES, IC+ (Bias: NP2, Referenz: NP2):
Pamela beneidete Lola, weil sie viel mehr Geld verdiente.
- (4b) ES, IC- (Bias: NP2, Referenz: NP1):
Pamela beneidete Lola, weil sie viel weniger Geld verdiente.

Damit jedoch noch nicht genug: Aus Gründen, die im Laufe der Arbeit noch ausführlich erläutert werden, wurden für jeden Stimulussatz zwei Versionen erstellt. Die erste Version war mit dem oben dargestellten Schema identisch: Der Nebensatz enthielt lediglich eine vier Wörter umfassende Verbalphrase und das ganze Satzgefüge hatte insgesamt 9 Wörter. Diese Länge ist ungefähr vergleichbar zur Länge der Sätze, wie sie von Caramazza et al. (1977) und auch in vielen nachfolgenden Untersuchungen eingesetzt wurden (Ehrlich, 1980; Vonk, 1984 etc.).

Die verwendete zweite Version stellte eine Erweiterung der ersten Version dar, wobei im Nebensatz zusätzlich noch eine Präpositionalphrase (zwischen Pronomen und Verbalphrase) eingefügt wurde. Diese neue Präpositionalphrase war in beiden Satzfortführungen identisch, umfasste ebenfalls vier Wörter und war variabel nach dem in (5) dargestellten Schema aufgebaut.

(5) Schema NS-PP: [P (Det) (Adj) N]_{PP}

Das heißt, die Präpositionalphrase bestand aus einer obligatorischen Präposition und einem Nomen. Hinzutreten konnten dann noch ein oder zwei Determinierer und/oder ein oder zwei Adjektive (z.B. *in der unvorhergesehenen Notlage*).

Die Einfügung der Präpositionalphrase hat insgesamt vier Effekte: a) Die Länge der Stimuli verlängert sich von 9 auf 13 Wörter; b) Die zusätzliche Präpositionalphrase führt durch die Einführung neuer Informationen zu einer stärkeren diskursiven Einbettung der Verbalphrase; c) Im Gegensatz zur Version ohne Präpositionalphrase, in der das Pronomen direkt nach seiner Einführung disambiguiert wird, gibt es nun eine „neutrale“ Phrase im Satz, in der die Referenz des Pronomens noch nicht durch pragmatische Faktoren determiniert wird, sondern vielmehr „in der Schwebe“ bleibt; d) Durch die Tatsache, dass die Präpositionalphrase bei beiden Satzfortführungen gleich war, minimiert sich die Wahrscheinlichkeit, dass (frühe) Effekte des IC-Cues auf Unterschiedlichkeiten des verbalen Materials in den beiden Fortführungen zurückgehen (vgl. hierzu auch Koornneef & van Berkum, 2006). All diese Effekte werden jedoch durch die zusätzliche Restriktion „erkaufte“, dass die Präpositionalphrase zu beiden Referenz-Lesarten passen muss und gleichzeitig für sich neutral bleibt. In (6) und (7) sind die obigen Beispiele mit der zusätzlichen PP-Einfügung dargestellt:

(6a) SE, IC+ (Bias: NP1, Referenz: NP1):

Rabea begeisterte Ines, weil sie durch die neue Gesangstechnik gewaltige Fortschritte gemacht hatte.

(6b) SE, IC- (Bias: NP1, Referenz: NP2):

Rabea begeisterte Ines, weil sie durch die neue Gesangstechnik schnell zu beeindrucken war.

(7a) ES, IC+ (Bias: NP2, Referenz: NP2):

Pamela beneidete Lola, weil sie trotz gleicher beruflicher Tätigkeit viel mehr Geld verdiente.

(7b) ES, IC- (Bias: NP2, Referenz: NP1):

Pamela beneidete Lola, weil sie trotz gleicher beruflicher Tätigkeit viel weniger Geld verdiente.

Der bisherige Stand lässt sich folgendermaßen zusammenfassen: Für jede der vier Verbgruppen, die sich aus der Kombination von *Verb-Bias* (2: NP1 vs. NP2) und *Verbtyp* (2: Handlung vs. Zustand) ergeben, wurde eine Satzversion erstellt, bei der sich das Pronomen aus pragmatischen Gründen entweder auf das Subjekt oder Objekt des Hauptsatzes bezieht (2 *IC-Cue*: + vs. -). Zusätzlich konnte jeder Satz entweder *mit* oder *ohne* zusätzliche Präpositionalphrase stehen (2 *Satzversion*: PP+ vs. PP-).

Was bis jetzt noch fehlt, ist der Gender-Cue. Dieser ist im Vergleich zur Kausalitätsmanipulation vergleichsweise einfach zu operationalisieren, da hierfür einfach nur die Namen der im Hauptsatz genannten Aktanten geändert werden müssen. Das geschah wie folgt: Insgesamt gibt es 160 Stimuli, in denen jeweils zwei Aktanten erwähnt werden. Deswegen wurden

in einem ersten Schritt 320 gängige deutsche Vornamen (160 männlich, 160 weiblich) gesammelt, indem verschiedene sozialpsychologische Studien zur Wahrnehmung von Vornamen (Rudolph & Spörrle, 1999; Rudolph, Böhm & Lummer, 2006) und diverse Namens-Wörterbücher (z.B. Duden, 2002) gesichtet wurden. Danach wurde den 160 Stimulussätzen zufällig ein „Geschlecht“ zugewiesen. Das heißt, die Hälfte der Sätze waren „männliche“ Stimuli und die andere Hälfte „weibliche“. „Männliche“ Stimuli verwendeten immer das Personalpronomen *er* und enthielten im nicht-gender-disambiguierten Fall zwei männlichen Aktanten. „Weibliche“ Stimuli verwendeten hingegen immer das Personalpronomen *sie* und enthielten in der nicht-gender-disambiguierten Bedingung zwei weibliche Vornamen. Diese feste (aber nicht konfundierende) Zuweisung eines bestimmten Personalpronomens zu einem der Sätze war deshalb notwendig, weil das Personalpronomen *sie* im Deutschen (ebenso wie im Niederländischen und im Gegensatz zum Englischen) ambig ist: Es kann sowohl auf eine einzelne weibliche Person als auch auf zwei (oder mehr) Personen (egal welchen Geschlechts) zusammen referieren. Obwohl diese Ambiguität spätestens aufgrund der obligatorischen Person-Numerus-Genus-Kongruenz von Subjekt und Prädikat im Deutschen satzfinal aufgelöst wird, ist nicht auszuschließen, dass diese lokale Unsicherheit die pronominale Auflösung zumindest temporär beeinflusst.

Für jeden der insgesamt 320 Stimulussätze (160 Verben x 2 IC-Cue-Bedingungen) wurden wiederum zwei unterschiedliche Fassungen erstellt; eine, in denen beide Aktanten das gleiche Geschlecht haben (Gender-Cue-), und eine andere, bei der der jeweilige Nicht-Referent dem anderen Geschlecht angehört (Gender-Cue+). Das heißt, in dieser Bedingung waren zur Auflösung des Pronomens keine weit reichenden pragmatischen Schlussfolgerungsprozesse notwendig, sondern der Referent des Pronomens konnte allein aufgrund der morpho-syntaktisch eindeutigen Genus-Numerus-Kongruenz zwischen Pronomen und Referent identifiziert werden (vgl. (8), eine vollständige Liste aller Stimulussätze findet sich in Appendix B).

(8a) Verbtyp ES, IC+, Gender-:

Pamela beneidete Lola, weil sie trotz gleicher beruflicher Tätigkeit viel mehr Geld bekam.

(8b) ES, IC+, Gender+

Klaus beneidete Lola, weil sie trotz gleicher beruflicher Tätigkeit viel mehr Geld bekam.

(8c) Verbtyp ES, IC-, Gender-:

Pamela beneidete Lola, weil sie trotz gleicher beruflicher Tätigkeit viel weniger Geld verdiente.

(8d) Verbtyp ES, IC-, Gender+:

Pamela beneidete Heiner, weil sie trotz gleicher beruflicher Tätigkeit viel weniger Geld verdiente.

Damit sind alle vier theoretisch relevanten Experimentalbedingungen komplett realisiert. Sie ergeben sich aus der Kombination von drei sprachlichen Merkmalen, deren Beziehung hier noch einmal überblicksartig dargestellt werden soll (vgl. Tab. 5.3).

Gender-Cue	IC-Cue	<i>Verb-Bias</i> (lexikalisch-semantisch)	<i>Gender-Kongruenz</i> (morpho-syntaktisch)	<i>NS-Fortsetzung</i> (pragmatisch)
+	+	NP1	NP1	NP1
+	-	NP1	NP2	NP2
-	+	NP1	NP1/NP2	NP1
-	-	NP1	NP1/NP2	NP2
+	+	NP2	NP2	NP2
+	-	NP2	NP1	NP1
-	+	NP2	NP1/NP2	NP2
-	-	NP2	NP1/NP2	NP1

Tab. 5.3: Überblick über die vier Hauptuntersuchungsbedingungen.

Die vier Hauptuntersuchungsbedingungen sind in den ersten beiden Spalten abgetragen, und zwar zweimal, zunächst für die NP1-Verben, dann für die NP2-Verben. Die implizite Kausalitätsrichtung des Verbs (*Verb-Bias*) ist das erste Merkmal, das für die Experimentalbedingungen eine Rolle spielt. Es ist ein lexikalisch-semantisches Verbmerkmal, das entweder dem Hauptsatz-Subjekt oder dem Hauptsatz-Objekt kausale Salienz verleiht. Das zweite Merkmal, die *Gender-Kongruenz* von Referent und Pronomen ist morpho-syntaktischer Natur. Es kann entweder eine referenzielle Beziehung des Pronomens zum Hauptsatz-Subjekt oder zum Hauptsatz-Objekt etablieren. Im Gegensatz zum Verb-Bias ist dieses Merkmal ausgesprochen mächtig. Seine Verletzung führt zu einer klar ungrammatischen Struktur. Allerdings ist es im Fall geschlechts-homogener Aktanten uninformativ (was durch den Eintrag „NP1/NP2“ gekennzeichnet ist). Das dritte Merkmal, das für das Auflösungsverhalten relevant ist, ist dann die *Fortführung des Nebensatzes*. Aufgrund von pragmatischen Informationen ist eine der beiden Auflösungsmöglichkeiten vor dem Hintergrund unseres Weltwissens wahrscheinlicher oder plausibler. In Bezug auf seine Stärke ist dieses Merkmal ungefähr in der Mitte zwischen lexikalischem Bias und morpho-syntaktischer Gender-Kongruenz angesiedelt: Seine Verletzung führt zwar nicht direkt zu einem ungrammatischen Satz, jedoch zu einer markierten, meist sogar zu einer nicht-akzeptablen Struktur (s. *Lisa schüttelte Peter, weil sie schnarchte.*). Wie in der Tabelle zu sehen, gehen Gender- und Pragmatik-Informationen immer in die gleiche Richtung und werden nie gegeneinander ausgespielt. Denn zur Verarbeitung solcher Sätze wären dann u.U. sehr spezielle Verarbeitungsprozesse notwendig. Auch würde die Untersuchung durch die Verwendung von pragmatisch defektiven

Sätzen sehr unglaubwürdig, weil sich die Vpn dann auf gar keinen Auflösungs-Cue mehr verlassen könnten. Allerdings können Gender-Cue und Fortsetzung beide gegen oder mit der impliziten Kausalitätsrichtung des Verbs wirken. Dies ist die Manipulation des IC-Bias.

5.2.2 Voruntersuchung II: Plausibilität und Stimulusqualität

In einer zweiten Voruntersuchung sollte überprüft werden, ob sich die finalen Stimulusätze in ihrer Verständlichkeit und Plausibilität voneinander unterscheiden. Diese Prüfung ist deshalb entscheidend, weil zumindest einer der beiden experimentellen Faktoren, nämlich der IC-Cue, mit den beiden unterschiedlichen Itemfassungen konfundiert ist. Einige Autoren (z.B. Garnham, 2001) weisen berechtigterweise darauf hin, dass dies ein Einfalltor für systematische Verzerrungen ist. Denn wenn es ein Charakteristikum ic-inkompatibler Sätze ist, dass ihre Fortsetzung weniger zur Semantik des Verbs passt, dann erscheinen sie unplausibler und man weiß nicht mehr genau, warum diese Sätze schlechter bzw. langsamer verarbeitet werden: Lag es an der Inkompatibilität des IC-Cues oder an der Unplausibilität?

Allerdings ist die Annahme, dass eine kausalitäts-inkompatible Fortsetzung zwangsläufig zu weniger plausiblen oder weniger verständlichen Sätzen führt, nicht zwingend. Denn es ist klar, dass IC-Informationen immer durch andere sprachliche Informationen „überstimmt“ werden können: Wenn es keine Möglichkeit gäbe, durch zusätzliche sprachliche Informationen unserem impliziten oder expliziten Weltwissen zu widersprechen, dann ließen sich unplausible Sachverhalte sprachlich gar nicht ausdrücken. Deshalb muss eine kausalitätsinkongruente Fortführung, auch wenn sie unerwartet ist, nicht unbedingt unplausibel sein. Das bedeutet, dass eine ic-inkompatible Fortsetzung u.U. begründungsbedürftiger, voraussetzungsreicher und inferenz-intensiver ist, nicht aber „unplausibler“. Ic-inkompatible Sätze können aus mindestens zwei Gründen schwerer verarbeitet werden: Zum einen, weil bei diesen Sätzen mehr Informationen verarbeitet werden müssen oder aber kompliziertere Schlussfolgerungsprozesse notwendig sind. Zum anderen, weil es schwieriger ist, unter den strengen Bedingungen des Itemgenerierungsschemas eine ic-inkompatible Fortsetzung zu formulieren, die ähnlich „zwingend“ ist wie die kompatible Version.

Ersteres Problem ist gewissermaßen sachimmanent und lässt sich kaum vermeiden, wenn man die IC untersucht. Es lässt sich höchstens minimieren. In der Tat ist davon auszugehen, dass ein Großteil des IC-Effekts mehr oder weniger direkt auf solche oder ähnliche Verarbeitungsprozesse zurückgeführt werden kann, die zur „normalen“ Satzverarbeitung hinzutreten. Genau aus diesem Grund wurde ja behauptet, dass die implizite Verbkausalität ein sprachliches Merkmal ist, welches die erklärungs-basierte Verarbeitung beeinflusst. Proble-

matisch ist hingegen der zweite Fall, weil hier die Verarbeitungsschwierigkeit eben nicht sach-, sondern vielmehr methodenimmanent ist. Das heißt, u.U. läuft man durch bestimmte experimentelle Kontrollmaßnahmen Gefahr, sprachliche Artefakte zu generieren. Diese gilt es möglichst zu vermeiden, indem nur solche Fortführungen verwendet werden, bei denen die erschwerte Verarbeitung lediglich (oder zumindest größtenteils) auf die mit ic-inkompatiblen Nebensätzen verbundenen, zusätzlichen Verarbeitungsprozesse zurückgeführt werden kann.

Allerdings ist es schwierig, die beiden Aspekte empirisch trennscharf voneinander zu erfassen. Wenn man Personen z.B. fragt, wie „plausibel“, „natürlich“ oder „sinnhaft“ ein Satz ist, dann ist wahrscheinlich, dass sich in den Antworten beide Aspekte miteinander vermengen. Denn ic-inkompatible Fortsetzungen sind schwieriger als ic-kompatible. Und wenn man „plausibel“ im Sinne von „unerwartet“ versteht, dann stimmt auch dies, weil ja vom Standard-Attributionsmuster abgewichen wird. Aus diesen Gründen wurde zur Prüfung der „Qualität“ der ic-kompatiblen und -inkompatiblen Fortsetzungen nicht auf eine breite studentische Stichprobe zurückgegriffen, welche die Plausibilität der Sätze beurteilt (vgl. z.B. Stewart et al., 2000). Denn hier ist die Gefahr, dass beide Plausibilitäts-Aspekte miteinander konfundiert werden, sogar besonders hoch.

Stattdessen wurde in der vorliegenden Studie auf das Instrumentarium der Expertenbefragung zurückgegriffen: Zwei fortgeschrittene Studierende der Psychologie, die selbst experimentell zu sprachpsychologischen Themen forschen und zudem in ausführlichen Gesprächen über die spezifische Problematik bei der Erfassung der impliziten Verbkausalität informiert wurden, wurden gebeten, die Qualität der Satzfortsetzungen in einem Rating zu beurteilen. Obwohl sie beide damit – zwangsläufig – über das grundsätzliche Anliegen der vorliegenden Studie informiert waren, war ihnen unbekannt, zu welcher Verbklasse ein bestimmtes Verb gehörte und ob es sich bei dem zu beurteilenden Satz um die ic-kompatible oder -inkompatible Version handelte. Einer der Studierenden beurteilte die Fassung ohne zusätzliche Präpositionalphrase, der andere die Fassung mit Präpositionalphrase. Ich selbst habe zu beiden Fassungen ein zusätzliches, unabhängiges Rating angefertigt, sodass sich die Intersubjektivität der Beurteilungen quantifizieren und beurteilen lässt.

Das Vorgehen war dabei wie folgt: Den Studierenden wurden zwei Listen mit jeweils allen 160 Stimulussätzen vorgelegt. Bei allen Sätzen wurden die gleichen beiden männlichen Aktanten (*Peter* und *Franz*) verwendet, d.h. die Beurteilung erfolgte in Abwesenheit eines erleichternden Gender-Cues und war deshalb besonders schwer. Die erste Liste enthielt alle 160 Stimulussätze, bei denen sich das Pronomen auf das Subjekt bezog (was entsprechend indiziert war). In der zweiten Liste war der intendierte Referent hingegen der Hauptsatz-

Objekt. Beide Listen wurden nacheinander und unabhängig voneinander bearbeitet. Auf einer fünfstufigen Ratingskala (von 1: *sehr schlecht* über 3: *akzeptabel* bis 5: *sehr gut*) sollte global beurteilt werden, wie „gut“ die Fortsetzung im Sinne der zweiten oben explizierten, rein methodisch bedingten Problematik ist. Als relevante Aspekte zur Beurteilung der Fortsetzungsgüte wurden dabei folgende Punkte genannt: a) die „Eindeutigkeit“ der Fortsetzung (d.h. wie leicht sich bei der gewählten Fortsetzung auch die nicht-intendierte Referenz etablieren lässt), b) die „Stimmigkeit“ der Fortsetzung (d.h. wie gut die gewählte Begründung zur gewählten Auflösung passt) und c) die „Komplexität“ der Fortsetzung (d.h. wie komplex und voraussetzungsreich die zur Etablierung der intendierten Referenz notwendigen Schlussprozesse sind).

An dieser Explikation ist zu erkennen, dass wirklich nur eine Erfassung der methodischen Verzerrungsgefahr angestrebt wurde. Obwohl dadurch die Gefahr einer Konfundierung von allgemeiner Verarbeitungsschwierigkeit und schlechter Itemformulierung minimiert war, können die beiden Aspekte allerdings auch hier nicht vollständig voneinander getrennt werden. Ein Grund hierfür ist, dass beide Aspekte selbstverständlich miteinander korreliert sind, wenn auch nicht perfekt. Da es ausgesprochen schwer ist, den holistischen Gesamteindruck der „Passung“ der intendierten Fortsetzung auf bestimmte Teilaspekte zurückzuführen, wurde auch von einer separaten Erfassung der verschiedenen Beurteilungsaspekte abgesehen und nur ein hochinferentes Globalurteil abgefragt. Dieses sollte die verschiedenen Merkmale und ihre Interaktion bündeln und synthetisieren. Deshalb fällt es im konkreten Fall wahrscheinlich auch reliabler aus als die Summe von künstlich voneinander getrennten Einzelaspekten.

Dafür spricht zumindest, dass die Urteilsübereinstimmung insgesamt sehr gut ausfiel. Für die Items ohne zusätzliche Präpositionalphrase lag der Intraklassenkoeffizient (ICC) bei $r = .749$ (Subjekt-Referenz) bzw. $r = .764$ (Objekt-Referenz). Die Übereinstimmung bei den Items mit zusätzlicher Präpositionalphrase war sogar noch höher (Subjekt-Referenz: $r = .881$; Objekt-Referenz: $r = .850$). Zur Reliabilitätssteigerung wurden die Ratings anschließend aggregiert.

Zunächst einmal ist festzuhalten, dass die Ratings mit einem Gesamt-Mittelwert von $M = 3.83$ ($SE_M = .06$) in beiden Bedingungen stabil über dem Skalenmittelpunkt (3: *akzeptabel*) liegen. Das spricht dafür, dass die Itemgenerierung insgesamt erfolgreich war und die Fortsetzungen durchschnittlich als akzeptabel bis gut gewertet werden. In einem nächsten Schritt wurde dann überprüft, ob es einen systematischen Unterschied zwischen den Fortführungen mit und ohne Präpositionalphrase gibt. Tab. 5.4 zeigt die Werte der beiden Ratings für die beiden Bedingungen einzeln.

Präpositionalphrase	Auflösungsrichtung			
	Subjekt-Referenz		Objekt-Referenz	
	M	SE _M	M	SE _M
ohne PP	3.87	.06	3.74	.07
mit PP	3.83	.08	3.88	.07

Tab. 5.4: Ratings der Fortsetzung in Abhängigkeit von der Auflösungsrichtung und der Itemformulierung.

Ein messwiederholte *by-items*-Varianzanalyse mit *Itemformulierung* als Faktor (2: *mit* vs. *ohne Präpositionalphrase*) ergab keinen Unterschied für die Subjekt-Ratings ($F < 1$), wohl aber einen für die Objekt-Ratings ($F(1,159) = 4.09, p < .045, \eta_p^2 = .025$), der jedoch numerisch ausgesprochen klein ausfällt (Differenz von 0.14 Skalenpunkten). (Die Power dieser wie auch aller nachfolgenden Varianzanalysen war wiederum hoch genug, um selbst kleine Effekte ($f^2 = .02$) mit einer Wahrscheinlichkeit von $1-\beta = 5\%$ abzusichern.) Wie erwartet sind die Satzfortsetzungen mit Präpositionalphrase etwas eindeutiger – wahrscheinlich aufgrund der zusätzlichen Informationen, die dort gegeben werden. Obwohl also evtl. ein leichter Eindeutigkeitsvorteil durch die Einfügung der Präpositionalphrase zu verzeichnen ist, fällt dieser eher schwach aus. Ziemlich sicher ist jedoch, dass es durch die Einfügung der Präpositionalphrase zu keiner Schwächung der Fortsetzungsplausibilität kommt. Aufgrund des nur sehr geringen Effekts der Präpositionalphrase, wurden die Ratings beider Itemversionen nachfolgend aus Darstellungsgründen zusammengefasst (wobei hinreichende Homogenität gewährleistet ist: Subjekt-Referenz: $r(\text{ICC}) = .741$; Objekt-Referenz: $r(\text{ICC}) = .710$).

In einem weiteren Schritt wurde überprüft, ob es einen systematischen Unterschied zwischen ic-kompatiblen und -inkompatiblen Items gibt. Hierfür wurde eine zweifaktorielle *by-items*-Mischvarianzanalyse berechnet, mit den Faktoren *Referenz* (2: Subjekt vs. Objekt, *within*) und *Bias* (2: NP1 vs. NP2, *between*). Abb. 5.3 zeigt die entsprechenden Ergebnisse.

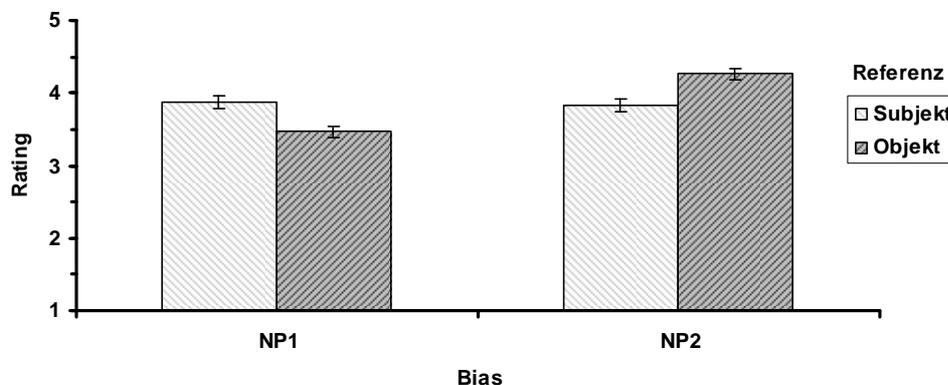


Abb. 5.3: Ratings der Fortsetzung in Abhängigkeit von der Auflösungsrichtung und dem IC-Bias.

Der Haupteffekt für den Faktor *Referenz* erwies sich als nicht signifikant ($F < 1$), aber es gab einen Haupteffekt für den Verb-Bias ($F(1,159) = 12.07, p < .005, \eta_p^2 = .071$): Verben mit NP1-Bias ($M = 3.67, SE_M = .06$) wurden grundsätzlich als etwas weniger passend beurteilt als Verben mit NP2-Bias ($M = 3.99, SE_M = .06$). Entscheidend ist jedoch v.a. die Interaktion der beiden Faktoren ($F(1,159) = 21.10, p < .001, \eta_p^2 = .118$): Während bei NP1-Verben die (ic-kongruenten) Subjekt-Fortsetzungen als passender beurteilt wurden als die Objekt-Fortsetzungen (Δ Rating: 0.40; $F(1,159) = 12.61, p < .001, \eta_p^2 = .074$), zeigte sich für die NP2-Verben genau das gegenteilige Muster (Δ Rating: -0.33; $F(1,159) = 8.67, p < .004, \eta_p^2 = .052$). Damit schlug sich in den Ratings deutlich der allgemeine IC-Effekt nieder.

Abschließend wurde überprüft, ob dieses Muster für Subjekt- und Objekt-Fortsetzungen unterschiedlich ausfällt. Hierfür wurde die Analyse um den Faktor *Verbtyp* (2: *Handlung* vs. *Zustand*) erweitert. Abb. 5.4 zeigt die entsprechenden Ergebnisse als Effekte des Referenz-Bias dargestellt. Positive Werte stehen für einen Subjekt-Bias, negative für einen Objekt-Bias.

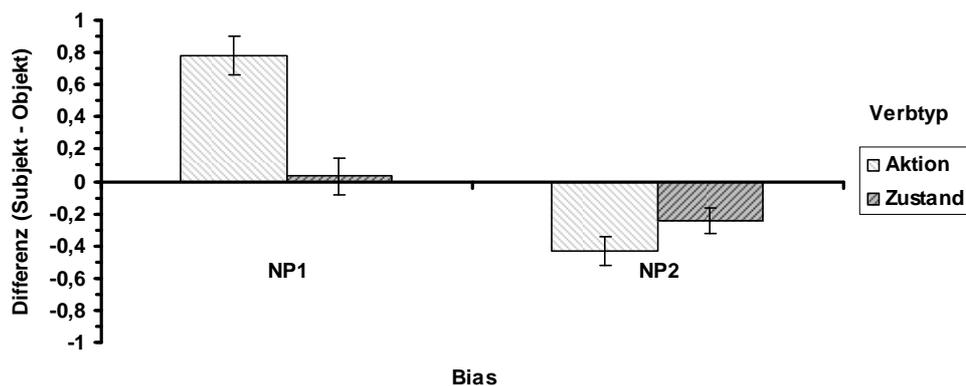


Abb. 5.4: Referenz-Bias in Abhängigkeit vom IC-Bias und dem Verbtyp.

Zunächst wurde das bestehende Ergebnismuster durch einem weiteren Haupteffekt für den Verbtyp ($F(1,159) = 16.49, p < .001, \eta_p^2 = .096$) ergänzt: Die Fortsetzungen für Zustandsverben ($M = 4.01, SE_M = .06$) wurden allgemein besser bewertet als die der Handlungsverben ($M = 3.66, SE_M = .06$). Die Interaktion der Faktoren *Verbtyp* und *Referenz* war knapp nicht signifikant ($F(1,159) = 3.49, p = .066$), dafür aber die Dreifachinteraktion ($F(1,159) = 9.09, p < .003, \eta_p^2 = .055$): Geplante Einzelvergleiche zeigten, dass der Unterschied zwischen ic-kongruenten und -inkongruenten Items lediglich für die Handlungsverben besteht (AP: Δ Rating: 0.78; $F(1,159) = 25.25, p < .001, \eta_p^2 = .139$; AE: Δ Rating: -0.43; $F(1,159) = 7.47, p < .007, \eta_p^2 = .046$), nicht jedoch für die Zustandsverben (SE: Δ Rating: 0.03; $F(1,159) < 1, p = .872, \eta_p^2 < .001$; ES: Δ Rating: -0.24; $F(1,159) = 2.46, p = .119, \eta_p^2 = .016$).

Das Ergebnismuster lässt sich so zusammenfassen: Trotz aller Bemühungen, möglichst ausgeglichene Satzfortsetzungen herzustellen, gibt es einen Unterschied in den Ratings für ic-kompatible und -inkompatible Items. Dies gilt v.a. für die Handlungsverben; bei den Zustandsverben werden die Fortsetzungen ähnlicher beurteilt. Allerdings ist offen, wie man diesen Befund interpretieren sollte: Denn auch wenn klar konstatiert werden muss, dass nicht ausgeschlossen werden kann, dass die Effekte des IC-Cues auf Unterschiede in der Plausibilität und Verständlichkeit zurückführbar sind, so ist dies doch eher unwahrscheinlich.

Denn zunächst ist festzuhalten, dass die Unterschiede numerisch vergleichsweise klein ausfallen: Der mittlere Unterschied zwischen den Subjekt- und den Objekt-Ratings liegt bei ca. 0.4 Skalenpunkten, umfasst also ca. 10 % der Gesamtskala. Darüber hinaus ist unklar, ob ein Teil des beobachteten Plausibilitätsunterschieds nicht auf die allgemeine Verarbeitungsschwierigkeiten ic-inkompatibler Verben zurückgeführt werden muss. Auch wenn durch Art und Aufbau der Untersuchung eine Konfundierung der beiden eingangs definierten Aspekte der Plausibilitätsbeurteilung möglichst vermieden werden sollte, ist nicht auszuschließen, dass diese durch den holistischen Charakter des Ratings doch in das Urteil eingeflossen ist.

Darüber hinaus liegen die Mittelwerte für die einzelnen Bedingungen immer stabil im akzeptablen bis guten Bereich. Um zu überprüfen, dass dies nicht nur für die Aggregat-Ebene, sondern auch für die Ebene der einzelnen Items gilt, wurde eine Reanalyse der Ergebnisse durchgeführt. Dabei wurden die Items danach aufgeteilt, ob sie einen akzeptablen Ratingwert (> 3) aufwiesen oder nicht. Tab. 5.5 zeigt die Anzahl akzeptabler Bewertungen der Subjekt- und der Objekt-Ratings für die vier Itemtypen. Lediglich für die Subjekt-Ratings der AP-Verben ist die Anzahl von nicht-akzeptablen Ratings ($N = 12$) bedenklich. Zwei separate χ^2 -Test für die Subjekt- und die Objekt-Ratings ergeben jedoch keine signifikanten Unterschiede zwischen den vier Verbtypen (beide $\chi^2(1) < 1$).

Rating			Verbtyp		Gesamt
			Handlung	Zustand	
Subjekt	Bias	NP1	39	36	75
		NP2	32	36	68
	Gesamt		71	72	143
Objekt	Bias	NP1	28	39	67
		NP2	36	38	74
	Gesamt		64	77	141

Abb. 5.5: Anzahl akzeptabel Items für die vier Verbtypen.

Die Werte machen ebenfalls deutlich, dass der Anteil als akzeptabel bewerteter Items in allen vier Bedingungen ausgesprochen hoch ist. Für Zustandsverben gibt es überhaupt keine Unterschiede, bei den Handlungsverben sind v.a. die Objekt-Ratings der AP-Verben problematisch.

6. Experiment Ia

6.1 Einleitung

Experiment I verfolgte mehrere Zielsetzungen gleichzeitig. Zunächst sollte sichergestellt werden, dass sich mit dem neu konstruierten Stimulusmaterial und der eingesetzten Methodik überhaupt das in der Literatur hinlänglich bekannte Effektmuster zeigt, wie es für den IC- und den Gender-Effekt typisch ist. Das ist nicht unbedingt selbstverständlich, da – wie in Kap. 2 ausführlich dargelegt – für beide Cues in der Literatur unterschiedliche Effektkonstellationen berichtet werden. Dies betrifft v.a. die Interaktion der beiden Cues. Die divergierenden Ergebnismuster können zwar häufig auf den eingesetzten Aufgabentyp und das generelle Untersuchungs-Setting zurückgeführt werden, allerdings berichten auch Studien, die einen sehr ähnlichen Aufbau haben, manchmal leicht divergierende Ergebnisse.

Das eingesetzte Stimulusmaterial unterschied sich zwar in seinem Aufbau nur unerheblich von anderen Studien zum IC-Effekt, benutzte aber – und das ist die eigentliche Besonderheit – die deutsche Sprache. Meines Wissens liegen noch keine online-Experimente zum IC-Effekt für das Deutsche vor, sondern lediglich offline-Studien zur Kausalattribution bei IC-Verben (vgl. z.B. Fiedler & Semin, 1988; Rudolph & Försterling, 1997b; Rudolph & von Hecker, 1997; 2006) und eine eher informell durchgeführte EKP-Studie aus der Linguistik (Härtel, 2001). Es ist jedoch nicht von vorneherein klar, dass sich die v.a. beim Englischen nachgewiesenen online-Effekte problemlos auf das Deutsche übertragen lassen. Denn wie bereits mehrfach angesprochen, ist das Deutsche im Vergleich zum Englischen eine morphologisch wesentlich reichhaltigere Sprache, was sich v.a. auf die Verarbeitung von Genus-Informationen auswirken dürfte. Ein anderer struktureller Unterschied, der für die online-Verarbeitung bedeutsam ist, besteht darin, dass das Deutsche im Gegensatz zum Englischen eine SOV-Sprache ist, bei der in Nebensätzen das Verb satzfinal steht. Dies ist v.a. für den IC-Effekt relevant, weil der IC-Kongruenz-Effekt im Wesentlichen aus der fehlenden Passung von Hauptsatz- und Nebensatz-Verb resultiert (vgl. Kap. 2). Wenn dieses erst später im Satz steht, dann kann es erst entsprechend spät zu einer Disambiguierung kommen.

Da bereits das Stimulusmaterial von den vorliegenden Studien abwich, wurde in Bezug auf die eingesetzte Aufgabe ein größtmöglicher Anschluss gewählt. Speziell fiel die Wahl auf die bereits in der Originalstudie von Caramazza et al. (1977) eingesetzte Auflösungsaufgabe. Allerdings sollten die Vpn den Referenten des Pronomens nicht dadurch identifizieren, dass sie seinen Namen aussprachen (Naming-Aufgabe), sondern dadurch, dass sie eine korrespondierende Taste drückten. Auflösungsaufgaben liefern lediglich ein globales Maß der Auflö-

sungsschwierigkeit und erlauben keine temporale Differenzierung der Effekte im Laufe der Satzverarbeitung. Sie vermitteln eher einen holistischen Gesamteinblick in den Auflösungsprozess und bieten sich als Einstieg in eine weiterführende Prozessanalyse besonders an. Ergänzend zur pronominalen Auflösungsaufgabe wurde eine Verifikationsaufgabe eingesetzt. Das heißt, im Anschluss an die Satzrezeption wurde eine Aussage präsentiert und die Vpn mussten entscheiden, ob diese mit dem zuvor dargebotenen Satz übereinstimmt oder nicht. Ein Vergleich dieser beiden Aufgabentypen ist interessant, weil er Aufschluss über die Persistenz der Effekte gibt, die evtl. über die eigentliche Auflösung hinauswirken.

Vonk (1985b) hatte für die beiden Aufgaben divergierende Ergebnismuster gefunden, die sie auf unterschiedliche Aufgabencharakteristika zurückführte. Allerdings ist ein direkter Vergleich mit den Ergebnissen von Vonk nicht möglich. Bei Vonk wurde die Verifikationsaufgabe einzeln eingesetzt, während sie in der vorliegenden Studie nach der Auflösung zu bearbeiten war. Es ist davon auszugehen, dass sie in diesem Fall nicht mehr Aspekte des Auflösungsprozesses selbst abbildet, sondern eher eine Phase der Verstehenskonsolidierung. Für diese sollten v.a. hierarchie-hohe Integrations-Prozesse relevant sein, während sich gedächtnis-basierte Fokus-Prozesse hier nicht mehr auswirken können.

Um eine möglichst große Nähe zur Untersuchung von Caramazza et al. (1977) herzustellen, wurde auch das Stimulusmaterial angepasst. Denn Caramazza et al. verwendeten sehr kurze Satzergänzungen. Das heißt, nach dem Pronomen folgten lediglich 3-4 Wörter, die den Satz disambiguierten und entweder eine kausalitäts-kongruente oder -inkongruente Lesart etablierten (*Ronald scolded Joue because he was annoying; Helen punished Cathy because she confessed to shoplifting* etc.). Es ist jedoch davon auszugehen, dass die Länge des Folgesatzes einen Einfluss auf die Auflösungsdynamik hat (Kap. 7). Um keine überflüssigen Abweichungen zum ursprünglichen Stimulusmaterial einzuführen, wurde in Experiment Ia die Stimulusversion ohne Präpositionalphrase verwendet (vgl. Kap. 5).

Eine weitere Neuerung von Experiment Ia war der Einsatz der RT-Verteilungsanalyse als Auswertungsverfahren für die Auflösungs- bzw. Antwortzeiten. Wie in Kap. 3 dargestellt, standen mit den verschiedenen Parametern der Weibull-Verteilung differenzielle Maße zur Erfassung hierarchie-niedriger, gedächtnis-basierter und hierarchie-hoher, erklärungs-basierter Prozesse zur Verfügung. Durch ihre Betrachtung lassen sich differenzierte Aussagen zur prozessualen Zusammensetzung des IC- und des Gender-Effekts machen. Auch deswegen ist es wichtig, die eingesetzte Aufgabe möglichst konstant zu halten. Denn dann lassen sich die über die Analyse mittlerer RTn hinausgehenden Ergebnisse besser interpretieren. Auch ist

wichtig, die RT-Verteilungsanalyse zunächst an einer datenanalytisch vergleichsweise einfache Globalaufgabe zu erproben.

Durch die separate Erfassung gedächtnis- und erklärungs-basierter Prozesse ist es dann möglich, spezifische Interaktionen von IC- und Gender-Cue zu prüfen. Denn in Kap. 2 wurde dafür argumentiert, dass die herkömmliche Null-Interaktion dadurch zustande kommt, dass unterschiedliche Teilprozesse des IC-Effekts differenziell mit dem Gender-Cue interagieren. Nur wenn ein eindeutiger Gender-Cue vorliegt, können gedächtnis-basierte IC-Fokus-Effekte auftreten. Gleichzeitig wird der IC-Integrations-Effekt in dieser Bedingung schwächer ausfallen, weil die gedächtnis-basierten Prozesse in diesem Fall den finalen Integrationsaufwand reduzieren. Wenn hingegen keine eindeutigen Gender-Information vorhanden sind, die das Pronomen disambiguieren, dann kann keine frühe gedächtnis-basierte Verarbeitung stattfinden. Stattdessen muss die Kongruenz allein satzfinal hergestellt werden, wodurch der IC-Integrations-Effekt in dieser Bedingung stärker ausfällt.

In beiden Fällen stehen die beiden Teilprozesse in einer trade-off-Relation zueinander und mitteln sich in einer undifferenzierten Analyse gegenseitig aus: In beiden Fällen kommt es zu Verarbeitungsschwierigkeiten, allerdings basieren diese auf jeweils unterschiedlichen Prozessen. Erst durch eine unabhängige Erfassung der beiden Verarbeitungsprozesse durch die RT-Verteilungsanalyse lässt sich diese Hypothese prüfen. Das beschriebene Ergebnismuster wäre eine starke Evidenz für eine dynamische Interaktion gedächtnis- und erklärungs-basierter Prozesse während der pronominalen Auflösung.

Zusammenfassend ergaben sich damit folgende Vorhersagen für die vorliegende Studie. In Bezug auf die traditionelle Auswertung wurde erwartet, dass sich das gleiche bzw. ein ähnliches Ergebnismuster ergibt wie für andere Studien zum IC-Effekt, die sich einer pronominalen Auflösungsaufgabe bedienen (insbesondere Caramazza et al., 1977, und Vonk, 1984). Dies gilt sowohl für das Verstehensprodukt, also die Akkuratheit der pronominalen Auflösung, als auch für den Verstehensprozess, also die Zeit, die zur Auflösung benötigt wird. Durch die differenzierende RT-Verteilungsanalyse kann diese Vorhersagemuster noch weiter präzisiert werden. Dies macht Tab. 6.1 deutlich, die bereits am Ende von Kap. 2 erläutert wurde und hier nur noch einmal abgebildet ist:

<i>Effekt</i>	<i>Teileffekt</i>	<i>Parameter</i>	<i>Interaktion Gender-Cue</i>
Gender	Confirmation	Shift	---
	Disengagement	Scale	---
IC	Fokus	Shift	Stärker, wenn disambiguiert
	Integration	Scale	Stärker, wenn nicht disambiguiert

Tab. 6.1: Vorhersagen für die verschiedenen Teileffekte des Gender- und des IC-Manipulation.

Bezüglich des Gender-Cues wurde erwartet, dass sowohl hierarchie-hohe als auch hierarchie-niedrige Prozesse an ihm beteiligt sind und sich zwei deutliche Teileffekte unterscheiden lassen, nämlich einmal ein primär gedächtnis-basierter *confirmation*-Effekt und zum anderen ein erklärungs-basierter *disengagement*-Effekt. Beim IC-Cue ist dies tendenziell ähnlich, da hier ein primär gedächtnis-basierter IC-Fokus-Effekt und ein erklärungs-basierter IC-Integrations-Effekt unterschieden werden. Allerdings sollten beide Effekte nicht gleichzeitig auftreten, da erwartet wird, dass der IC-Fokus-Effekt nur auftritt, wenn das Pronomen gleichzeitig gender-disambiguiert ist. Der IC-Integrations-Effekt ist zwar in beiden Gender-Bedingungen vorhanden, sollte aber im nicht-disambiguierten Fall stärker ausfallen. darüber hinaus sollten die hierarchie-hohen, nicht jedoch die hierarchie-niedrigen Effekte zusätzlich in der nachfolgenden Verifikationsaufgabe persistieren. Da Experiment I auf einer globalen Auflösungsaufgabe basiert, lässt sich hier noch keine weitergehende Unterscheidung zwischen frühen und späten Prozessen vornehmen.

6.2 Methode

6.2.1 Stichprobenbeschreibung und -bereinigung

Allgemeine Angaben. An Experiment Ia nahmen insgesamt $N = 24$ Versuchspersonen (16 Frauen, 8 Männer) teil. Das durchschnittliche Alter lag bei $M = 24.6$ Jahren ($SD = 4.7$). Es handelte sich hauptsächlich um Studierende der Psychologie (durchschnittlich im 6. Semester), die für ihre Teilnahme eine Versuchspersonenstunde oder eine Vergütung von 7,-€ erhielten. Alle Teilnehmer/innen berichteten, dass sie normal- oder korrigiert normalsichtig seien, und gaben Deutsch als Muttersprache an.

ELVES. Zur Überprüfung des allgemeinen Lesefähigkeits-Niveaus der Stichprobe und zur Stichprobenbereinigung wurden die drei hierarchieniedrigen Subtests *Satzverifikation*, *Sinnhaftigkeit Sätze* und *Sinnhaftigkeit Satzfolgen* des Lesefähigkeitsdiagnostikums ELVES erhoben (vgl. Richter & van Holt, 2005). Die Stichprobe wies für alle eingesetzten Subtests Werte auf, wie sie für studentische Stichproben typisch sind. Die entsprechenden Mittelwerte

lagen für den Subtest *Satzverifikation* bei $M = 16.6$ ($SD = 4.0$), für den Subtest *Sinnhaftigkeit Sätze* bei $M = 16.0$ ($SD = 4.1$) und für den Subtest *Sinnhaftigkeit Satzfolgen* bei $M = 15.9$ ($SD = 4.4$). Die drei Subskalen korrelierten hoch miteinander und wurden deswegen zu einem gemeinsamen Index zusammengefasst (Cronbachs $\alpha = .94$; $M = 15.9$, $SD = 3.9$). Eine Versuchsperson wies für alle drei ELVES-Subskalen einen erhöhten multivariaten Leverage-Wert auf ($h_{ii} = .61$, $p < .01$; Cohen, Cohen, West & Aiken, 2003, ch. 10), der darauf zurückgeführt werden konnte, dass sie insgesamt überdurchschnittlich abschnitt. Da ihr Antwortmuster jedoch sonst keine Auffälligkeiten aufwies, wurde die Vpn nicht von der Analyse ausgeschlossen.

Geschwindigkeit und Akkuratheit. Die mittlere Reaktionszeit und durchschnittliche Akkuratheit in den vier Untersuchungsbedingungen wurden pro Person zu einem Vektor mit insgesamt acht Werten zusammengefasst, für den dann ein multivariater Leverage-Wert berechnet wurde. Zwei Versuchspersonen erwiesen sich dabei als auffällig ($h_{ii} > .7$, $p < .01$). Eine Inspektion ihrer Antwortmuster zeigte, dass sie sehr niedrige Akkuratheits-Werte um 50 % hatten (auch für die Sätze mit einer grammatikalisch eindeutigen Lesart!). Da sie die Untersuchung offensichtlich instruktionswidrig bearbeitet hatten, wurden sie von allen Analysen ausgeschlossen.

6.2.2 Material

Sätze zur pronominalen Auflösung. Das in allen Experimenten eingesetzte Stimulusmaterial und seine Validierung wurden bereits ausführlich in Kap. 5 besprochen. In Experiment Ia wurde dabei die Fassung ohne Präpositionalphrase verwendet, d.h. alle Sätze folgten dem in (1) dargestellten Schema:

(1a) Schema: [NP1 IC-Verb NP2]_{HS-VP}, [weil] [Pro] [[(Det) (Adj/Adv) N]_{NS-NP} V (Mod) (Aux)]_{NS-VP}

(1b) Bsp.: [Dominik_{NP1} nervte_{IC-V} Daniela_{NP2}]_{HS-VP}, weil_{Comp} er_{Pro} [keine_{Det} Ruhe_N geben_V wollte_{Mod}]_{NS-VP}

Es wurden alle 160 Stimulussätze verwendet. Sie wurden isoliert dargeboten und waren nicht in einen satzübergreifenden Diskurskontext eingebunden. Es wurden keine weiteren Filler-Items eingesetzt. Achtzig der verwendeten Verben waren NP1- und 80 waren NP2-Verben. Innerhalb dieser beiden Klassen war wiederum jeweils die eine Hälfte Zustands- und die andere Hälfte Aktionsverben. Insgesamt gab es also jeweils 40 Verben von den vier Typen SE, ES, AP und AE (vgl. Kap. 5). Die 40 Stimuli einer Verbkategorie wurden per Zufall in vier separate Listen aufgeteilt, die wiederum einer der vier Untersuchungsbedingungen zugewiesen wurden (vgl. (2a-d) für das in (1) gegebene Beispiel).

- (2a) GenMinKonPlus: Dominik nervte Ralf, weil er keine Ruhe geben wollte.
- (2b) GenMinKonMin: Dominik nervte Ralf, weil er das Thema leid war.
- (2c) GenPlusKonPlus: Dominik nervte Daniela, weil er keine Ruhe geben wollte.
- (2d) GenPlusKonMin: Ulla nervte Ralf, weil er das Thema leid war.

Die Zuweisung der Listen zu den vier Untersuchungsbedingungen wurde dabei durch ein Lateinisches Quadrat ausbalanciert und über einen vierstufigen Listen-Faktor gesteuert, der zwischen den Versuchspersonen variiert wurde. Jede Vpn sah also jedes Item nur einmal und kein Item wurde in mehr als einer Untersuchungsbedingung dargeboten.

Verifikationsaussagen. Zu jedem Satz wurde eine Aussage konstruiert, die von den Versuchspersonen nach der Auflösung des Pronomens zu verifizieren war. Bei diesen Aussagen handelte es sich um eine reduzierte und umgestellte Fassung des zuvor dargebotenen Originalsatzes, bei dem die im Nebensatz gegebene Verbalphrase mit einem der beiden Aktanten als Subjekt kombiniert wurde. Die Verifikationsaussagen folgten dem in (3) gegebenen Schema und konnten durch die Einsetzung des „korrekten“ oder „inkorrekten“ Aktanten entweder richtig oder falsch sein.

- (3a) Schema: [Frage-NP] [Frage-VP].
- (3b) Dominik wollte keine Ruhe geben. (GenMinKonPlus: korrekte Version)
- (3c) Ralf wollte keine Ruhe geben. (GenMinKonPlus: inkorrekte Version)

Da die Verbalphrase des Auflösungs-Satzes und der zu verifizierenden Aussage für jede Vpn jeweils identisch waren, handelte es sich im Prinzip um eine pronominal aufgelöste Variante des Nebensatzes, die entweder mit der intendierten Lesart übereinstimmte oder nicht.

Jede Vpn bearbeitete alle 160 Verifikationsaussagen, wobei diese gleichteilig in der korrekten und der inkorrekten Version präsentiert wurden. Orthogonal hierzu wurde in gleich vielen Sätzen der erste und der zweite Aktant als Subjekt verwendet. Die verschiedenen Versionen der Verifikationsaussage wurden mittels eines einfachen Lateinischen Quadrats mit den vier Itemversionen koordiniert und durch den Listen-Faktor den vier Itemlisten zugewiesen.

6.2.3 Untersuchungsablauf

Die Untersuchung wurde in Kleingruppen von zwei bis fünf Personen computergestützt durchgeführt. Als Versuchssteuerungssoftware wurde das Programm *E-Prime 1.1* verwendet (Schneider, Eschman & Zuccolotto, 2002), das auf PCs betrieben wurde, an die ein 17-Zoll-Monitor angeschlossen war (Auflösung: 1024 x 768).

Die Vpn wurden vom Versuchsleiter begrüßt und an nahmen an durch Trennwände voneinander abgegrenzten Einzelarbeitsplätzen Platz. Sie wurden instruiert, die nachfolgenden Sätze so schnell wie möglich zu lesen und anzugeben, welcher der beiden im Hauptsatz genannten Aktanten mit dem Pronomen des Nebensatzes gemeint ist. Ihnen standen dafür die beiden Antworttasten „D“ und „K“ auf einer Standard-Computertastatur zur Verfügung, wobei die Taste „D“ jeweils dem ersten Aktanten (Subjekt) des Hauptsatzes entsprach und die Taste „K“ dem zweiten Aktanten in der Objektposition. Die Zuordnung der beiden Aktanten zu den Antworttasten war jeweils auf dem Bildschirm nochmals angegeben. Die Vpn wurden angewiesen, die Sätze so schnell wie möglich zu lesen. Sie sollten den Satz auch verstanden haben, da dies im Anschluss an jeden Satz durch eine weitere Aufgabe überprüft werde.

Der Ablauf war dabei für jedes Item so aufgebaut, dass der Satz zunächst durch einen Druck auf die Leertaste angefordert werden musste. Darauf erschien zunächst für 500 ms ein Fixationskreuz, und danach der Satz in schwarzen Buchstaben (Courier, Fett, 18pt) in der Mitte eines dunkelgrünen Bildschirms in einem Fenster mit weißem Hintergrund. Der Textfluss war so umgebrochen, dass Haupt- und Nebensatz jeweils in einer eigenen Zeile, aber gleichzeitig erschienen. Sobald die Vp eine der beiden Antworttasten drückte, verschwand der gesamte Satz. Um eine möglichst zügige Bearbeitung der Aufgabe zu gewährleisten, wurde bei Lesezeiten über 5 s ein roter Bildschirm von 1 s Dauer gezeigt, der die Vp aufforderte, den Satz schneller zu bearbeiten. Falls eine Vp „falsch“ geantwortet hatte, wurde sie darüber hinaus auf einem weiteren Bildschirm von 1 s daran erinnert, möglichst keine Fehler zu machen. Anschließend wurde dann die zu verifizierende Aussage in dem gleichen Fenster dargeboten wie der Originalsatz. Auch hier verschwand die Aussage nach Abgabe der Antwort und die Vpn wurden bei Bearbeitungszeiten über 2 s bzw. bei falscher Beantwortung daran erinnert, die Aufgabe schneller und/oder gründlicher zu bearbeiten.

Die Vpn hatten die Möglichkeit, sich an diesen Ablauf durch die Bearbeitung von vier Beispielitems zu gewöhnen. Durch einen Erinnerungs-Screen abgetrennt, folgte dann der Hauptdurchgang, der mit 20 Eisbrecher-Items begann, die nicht ausgewertet wurden, sondern lediglich die Aufgabe hatten, evtl. Übungseffekte (die üblicherweise am Anfang einer Untersuchung besonders stark ausfallen) zu minimieren. Direkt im Anschluss und ohne dass den Vpn der Übergang kenntlich gemacht wurde, wurden dann die 160 Experimentalitems in zufälliger Reihenfolge ohne Pause dargeboten. Im Anschluss bearbeiteten die Vpn die drei Subtests *Satzverifikation*, *Sinnhaftigkeit Sätze* und *Sinnhaftigkeit Satzfolgen* des Lesefähigkeits-Diagnostikums *ELVES* und beantworteten noch einige allgemeine Fragen zu ihrer Person. Insgesamt dauerte die Untersuchung ca. 45 Minuten.

6.2.4 Design und Abhängige Variablen

Das „Kerndesign“ der Untersuchung war ein $2(\text{Gender-Cue: vorhanden vs. nicht vorhanden}) \times 2(\text{IC-Cue: vorhanden vs. nicht vorhanden})$ -Design. Beide Faktoren waren messwiederholt, sowohl über Versuchspersonen (da jede Vpn jede Stimuluskombination bearbeitete) als auch über Items (da jedes Item in jeder Stimuluskombination präsentiert wurde). Alle Merkmale der einzelnen Stimulustypen (z.B. NP1- vs. NP2-Verb, Zustands- vs. Aktionsverb, Genus-Version etc.) waren nur messwiederholt über Versuchspersonen, nicht aber über Items. Das einzige Merkmal, das zwischen den Vpn variiert wurde, war der Listen-Faktor, der steuerte, in welcher Untersuchungsbedingung welche Items präsentiert wurden. Da sowohl Itemmerkmale als auch Listen-Faktor für die vorliegende Fragestellung von vergleichsweise randständigem Interesse sind, fungieren sie lediglich als Kontrollfaktoren, und werden nicht berichtet.

Die zentrale Abhängige Variable war die zur Auflösung des Pronomens benötigte Antwortlatenz bzw. – genauer – die Verteilung dieser Antwortlatenzen. Die „Akkuratheit“ der gewählten Auflösung wurde auch aufgezeichnet; sie war jedoch eher sekundär zur Beurteilung der Qualität der abgegebenen Reaktionszeit wichtig und um den *speed-accuracy*-Tradeoff einschätzen zu können. Bei der Verifikationsaufgabe wurden ebenfalls die Akkuratheit und die Geschwindigkeit der Beantwortung erhoben. Die Verifikationsaufgabe war jedoch insgesamt von eher nachgeordnetem Interesse, da ihr Einsatz v.a. durchführungstechnisch motiviert war. Allerdings konnte durch sie überprüft werden, ob die hierarchie-hohen Teilprozesse auch in einer nachgeschalteten Aufgabe zur Verstehenskonsolidierung persistieren.

6.3 Ergebnisse

6.3.1 Pronominale Auflösungsaufgabe

6.3.1.1 Akkuratheit: Anteil intendierter pronominaler Auflösungen

Ergebnisse. Abb. 6.1 zeigt die „Akkuratheit“ (Anteil von Antworten in der intendierten Lesart) der von den Vpn gewählten pronominalen Auflösung für die vier Untersuchungsbedingungen.

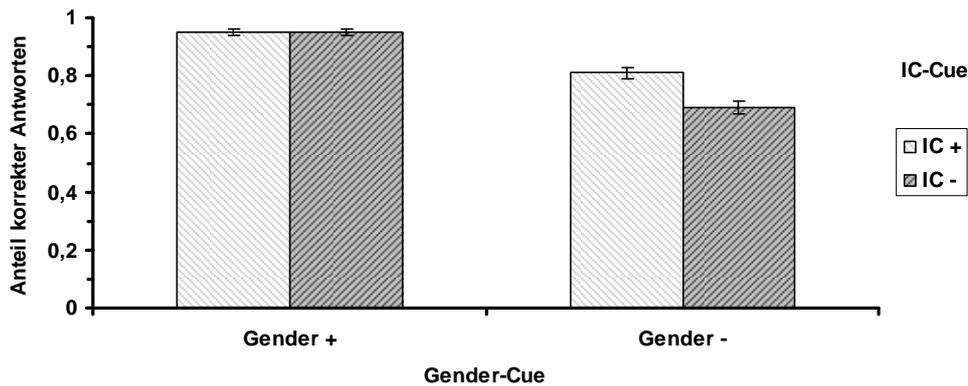


Abb. 6.1: Anteil intendierter pronominaler Auflösungen in den vier Untersuchungsbedingungen.

Eine messwiederholte ANOVA mit den Faktoren *Gender-Cue* (2: *vorhanden* vs. *nicht vorhanden*) und *IC-Cue* (2: *vorhanden* vs. *nicht-vorhanden*) ergab zunächst – wie nicht anders zu erwarten – einen sehr deutlichen Unterschied zwischen gender-disambiguierten und nicht-gender-disambiguierten Sätzen ($F(1,21) = 223.36, p < .001, \eta_p^2 = .914$; $F(1,159) = 117.89, p < .001, \eta_p^2 = .426$), da die disambiguierten Sätze wegen ihrer grammatikalisch eindeutig festgelegten Lesart deutlich „korrekter“ beantwortet wurden. Darüber hinaus gab es auch einen deutlichen Haupteffekt für die implizite Verbkausalität ($F(1,21) = 35.32, p < .001, \eta_p^2 = .627$; $F(1,159) = 142.55, p < .001, \eta_p^2 = .473$), der jedoch durch den signifikanten Interaktionsterm ($F(1,21) = 23.07, p < .001, \eta_p^2 = .523$; $F(1,159) = 119.77, p < .001, \eta_p^2 = .430$) noch weiter qualifiziert wurde: Wie in Abb. 6.1 deutlich zu erkennen, wirkte sich die Verbkausalität nur dann auf das Verstehensprodukt aus, wenn der Gender-Cue beide Interpretationen überhaupt grammatikalisch zuließ ($F(1,21) = 34.84, p < .001, \eta_p^2 = .624$; $F(1,159) = 168.64, p < .001, \eta_p^2 = .515$). Wenn die Auflösung eines Pronomens jedoch bereits durch den Gender-Cue festgelegt wurde, dann gab es keinen Kausalitäts-Effekt (beide $F < 1$).

Arcussinus-Transformation. Um zu überprüfen, ob das berichtete Ergebnismuster evtl. auf eine Verletzung der Anwendungsvoraussetzungen der ANOVA zurückgeführt werden kann, wurde eine zusätzliche Analyse mit arcussinus-transformierten Hit-Rates durchgeführt. Eine Verletzung der Normalverteilungsvoraussetzung der Residuen ist bei der Verwendung von Anteilswerten praktisch zwangsläufig gegeben, weil diese definitionsgemäß durch die Werte 0 und 1 zweiseitig begrenzt sind. Das Ergebnismuster dieser Analyse stimmt vollständig mit dem eben berichteten überein, weswegen es hier nicht berichtet wird (vgl. Appendix C.1.1).

Zusammenfassung. Für die die Variable, die das konkrete Produkt des Verstehensprozesses abbildet, nämlich die gewählte pronominaler Auflösung, ergab sich ein sehr eindeutiges Bild: Wenn für das Pronomen im Nebensatz nur ein grammatikalisch korrekter Referent vor-

handen ist, dann wurde dieser mit hoher Sicherheit (von ca. 95%) auch gewählt. Wenn jedoch beide im Hauptsatz genannten Personen als potenzielle Referenten in Frage kamen, dann erwies sich die IC als ein relevanter Einflussfaktor: Stimmt die im Nebensatz gegebene Begründung mit der durch das Verb induzierten Auflösungserwartung überein, dann wurde zu 85 % die „korrekte“ (d.h. intendierte) Auflösung gewählt. Wenn hingegen die im Nebensatz gegebene Begründung erwartungskonträr war, dann sank die Trefferrate um weitere 10 % ab.

Obwohl dieses Effektmuster sowohl mit den entsprechenden Werten der Voruntersuchung als auch mit denen anderer IC-Studien mit einer Auflösungsaufgabe (Caramazza et al., 1977; Ehrlich; 1980; Vonk, 1984) übereinstimmte, ist auffällig, dass die „Fehler“-Rate in den nicht-gender-disambiguierten Untersuchungsbedingungen (20 %) doch substantiell war. Sowohl Caramazza et al. (1977) als auch Vonk (1984; 1985b; hier wurden allerdings ausschließlich kausalitätskongruente Sätze vom Typ IC+ verwendet) berichteten niedrigere Fehlerraten (um die 10 %), auch für die ambigen Sätze. In der Studie von Ehrlich (1980) war die mittlere Fehlerrate (ebenfalls nur für kausalitätskongruente Sätze!) sogar höher (22 % für die *weil*-Konjunktion und 37 % für die *aber*-Konjunktion). Bei Stewart et al. (2000, Exp. 2) fanden sich ähnlich hohe Werte (22 % in der kongruenten und 35 % in der inkongruenten Bedingung).

Obwohl die hier berichteten „Fehler“-Raten also noch vergleichsweise moderat ausfallen, sollen trotzdem einige Vorsichtsmaßnahmen getroffen werden, um sicher zu gehen, dass die nachfolgend berichteten Reaktionszeitergebnisse nicht artifiziell verzerrt sind. Dies könnte dadurch zustande kommen, dass durch die Einbeziehung auch der falsch beantworteten Items zusätzliche, aufgabenfremde Bearbeitungsstrategien überproportional in den Daten vertreten sind. Alle Analysen werden deshalb im Folgenden doppelt durchgeführt: Einmal mit allen Antworten – egal ob falsch oder richtig –, um eine möglichst große Datenmenge für die RT-Verteilungsanalyse zur Verfügung zu haben; das andere mal sollen jedoch nur die „korrekten“ Antworten in die Analyse einbezogen werden, um zu überprüfen, ob sich das Antwortmuster durch den Ausschluss dieser Datenpunkte verändert.

6.3.1.2 Mittlere Reaktionszeit

In Kap. 3 wurde dafür argumentiert, dass eine Betrachtung allein der mittleren RT ausgesprochen irreführend sein kann. Dennoch soll im Folgenden dargestellt werden, welche Ergebnisse sich bei einer konventionellen (d.h. rein mittelwerts-basiert) Auswertung ergeben.

Datenbereinigung. Es wurde nur eine sehr sparsame Bereinigung auf Ebene der Einzelantworten vorgenommen. Als unterer Cut-off-Wert wurde für die pronominale Auflösung ein

Wert von 1000 ms angesetzt, da es sehr unwahrscheinlich ist, dass Items mit Antwortlatenzen unter diesem Wert adäquat bearbeitet wurden. Keine Antwort wies eine derart kurze Bearbeitungszeit auf. Weiterhin wurden alle Werte, die 4 Standardabweichungen vom bedingungs-spezifischen Mittelwert einer Person abwichen, durch diesen ersetzt. Diese Maßnahme führte insgesamt zur Ersetzung von ca. 0.2 % aller Datenpunkte.

Ergebnisse. Abb. 6.2 zeigt die mittlere Reaktionszeit bis zur Auflösung der pronominalen Anapher in den vier Untersuchungsbedingungen.

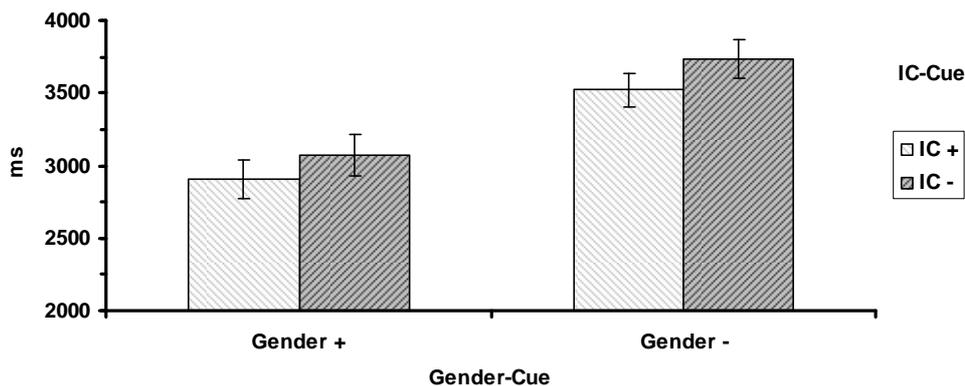


Abb. 6.2: Mittlere RT für die pronominale Auflösung in den vier Untersuchungsbedingungen.

Eine messwiederholte ANOVA zeigte zunächst einen äußerst varianzstarken Haupteffekt für den Faktor *Gender-Cue* ($F(1,21) = 64.73, p < .001, \eta_p^2 = .755; F(1,159) = 387.28, p < .001, \eta_p^2 = .709$): Sätze mit zwei gleichgeschlechtlichen potenziellen Referenten für das Pronomen benötigten ca. 650 ms länger, um erfolgreich aufgelöst zu werden, als Sätze mit nur einem grammatikalisch möglichen Referenten. Darüber hinaus gab es auch einen Effekt für den Faktor *IC-Cue* ($F(1,21) = 46.21, p < .001, \eta_p^2 = .688; F(1,159) = 16.95, p < .001, \eta_p^2 = .096$): Sätze, bei denen die explizit gegebene Begründung mit der impliziten Verbkausalität kongruent war, wurden ca. 180 ms schneller verarbeitet. Die Interaktion beider Faktoren überschritt nicht das geforderte Signifikanzniveau ($F(1,21) = 1.82, p = .192, \eta_p^2 = .080; F(1,159) = 1.40, p = .239, \eta_p^2 = .009$). Zwar war der Kausalitätseffekt in der ambigen Bedingung ein wenig stärker ausgeprägt ($F(1,21) = 41.49, p < .001, \eta_p^2 = .664; F(1,159) = 13.91, p < .001, \eta_p^2 = .080$). Er trat jedoch auch auf, wenn ein Pronomen bereits gender-disambiguiert war ($F(1,21) = 10.92, p < .003, \eta_p^2 = .342; F(1,159) = 6.47, p < .012, \eta_p^2 = .039$).

Re-Analyse der „korrekten“ Antworten. Um zu überprüfen, ob das berichtete Effektmuster evtl. durch Unterschiede zwischen falsch und richtig beantworteten Items moderiert wurde, wurde die eben berichtete Analyse nochmals nur mit den „korrekt“ beantworteten

Items durchgeführt. Das berichtete Ergebnismuster konnte repliziert werden und ist vollständig in Appendix C.1.2 einsehbar.

Zusammenfassung. Die mittlere Reaktionszeit zur Auflösung der pronominalen Anapher zeigte das für Studien zum IC-Effekt mit Auflösungsaufgabe typische Muster (z.B. Caramazza et al., 1977; Vonk, 1985b; im Überblick: Garnham, 2001): Einen starken Effekt für den Gender-Cue und einen etwas schwächeren für den Kausalitäts-Cue, wobei beide Faktoren nicht miteinander interagierten. Auch numerisch ergaben sich nahezu identische Ergebnisse wie in den bisher durchgeführten Studien: So war der Gender-Effekt bei Caramazza et al. (1977) ca. 590 ms groß, der Kausalitätseffekt 230 ms. Die von Vonk (1985b) berichteten Effektgrößen (650 ms für den Gender-Cue und 150 ms für den IC-Cue) korrespondieren ebenfalls mit dem in der vorliegenden Untersuchung gefundenen Effektmuster. Bedeutsam ist v.a., dass sich die IC auch dann auf die zur pronominalen Auflösung benötigte Zeit auswirkt, wenn das Pronomen durch den Gender-Cue bereits vollständig determiniert ist.

Nachdem so das für Studien zum IC-Effekt typische Effektmuster qualitativ und quantitativ vollständig repliziert werden konnte, folgt nun der für die vorliegende Fragestellung interessantere Aspekt, nämlich welche kognitiven Verarbeitungsprozesse für die beiden Effekte verantwortlich sind. Insbesondere stellt sich die Frage, ob sich für die beiden experimentellen Manipulationen – Gender- und IC-Cue – unterschiedliche kognitive Teilprozessstypen finden lassen, die sich differenziell in bestimmten Momenten der Reaktionszeitverteilungen niederschlagen.

6.3.1.3 Reaktionszeitverteilung: Weibull-Parameter

Parameterschätzung. Für jede Versuchsperson fiel in jeder der vier Untersuchungsbedingungen eine eigene Reaktionszeitverteilung an, die auf jeweils 40 Datenpunkten basierte. An jede dieser 88 (= 22 x 4) beobachteten Reaktionszeitverteilungen wurde eine eigene Weibull-Verteilung angepasst, indem möglichst passende shift-, scale- und shape-Parameter gesucht wurden (vgl. Kap. 3). Die Parameterschätzung erfolgte mittels des Hierarchical Bayes-Verfahrens von Rouder et al. (2003), bei dem optimale Bayes-Schätzer mittels MCMC-Simulationen gesucht werden. Als Startwert für den scale-Parameter wurde ein Wert von $\lambda = \theta^{\beta} = 1$ angesetzt. Die verwendeten Parameter der beiden Prior-Verteilungen des scale-Parameters waren $a_1 = 2.0$, $b_1 = 0.1$, $a_2 = 2.0$ und $b_2 = 0.05$.

Modellgüte. Die Passung der Weibull-Verteilung für die einzelnen RT-Verteilungen wurde mittels des in Kap. 3.5 beschriebenen Chi-Quadrat-Verfahrens beurteilt. Der Fit war insgesamt sehr zufrieden stellend: Der mittlere Chi-Quadrat-Wert lag bei $\chi^2(3) = 4.82$, $p > .30$

und nur drei der insgesamt 88 Einzelverteilungen (= 3.4 %) mussten auf dem 1 %-Niveau zurückgewiesen werden.

Durchschnittliche Parameterausprägung. Abb. 6.3 zeigt die drei Weibull-Parameter über alle Untersuchungsbedingungen gemittelt. Dabei ist zu beachten, dass der shift- und scale-Parameter grundsätzlich in der Original-Metrik der Verteilung vorliegen (in diesem Fall also Millisekunden), der shape-Parameter jedoch eine eigene Metrik hat (ohne direkte inhaltliche Bedeutung). Da der shape-Parameter bei Reaktionszeit-Verteilungen typischer Weise Werte zwischen 1 und 4 annimmt, werden der shift- und der scale-Parameter in Sekunden angegeben, damit alle Parameter eine ähnliche Größenordnung haben und sich besser vergleichen lassen.

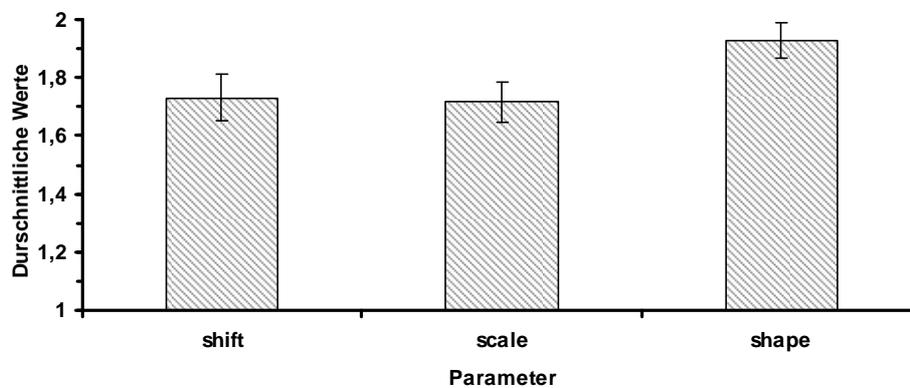


Abb. 6.3: Durchschnittliche Weibull-Parameter der RT-Verteilungen der pronominalen Auflösung.

Der durchschnittliche shift-Parameter lag bei 1.73 s, der durchschnittliche scale-Parameter bei 1.72 s und der durchschnittliche shape-Parameter bei 1.93. Der shape-Parameter hatte damit einen Wert, wie er für RT-Experimente typisch ist (vgl. Logan, 1992). Scale- und shift-Parameter können nicht mit denen anderer RT-Experimente verglichen werden, weil sie je nach eingesetzter Aufgabe unterschiedlich hoch ausfallen. Die Relation des scale- zum shape-Parameter (ca. 1:1) war jedoch typisch für Aufgabentypen mittleren Schwierigkeits- und Übungsgrades. Auch entsprach der shift-Parameter erwartungsgemäß ungefähr der Hälfte des Mittelwertes (vgl. Rouder, 2005a).

Jeder diese drei Parameter wurde nun separat mittels einer messwiederholten ANOVA mit den beiden Faktoren *Gender-Cue* (2: *vorhanden* vs. *nicht vorhanden*) und *IC-Cue* (2: *vorhanden* vs. *nicht vorhanden*) auf Unterschiede zwischen den einzelnen Untersuchungsbedingungen geprüft. Dabei ergaben sich folgende Ergebnisse:

shift-Parameter. Es gab einen starken Haupteffekt für den Faktor *Gender-Cue* ($F(1,21) = 23.20, p < .001, \eta_p^2 = .525$). Wie Abb. 6.4 zeigt, lag der shift-Parameter für Sätze mit zwei gleichgeschlechtlichen Personen ungefähr 280 ms höher als für Sätze, bei denen nur eine Per-

son mit dem Pronomen gender-kongruent war. Der Haupteffekt für den IC-Cue ($F < 1$) und die Interaktion der beiden Faktoren ($F(1,21) = 1.66, p = .211$) waren nicht signifikant.

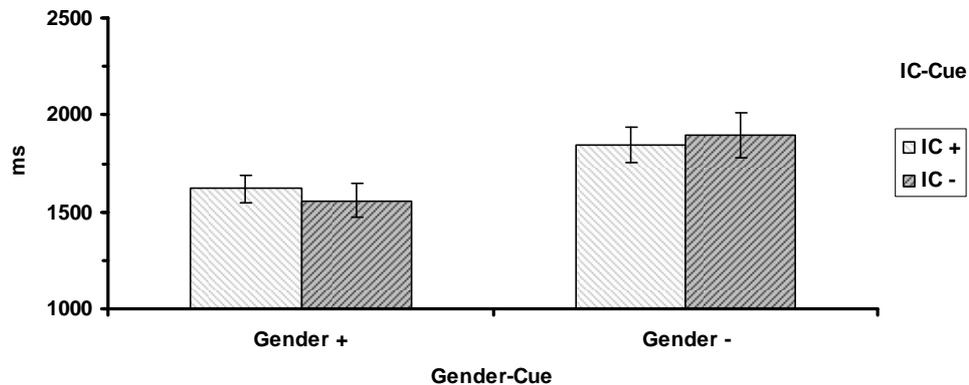


Abb. 6.4: Ausprägung des shift-Parameters der RT-Verteilungen der pronominalen Auflösung.

scale-Parameter. Auch für den scale-Parameter gab es einen Haupteffekt des Faktors *Gender-Cue* ($F(1,21) = 49.72, p < .001, \eta_p^2 = .703$): Reaktionszeitverteilungen für die Sätze, bei denen das Pronomen gender-ambig war, wiesen einen um ca. 390 ms höheren scale-Parameter auf als Sätze mit disambiguierten Pronomen (vgl. Abb. 6.5). Auch der Haupteffekt des Faktors *IC-Cue* war signifikant ($F(1,21) = 25.26, p < .001, \eta_p^2 = .546$): In den Untersuchungsbedingungen, in denen die IC und die explizite Auflösung übereinstimmten, fiel der scale-Parameter um ca. 200 ms niedriger aus. Die Interaktion der beiden Faktoren war nicht signifikant ($F < 1$). Das bedeutet, dass der IC-Cue auch dann den scale-Parameter beeinflusste, wenn der Satz bereits gender-disambiguiert war.

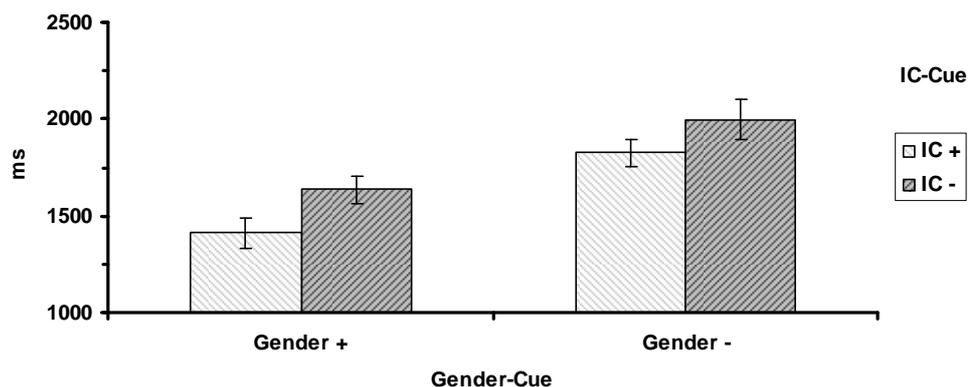


Abb. 6.5: Ausprägung des scale-Parameters der RT-Verteilungen der pronominalen Auflösung.

shape-Parameter. Für den shape-Parameter der Weibull-Verteilung gab es einen sehr schwachen, aber knapp signifikanten Effekt des *Gender-Cues* ($F(1,21) = 4.40, p < .048, \eta_p^2 = .173$). Zusätzlich gab es einen marginal signifikanten Effekt des *IC-Cues* ($F(1,21) = 4.12, p = .055, \eta_p^2 = .164$). Beide Effekte waren jedoch – speziell im Vergleich zu den varianzstarken Effekten dieser Manipulation auf den shift- und scale-Parameter – mit einer Differenz von 0.1

eher schwach ausgeprägt Die Interaktion der beiden Faktoren war nicht signifikant ($F(1,21) = 1.27, p = .272$) (vgl. Abb. 6.6).

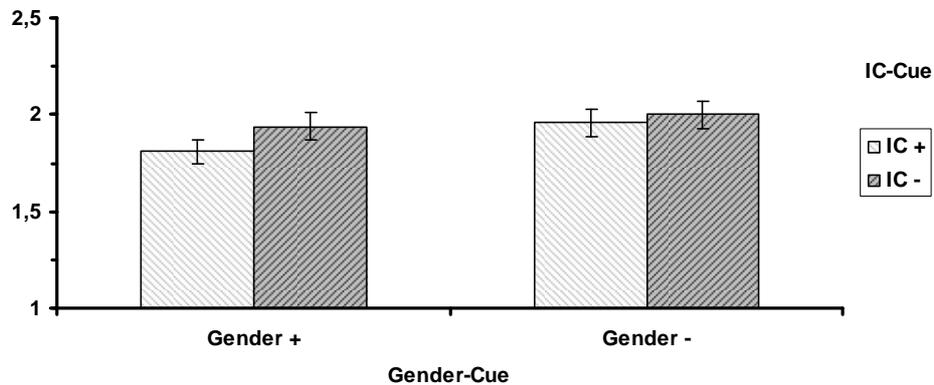


Abb. 6.6: Ausprägung des shape-Parameters der RT-Verteilungen der pronominalen Auflösung.

Re-Analyse der „korrekten“ Antworten. Um sicher zu gehen, dass das berichtete Ergebnismuster nicht durch aufgabenfremde Verarbeitungsstrategien oder ressourcenintensive Re-Analyse-Prozesse verzerrt war, wurden alle berichteten Analysen nur mit den „korrekt“ beantworteten Items einer Person wiederholt. Effektmuster, Einzeleffekte und Effektgrößen blieben vollständig erhalten bzw. zeigten sich noch prägnanter. Ein vollständiger Bericht der Ergebnisse findet sich in Appendix C.1.3.

Überblick Effekte. Da die beiden Cues nie miteinander interagierten, lässt sich ihr Einfluss auf die drei Parameter der Weibull-Verteilung getrennt voneinander darstellen. Abb. 6.7 zeigt die Effekte der Gender-Manipulation auf die drei Parameter der Weibull-Verteilung (wobei shift- und scale-Parameter zwecks Vergleichbarkeit wieder in Sekunden angegeben sind). Es ist deutlich zu erkennen, dass sich der Verarbeitungsvorteil für Sätze, in denen nur einer der beiden Referenten mit einem Pronomen gender-kompatibel war, sowohl in einer Reduktion des shift- als auch des scale-Parameter niederschlug. Auf den shape-Parameter hatte die Gender-Manipulation auch einen leichten Effekt, der aber relativ gering war.

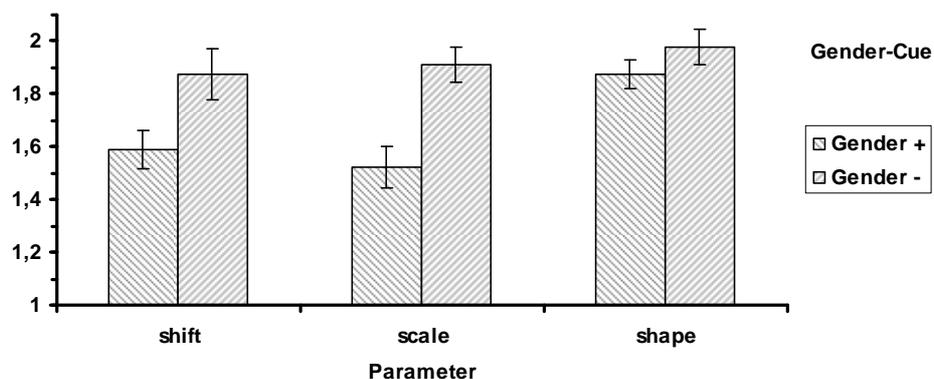


Abb. 6.7: Effekte des Gender-Cues auf die Weibull-Parameter der RT-Verteilungen der pronominalen Auflösung.

Abb. 6.8 zeigt die Effekte der Kausalitäts-Manipulation auf die drei Parameter der Weibull-Verteilung. Es ist klar erkennbar, dass die Kausalitäts-Manipulation selektiv nur den scale-Parameter der Weibull-Verteilung beeinflusste.

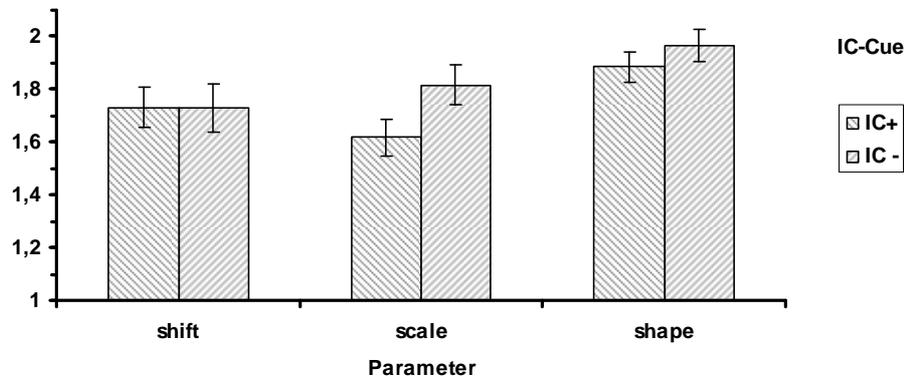


Abb. 6.8: Effekte des IC-Cues auf die Weibull-Parameter der RT-Verteilungen der pronominalen Auflösung.

Verteilungen. Was die dargestellten Veränderungen in den Weibull-Parametern für die Form der RT-Verteilung bedeuten, lässt sich besser erläutern, wenn man sich die konkreten Kurven ansieht, die sich aus ihnen ergeben (vgl. Abb. 6.9).

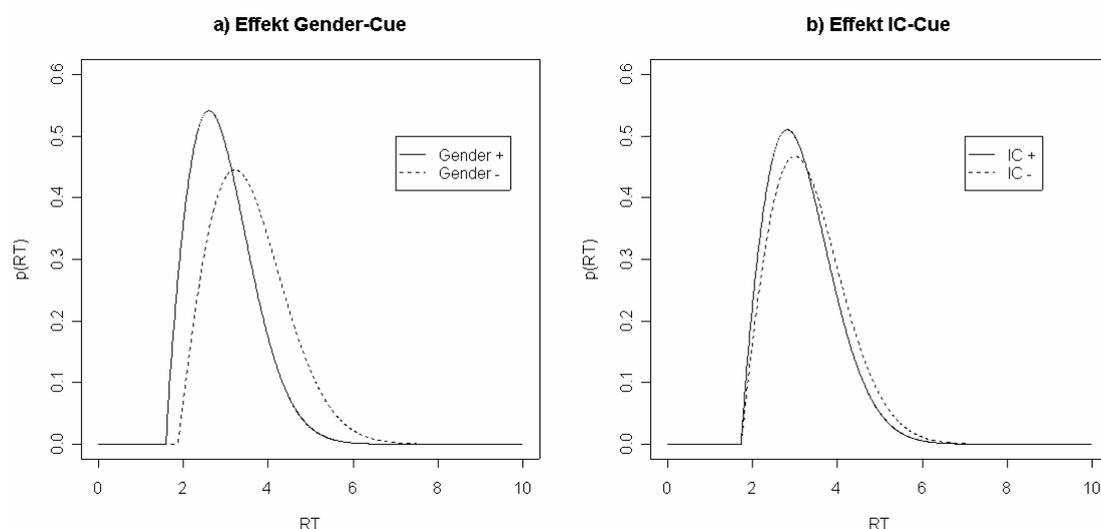


Abb. 6.9: Unterschiede zwischen den geschätzten Weibull-Verteilungen für die Manipulation a) des Gender-Cues und b) des IC-Cues.

Abb. 6.9 a) zeigt den Effekt der Gender-Manipulation auf die Verteilung der Antwortzeiten: Man kann deutlich sehen, dass die gesamte Verteilung in der Bedingung *Gender-* nach rechts rutschte (Effekt für den shift-Parameter). Das bedeutet, dass bei Sätzen, in denen zwei gender-kompatibler Referenten zur Verfügung standen, die minimale Antwortzeit länger ausfiel. Gleichzeitig kann man aber auch sehen, dass sich durch die Gender-Manipulation der Masseschwerpunkt der Verteilung deutlich verschob (Effekt scale-Parameter): Die Kurve in der Bedingung *Gender-* hat einen deutlich niedrigeren „Gipfel“ und ist insgesamt „breiter“

als in der Bedingung *Gender+*. Das bedeutet, dass ein Teil der Reaktionsverzögerung des Gender-Cues dadurch zustande kommt, dass das Antwortverhalten insgesamt „variabler“ ausfällt und deswegen ein größerer Teil der Antworten zu einem späteren Zeitpunkt ausgegeben wurden. Die Form der Verteilung veränderte sich jedoch nicht sehr stark, da die Verteilung in beiden Bedingungen deutlich linkssteil bzw. rechtsschief ist. Die Verteilung in der Bedingung *Gender-* war ein wenig symmetrischer; allerdings war dieser Unterschied nicht sehr groß.

Abb. 6.9 b) zeigt den entsprechenden Effekt der Kausalitäts-Manipulation auf die Verteilung der Antwortzeiten. Zwei Dinge sind im Vergleich zu den Effekten des Gender-Cues auffällig: Zum einen ist deutlich erkennbar, dass sich eine Manipulation des IC-Cues nur auf die Variabilität der Antwort-Verteilung auswirkte und nicht auf den Zeitpunkt der frühestmöglichen Antwortgenerierung. Und zweitens ist klar ersichtlich, dass der Effekt der impliziten Verbkausalität kleiner ausfiel als die Auswirkungen der Gender-Manipulation, die zu einer massiven Umstrukturierung der RT-Verteilung führten.

6.3.1.4 Diskussion pronominaler Auflösungsaufgabe

Die Analyse der Akkuratheit und der mittleren Reaktionszeiten der pronominalen Auflösungsaufgabe hat gezeigt, dass das Effektmuster von Experiment Ia eine große Ähnlichkeit zu bereits vorliegenden Studien zum IC-Effekt aufweist. Darüber hinaus sind v.a. zwei Aspekte relevant, die sich aus der Betrachtung der Reaktionszeitverteilung ergeben:

Zum einen legte die Analyse der Reaktionszeitverteilungen nahe, dass der in der Mittelwertanalyse so homogen erscheinende Effekt des Gender-Cues in der Tat aus zwei unabhängigen Teileffekten besteht, die auf unterschiedlichen Verarbeitungsprozessen basieren. Der eine Teilprozess beeinflusst v.a. den shift-Parameter der RT-Verteilung und ist damit eher ein Reflex früher und peripherer Verarbeitungsprozesse. Wenn die in dieser Arbeit vorgeschlagene Interpretation des shift-Parameters gültig ist, dann handelt es sich dabei um einen primär gedächtnis-getriebenen *bonding-* bzw. *confirmation-*Prozess (sensu Rigalleau et al., 2004). Der andere Teileffekt der Gender-Manipulation beeinflusst hingegen den scale-Parameter. Es ist wahrscheinlich, dass dieser Parameter von späten, zentralen Verarbeitungsprozessen bedingt wird, die höhere kognitive Teilfähigkeiten involvieren. Wenn die in dieser Arbeit vertretene Interpretation des scale-Parameters gültig ist, dann handelt es sich hierbei um Integrations- oder *disengagement-*Prozesse.

Im Gegensatz hierzu – und das ist das zweite wichtige Ergebnis der RT-Verteilungsanalyse – beeinflusste der IC-Cue ausschließlich den scale-Parameter der Weibull-Verteilung. Die beiden Cues scheinen also in der Tat selektiv und differenziell unterschiedli-

che Komponenten des pronominalen Auflösungsprozesses anzusprechen: Während der Gender-Cue sowohl den shift- als auch den scale-Parameter beeinflusst, verändert der IC-Cue ausschließlich den scale-Parameter. Und unter der Voraussetzung, dass die Parameter unterschiedliche kognitive Prozesse abbilden, spricht dies dafür, dass für den Gender-Effekt sowohl Bindungs- als auch Integrationsprozesse relevant sind, während der IC-Effekt im vorliegenden Fall primär von Integrationsprozessen getragen wird.

6.3.2 Verifikationsaufgabe

6.3.2.1 Akkuratheit: Anteil korrekt verifizierter Aussagen

Ergebnisse. Abb. 6.10 zeigt die Akkuratheit (Hit-Rate) für die Beurteilung der Aussagen, die nach jedem Satz verifiziert werden mussten.

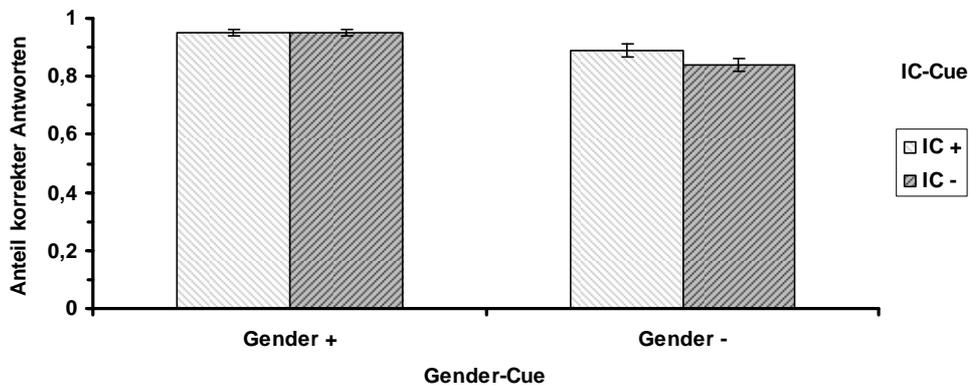


Abb. 6.10: Anteil korrekt verifizierter Aussagen.

Es zeigte sich ein recht ähnliches Bild wie für die Akkuratheit der pronominalen Auflösung, und entsprechend ergab eine messwiederholte ANOVA auch eine ähnliche Effektkonstellation: Einen starken Haupteffekt für den Gender-Cue ($F(1,21) = 60.21, p < .001, \eta_p^2 = .741$; $F(1,159) = 60.78, p < .001, \eta_p^2 = .277$) und einen schwachen Haupteffekt für den IC-Cue ($F(1,21) = 4.74, p < .041, \eta_p^2 = .184$; $F(1,159) = 3.87, p < .051, \eta_p^2 = .024$), der durch einen signifikanten Interaktionsterm ($F(1,21) = 9.06, p < .007, \eta_p^2 = .301$; $F(1,159) = 3.89, p < .050, \eta_p^2 = .024$) auf die nicht gender-disambiguierten Sätze eingegrenzt wurde (disambiguiert: $F(1,21) = 8.62, p < .008, \eta_p^2 = .291$; $F(1,159) = 4.87, p < .029, \eta_p^2 = .030$; nicht-disambiguiert: beide $F < 1$).

Arcussinus-Transformation. Wieder wurde mittels Arcussinus-Transformation geprüft, ob das berichtete Ergebnismuster auch unter realistischeren Verteilungsannahmen stabil war. Es ergaben sich keine auffälligen Abweichungen im Ergebnismuster, das vollständig in Appendix C.1.4 eingesehen werden kann.

Zusammenfassung. Für die Korrektheit der Verifikationsaufgabe ergab sich das charakteristische Bild, dass der morpho-syntaktisch eindeutige Gender-Cue die Satzverarbeitung dramatisch verbessert – gemessen als Fähigkeit zu entscheiden, ob eine Aussage mit einem vorherigen Satz übereinstimmt. Wenn hingegen kein eindeutiger Hinweisreiz vorhanden war, dann ist auch die Kongruenz der IC zur explizit gegebenen Begründung bedeutsam.

Grundsätzlich ist das durchschnittliche Korrektheitsniveau mit ca. 90% höher als für die pronominale Auflösung. Einer der Gründe hierfür ist, dass die Vpn durch die Rückmeldung zur pronominalen Auflösung weitere Informationen zur Bearbeitung der Verifikationsaufgabe bekommen haben und ihre Antwort evtl. darauf abstimmten. Ein anderer Grund für das verbesserte Abschneiden bei der Verifikationsaufgabe ist jedoch, dass diese Art von Aufgabe wesentlich einfacher zu bearbeiten ist, weil entscheidende Schlüsselinformationen des Satzes in der zu verifizierenden Aussage noch einmal genannt werden. Auch waren die Verifikationsaussagen wesentlich kürzer und einfacher als die Originalaussagen. Obwohl dieses Ergebnismuster qualitativ mit dem anderer Studien übereinstimmt, kann diese Übereinstimmung nicht genau beurteilt werden, da diese meist keine präzisen Akkuratheitswerte berichten (Vonk, 1985a; 1985b; Garnham et al., 1992).

6.3.2.2 Mittlere Reaktionszeiten

Datenbereinigung. Für die Datenbereinigung auf Itemebene wurde als unter Cut-off-Wert ein Wert von 200 ms angesetzt. Keine Antwort wies eine derart kurze Reaktionszeit auf. Weiterhin wurden alle Werte, die 4 Standardabweichungen vom bedingungsspezifischen Mittelwert einer Person abwichen, durch diesen ersetzt. Diese Maßnahme führte insgesamt zur Ersetzung von ca. 0.3 % aller Datenpunkte.

Ergebnisse. Abb. 6.11 zeigt die mittlere Zeit, die benötigt wurde, um die nach jedem Item dargebotene Aussage zu verifizieren.

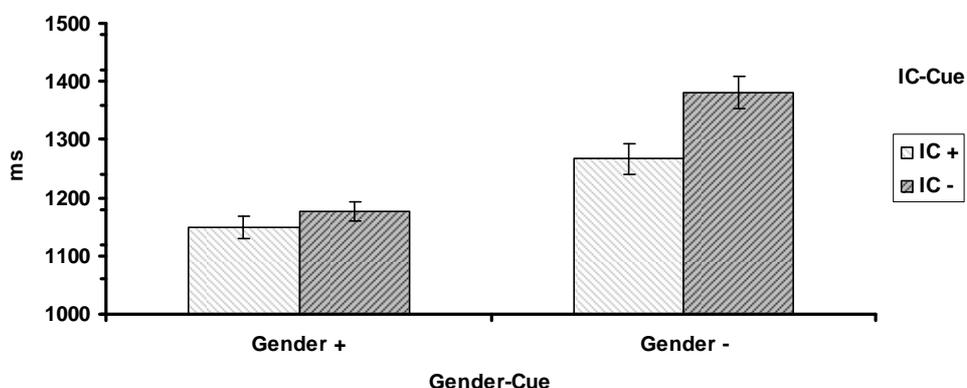


Abb. 6.11: Mittlere RT für die Verifikationsaufgabe in den vier Untersuchungsbedingungen.

Wieder war ein varianzstarker Haupteffekt für den Faktor *Gender-Cue* ($F(1,21) = 38.23, p < .001, \eta_p^2 = .645$; $F(1,159) = 68.00, p < .001, \eta_p^2 = .300$) besonders dominant. Sätze, in denen nur ein vom Geschlecht passender Referent zur Verfügung stand, wurden ca. 160 ms schneller verifiziert. Aber auch für der Verifikationsaufgabe gab es wieder einen Effekt des Faktors *IC-Cue* ($F(1,21) = 14.25, p < .001, \eta_p^2 = .404$; $F(1,159) = 6.09, p < .015, \eta_p^2 = .037$). Sätze, bei denen explizite Begründung und implizite Verbkausalität kongruent waren, wurden ca. 70 ms schneller verarbeitet. Die Interaktion beider Faktoren war lediglich marginal signifikant ($F(1,21) = 3.90, p = .062, \eta_p^2 = .139$; $F(1,159) = 3.03, p = .084, \eta_p^2 = .019$), deutet sich aber in der Mittelwertskonstellation an: Ähnlich wie bei der Akkuratheit der Antworten wirkt sich der IC-Cue nur aus, wenn ein Pronomen nicht bereits gender-disambiguiert war (EHE IC-Cue: $F(1,21) = 10.08, p < .005, \eta_p^2 = .324$; $F(1,159) = 7.30, p < .008, \eta_p^2 = .044$), nicht jedoch, wenn nur eine Lesart des Pronomens grammatikalisch möglich war (beide $F < 1$).

Re-Analyse der „korrekten“ Antworten. Um zu überprüfen, ob das berichtete Effektmuster evtl. durch Unterschiede zwischen falsch und richtig beantworteten Items moderiert wurde, wurde die eben berichtete Analyse nochmals nur mit den korrekt beantworteten Items gerechnet. Das Ergebnismuster konnte vollständig repliziert werden. Ein vollständiger Bericht der Analyse findet sich in Appendix C.1.5.

Zusammenfassung. Die vielleicht größte Auffälligkeit für die mittleren RTn für die Verifikationsaufgabe ist, dass die Interaktion der beiden Faktoren – im Gegensatz zur Reaktionszeit bei der pronominalen Auflösung – nun relevant zur Beurteilung der Ergebnisse ist: Der Effekt des IC-Cues ist nur bzw. stärker wirksam bei Sätzen, die nicht bereits durch den Gender-Cue disambiguiert wurden. Dieser Befund steht in partiellem Gegensatz zu den Ergebnissen anderer Studien zur impliziten Verbkausalität, die eine Verifikationsaufgabe verwendeten (z.B. Vonk, 1985a; 1985b). Bei diesen zeigte sich ein Ergebnismuster, wie es hier für die pronominalen Auflösungsaufgabe berichtet wurde, da bei dieser die Interaktion der beiden Faktoren ebenfalls nicht relevant ist.

6.3.2.3 Reaktionszeitverteilung: Weibull-Parameter

Parameterschätzung. Das Vorgehen bei der Parameterschätzung entsprach dem in Kap. 3.4 skizzierten Vorgehen. Als Startwert für den scale-Parameter wurde wiederum ein Wert von $\lambda = \theta^b = 1$ angesetzt. Die Parameter der beiden Prior-Verteilungen des scale-Parameters waren $a_1 = 2.0, b_1 = 0.1, a_2 = 2.0$ und $b_2 = 0.2$.

Modellgüte. Die Passung der durch die Parameterschätzungen generierten, einzelnen RT-Verteilungen wurde wiederum mittels des in Kap. 3.5 beschriebenen Chi-Quadrat-

Verfahrens beurteilt. Der Fit fiel dabei mit einem mittleren Chi-Quadrat-Wert von $\chi^2(3) = 4.98$, $p > .29$ vergleichbar aus wie für die RT-Verteilungen der pronominalen Auflösung. Der Fit von 6 der insgesamt 88 Einzelverteilungen (= 6.8%) war nicht gänzlich zufrieden stellend, weil der entsprechende Chi-Quadrat-Wert auf dem 1 %-Niveau signifikant war.

Durchschnittliche Parameterausprägung. Abb. 6.12 zeigt die drei Weibull-Parameter der Reaktionszeit-Verteilung der Verifikationsaufgabe und ihre Beziehungen zueinander (wobei shift- und scale-Parameter zwecks Vergleichbarkeit wieder in Sekunden angegeben sind).

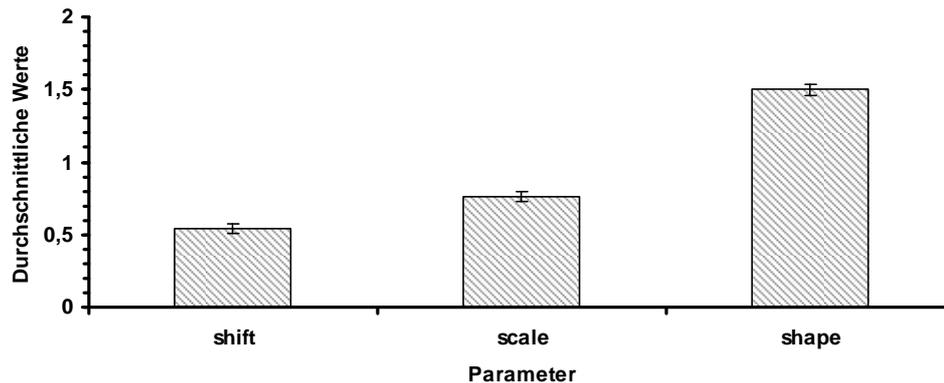


Abb. 6.12: Durchschnittliche Weibull-Parameter der RT-Verteilungen der Verifikationsaufgabe.

Der durchschnittliche shift-Parameter lag bei 0.54 s, der durchschnittliche scale-Parameter bei 0.77 s und der durchschnittliche shape-Parameter nahm den Wert 1.50 an. Shift- und scale-Parameter waren verständlicher Weise wesentlich niedriger als bei der pronominalen Auflösung, aber wiederum ungefähr gleich groß. Auch der shape-Parameter fiel etwas niedriger aus als der korrespondierende Parameter bei der pronominalen Auflösung; d.h. die Verteilungen waren etwas schief. Der deutliche Unterschied im shape-Parameter ist dabei u.U. ein Hinweis darauf, dass der pronominalen Auflösungs- und der Verifikationsaufgabe unterschiedliche kognitive Prozesse bzw. funktionale Architekturen zugrunde liegen (vgl. Rouder et al., 2003).

Diese drei Parameter wurden nun jeweils separat mittels einer messwiederholten ANOVA mit den beiden Faktoren *Gender-Cue* (2: *vorhanden* vs. *nicht vorhanden*) und *IC-Cue* (2: *vorhanden* vs. *nicht vorhanden*) auf Unterschiede geprüft. Dabei ergaben sich folgende Ergebnisse:

shift-Parameter. Für den shift-Parameter gab es keine signifikanten Effekte der experimentellen Faktoren (alle $F < 1$), lediglich für den IC-Cue deutete sich ein leichter Effekt an, der jedoch nicht das geforderte Signifikanzniveau erreichte ($F(1,21) = 2.66$, $p = .118$, $\eta_p^2 = .112$).

scale-Parameter. Für den scale-Parameter gab es einen Haupteffekt des Faktor *Gender-Cue* ($F(1,21) = 85.58, p < .001, \eta_p^2 = .803$): Reaktionszeitverteilungen für Sätze, bei denen das Pronomen nicht gender-disambiguiert war, wiesen einen um ca. 160 ms höheren scale-Parameter auf als Sätze mit disambiguierten Pronomen (vgl. Abb. 6.13). Der Haupteffekt des Faktors *IC-Cue* war ebenfalls signifikant ($F(1,21) = 4.59, p < .044, \eta_p^2 = .179$), aber ein signifikanter Interaktionsterm ($F(1,21) = 6.94, p < .016, \eta_p^2 = .248$) indizierte, dass der Effekt des IC-Cues nur in der Bedingung *Gender-* vorhanden war (ca. 120 ms; $F(1,21) = 8.72, p < .008, \eta_p^2 = .293$) und in der Bedingung *Gender+* nicht signifikant war ($F < 1$). Das bedeutet, dass der IC-Cue nur dann den scale-Parameter beeinflusste, wenn der Satz nicht bereits gender-disambiguiert war.

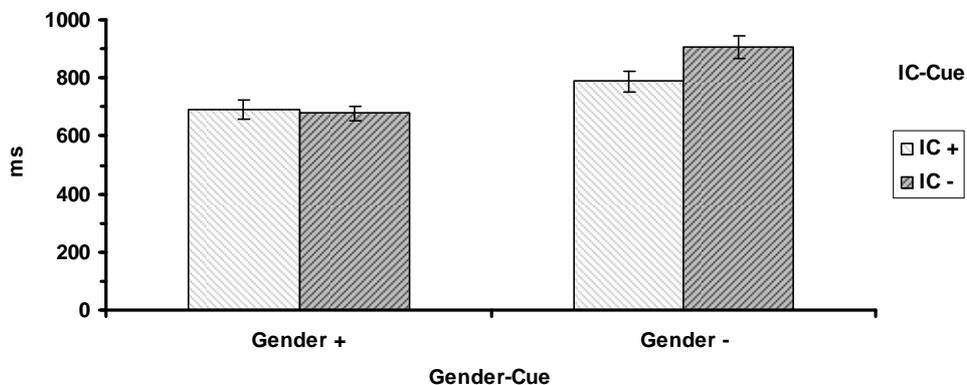


Abb. 6.13: Ausprägung des scale-Parameters der RT-Verteilungen der Verifikationsaufgabe.

shape-Parameter. Für den shape-Parameter gab es keine signifikanten Effekte (alle $F < 1$). Lediglich für den Gender-Cue deutete sich ein schwacher Effekt an, der allerdings ebenfalls nicht signifikant wurde ($F(1,21) = 2.32, p = .143, \eta_p^2 = .099$).

Re-Analyse der „korrekten“ Antworten. Wieder wurden alle Analysen nur mit den korrekt beantworteten Items wiederholt. Ein vollständiger Bericht der Ergebnisse findet sich in Appendix C.1.6. Das Ergebnismuster blieb im Wesentlichen identisch. Zwei kleinere Abweichung sind evtl. bedeutsam: Zum einen erreichte die Interaktion der beiden Faktoren nun nicht mehr das geforderte Signifikanzniveau und zum anderen war der Gender-Effekt für den shape-Parameter etwas deutlicher ausgeprägt.

Zusammenfassung. Zwei Aspekte des Ergebnismusters sind besonders bedeutsam: Zum einen gab es auch für die Weibull-Parameter eine Interaktion der beiden Cues, da der Effekt des IC-Cues nur in der ambigen Bedingung auftrat. Dies steht in einem charakteristischen Kontrast zur pronominalen Auflösungsaufgabe, bei der diese Interaktion für keinen der Parameter zu finden war. Zum anderen sind die Effekte der beiden Faktoren allein im scale-Parameter zu verorten, während es keine Effekte für den shift- und den shape-Parameter gibt.

6.3.2.4 Diskussion Verifikationsaufgabe

Die vorliegenden Ergebnisse stimmen bezüglich der Akkuratheit der Verifikation mit denen anderer Studien überein, die eine Verifikationsaufgabe verwendeten. Hier zeigte sich auch das gleiche Ergebnismuster wie bei der pronominalen Auflösungsaufgabe. Allerdings ergaben sich sowohl für die Analyse der mittleren RTn als auch für die RT-Verteilungsanalyse bedeutsame Unterschiede: Die Kongruenz von expliziter und impliziter Kausalität beeinflusste den Bearbeitungsprozess nur, wenn ein Satz nicht bereits durch Gender-Informationen disambiguiert wurde. Dieser Befund steht im Gegensatz zu den Ergebnissen der beiden Studien von Vonk (1985a, 1985b), die – ähnlich wie hier für die pronominale Auflösung – jeweils keine bedeutsame Interaktion fand.

Allerdings wurde in den Studien von Vonk die Verifikationsaufgabe auch *statt* der pronominalen Auflösungsaufgabe eingesetzt. Im Gegensatz hierzu folgte die Verifikationsaufgabe in der vorliegenden Studie der pronominalen Auflösung, was bedeutende Implikationen für die anzusetzenden Verarbeitungsprozesse hat. So ist zum einen davon auszugehen, dass zum Zeitpunkt der Verifikation der eigentliche Auflösungsprozess bereits stattgefunden hat und die Verifikationsaufgabe eher einen Prozess der Verstehenskonsolidierung abbildet, der das Verstehensprodukt nicht aufbaut, sondern lediglich *überprüft*. Und zum anderen ist klar, dass allein schon wegen der zeitlichen Verzögerung zwischen der Stimuluspräsentation und Verifikationsaufgabe sämtliche gedächtnis-basierten Prozesse keine relevante Rolle mehr spielen können, weil diese auf Informationen zurückgreifen, die zu diesem Zeitpunkt entweder bereits verfallen sind oder aber in ein höherwertiges Verstehensprodukt integriert wurden. Es ist deswegen ein für die Konstruktvalidität der RT-Verteilungsparameter ganz entscheidender Befund, dass alle Verifikations-Effekte allein vom scale-Parameter getragen wurden.

Beide Punkte zusammengenommen implizieren, dass bei der hier eingesetzten, nachgeschalteten Verifikationsaufgabe nur vergleichsweise hohe kognitive Integrationsprozesse involviert sind, die das bereits gebildete Verstehensprodukt erneut auf seine Richtigkeit überprüfen. Sie involviert also eine Art von Check-Mechanismus, der offensichtlich früher beendet werden kann, wenn ein Satz bereits durch Gender-Informationen disambiguiert ist. Allerdings kann man dem vorliegenden Ergebnismuster nicht entnehmen, was diese frühere Beendigung veranlasst: Entweder wird im Falle einer rein grammatikalischen Übereinstimmung die diskursive Kohärenz nicht mehr überprüft oder diese Überprüfung kann im Falle einer vorherigen gender-Disambiguierung einfach schneller vorgenommen werden.

6.4 Diskussion Experiment Ia

Als Ergebnis von Experiment Ia lassen sich mehrere Aspekte festhalten. Zunächst einmal kann festgestellt werden, dass sich in Bezug auf die Betrachtung der mittleren RTn das bekannte Ergebnismuster, wie es sowohl für den IC- als auch den Gender-Effekt in der Literatur berichtet wird, vollkommen repliziert werden konnte: Der Gender-Cue führte erwartungsgemäß zu einer starken Verkürzung der pronominalen Auflösungszeit ebenso wie der IC-Cue, der jedoch einen etwas schwächeren Effekt hatte. Entscheidend ist v.a., dass beide Effekte nicht miteinander interagierten, d.h. der IC-Cue war unabhängig vom Gender-Cue auch dann wirksam, wenn das Pronomen disambiguiert war.

Es ist unwahrscheinlich, dass die schnellere Bearbeitung im Falle einer Gender-Disambiguierung lediglich dadurch zustande gekommen ist, dass die Vpn einfach nach dem Pronomen nicht mehr weiter gelesen haben. Eine solche Rezeptionsstrategie hätte zu erhöhten Fehlerraten in der anschließenden Verifikationsaufgabe führen müssen, die spezifisch nach den Inhalten des Nebensatzes fragte. Das Gegenteil war jedoch der Fall: In dieser Bedingung fanden sich sogar besonders hohe Akkuratheitswerte weit über 90 %. In Experiment Ib soll diese Alternativerklärung jedoch noch weiter geprüft werden.

Nachdem sichergestellt war, dass die herkömmliche Analyse zu den gleichen Ergebnissen führt wie bei den bisherigen Studien, konnte in einem nachfolgenden Schritt eine differenziertere Analyse der Reaktionszeitverteilungen vorgenommen werden. Dabei muss aus der primär methodisch-statistischen Perspektive zunächst einmal festgehalten werden, dass sich das eingesetzte Hierarchical Bayes-Verfahren von Rouder et al. (2003) hervorragend bewährt hat: Die MCMC-Ketten der Parameterschätzung hatten gute Konvergenzeigenschaften, und die resultierenden Weibull-Verteilungen wiesen einen guten Fit zu den empirisch beobachteten Reaktionszeitverteilungen auf. Zur Validität der Parameter lassen sich allerdings an dieser Stelle kaum Aussagen machen. Es kann aber festgehalten werden, dass sich keine Effekte ergeben haben, die ihr widersprechen. Dies wäre insbesondere der Fall gewesen, wenn sich der IC-Effekt allein im shift-Parameter niedergeschlagen hätte. Dies wäre theoretisch unplausibel gewesen. Dies liegt nicht daran, dass das Verfahren nicht sensitiv genug ist, um überhaupt einen shift-Effekt zu erfassen. Dies lässt sich daran erkennen, dass es für den Gender-Cue auch shift-Effekte gab. Die Verteilungsparameter scheinen also in der Lage zu sein, unterschiedliche Verarbeitungsaspekte differenziell und unabhängig voneinander abzubilden.

Auf der Grundlage der differenzierteren Analyse der Verteilungsparameter in den einzelnen Bedingungen können einige weitere Aussagen gemacht werden, die aufgrund der Basis

einer mittelwerts-basierten Analyse nicht möglich wären. In Bezug auf den Gender-Effekt kann zunächst festgehalten werden, dass es in der Tat Evidenz dafür gibt, dass er sich aus zwei unterschiedlichen Teileffekten ergibt, die ungefähr in gleichen Anteilen zu ihm beitragen. Ein Teileffekt wird dabei vom shift-Parameter der Weibull-Verteilung getragen, der vergleichsweise hierarchie-niedrige, gedächtnis-basierte Prozesse abbilden soll. Wie in Kap. 2 dargestellt, handelt es sich dabei wahrscheinlich um einen (frühen) *confirmation*-Effekt, der in der *bonding*-Phase der pronominalen Auflösung die Passung der Gender-Relation von Pronomen und Antezedens evaluiert und einen entsprechenden *bond* etabliert. Offensichtlich kann dies einfacher bzw. schneller erfolgen, wenn lediglich ein gender-kongruenter Referent vorhanden ist. Der andere Teileffekt des Gender-Cues wird hingegen vom scale-Parameter (und zu geringen Teilen auch vom shape-Parameter) getragen, die im Rahmen der vorgestellten Zwei-Prozess-Interpretation der Weibull-Parameter hierarchie-hohe, erklärungs-basierte Teilprozesse der pronominalen Auflösung erfassen. Unter Rückgriff auf die Diskussion in Kap. 2 lässt sich dieser Teileffekt als (später) *disengagement*-Effekt identifizieren, welcher vornehmlich die finale Auflösung des Pronomens beeinflusst. Die vorliegenden Daten erlauben allerdings keine Rückschlüsse darüber, auf welchen Prozessen der Effekt genau beruht: Entweder ist er ein Konsequenz der Tatsache, dass durch die Disambiguierung in der *resolution*-Phase weniger *bonds* evaluiert werden müssen, was entsprechend weniger Zeit in Anspruch nimmt. Auf der anderen Seite wäre aber möglich, dass die Gender-Disambiguierung zu einer distinkteren Repräsentation der Aktanten führt, die sich deswegen leichter in das finale Diskursmodell integrieren lässt (vgl. Garnham et al., 1992).

Für den IC-Cue ergab sich das überraschende Ergebnis, dass er alleine im scale-Parameter verortet war und keine hierarchie-niedrigen Fokus-Effekte gefunden werden konnte. Dies scheint zunächst für die in Kap. 2 diskutierte Integrations-Hypothese zu sprechen (vgl. Stewart et al., 2000). Diese geht davon aus, dass der IC-Effekt allein auf einer erschwerten Integration ic-inkompatibler Informationen beruht und nicht auf eine frühe Refokussierung zurückgeführt werden kann. Diese hätte sich im shift-Parameter zeigen müssen, der die hierarchie-niedrigen Prozesse abbildet. Der IC-Cue beeinflusste jedoch in allen Bedingungen immer nur den scale-Parameter, der sich auf die hierarchie-hohen, gedächtnis-basierten Prozesse bezieht.

Die Tatsache, dass sich der IC-Effekt erwartungskonträr lediglich im scale-Parameter fand, steht in einem engen Zusammenhang mit dem Befund, dass die beiden Faktoren nicht miteinander interagierten. Denn für den Fokus-Effekt wurde angenommen, dass er nur in der disambiguierten Bedingung auftritt, weil nur dann die Inkongruenz des Nebensatzes auch früh

erkennbar ist. Der IC-Integrations-Effekt sollte hingegen entsprechend schwächer ausfallen. In der ambigen Bedingung sollte es hingegen zu keinen frühen Fokus-Effekten kommen, was von einem entsprechend stärker ausgeprägten Integrations-Effekt kompensiert werden muss. Wie gesehen, hat sich dieses Interaktionsmuster empirisch nicht gezeigt: Weder gab es einen shift-Effekt für den IC-Cue in der disambiguierten Bedingung, noch fiel der scale-Effekt in der ambigen Bedingung stärker aus. Während die Abwesenheit des IC-Fokus-Effekts im vorliegenden Datensatz kaum in Frage gestellt werden kann, gibt es allerdings Anlass zur Annahme, dass der IC-Kongruenz-Effekt in der ambigen Bedingung stärker ausfällt. Hierfür sprechen insbesondere die Reaktionszeit-Daten der Verifikationsaufgabe. Dort fanden sich Effekte des IC-Cues allein in der ambigen Bedingung.

Dabei spricht einiges dafür, dass in der vorliegenden Untersuchung mit der Verifikationsaufgabe im Wesentlichen ein Aspekt der Verstehenskonsolidierung erfasst wird, der lediglich vom *resolution*-Anteil der beiden Cue-Effekte beeinflusst wird. Gedächtnis-basierte Prozesse können sich hingegen schon allein aufgrund des zeitlichen Abstandes hier nicht mehr auswirken. Deswegen spricht es für die Konstruktvalidität der RT-Verteilungsanalyse, dass alle Verifikations-Effekte nur vom scale-Parameter getragen wurden. Denn dieser bildet im Rahmen der Zwei-Prozess-Interpretation der Weibull-Parameter v.a. erklärungs-basierte Integrationsprozesse ab.

Zusammenfassend lässt sich das Ergebnismuster von Experiment Ia wie in Tab. 6.2 beschreiben:

<i>Effekt</i>	<i>Teileffekt</i>	<i>Pronominale Auflösung</i>	<i>Verifikation</i>
Gender	Confirmation	✓	---
	Disengagement	✓	✓
IC	Fokus	---	---
	Integration	✓	✓

Tab. 6.2: Effekte in Experiment Ia für die pronominale Auflösungs- und Verifikationsaufgabe.

Beim Gender-Effekt liegt Evidenz dafür vor, dass an ihm sowohl ein hierarchie-niedriger *confirmation*-Effekt als auch ein hierarchie-hoher *disengagement*-Effekt beteiligt ist. Dabei persistiert der *disengagement*-Effekt wie vorhergesagt auch noch in der nachfolgenden Verifikationsaufgabe. Der *confirmation*-Anteil lässt sich hingegen erwartungsgemäß nicht mehr nachweisen. Beim IC-Effekt liegt Evidenz für die Existenz einer hierarchie-hohen Integrations-Komponente vor. Da es jedoch keine Interaktion der beiden Cues gab, müssen hier Einschränkungen gemacht werden: Denn es wäre zu erwarten gewesen, dass der Integrations-effekt in der ambigen Bedingung stärker ausfällt. Genau dieses Muster zeigte sich jedoch erst

bei der Verifikationsaufgabe. Der entscheidende Punkt ist also, dass in Experiment Ia erwartungskonträr kein gedächtnis-basierter IC-Fokus-Effekt nachgewiesen werden konnte. Dieser hätte in der disambiguierten Bedingung zu einer frühen Refokussierung führen sollen, was jedoch offensichtlich nicht passiert ist. Wenn es jedoch nicht zu einer frühen Refokussierung kommt, dann ist auch die spätere Integration nicht erleichtert. Die alles entscheidende Frage ist deshalb: Warum gab es in diesem Experiment keinen IC-Fokus-Effekt? Dieses Problem soll im Rahmen von Experiment Ib angegangen werden, in dem das Stimulusmaterial so verändert wird, dass frühe Fokussierungs-Effekte wahrscheinlicher sind.

7. Experiment Ib

7.1 Einleitung

In Experiment Ia konnten bereits einige Annahmen über das Verhältnis gedächtnis- und erklärungs-basierter Verarbeitungsprozesse bestätigt werden: Sie sind beide am pronominalen Auflösungsprozess beteiligt und komplettieren sich gegenseitig. Allerdings kam es zu dem erwartungskonträren Ergebnis, dass der IC-Effekt ein reiner Integrations-Effekt ist und allein von hierarchie-hohen Prozessen getragen wird. Es fand sich kein früher IC-Fokus-Effekt, der mit hierarchie-niedrigen Mitteln den Auflösungsprozess vorbereitet, wenn die Inkongruenz früh signalisiert wird.

Diese Tatsache ist misslich, weil deswegen eine wichtige Vorhersage des in Kap. 1 entwickelten Rahmenmodells der pronominalen Auflösung nicht überprüft werden konnte. Denn dort wurde nicht nur die Auffassung vertreten, dass gedächtnis- und erklärungs-basierte Prozesse gleichzeitig, aber unabhängig voneinander an der pronominalen Auflösung beteiligt sind, sondern dass sie sich aufeinander beziehen und gegenseitig ergänzen. Entsprechend wurde in Kap. 2 für den IC-Effekt eine Trade-off-Relation des IC-Fokus- und des IC-Integrations-Effekts postuliert: Wenn der frühe gedächtnis-basierte IC-Fokus-Effekt die finale Auflösung bereits vorbereiten kann, dann vereinfacht sich die erklärungs-basierte Integration.

Man muss daher fragen, warum der IC-Fokus-Effekt in Experiment Ia ausgeblieben ist. Eine Möglichkeit ist, dass es ihn „einfach nicht gibt“, wie dies die Integrations-Hypothese behauptet (vgl. z.B. Stewart et al., 2000). Allerdings ist diese Option theoretisch und empirisch eher unplausibel: Wie in Kap. 2 ausführlich besprochen, wurde der IC-Fokus-Effekt in mehreren Studien unabhängig voneinander belegt. Auch lässt er sich direkt aus der in Kap. 2.1.2 entwickelten ereignisstrukturellen Rekonstruktion des IC-Effekts herleiten. Eine weitere Möglichkeit für den fehlenden IC-Fokus-Effekt in Experiment Ia ist, dass das verwendete Auswertungsverfahren nicht sensitiv genug war, um ihn zu entdecken. Allerdings ist auch dies eher unwahrscheinlich, weil für den Gender-Cue sehr wohl shift-Effekt nachgewiesen werden konnten.

Ich halte im vorliegenden Fall für wahrscheinlicher, dass Eigenschaften des Stimulusmaterials dafür verantwortlich sind, dass kein IC-Fokus-Effekt beobachtet werden konnte. Diese Vermutung erhärtet sich, wenn man sich die Studien, die einen IC-Fokus-Effekt bisher berichtet haben, etwas näher betrachtet. Denn dann zeigt sich, dass diese Studien entweder mit einer Einbettung der Stimulus-Sätze in einen Diskurskontext gearbeitet haben (z.B. McKoon et al., 1993; Koornneef & van Berkum, 2006) und/oder längere Stimulussätze ver-

wendeten (McDonald & MacWhinney, 1995; Long & de Ley, 2000; Koornneef & van Berkum, 2006). Warum aber sollte das einen Unterschied machen?

In dem Fall, bei dem der Stimulus-Satz in einen diskursiven Kontext eingebettet ist, liegt die Erklärung m.E. auf der Hand: Im Gegensatz zu isolierten Sätzen kommt es bei Diskursen zu einer wesentlich stärkeren Voraktivierung von Szenario-Informationen, weil diese bereits in den Sätzen vor dem target-Satz instantiiert werden können. Bei Koornneef und van Berkum (2006) bekommt man z.B. schon vor der Rezeption des IC-Verbs mitgeteilt, dass es sich um einen Verkehrsunfall handelt, an dem zwei Aktanten beteiligt sind. Dieses Szenario besitzt für sich bereits eine gewisse Binnenstruktur, die bestimmte Ereignisse umfasst und ihre Relationen zueinander spezifiziert. Darüber hinaus werden nach dem SMF-Ansatz von Sanford und Garrod (1998) die Aktanten direkt den verschiedenen Ereignissen des Szenarios zugewiesen. Zwei dominante Rollen des Verkehrsunfall-Skripts sind z.B. SCHULDIGER und OPFER, die wiederum mit der Rollenverteilung in bestimmten Sub-Ereignissen des Szenarios (x fährt y hinten auf) korreliert sind. Wenn nun ein IC-Verb (z.B. *entschuldigen* bei Koornneef & van Berkum, 2006) eingeführt wird, dann können die Aktanten direkt den Rollen (und damit auch der Reihenfolge der Ereignisse) zugewiesen werden, weil das gesamte Szenario-Wissen für ein erfolgreiches Mapping vorhanden ist. Wenn eine szenario-inkongruente Fortsetzung folgt (*Linda entschuldigte sich bei David, weil er...*), dann wird diese leicht und sicher entdeckt und kann entsprechend auch wieder behoben werden. Bei isolierten Sätzen ist dies hingegen sehr viel schwieriger. So sind z.B. in dem Satz *Sarah questioned Timothy because he hadn't told the truth.* (vgl. Caramazza et al., 1977) zum Zeitpunkt der Etablierung der ic-inkompatiblen Fortsetzung (*he*) noch kaum szenario-relevante Informationen vorhanden. Deshalb ist die durch das IC-Verb vorgegebene Auflösungstendenz auch nicht ähnlich zwingend.

An dieser Stelle ergibt sich die aus theoretischer Perspektive äußerst interessante Konstellation, dass sich die semantische Interpretation eines Verbs auf der Satzebene ggf. durch die Einbettung in einen weiteren Diskurskontext verändern kann: Je nachdem in welchem Kontext ein Verb eingebettet ist, kann sein ereignisstruktureller Rahmen mehr oder weniger relevant oder sogar verschoben sein (vgl. Kap. 2.1.2). Dies ist eine weitere Möglichkeit, wie erklärungs- und gedächtnis-basierte Prozesse in der Sprachverarbeitung miteinander interagieren können: Durch die erklärungs-basierte Verarbeitung von Kontextinformationen werden evtl. gedächtnis-basierte Prozesse überhaupt erst ermöglicht, weil sie auf top-down zur Verfügung gestelltes Wissen zurückgreifen können.

In dem Fall, in dem die Stimulussätze nicht in einen Diskurskontext eingebettet sind, aber längere Sätze zu bearbeiten sind, würde ich eine sehr ähnliche Erklärungsstruktur ansetzen. Allerdings findet hier die Elaboration der Diskursinformationen nicht *vor* der Präsentation des *target*-Satzes statt, sondern erst *nach* der Verarbeitung des Pronomens. Machen wir uns das an einem Beispiel klar: In einem Satz wie *Linda entschuldigt sich bei David, weil...* erwartet man eigentlich ein Ereignis, das von Linda ausgegangen ist (und damit das Pronomen *sie*). Wenn erwartungskonträr eine Fortsetzung über *David* folgt (wie *Linda entschuldigte sich bei David, weil er unschuldig war.*, vgl. Koornneef & van Berkum, 2006), dann wird diese Inkongruenz direkt durch die Fortsetzung aufgelöst und man hat kaum Zeit, sich über die Inkongruenz zu wundern. In einem langen inkongruenten Stimulus-Satz wie *Linda entschuldigte sich bei David, weil er nach Meinung aller anwesenden Zeugen nicht Schuld war.* ist dies anders. Hier aktiviert die zusätzliche Phrase *nach Meinung aller anwesenden Zeugen* zusätzliche Szenario-Informationen, welche die Inkonsistenz salienter machen, bis sie schließlich erst ganz am Ende aufgelöst wird. Das heißt, in elaborierten, aber inkonsistenten Fortsetzungen wird das Szenario durch die ic-inkongruente Fortsetzung instabil und inkohärent. In den kurzen, nicht-elaborierten Versionen bleibt dieser Effekt aus, weil dem Sprachsystem sowohl die notwendigen Informationen als auch die notwendige Zeit zur Verarbeitung der Inkonsistenz fehlt.

Da der Grundmechanismus in beiden Fällen der gleiche ist, ist davon auszugehen, dass sowohl im Fall der diskursiven Ergänzung als auch im Fall einer Verlängerung der Stimulussätze ein IC-Fokus-Effekt entsteht, weil zusätzliche szenario-elaborierende Informationen eingeführt und verarbeitet werden können. Dieser führt dann in einem nächsten Schritt zur beschriebenen Trade-off-Relation mit dem IC-Integrations-Effekt. In Kap. 5 wurde bereits darauf hingewiesen, dass es durch die zur Auswertung benötigte Stimulusmenge nicht möglich war, mit diskursiv gerahmten Kontexten zu arbeiten. Allerdings wurden ja bereits im Vorfeld der Untersuchung zu jedem Stimulussatz zwei Itemversionen konstruiert. Sie unterschieden sich lediglich durch die Einfügung einer präpositionalen Ergänzungsphrase, welche die IC-Kongruenz der Stimuli jedoch nicht gefährdet. In Experiment Ib soll nun die Itemversion mit zusätzlicher Präpositionalphrase verwendet werden. Die Vorhersage ist dabei, dass sich das gleiche Ergebnismuster zeigt wie in Experiment Ia, jedoch zusätzlich auch ein IC-Fokus-Effekt entsteht. Dieser sollte ausschließlich in der gender-disambiguierten Bedingung auftreten und selektiv nur den shift-Parameter beeinflussen. In der ambigen Bedingung sollte es diesen Effekt hingegen nicht geben. Dafür sollte in dieser Bedingung der IC-Integrations-

Effekt umso stärker ausfallen, weil die Inkongruenz nicht früh entdeckt werden kann und der Auflösungsmechanismus plötzlich mit einer inkonsistenten Fortsetzung konfrontiert ist.

In Experiment Ib wurde jedoch noch eine weitere, weniger zentrale Veränderung vorgenommen, die sich auf die Verifikationsaufgabe bezieht. In der Diskussion zu Experiment Ia wurde bereits darauf hingewiesen, dass es eine Alternativerklärung für den Erleichterungseffekt in den gender-disambiguierten Bedingungen gibt. Diese geht davon aus, dass disambiguierte Sätze gar nicht bis zum Ende rezipiert werden, sondern sozusagen bereits vorher „weggedrückt“ werden. Es wurde bereits darauf hingewiesen, dass dies eher unwahrscheinlich erscheint, weil sich in der anschließenden Verifikationsaufgabe für diese Bedingungen keine höheren Fehlerwerte ergaben. Im Gegenteil: Gerade in den disambiguierten Bedingungen wurde besonders korrekt und schnell geantwortet. Dieses Argument ist jedoch vergleichsweise indirekt, da es ja ebenso möglich, dass sich im disambiguierten Fall schneller inferieren lässt, ob die gezeigte Aussage wahr oder falsch ist. In Experiment Ib wurde deswegen eine neue Art von Verifikationsstimuli als eine Art Sicherheitsmaßnahme eingeführt: In dieser Version wurde nicht nur eine falsche Zuordnung der Aktanten zu dem im Nebensatz spezifizierten Ereignissen präsentiert, sondern die Nebensatz-Proposition als Ganzes ausgetauscht. Das heißt, auf einen Satz wie *Linda entschuldigte sich bei David, weil er unschuldig war*. folgte manchmal eine Aussage wie *David war müde*. Wenn die Vpn in der disambiguierten Bedingung die Sätze gar nicht zu Ende lesen, dann sollten gerade in dieser Bedingung besonders viele Fehler gemacht werden.

7.2 Methode

7.2.1 Stichprobenbeschreibung und -bereinigung

Allgemeine Angaben. An Experiment Ib nahmen insgesamt $N = 27$ Versuchspersonen (25 Frauen, 2 Männer) teil. Keine der Vpn hatte bereits an Experiment Ia teilgenommen. Das durchschnittliche Alter lag bei $M = 23.0$ Jahren ($SD = 4.4$). Es handelte sich hauptsächlich um Studierende der Psychologie (durchschnittlich im 4. Semester), die für ihre Teilnahme eine Versuchspersonenstunde oder eine Vergütung von 7,-€ erhielten. Alle Teilnehmer/innen berichteten, dass sie normal- oder korrigiert normalsichtig sind, und gaben Deutsch als Muttersprache an. Eine Person gab an, über eine nicht-korrigierbare Sehschwäche zu verfügen. Sie wurde deshalb von allen weiteren Analysen ausgeschlossen.

ELVES. Insgesamt wies die Stichprobe wiederum in allen drei eingesetzten hierarchie-niedrigen Subtests von ELVES ein durchschnittliches und unauffälliges Ergebnis auf. Der Mittelwert für den Subtest *Satzverifikation* betrug $M = 16.0$ ($SD = 3.7$), für den Subtest *Sinn-*

haftigkeit Sätze lag er bei $M = 15.8$ ($SD = 4.2$) und für den Subtest *Sinnhaftigkeit Satzfolgen* bei $M = 16.3$ ($SD = 3.8$). Die drei Subskalen korrelierten hoch miteinander und wurden deswegen zu einem gemeinsamen Index zusammengefasst (Cronbachs $\alpha = .86$; $M = 15.9$, $SD = 3.5$). Für eine Person konnte kein Wert für den Subtest *Sinnhaftigkeit Satzfolgen* bestimmt werden, weil sie zu häufig unter dem angesetzten Cut-off-Wert von 500 ms geantwortet hatte. Diese Person wurde deshalb von allen weiteren Analysen ausgeschlossen. Eine Versuchsperson wies für alle drei ELVES-Subskalen einen erhöhten multivariaten Leverage-Wert auf ($h_{ii} = .37$, $p < .01$; vgl. Cohen et al., 2003), der darauf zurückgeführt werden konnte, dass sie in zwei der Subtests unterdurchschnittliche Werte aufwies. Da ihr Antwortmuster jedoch ansonsten unauffällig war, wurde sie nicht von der Analyse ausgeschlossen.

Geschwindigkeit und Akkuratheit. Die mittlere Reaktionszeit und durchschnittliche Akkuratheit in den vier Untersuchungsbedingungen wurden pro Person zu einem Vektor mit insgesamt acht Werten zusammengefasst, für den dann ein multivariater Leverage-Wert bestimmt wurde. Zwei Vpn wiesen problematische Werte auf ($h_{ii} > .70$, $p < .01$). Eine Inspektion ihrer Antwortmuster zeigte, dass diese Personen sehr viele Fehler machten und teilweise sehr schnell antworteten: Ihr durchschnittlicher Akkuratheitswert lag bei 82 %, in der nicht gender-disambiguierten Bedingung sogar nur bei 63 %. Da diese Vpn offensichtlich instruktionswidrig gearbeitet hatten, wurden sie von allen weiteren Analysen ausgeschlossen.

7.2.2 Material

Sätze zur pronominalen Auflösung. Für eine genaue Beschreibung des eingesetzten Stimulusmaterials sei an dieser Stelle wieder auf Kap. 5 verwiesen. Im Gegensatz zu Experiment Ia kam diesmal die Fassung mit Präpositionalphrase zum Einsatz, d.h. in alle Sätze wurde direkt hinter dem Pronomen eine zusätzliche Präpositionalphrase eingebaut, die in allen Itemversionen identisch blieb und so eine gewisse Konstanz gewährleistete. Daraus ergibt sich dann das folgende Schema (vgl. (1) und das Beispiel in (2)):

(1) [NP1 IC-Verb NP2]_{HS-VP}, [weil] [Pro] [P (Det) (Adj) N]_{PP} [[(Det) (Adj/Adv) N]_{NS-NP} V (Mod) (Aux)]_{NS-VP}

(2) [Dominik nervte Daniela]_{HS}, [weil er]_{Pro} [nach der langen Diskussion]_{PP} [keine Ruhe geben wollte]_{VP}.

Es wurden alle 160 Stimulussätze verwendet, die sich gleichteilig sowohl aus NP1- bzw. NP2-Verben als auch Zustands- bzw. Handlungsverben zusammensetzten (vgl. Kap. 5). Filler-Items kamen nicht zum Einsatz. Für alle Items wurden vier Versionen hergestellt, die jeweils den vier Untersuchungsbedingungen entsprachen (mit/ohne Gender-Cue bzw. mit/ohne IC-Cue). Die Items wurden mittels der bereits in Kap. 6.2 dargestellten 16 Itemlisten (bzw. den 4 „Super“-Listen, die jeweils vier Itemlisten der verschiedenen Stimulustypen in sich ver-

einten) mit den Untersuchungsbedingungen koordiniert, wobei die Zuweisung durch ein Lateinisches Quadrat mit einem Listen-Faktor ausbalanciert wurde.

Verifikationsaussagen. Zu jedem Satz wurde eine Aussage konstruiert, die von den Vpn nach der Auflösung des Pronomens zu verifizieren war. Das Konstruktionsprinzip war dabei das gleiche wie bei Experiment Ia. Das heißt, im Prinzip handelte es sich bei den zu verifizierenden Aussagen um pronominal aufgelöste Versionen des Nebensatzes. Dieses Satzfragment enthielt nicht die neu eingefügte Präpositionalphrase. Im Vergleich zu Experiment Ia divergierten Auflösungs- und Verifikationsaussage also.

Wieder wurde zu jeder Verifikationsaussage eine korrekte und eine inkorrekte Version hergestellt, die Verifikationsleistung wurde sowohl über die Akzeption korrekter als auch die Ablehnung inkorrektur Aussagen definiert. In Experiment Ib wurde jedoch noch eine dritte Version der Verifikationsaussage hergestellt, die als eine Art Kontrollbedingung fungierte: In dieser so genannten „falschen“ Version wurde die Verbalphrase des Nebensatzes ausgetauscht, d.h. die Semantik von Originalsatz und Verifikationsaussage war schlichtweg unterschiedlich. Da es jedoch vorkommen kann, dass die Verifikationsaussage trotzdem mit dem Originalsatz kompatibel ist, wurde in der Instruktion entsprechend darauf hingewiesen, dass es wirklich nur darum geht, ob die zu verifizierende Aussage im vorhergehenden Satz vorkam ist. Die sowieso schon schwammige Grenze zwischen Verifikations- und Rekognitionsaufgabe wird damit noch unklarer. In (3) sind die drei Versionen der Verifikationsaussage des oben genannten Beispielsatzes einander gegenübergestellt:

- (3a) korrekt Version: Dominik wollte keine Ruhe geben.
- (3b) inkorrekte Version: Daniela wollte keine Ruhe geben.
- (3c) „falsche“ Version: Dominik war das Thema Leid.

Jede Vpn bearbeitete alle 160 Verifikationsaussagen. Ein Viertel aller Aussagen wurde dabei in der „falschen“ Version dargeboten, die restlichen 120 Aussagen gleichmäßig in der korrekten und in der inkorrekten Version. Es wurde darauf geachtet, dass über alle Sätze hinweg gleich häufig der erste und der zweite Aktant als Subjekt der Verifikationsaussage verwendet wurden. Die Zuordnung der einzelnen Items zu den verschiedenen Versionen erfolgte wiederum mittels der 16 Itemlisten, die durch ein Lateinisches Quadrates ausbalanciert wurden. Im Gegensatz zur Zuweisung bei der pronominalen Auflösungsaufgabe war es durch die Vielzahl der Kombinationsmöglichkeiten von Untersuchungsbedingung, Formulierungsversion der Verifikationsaussage und den verschiedenen Itemmerkmalen nicht möglich, jede der vier Zellen der Untersuchung gleich häufig zu besetzen. Es wurde jedoch darauf geachtet, dass für jede Zelle des Grunddesigns mindestens 20 Aussagen zur Verfügung standen.

7.2.3 Untersuchungsablauf

Abgesehen von der Tatsache, dass unterschiedliche Materialien verwendet wurden, war der Untersuchungsablauf vollkommen identisch aufgebaut wie Experiment Ia. Durch die Präpositionalphrase waren die einzelnen Items ein wenig länger, weshalb der Text bei der Darbietung der Aussagen wie folgt umgebrochen wurde: [Subjekt] [Verb] [Objekt], // [weil] [Pro] [PP] // [VP]. Da die Einfügung der Präpositionalphrase natürlich auch die Lesezeit eines Items beeinflusst, wurde die maximale Lesezeit, nach der eine Rückmeldung schneller zu lesen dargeboten wurde, auf 6 Sekunden angehoben. Wegen der umfangreicheren Stimuli dauerte die Untersuchung auch ca. 5 Minuten länger.

7.2.4 Design und Abhängige Variablen

Grunddesign, ergänzende Faktoren sowie Abhängige Variablen waren die gleichen wie in Experiment Ia. Auch die Relation der beiden eingesetzten Aufgabentypen sowie die Relevanz von Korrektheits- und Geschwindigkeitsmaßen zur Beantwortung der zugrunde liegenden Fragestellung waren ähnlich einzuschätzen.

7.3 Ergebnisse

7.3.1 Pronominale Auflösungsaufgabe

7.3.1.1 Akkuratheit: Anteil intendierter pronominaler Auflösungen

Ergebnisse. Abb. 7.1 zeigt die „Akkuratheit“ (Anteil von Antworten in der intendierten Lesart) der von den Vpn gewählten pronominalen Auflösung für die vier Untersuchungsbedingungen.

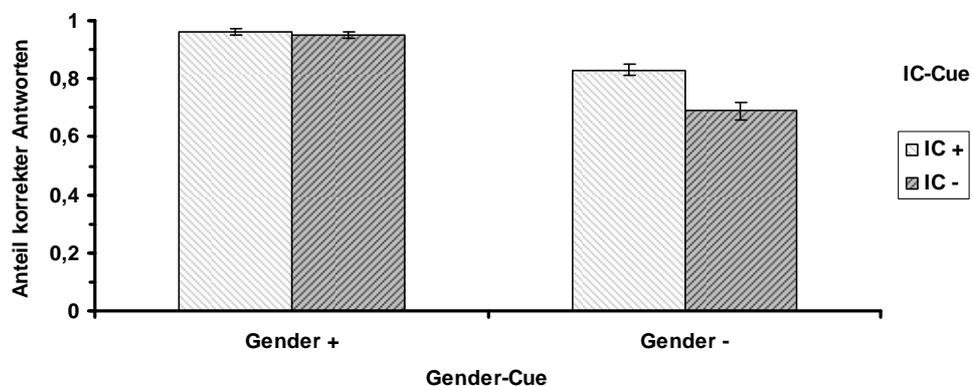


Abb. 7.1: Anteil intendierter pronominaler Auflösungen in den vier Untersuchungsbedingungen.

Eine messwiederholte ANOVA mit den Faktoren *Gender-Cue* (2: vorhanden vs. nicht vorhanden) und *IC-Cue* (2: vorhanden vs. nicht-vorhanden) ergab einen sehr deutlichen Unterschied zwischen gender-disambiguierten und nicht-gender-disambiguierten Sätzen ($F(1,22) = 107.30, p < .001, \eta_p^2 = .830$; $F(1,159) = 202.11, p < .001, \eta_p^2 = .560$), da die disambiguierten Sätze um ca. 20 % korrekter beantwortet wurden. Desgleichen gab es einen Haupteffekt für die implizite Verbkausalität ($F(1,22) = 77.99, p < .001, \eta_p^2 = .780$; $F(1,159) = 40.33, p < .001, \eta_p^2 = .202$), der jedoch durch den signifikanten Interaktionsterm ($F(1,22) = 32.46, p < .001, \eta_p^2 = .596$; $F(1,159) = 29.95, p < .001, \eta_p^2 = .158$) noch weiter qualifiziert wurde: Wie auch in Abb. 7.1 deutlich zu erkennen, wirkte sich die Verbkausalität nur dann auf das Antwortverhalten aus (ca. 14 %), wenn der Gender-Cue beide Interpretationen grammatikalisch zuließ ($F(1,22) = 58.99, p < .001, \eta_p^2 = .728$; $F(1,159) = 42.48, p < .001, \eta_p^2 = .211$). Wenn die Auflösung eines Pronomens jedoch durch den Gender-Cue festgelegt war, war dieser Effekt nicht signifikant (ca. 2%; $F(1,22) = 3.79, p = .064, \eta_p^2 = .147$; $F(1,159) = 2.29, p = .132, \eta_p^2 = .014$).

Arcussinus-Transformation. Um eine Normalverteilung der Residuen zu gewährleisten, wurde die Analyse nochmals mit arcussinus-transformierten Werten durchgeführt. Die vollständigen Ergebnisse sind in Appendix C.2.1 einsehbar. Das Ergebnismuster war jedoch vollkommen identisch zu dem berichteten.

Zusammenfassung. Das Ergebnismuster für das konkrete Verstehensprodukt war absolut identisch zu dem der pronominalen Auflösung in Experiment Ia: Es liegt ein starker Verarbeitungsvorteil für gender-disambiguierte Pronomen vor und ein schwächerer Effekt des IC-Cues, wenn die Referenz des Pronomens nicht durch die morphosyntaktischen Informationen determiniert war. Das übergreifende Akkuratheitsniveau von ca. 86% war ebenfalls ungefähr vergleichbar mit Experiment Ia, d.h. die Einfügung der Präpositionalphrase hat das finale Verstehensprodukt weder wesentlich verbessert noch deutlich erschwert.

7.3.1.2 Mittlere Reaktionszeit

Datenbereinigung. Für die pronominale Auflösung wurde aufgrund des verlängerten Itemmaterials ein Wert von 1500 ms als unterer Cut-off-Wert angesetzt. Kein Datenpunkt fiel unter diesen kritischen Grenzwert. Weiterhin wurden alle Werte, die 4 Standardabweichungen vom bedingungsspezifischen Mittelwert einer Person abwichen, durch diesen ersetzt. Diese Maßnahme führte zu einer Ersetzung von ca. 0.2 % aller Datenpunkte.

Ergebnisse. Abb. 7.2 zeigt die mittlere Reaktionszeit bis zur Auflösung der pronominalen Anapher in den vier Untersuchungsbedingungen.

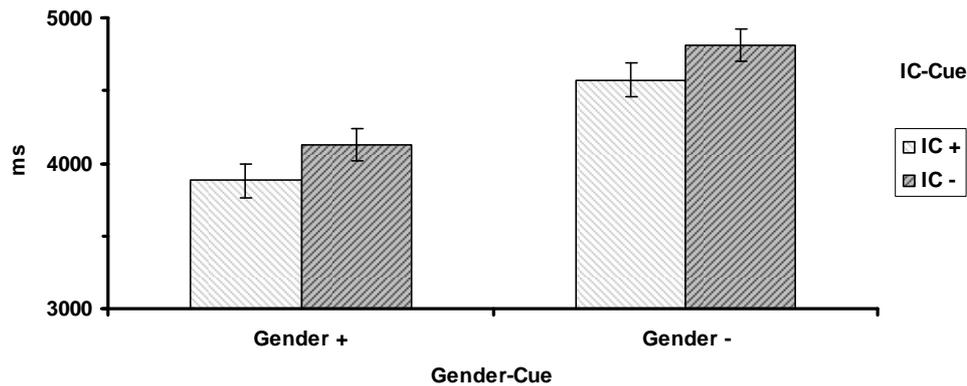


Abb. 7.2: Mittlere RT für die pronominale Auflösung in den vier Untersuchungsbedingungen.

Die messwiederholte ANOVA zeigte einen äußerst varianzstarken Haupteffekt für den Faktor *Gender-Cue* ($F(1,22) = 93.44, p < .001, \eta_p^2 = .809$; $F(1,159) = 385.40, p < .001, \eta_p^2 = .708$): Sätze mit zwei gleichgeschlechtlichen Referenten für das Pronomen brauchten ca. 690 ms länger, um erfolgreich aufgelöst zu werden, als Sätze mit nur einem grammatikalisch möglichen Referenten. Darüber hinaus gab es jedoch auch einen Effekt für den Faktor *IC-Cue* ($F(1,22) = 49.94, p < .001, \eta_p^2 = .694$; $F(1,159) = 36.12, p < .001, \eta_p^2 = .185$): Sätze, bei denen die explizit gegebene Begründung mit der IC kongruent war, wurden ca. 250 ms schneller verarbeitet. Die Interaktion beider Faktoren war hingegen nicht signifikant ($F(1,22) < 1, p = .917, \eta_p^2 = .001$; $F(1,159) < 1, p = .953, \eta_p^2 = .001$), d.h. der Kausalitätseffekt war in beiden Gender-Cue-Bedingungen gleich stark ausgeprägt.

Re-Analyse der „korrekten“ Antworten. Um zu überprüfen, ob das berichtete Effektmuster evtl. durch Unterschiede zwischen falsch und richtig beantworteten Items moderiert wurde, wurde die Analyse nochmals nur mit den „korrekt“ beantworteten Items durchgeführt. Das Ergebnismuster (s. Appendix C.2.2) blieb identisch erhalten; beide Effekte waren sogar noch stärker ausgeprägt.

Zusammenfassung. Die Analyse der mittleren Reaktionszeiten erbrachte ein sehr ähnliches Muster wie in Experiment Ia: Beide Cues hatten einen deutlichen Effekt auf die Zeit zur Auflösung des Pronomens, interagierten aber nicht miteinander. Wieder war der Gender-Cue der varianzstärkere der beiden Effekte, allerdings war der IC-Cue im Vergleich zu Experiment Ia deutlich stärker geworden (um ca. 40%). Die Stärke des Gender-Cues stieg zwar auch ein wenig an, allerdings wesentlich schwächer (ca. 6%). Die mittleren Reaktionszeiten waren um ca. 1000 ms länger als bei den vergleichbaren Sätzen ohne Präpositionalphrase aus Experiment Ia. Aufgrund der Tatsache, dass jede Präpositionalphrase vier Wörter umfasste, ist dies ein durchaus realistischer Wert, wenn man eine durchschnittliche Lesezeit von 250 ms pro Wort ansetzt (vgl. z.B. Just & Carpenter, 1980).

7.3.1.3 Reaktionszeitverteilung: Weibull-Parameter

Parameterschätzung. Das Vorgehen bei der Parameterschätzung entsprach dem in Kap. 3.4 skizzierten Vorgehen. Als Startwert für den scale-Parameter wurde ein Wert von $\lambda = \theta^{\beta} = 0.5$ angesetzt. Die verwendeten Parameter der beiden Prior-Verteilungen des scale-Parameters waren $a_1 = 2.0$, $b_1 = 0.1$, $a_2 = 2.0$ und $b_2 = 0.03$.

Modellgüte. Die Passung der Weibull-Verteilung für die einzelnen RT-Verteilungen wurde mittels des in Kap. 3.5 beschriebenen Chi-Quadrat-Verfahrens beurteilt. Der Fit war insgesamt zufrieden stellend: Der mittlere Chi-Quadrat-Wert lag bei $\chi^2(3) = 5.05$, $p < .29$. Vier der insgesamt 92 (= 23 x 4) Einzelverteilungen (= 4.3 %) waren auf dem 1%-Prozent-Niveau signifikant.

Durchschnittliche Parameterausprägung. Abb. 7.3 zeigt die drei Weibull-Parameter der Reaktionszeit-Verteilung und ihre Beziehungen zueinander (shift- und scale-Parameter in Sekunden).

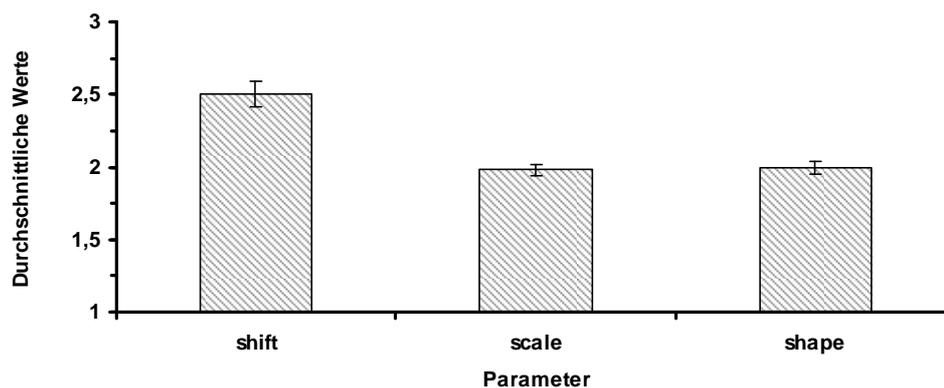


Abb. 7.3: Durchschnittliche Weibull-Parameter der RT-Verteilungen der pronominalen Auflösung.

Der durchschnittliche shift-Parameter lag bei 2.52 s, der durchschnittliche scale-Parameter bei 2.02 s und der durchschnittliche shape-Parameter bei 1.97. Wie zu erwarten, schlug sich also die unterschiedliche Länge des Stimulus-Materials hauptsächlich im shift-Parameter nieder, der um ca. 0.7 s höher ausfiel als in Experiment Ia. Für den scale- und den shape-Parameter ergaben sich hingegen fast identische Werte wie in Experiment Ia.

Diese drei Parameter wurden jeweils mittels einer messwiederholten ANOVA mit den beiden Faktoren *Gender-Cue* (2: *vorhanden* vs. *nicht vorhanden*) und *IC-Cue* (2: *vorhanden* vs. *nicht vorhanden*) auf Unterschiede geprüft. Dabei ergaben sich folgende Ergebnisse:

shift-Parameter. Es gab einen starken Haupteffekt für den Faktor *Gender-Cue* ($F(1,22) = 50.86$, $p < .001$, $\eta_p^2 = .698$): Der shift-Parameter für Sätze mit zwei gleichgeschlechtlichen Personen war ca. 380 ms höher als bei Sätzen, bei denen nur eine Person mit dem Pronomen gender-kongruent war. Der Haupteffekt für den IC-Cue war ebenfalls signifikant ($F(1,22) =$

9.25, $p < .006$, $\eta_p^2 = .296$): Sätze mit ic-kongruenten Begründungen hatten einen um ca. 100 ms niedrigeren shift-Parameter. Im Gegensatz zu Experiment Ia war die Interaktion der beiden Faktoren ebenfalls signifikant ($F(1,22) = 6.78$, $p < .016$, $\eta_p^2 = .236$): Wenn ein Satz bereits durch den Gender-Cue disambiguiert war, dann gab es einen substantiell positiven Effekt (von ca. 200 ms) des Kausalitäts-Cues ($F(1,22) = 17.14$, $p < .001$, $\eta_p^2 = .438$), während dieser Effekt in der Bedingung *Gender-* nicht signifikant war ($F < 1$; vgl. Abb. 7.4).

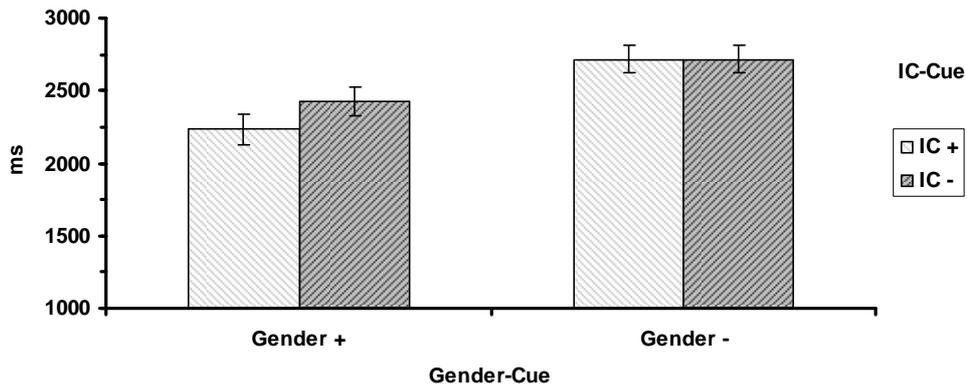


Abb. 7.4: Ausprägung des shift-Parameters der RT-Verteilungen der pronominalen Auflösung.

scale-Parameter. Für den scale-Parameter gab es einen Haupteffekt des Faktors *Gender-Cue* ($F(1,22) = 28.88$, $p < .001$, $\eta_p^2 = .568$): Reaktionszeitverteilungen für die Sätze, bei denen das Pronomen ambig war, wiesen einen um ca. 330 ms höheren scale-Parameter auf als Sätze mit disambiguierten Pronomen. Gleichzeitig gab es einen Haupteffekt für den Faktor *IC-Cue* ($F(1,22) = 13.50$, $p < .001$, $\eta_p^2 = .380$): In den Untersuchungsbedingungen, in denen die implizite Verbkausalität und die explizite Auflösung übereinstimmten, fiel der scale-Parameter um ca. 170 ms niedriger aus. Die Interaktion der beiden Faktoren war ebenfalls signifikant ($F(1,22) = 6.90$, $p < .015$, $\eta_p^2 = .239$): Wie in Abb. 7.5 zu erkennen, hatte der IC-Cue in der Bedingung *Gender-* einen Effekt von ca. 270 ms ($F(1,22) = 24.22$, $p < .001$, $\eta_p^2 = .524$). In der Bedingung *Gender+* gab es zwar auch einen Effekt des IC-Cues von ca. 70 ms, der jedoch nicht signifikant war ($F < 1.2$, $p = .295$, $\eta_p^2 = .050$).

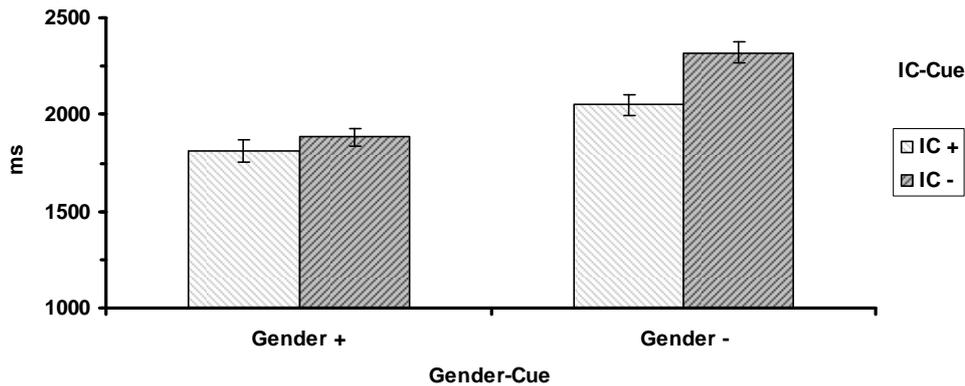


Abb. 7.5: Ausprägung des scale-Parameters der RT-Verteilungen der pronominalen Auflösung.

shape-Parameter. Kein Effekt war signifikant (alle $F < 1.3$).

Re-Analyse der „korrekten“ Antworten. Das berichtete Effektmuster bildete sich noch prägnanter ab, da nun auch der Interaktionsterm des scale-Parameters signifikant wurde. Eine andere Abweichung betrifft den shape-Parameter, für den es einen leichten Effekt des Gender-Cues gab. Ein vollständiger Bericht der Ergebnisse findet sich in Appendix C.2.3.

Überblick Effekte. Der entscheidende Befund der vorliegenden Analyse im Unterschied zu Experiment Ia war, dass die beiden Cues nicht unabhängig voneinander sind, sondern vielmehr interagierten. Insbesondere schien der Effekt des IC-Cues von der Ausprägung des Gender-Cues abhängig zu sein, weil er sich dann in unterschiedlichen Reaktionszeitkomponenten niederschlug. Diese differenziellen Effekte sollen abschließend noch einmal miteinander kontrastiert werden, damit das Muster noch deutlicher wird.

Abb. 7.6 zeigt die Effekte der Gender-Manipulation auf die drei Parameter der Weibull-Verteilung (shift- und scale-Parameter in Sekunden angegeben). Ähnlich wie in Experiment Ia ist deutlich zu erkennen, dass der Verarbeitungsvorteil für Sätze, in denen nur einer der beiden Referenten mit einem Pronomen gender-kompatibel war, sowohl durch eine Reduktion des shift- als auch eine Reduktion des scale-Parameters zustande kam. Der Teileffekt für den shift-Parameter war mit 380 ms etwas stärker ausgeprägt als der Teileffekt des scale-Parameters mit 330 ms. Auf den shape-Parameter hatte der Gender-Cue hingegen keinen wesentlichen Einfluss.

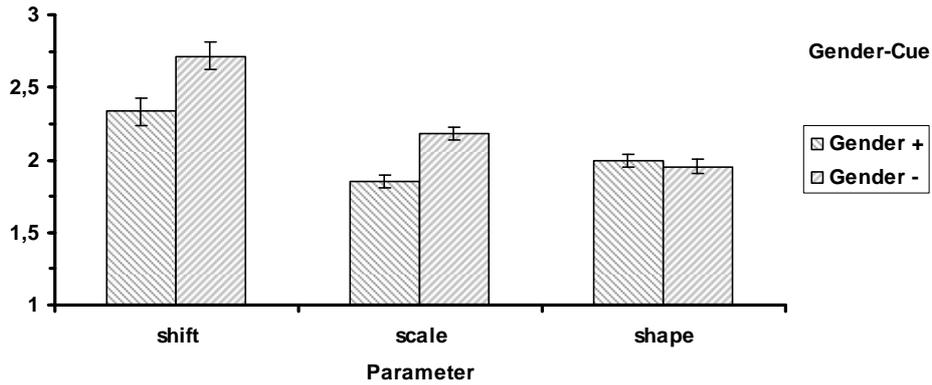


Abb. 7.6: Effekte des Gender-Cues auf die Weibull-Parameter der RT-Verteilungen der pronominalen Auflösung.

Abb. 7.7 zeigt die Effekte der Kausalitäts-Manipulation auf die drei Parameter der Weibull-Verteilung, wenn der Satz gender-ambig war. Genauso wie in Experiment Ia, beeinflusste der IC-Cue selektiv nur den scale-Parameter der Verteilung. Für den shift- und den shape-Parameter gab es hingegen keinen signifikanten Effekte.

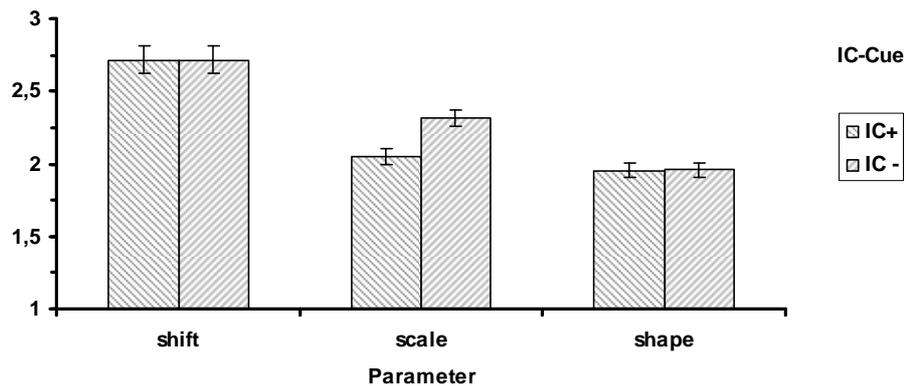


Abb. 7.7: Effekte des IC-Cues auf die Weibull-Parameter der RT-Verteilungen der pronominalen Auflösung in der Bedingung Gender-Cue -.

Abb. 7.8 zeigt die Effekte des IC-Cues auf die drei Parameter der Weibull-Verteilung, wenn im Satz gleichzeitig auch ein Gender-Cue vorhanden war. Im scharfen Kontrast zum Effekt des IC-Cues in der unaufgelösten Bedingung zeigte sich ein ähnliches Bild wie für den Gender-Cue: Es gab einen starken Effekt für den shift-Parameter von ca. 200 ms und einen schwächeren für den scale-Parameter von ca. 70 ms. Es gab keinen relevanten Effekt für den shape-Parameter.

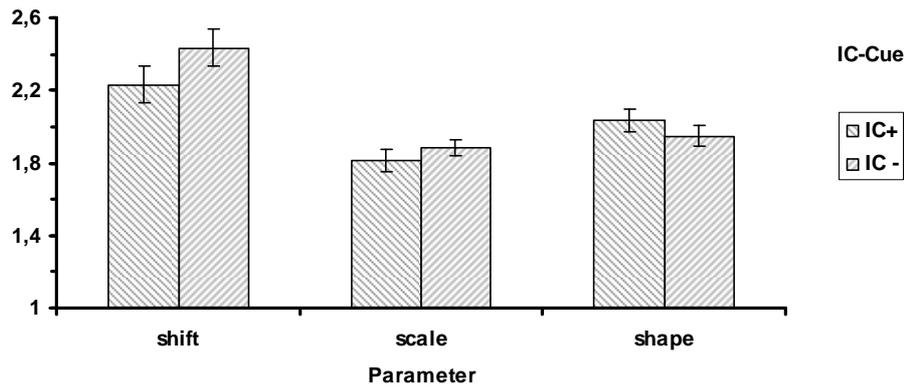


Abb. 7.8: Effekte des IC-Cues auf die Weibull-Parameter der RT-Verteilungen der pronominalen Auflösung in der Bedingung Gender-Cue +.

Verteilungen. Abb. 7.9 zeigt den Effekt der Gender-Manipulation auf die Verteilung der Antwortzeiten: Man kann deutlich sehen, dass die gesamte Verteilung in der Bedingung *Gender-* nach rechts rutschte (Effekt für den shift-Parameter). Gleichzeitig ist auch erkennbar, dass sich durch die Gender-Manipulation der Masseschwerpunkt der Verteilung verschob (Effekt auf dem scale-Parameter).

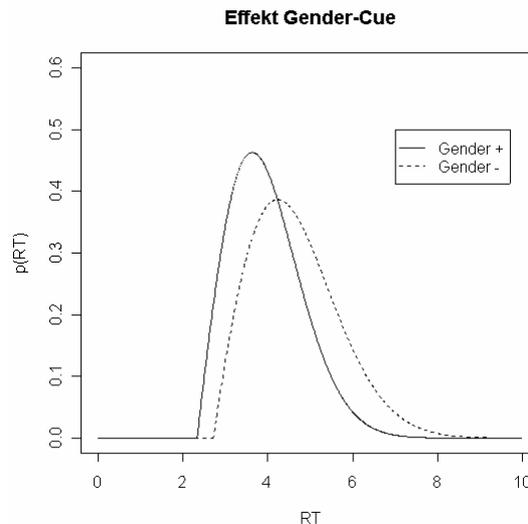


Abb. 7.9: Unterschiede zwischen den geschätzten Weibull-Verteilungen für die Manipulation des Gender-Cues.

Abb. 7.10 a) und b) zeigen die entsprechenden Effekte des IC-Cues auf die Verteilung der Antwortzeiten bei An- und Abwesenheit des Gender-Cues. In Teilabbildung a) wird deutlich, dass der Effekt des IC-Cues in der ambigen Bedingung allein im scale-Parameter lag, da beide Verteilungen die gleiche minimale Reaktionszeit aufwiesen, aber einen deutlich unterschiedlichen Masseschwerpunkt. In Teilabbildung b) ist hingegen erkennbar, dass die Minima der beiden Verteilungen deutlich divergierten. Gleichzeitig gab es aber auch ein Unterschied in der Höhe der „Spitzen“ der beiden Verteilungen, der allerdings geringer ausfiel als in der ambigen Bedingung.

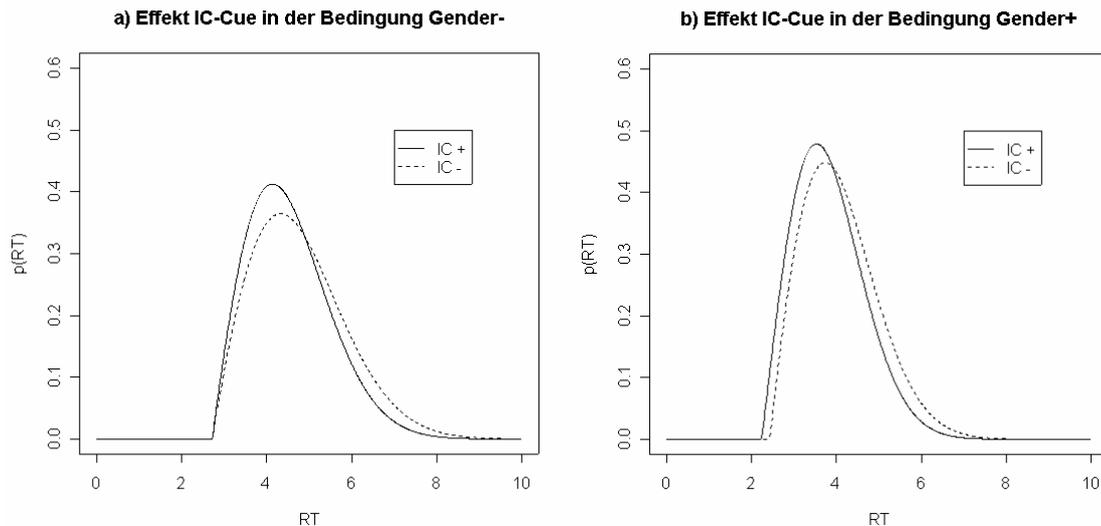


Abb. 7.10: Unterschiede zwischen den geschätzten Weibull-Verteilungen für die Manipulation des IC-Cues in a) der Bedingung Gender – und b) in der Bedingung Gender +.

7.3.1.4 Diskussion pronominale Auflösungsaufgabe

Insgesamt weist das Ergebnismuster für die pronominale Auflösungsaufgabe eine große Ähnlichkeit zu dem in Experiment Ia auf. Das gilt zum einen für die Akkuratheit der Auflösung, bei der es praktisch kein Unterschied zu Experiment Ia gibt. Dies ist deshalb erstaunlich, weil das Stimulusmaterial in Experiment Ib durch die Einfügung der Präpositionalphrase deutlich länger und damit auch komplexer geworden ist. Eigentlich wäre zu erwarten gewesen, dass die Akkuratheit der Auflösung durch diese zusätzliche Ressourcenbelastung leidet, und zwar v.a. in der Bedingung ohne expliziten Gender-Cue. Eine Erklärungsmöglichkeit dafür, dass dies hier nicht passierte, ist z.B., dass durch die eingefügte Präpositionalphrase zusätzliche Informationen verfügbar wurden, die die explizite Begründung des Nebensatzes in ihrer Plausibilität stärkten und deswegen zwingender machten. Damit könnte man zusätzlich noch die Tatsache erklären, dass der IC-Effekt an Stärke gewonnen hat.

Das eigentliche Novum von Experiment Ib ist jedoch sicherlich, dass IC- und Gender-Cue nun miteinander interagierten, und zwar dahingehend, dass der IC-Effekt in der disambiguierten Bedingung zusätzlich den shift-Parameter der Reaktionszeitverteilung beeinflusste. Wenn das Pronomen hingegen ambig war, dann war der IC-Effekt weiterhin nur im scale-Parameter verortet. Dieser Effekt war direkt auf die Einfügung der Präpositionalphrase zurückzuführen, weil dies die einzige relevante Veränderung war, die im Übergang von Experiment Ia zu Experiment Ib vorgenommen wurde. Wie in der Einleitung zu Experiment Ib dargestellt, führt diese dazu, dass in der Präpositionalphrase noch weitere szenario-relevante Informationen eingebracht werden, die den IC-Bias des Verbs verstärken und die Unplausibilität der erwartungskonträreren Fortsetzung salienter machen. Weiterhin wird durch diese Ein-

fügung so etwas wie eine „Unsicherheits-Strecke“ in der Satzmitte geschaffen, die überhaupt erst die notwendigen zeitlichen und informationellen Voraussetzungen zur Verfügung stellt, damit sich evtl. Refokussierungsprozesse zeigen können. Wenn hingegen die finalen, die Bedeutung des Pronomens letztendlich determinierenden Informationen direkt folgen, dann kann keine Refokussierung stattfinden oder aber sie wird von den zu diesem Zeitpunkt einsetzenden, höheren kognitiven Prozessen überlagert.

7.3.2 Verifikationsaufgabe

7.3.2.1 Akkuratheit: Anteil korrekt verifizierter Aussagen

Vergleich Aussagetypen. Zunächst wurden in einer separaten Analyse die verschiedenen Aussagetypen (korrekt vs. inkorrekt vs. falsch, vgl. Abschnitt 7.2) miteinander verglichen. Abb. 7.11 zeigt die Akkuratheitswerte für die drei Arten von Verifikationsaussagen.

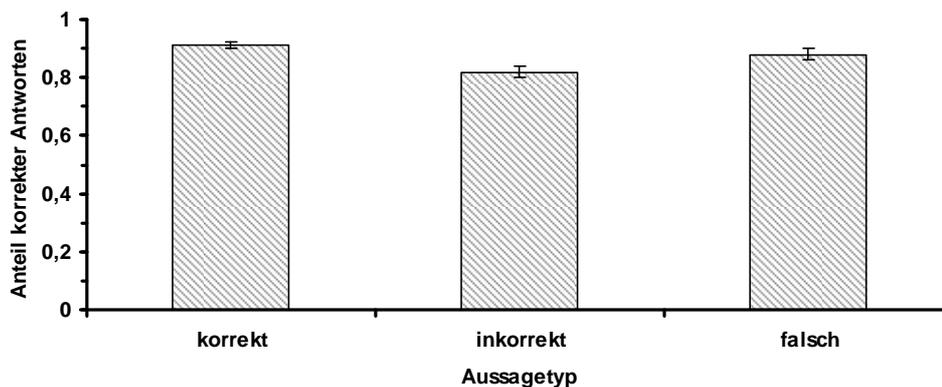


Abb. 7.11: Anteil korrekt verifizierter Aussagen nach Aussagetyp.

Eine einfaktorielle, messwiederholte ANOVA mit den drei Aussagetypen als Stufen ergab global, dass sich die drei Mittelwerte unterschieden ($F(2,44) = 7.72, p < .001, \eta_p^2 = .260$; $F(2,318) = 13.50, p < .001, \eta_p^2 = .078$). Anschließende Einzelvergleiche zeigten, dass sich der Mittelwert der inkorrekten Aussagen ($M = .82, SE_M = .02$) signifikant von den anderen Bedingungen absetzte ($p < .001$), während die Mittelwerte der korrekten ($M = .91, SE_M = .01$) und der falschen Bedingung ($M = .88, SE_M = .02$) nicht voneinander getrennt werden konnten ($p = .293$). Weder für die korrekten noch für die falschen Aussagen gab es einen Unterschied zwischen den gender-disambiguierten und den nicht-disambiguierten Itemversionen (alle $F < 2$).

Ergebnisse. Abb. 7.12 zeigt die Akkuratheit (Hit-Rate) für die Beurteilung der Aussagen, die nach jedem Satz verifiziert werden mussten.

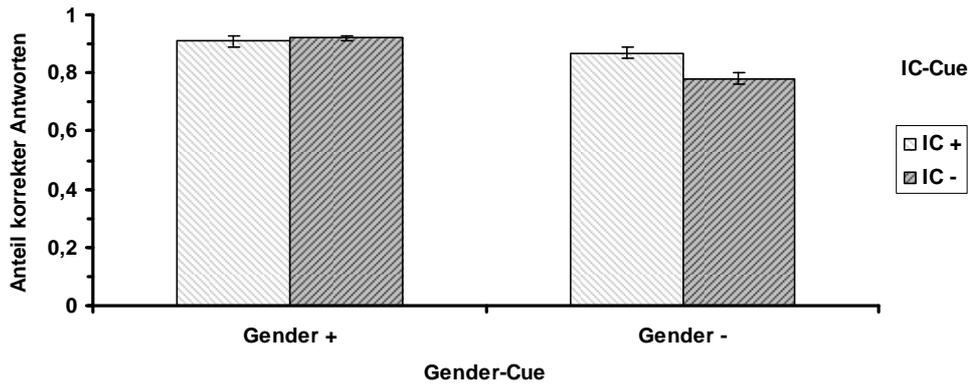


Abb. 7.12: Anteil korrekt verifizierter Aussagen nach Untersuchungsbedingung.

Eine messwiederholte ANOVA ergab einen starken Haupteffekt für den Faktor Gender-Cue ($F(1,22) = 59.30, p < .001, \eta_p^2 = .729$; $F(1,159) = 47.58, p < .001, \eta_p^2 = .230$): Gender-disambiguierte Sätze wiesen um 9 % höhere Verifikationswerte auf als ambige. Auch der Haupteffekt des IC-Cues war signifikant ($F(1,22) = 6.07, p < .022, \eta_p^2 = .216$; $F(1,159) = 10.39, p < .002, \eta_p^2 = .061$), wurde aber durch einen signifikanten Interaktionsterm ($F(1,22) = 21.44, p < .001, \eta_p^2 = .494$; $F(1,159) = 17.19, p < .001, \eta_p^2 = .098$) auf die nicht gender-disambiguierten Sätze eingegrenzt (disambiguiert: $F(1,22) = 21.95, p < .001, \eta_p^2 = .499$; $F(1,159) = 19.54, p < .001, \eta_p^2 = .073$; nicht-disambiguiert: beide $F < 1$). Der IC-Effekt war in der Bedingung *Gender-* ca. 10 % groß, in der Bedingung *Gender+* betrug er nur ca. 1%.

Arcussinus-Transformation. Durch Arcussinus-Transformation wurde geprüft, ob das berichtete Ergebnismuster stabil war. Es ergaben sich keine auffälligen Abweichungen im Ergebnismuster, das vollständig in Appendix C.2.4 eingesehen werden kann.

Zusammenfassung. Das Ergebnismuster für die Akkuratheit entspricht vollkommen dem für Experiment Ia berichteten: Beide Cues wirkten sich ungefähr gleich stark auf die Verifikationsleistung aus; der IC-Cue jedoch nur, wenn ein Satz ambig war. Das grundsätzliche Akkuratheitsniveau war ebenfalls mit dem von Experiment Ia vergleichbar.

Die Tatsache, dass es keinen Unterschied zwischen den korrekten und den falsch formulierten Verifikationsaussagen gab, kann als ein weiterer Hinweis dafür gewertet werden, dass die in den Experimenten Ia und Ib beobachteten Fehlerraten auf das im Vergleich zu anderen Reaktionszeit-Experimenten wesentlich komplexere Stimulusmaterial und allgemeine Charakteristika der Aufgabe zurückgeführt werden können. Der Vergleich mit den falschen Itemversionen, die eine ganz andere Semantik hatten, zeigt nämlich, dass eine Fehlerquote von 90 % offensichtlich die Obergrenze an Akkuratheit ist, die in einer Untersuchung wie der vorliegenden maximal erreicht werden kann. Diese Aussagen wurden genauso gründlich verifiziert wie die korrekten Aussagen.

7.3.2.2 Mittlere Reaktionszeiten

Datenbereinigung. Die Datenbereinigung auf der Itemebene war wiederum sehr sparsam. Als unterer Cut-off-Wert für die Verifikationsaufgabe wurde ein Wert von 200 ms angesetzt. Keine Antwort wies eine derart kurze Reaktionszeit auf. Weiterhin wurden alle Werte, die 4 Standardabweichungen vom bedingungsspezifischen Mittelwert einer Person abwichen, durch diesen ersetzt. Diese Maßnahme führte insgesamt zur Ersetzung von ca. 0.2 % aller Datenpunkte.

Vergleich Aussagetypen. Zunächst wurden in einer separaten Analyse die verschiedenen Aussagetypen (korrekt vs. inkorrekt vs. falsch, vgl. Abschnitt 7.2) miteinander verglichen. Abb. 7.13 zeigt die Reaktionszeit, die durchschnittlich für die drei Arten von Verifikationsaussagen benötigt wurde.

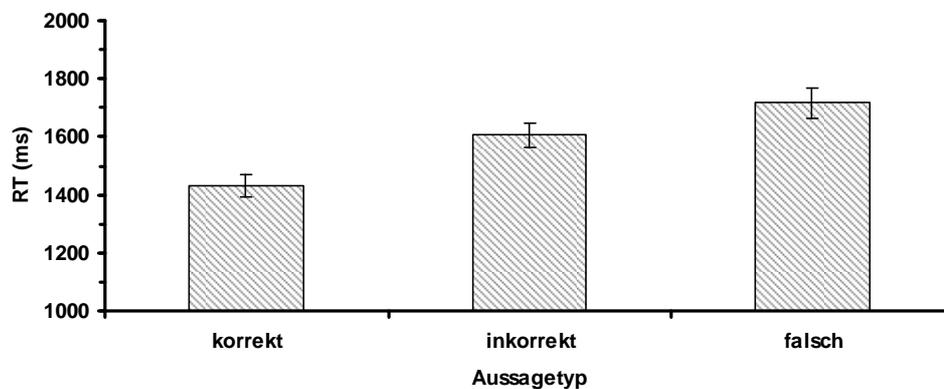


Abb. 7.13: Durchschnittliche Reaktionszeit für die verschiedenen Aussagetypen.

Eine einfaktorielle, messwiederholte ANOVA mit den drei Aussagetypen als Stufen ergab, dass sich die Mittelwerte unterschieden ($F(2,44) = 32.22, p < .001, \eta_p^2 = .260$; $F(2,318) = 13.50, p < .001, \eta_p^2 = .078$). Anschließende Einzelvergleiche zeigten, dass sich alle Mittelwerte signifikant voneinander absetzten. Da die Beantwortung der Itemversion „falsch“ deutlich am meisten Zeit beanspruchte, aber nicht für jede Vp über die vier Untersuchungsbedingungen balanciert werden konnte, wurden Antworten für diesen Itemtyp nicht in die folgenden Analysen einbezogen.

Vergleich Untersuchungsbedingungen. Abb. 7.14 zeigt die mittlere Zeit, die benötigt wurde, um die nach jedem Item dargebotene Aussage zu verifizieren.

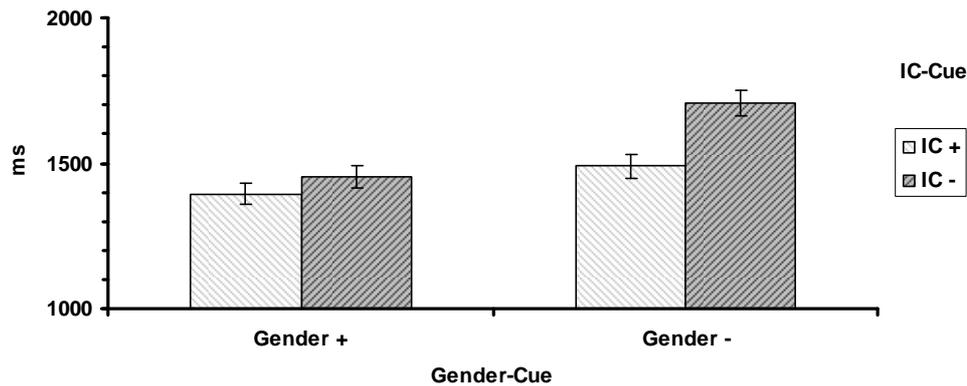


Abb. 7.14: Mittlere RT für die Verifikationsaufgabe in den vier Untersuchungsbedingungen.

Eine zweifaktorielle, messwiederholte ANOVA mit den beiden Faktoren *Gender-Cue* (2: vorhanden vs. nicht-vorhanden) und *IC-Cue* (2: vorhanden vs. nicht vorhanden) mit den über Personen bzw. über alle Items aggregierten RT-Mittelwerten ergab folgende Ergebnisse:

Wie schon für die Verifikationsaufgabe in Experiment Ia war ein varianzstarker Haupteffekt für den Faktor *Gender-Cue* ($F(1,22) = 53.38, p < .001, \eta_p^2 = .708; F(1,159) = 33.62, p < .001, \eta_p^2 = .175$) besonders dominant. Sätze, in denen nur ein vom Geschlecht passender Referent zur Verfügung stand, wurden ca. 170 ms schneller verifiziert. Zusätzlich gab es einen schwächeren, aber konsistenten Effekt für den Faktor *IC-Cue* ($F(1,22) = 37.50, p < .001, \eta_p^2 = .630; F(1,159) = 27.24, p < .001, \eta_p^2 = .146$). Sätze, bei denen explizite Begründung und implizite Verbkausalität kongruent waren, wurden ca. 140 ms schneller verarbeitet. Die Interaktion beider Faktoren war ebenfalls signifikant ($F(1,22) = 15.23, p < .001, \eta_p^2 = .409; F(1,159) = 23.59, p = .001, \eta_p^2 = .129$). Wie in Abb. 7.14 deutlich wird, gab es nur dann einen IC-Effekt, wenn ein Pronomen ambig war (nicht gender-disambiguiert: $F(1,22) = 36.10, p < .001, \eta_p^2 = .621; F(1,159) = 42.52, p < .001, \eta_p^2 = .211$; gender-disambiguiert: $F(1,22) = 5.78, p < .025, \eta_p^2 = .208; F(1,159) < 1, p = .574, \eta_p^2 = .002$).

Re-Analyse der „korrekten“ Antworten. Um zu überprüfen, ob das berichtete Effektmuster evtl. durch Unterschiede zwischen falsch und richtig beantworteten Items moderiert wurde, wurde die Analyse nochmals nur mit den korrekt beantworteten Items durchgeführt. Das berichtete Ergebnismuster konnte vollständig repliziert werden. Ein vollständiger Bericht des Ergebnismusters findet sich in Appendix C.2.5.

Zusammenfassung. Das Ergebnismuster war vollkommen parallel zu Experiment Ia. Es gab einen starken Effekt für den Gender-Cue, für den IC-Cue interagierten die beiden Faktoren miteinander und der Kausalitäts-Effekt konnte sich nur auswirken, wenn der Satz nicht bereits gender-disambiguiert war. Da die Ergebnisstruktur bereits ausführlich in Experiment Ia besprochen wurde, bedarf sie hier keiner ausführlichen Erläuterung mehr.

7.3.2.3 Reaktionszeitverteilung: Weibull-Parameter

Datenbasis, Parameterschätzung und Modellgüte. Durch den Ausschluss der „falschen“ Itemversion standen pro Untersuchungsbedingung lediglich 20 Items zur Verfügung. Von einer separaten Analyse nur der korrekten Antworten wurde deshalb abgesehen. Abgesehen von der geringen Itemzahl entsprach das Vorgehen bei der Parameterschätzung dem in Kap. 3.4 skizzierten Vorgehen. Als Startwert für den scale-Parameter wurde wiederum ein Wert von $\lambda = \theta^{\beta} = 2$ angesetzt. Die Parameter der beiden Prior-Verteilungen des scale-Parameters waren $a_1 = 2.0$, $b_1 = 0.1$, $a_2 = 2.0$ und $b_2 = 0.2$. Allerdings lässt sich das in Kap. 3.5 beschriebene Chi-Quadrat-Verfahren mit dieser geringen Itemzahl nicht anwenden, weswegen sich die Modellgüte nicht auf diese Weise beurteilen lässt.

Durchschnittliche Parameterausprägung. Abb. 7.15 zeigt die drei Weibull-Parameter der Reaktionszeit-Verteilung der Verifikationsaufgabe und ihre Beziehungen zueinander (shift- und scale-Parameter in Sekunden).

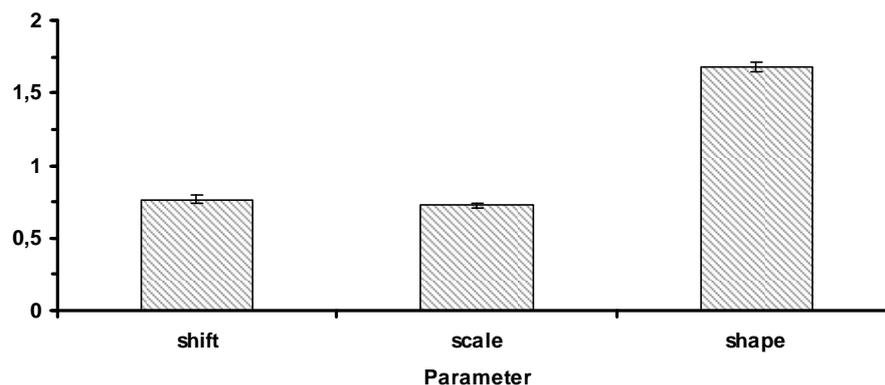


Abb. 7.15: Durchschnittliche Weibull-Parameter der RT-Verteilungen der Verifikationsaufgabe.

Der durchschnittliche shift-Parameter lag bei 0.76 s, der scale-Parameter bei 0.73 s und der shape-Parameter bei 1.69. Im Vergleich zur Verifikationsaufgabe von Experiment Ia war der shift-Parameter um ca. 0.2 s angestiegen, was darauf hinweist, dass die größere Komplexität des Itemmaterials sich auch in der Mindestantwortzeit der Verifikationsaufgabe niederschlug. Der shape-Parameter war im Vergleich zu Experiment Ia etwas erhöht, allerdings immer noch bedeutend niedriger als der shape-Parameter für die pronominale Auflösung. Die drei Parameter wurden wiederum auf Unterschiede hinsichtlich der experimentellen Manipulationen untersucht:

shift-Parameter. Es ergaben sich keine signifikanten Effekte (alle $F < 1.1$). Lediglich für den IC-Cue deutete sich ein schwacher Effekt an, der jedoch nicht signifikant war ($F(1,22) = 2.40$, $p = .136$, $\eta_p^2 = .098$).

scale-Parameter. Es gab einen starken Haupteffekt für den Faktor *Gender-Cue* ($F(1,22) = 67.04, p < .001, \eta_p^2 = .753$): Reaktionszeitverteilungen für die Sätze, bei denen das Pronomen nicht gender-disambiguiert war, hatten einen um ca. 175 ms höheren scale-Parameter als Sätze mit disambiguierten Pronomen. Der Haupteffekt für die Kausalitäts-Manipulation war nicht signifikant ($F(1,22) < 1, p = .658, \eta_p^2 = .009$), da er vom Gender-Cue moderiert wurde (Interaktions- $F(1,22) = 26.88, p < .001, \eta_p^2 = .550$): Wenn beide im Satz genannten Personen als Referenten möglich waren, dann wiesen ic-kongruente Sätze einen um ca. 100 ms niedrigeren scale-Parameter auf ($F(1,22) = 14.77, p < .001, \eta_p^2 = .402$). Wenn hingegen grammatisch sowieso nur ein möglicher Referent in Frage kam, dann brauchten ic-kongruente Sätze ca. 85 ms länger, um verarbeitet zu werden ($F(1,22) = 6.04, p < .022, \eta_p^2 = .216$; vgl. Abb. 7.16).

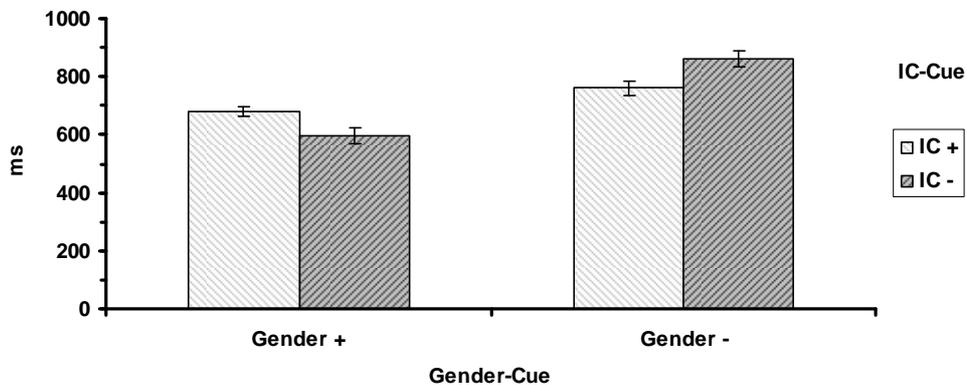


Abb. 7.16: Ausprägung des scale-Parameters der RT-Verteilungen der Verifikationsaufgabe.

shape-Parameter. Es gab keinen Haupteffekt für den Gender-Cue ($F(1,22) = 2.48, p = .130, \eta_p^2 = .101$). Hingegen war der Haupteffekt für den IC-Cue signifikant ($F(1,22) = 22.91, p < .001, \eta_p^2 = .510$), wurde allerdings durch einen signifikanten Interaktionsterm weiter qualifiziert ($F(1,22) = 24.07, p < .001, \eta_p^2 = .522$): Der IC-Cue hatte nur in der Bedingung *Gender+* einen Effekt ($F(1,22) = 41.53, p < .001, \eta_p^2 = .654$), der dafür mit einer Differenz von ca. .65 Skalenpunkten vergleichsweise groß ausfiel (vgl. Abb. 7.17). In der Bedingung *Gender-* war der Effekt des IC-Cues hingegen nicht signifikant ($F(1,22) < 1, p = .762$).

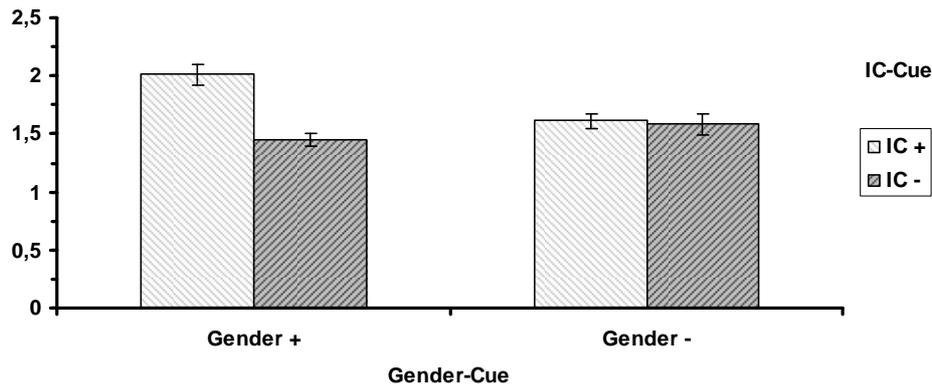


Abb. 7.17: Ausprägung des shape-Parameters der RT-Verteilungen der Verifikationsaufgabe.

Zusammenfassung. Für die Weibull-Parameter der Reaktionszeitverteilungen ergab sich prinzipiell ein ähnliches Bild wie in Experiment Ia: Alle entscheidenden Effekte betrafen lediglich den scale-Parameter. Auch hatte der IC-Cue keinen generellen Effekt auf den scale-Parameter, sondern nur, wenn der Satz ambig war. Auch die Größe der Effekte war identisch mit denen in Experiment Ia. Eine Abweichung war allerdings, dass eine der Untersuchungsbedingungen einen stark erhöhten shape-Parameter aufwies. Darüber hinaus ist erstaunlich, dass der IC-Cue in der ambigen Bedingung sogar einen leicht negativen Effekt auf den scale-Parameter hatte – ein Effekt, der sich nur schwer theoretisch interpretieren lässt. Evtl. ist dies ein methodischer Reflex auf den extrem erhöhten shape-Parameter.

7.3.2.4 Diskussion Verifikationsaufgabe

Die Ergebnisse für die Verifikationsaufgabe wiesen bei allen drei betrachteten Variablen (Akkuratheit, Mittlere Reaktionszeit, Reaktionszeitverteilungen) eine erstaunliche Parallelität zu Experiment Ia auf. Die Effekte der vorhergehenden Untersuchung konnten weitgehend repliziert werden. Neu war in Experiment Ib hingegen der zusätzlich konstruierte Aussagetype, bei dem nicht nur die Auflösung des Pronomens, sondern die gesamte Semantik des (Neben-)Satzes vertauscht wurde. Durch dessen Einbeziehung lassen sich zwei über den Erkenntnisstand von Experiment Ia hinausgehende Aussagen formulieren:

Zum einen ist die Tatsache von Bedeutung, dass sich die Akkuratheitswerte für den „korrekten“ und den „falschen“ Aussagetype nicht voneinander unterscheiden. Man kann sich deshalb vergleichsweise sicher sein, dass die Vpn die Sätze angemessen gelesen und verstanden haben. Ansonsten wären sie nicht in der Lage gewesen, Sätze mit einer falschen Semantik erfolgreich zurückzuweisen. Erhöhte Fehlerraten traten nur in dem Fall auf, in dem die semantische Inkonsistenz allein die Auflösung des Pronomens betraf. Dies ist nicht nur ein Hinweis darauf, dass pronominale Fehlauflösungen schwerer zu entdecken (bzw. zurückzuwei-

sen) sind, sondern spricht auch dafür, dass solche Fehler v.a. von persistierenden Integrationsprozessen verursacht werden. Andererseits machte eine Betrachtung der Beantwortung der „falschen“ Aussagen aber auch deutlich, dass die beobachteten Akkuratheitswerte in Experiment Ia und Ib offensichtlich das Maximum dessen sind, was bei der Komplexität des verwendeten Stimulus-Materials, dem induzierten Zeitdruck bei der Rezeption des Satzes sowie bei der generellen Länge der Untersuchung an Akkuratheit vernünftigerweise erwartet werden kann. Oder anders formuliert: Wenn Vpn schon in 10 % aller Fälle eine vollkommen andere Aussage mit dem Originalsatz verwechseln, dann darf man sich nicht wundern, wenn die gleiche Fehlerrate auch für die Detektion (in)korrekter pronominaler Auflösungen auftritt. Andererseits waren die Vpn aber grundsätzlich in der Lage, beide Unterscheidungen im Normalfall korrekt zu treffen (schließlich wurden die anderen 90 % der Fälle richtig beantwortet).

7.4 Diskussion Experiment Ib

Mehrere Punkte sind festzuhalten. Zunächst einmal konnten fast alle Effekte von Experiment Ia repliziert werden. Dies zeigt sich besonders auffällig für die mittelwerts-basierte Analyse, die wieder vollkommen mit dem in der Literatur berichteten Effektmuster übereinstimmte: Gender- und IC-Cue beeinflussten die pronominale Auflösung gleichzeitig, interagierten aber nicht miteinander.

Aber auch bei der RT-Verteilungsanalyse gibt es deutliche Schnittmengen zu Experiment Ia. Besonders deutlich ist das beim Gender-Effekt. Hier konnte der Befund bestätigt werden, dass die Wirkung des Gender-Cues aus zwei unabhängigen Teileffekten besteht, nämlich einer hierarchie-niedrigen *confirmation*- und einer hierarchie-hohen *disengagement*-Komponente, die beide gleichzeitig und gleich stark den Auflösungsprozess beeinflussen.

Beim IC-Effekt gab es allerdings eine bedeutende Abweichung zwischen Experiment Ia und Experiment Ib: Wenn die Sätze gender-disambiguiert waren, beeinflusste der IC-Cue nicht mehr den scale-, sondern vielmehr den shift-Parameter. Dies spricht dafür, dass in diesem Fall die inkongruente Kausalrelation bereits mit dem Pronomen entdeckt und ein hierarchie-niedriger Refokussierungsprozess zur Umstrukturierung der Diskursrepräsentation genutzt wurde. Bei den ambigen Sätzen bildete sich der IC-Effekt weiterhin nur im scale-Parameter ab. Dies spricht dafür, dass es sich hier um einen hierarchie-hohen Integrationseffekt handelt, der durch die kausale Inkongruenz von Hauptsatz- und Nebensatz-Proposition verursacht wird. Entsprechend fiel der Effekt in der disambiguierten Bedingung auch viel schwächer aus, da die späte Integration durch die skizzierte Trade-off-Relation zwischen früher Fokussierung und späterer Auflösung bereits vorbereitet war.

Dieses Ergebnismuster steht in direktem Kontrast zu Experiment Ia, in dem der IC-Cue in beiden Gender-Bedingungen lediglich den scale-Parameter beeinflusste. Eine gemeinsame Reanalyse beider Experimente mit den *within*-Faktoren *Gender-Cue*(*vorhanden* vs. *nicht vorhanden*) und *IC-Cue*(*vorhanden* vs. *nicht-vorhanden*) und dem *Experiment* (*Ia* vs. *Ib*) als Zwischengruppenfaktor ergab für den shift-Parameter eine signifikante Dreifachinteraktion ($F(1,43) = 7.25, p < .010, \eta_p^2 = .144$). Der IC-Cue führte lediglich in Experiment Ib in der gender-disambiguierten Bedingung zu niedrigeren shift-Werten (vgl. Abb. 7.18; alle anderen Interaktionen mit dem Experiment-Faktor waren hingegen nicht signifikant).

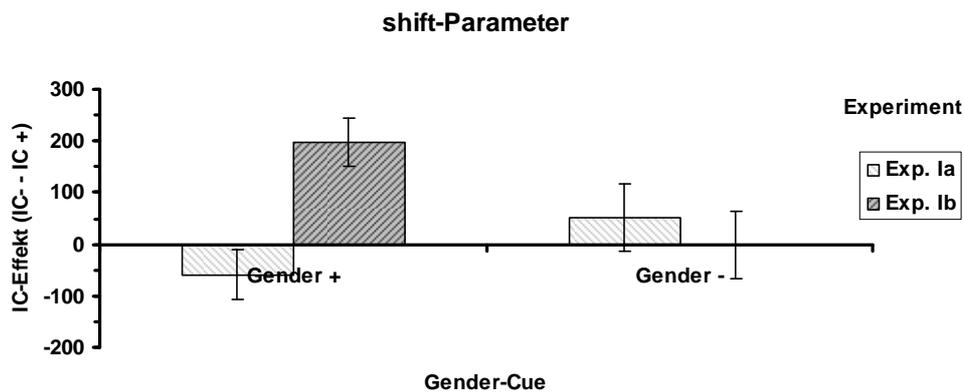


Abb. 7.18: Einfluss des IC-Cues auf den shift-Parameter in den beiden Gender-Bedingungen und den beiden Experiment Ia und Ib.

Eine ähnliche Dreifachinteraktion gab es auch für den scale-Parameter ($F(1,43) = 4.97, p < .031, \eta_p^2 = .104$), da der IC-Cue in Experiment Ib zwar in der nicht-disambiguierten Bedingung einen Einfluss hatte, nicht jedoch in der disambiguierten Bedingung (vgl. Abb. 7.19 alle anderen Interaktionen mit dem Experiment-Faktor waren wiederum nicht signifikant).

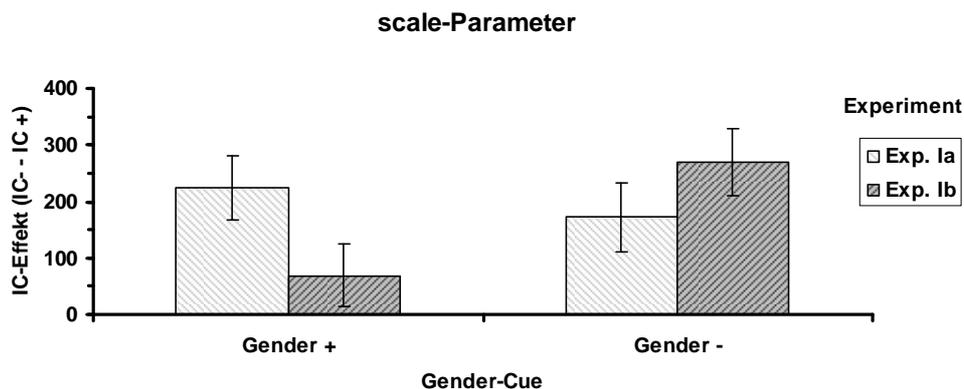


Abb. 7.19: Einfluss des IC-Cues auf den scale-Parameter in den beiden Gender-Bedingungen und den beiden Experiment Ia und Ib.

Damit konnte die postulierte Trade-off-Relation zwischen gedächtnis- und erklärungs-basierten Verarbeitungsprozessen in Bezug auf den IC-Effekt bei der pronominalen Auflösung nachgewiesen werden: Wenn hierarchie-niedrige Prozesse bereits ausreichend sind, um

ein adäquates Verstehensprodukt zu etablieren, dann nimmt komplementär der Einfluss erklärungs-basierter Prozesse ab, weil sie in diesem Fall nicht benötigt werden. Wenn hingegen keine basalen gedächtnis-basierten Prozesse zum Einsatz kommen können, tragen die erklärungs-basierten Prozesse die Hauptlast der Verarbeitung. Beide Prozesse dienen jedoch dem gleichen Ziel, nämlich ein möglichst adäquates Verstehensprodukt zu etablieren.

Die beschriebene Effektkonstellation macht auf eine Schwäche traditioneller Auswertungsmethoden aufmerksam: In der mittelwerts-basierten Analyse interagieren die Faktoren nicht miteinander. Das heißt, der IC-Cue beeinflusst die pronominale Auflösung auch dann, wenn er eigentlich überflüssig ist, weil das Pronomen bereits durch den Gender-Cue perfekt determiniert ist. Dieser Befund war im Rahmen der herkömmlichen Interpretation des IC-Effekts eher verwunderlich (vgl. Caramazza et al., 1977). Durch die differenziertere Analyse der beiden am IC-Effekt beteiligten Teileffekte lässt er sich hingegen theoretisch plausibel erklären: Zwar ist richtig, dass die IC sowohl im disambiguierten als auch im nicht-disambiguierten Fall den Auflösungsprozess beeinflusst, allerdings beruht der Effekt in beiden Fällen auf sehr unterschiedlichen kognitiven Prozessen. Im disambiguierten Fall sind vergleichsweise hierarchie-niedrige Refokussierungsprozesse für ihn verantwortlich, die bereits früh versuchen, die inkongruente Fortsetzung in die aktuelle Szenario-Struktur zu integrieren. Im nicht-disambiguierten Fall sind es hingegen eher hierarchie-hohe Integrationsprozesse, die durch inferentielle Zusatzaktivität eine kohärente Diskursrepräsentation wiederherstellen müssen. Wenn – wie in einer rein mittelwerts-basierten Analyse – nicht zwischen den verschiedenen Teilprozessen unterschieden wird, dann bleibt diese unterschiedliche Fundierung des Effektes unentdeckt und letztlich auch unerklärlich.

Da eigentlich alle bisher vorliegenden Befunde auf Aggregat-Analysen beruhen, kann damit aber auch der theoretisch und empirisch uneinheitliche Forschungsstand zum IC-Effekt erklärt werden. Wenn lediglich pauschal untersucht wird, ob eine bestimmte Variable überhaupt einen Effekt hat, ohne dessen Grundlage theoretisch zu spezifizieren und empirisch adäquat zu erfassen, dann ist zu erwarten, dass verschiedene Studien zu unterschiedlichen Ergebnissen kommen. Dabei lassen sich durchaus konkrete Moderator-Variablen benennen, die dafür verantwortlich sind, ob der IC-Effekt von eher gedächtnis- oder eher erklärungs-basierten Prozessen getrieben wird. Hierzu zählt insbesondere eine morpho-syntaktisch eindeutige Signalisierung der Inkongruenz sowie die Verfügbarkeit der Informationen, die zur Refokussierung notwendig sind. Wenn beides nicht gegeben ist, dann kann die Inkongruenz zwischen IC-Bias und Fortsetzung notgedrungen nur mittels ressourcen-intensiver Reanalyse-Prozesse am Ende des Verstehensprozesses vorgenommen werden. Wenn verschiedene Stu-

dien unterschiedliche Konstellationen in den genannten Kontextvariablen realisieren, dann finden sie selektiv entweder einen frühen IC-Fokus-Effekt oder einen späten IC-Kongruenz-Effekt, ohne sich des Unterschiedes bewusst zu sein.

Als ein wesentlicher Einflussfaktor für die Entstehung eines IC-Fokus-Effekts wurde die Verfügbarkeit relevanter Informationsbestände genannt, welche den Prozess der Refokussierung leiten können. Konkret kann das im Vergleich zu Experiment Ia konträre Ergebnismuster im vorliegenden Experiment auf die Einfügung der satzergänzenden Präpositionalphrase zurückgeführt werden. Es wurde bereits einleitend darauf hingewiesen, dass diese Einfügung mehrere Effekte auf einmal hat. Zum einen führt sie durch die Aktivierung weiterer Szenario-Informationen zu einer salienteren Signalisierung einer Inkongruenz (falls diese durch den Gender-Cue signalisiert wurde). Dadurch dass die durch das IC-Verb aufgespannte Ereignisstruktur noch weiter elaboriert wird, fällt die Inkonsistenz noch mehr auf und entsprechende Refokussierungs- bzw. Reanalyse-Prozesse können überhaupt erst eingeleitet werden. Zum zweiten stellt die Präpositionalphrase die notwendige „Verarbeitungsstrecke“ her, in der sich solche remedialen Maßnahmen auswirken können. Wenn (wie in den Versionen ohne Präpositionalphrase) die auflösende Verbalphrase direkt auf das inkongruente Pronomen folgt, dann wird die frühe Reanalyse durch die neuen Verbinformationen entweder überflüssig oder sofort von der satzfinalen Integration überlagert. Drittens stellt die Präpositionalphrase die notwendigen Informationsbestände sowohl für die Vorbereitung der finalen Integration als auch für die Refokussierungsprozesse zur Verfügung.

Es ist an der Zeit, sich dies an einem Beispiel klar zu machen. Wenn man einen Satz wie *Philip gratulierte Lothar, weil er in der schweren Klausur...* liest, dann instantiiert das IC-Verb *gratulieren* ein gewisses Basis-Szenario, dass u.a. die rudimentäre Ereignisstruktur umfasst, dass jemand zuerst etwas (besonderes: Distinktheit!) getan hat (1. Ereignis), was dann dazu führt (kausale Verbindung), dass ein anderer diesem nun gratuliert (2. Ereignis). Wenn nun der Kausalkonnektor *weil* verarbeitet wird, dann erwartet man Informationen über das Ereignis am Anfang der kausalen Ereigniskette; in diesem Fall konkret, was der erste Aktant wohl Tolles getan hat, dass man ihm nun gratuliert. Der entscheidende Punkt ist, dass die Präpositionalphrase (*in der schweren Klausur*) Informationen dazu liefert, wie das Basis-Szenario weiter ausgebaut werden kann. Entsprechend erwartet man nach der Präpositionalphrase auch eine Fortsetzung wie ... *eine gute Note hatte*. etc. Diese Fortsetzung würde nicht nur das GRATULIEREN-Skript erfolgreich komplettieren, sondern sich auch auf den Anfang der Kausalkette beziehen. Die Integration dieser Fortsetzung vereinfacht sich dadurch noch weiter, dass sie durch die zusätzlichen Informationen leichter evaluiert werden kann.

Folgt jedoch eine ic-inkongruente Fortsetzung wie ...*jede bessere Leistung bewunderte*. dann trifft das einen doppelt unvorbereitet: Zum einen, weil diese Information nicht ohne Zusatzinferenz ins GRATULIEREN-Skript und seine ereignisstrukturelle Kausalkette passt; zum anderen aber, weil man durch die Präpositionalphrase noch stärker der Überzeugung war, dass jetzt eigentlich eine schema-kongruente Fortsetzung folgen muss. Entsprechend zeigt sich in dieser Bedingung ein besonders starker Integrations-Effekt für ic-inkompatible Fortsetzungen. Dieser fällt sogar noch stärker aus, weil das Skript durch die zusätzliche Präpositionalphrase noch gestärkt wurde (aus dem direkten Vergleich von Experiment Ia und Ib ergibt sich ein zusätzlicher Verarbeitungsaufwand von ca. 100 ms; vgl. Abb. 7.19).

Interessant ist nun, was in den beiden gender-disambiguierten Bedingungen passiert. Bei einem Satz wie *Angela gratulierte Lothar, weil er in der schweren Klausur ...bereitet eine ic-kompatible Fortsetzung (...eine gute Note erzielte.)* überhaupt keine Probleme, weil alle Informationsquellen auf eine gemeinsame Lösung konvergieren. Wenn allerdings durch den Gender-Cue bereits früh signalisiert wird, dass eine Uminterpretation des Szenarios notwendig ist (*Lothar gratulierte Angela, weil er in der schweren Klausur...*), dann kann die Präpositionalphrase einem dabei helfen, eine solche Refokussierung vorzunehmen, und zwar gerade weil sie Informationen dafür bereitstellt, wie diese (bias-widrige) Erklärung aussehen könnte. Entsprechend zeigen sich hier zwar früh Irritationen, die jedoch schlussendlich zu einem besseren Satzverständnis führen. Der direkte Vergleich von Experiment Ia und Ib zeigt, dass dieser Erleichterungseffekt ca. 150 ms beträgt (vgl. Abb. 7.19).

Durch die reziproke Beziehung des IC-Fokus- und des IC-Integrations-Effekts liegt nun in der Tat Evidenz für eine starke Wechselwirkung gedächtnis- und erklärungs-basierter Verarbeitungsprozesse bei der pronominalen Auflösung vor: Zum einen bauen gedächtnis-basierte Prozesse wie der IC-Fokus-Effekt auf Wissensstrukturen auf, die top-down überhaupt erst einen adäquaten Erklärungsrahmen bereitstellen. Wenn dieser vorhanden ist, können die hierarchie-niedrige Prozesse zu einem Verstehensprodukt führen, das ressourcen-intensive Integrationsprozesse überflüssig macht. Das heißt, beide Prozessarten arbeiten nicht gegeneinander und auch nicht aneinander vorbei, sondern bedingen sich vielmehr gegenseitig, in dem sie gemeinsam eine möglichst kohärente Diskursrepräsentation etablieren.

Tab. 7.1 zeigt die Effekte von Experiment Ib noch einmal im Überblick:

<i>Effekt</i>	<i>Teileffekt</i>	<i>Pronominale Auflösung</i>	<i>Verifikation</i>
Gender	Confirmation	✓	---
	Disengagement	✓	✓
IC	Fokus	✓	---
	Integration	✓	✓

Tab. 7.1: Effekte in Experiment Ib für die pronominale Auflösungs- und Verifikationsaufgabe.

In Bezug auf den Gender-Effekt gibt es Evidenz für den hierarchie-niedrigen *confirmation*-Effekt und den hierarchie-hohen *disengagement*-Effekt. Letzterer persistiert auch noch in der anschließenden Verifikationsaufgabe. Beim IC-Cue gibt es diesmal auch Evidenz für einen hierarchie-niedrigen Fokus-Effekt, der erwartungsgemäß lediglich in der disambiguierten Bedingung auftritt. Gleichzeitig zeigt sich auch stark der hierarchie-hohe Integrations-Effekt, der in der nicht-disambiguierten Bedingung stärker auftritt. Er persistiert ebenfalls noch in bei der späteren Verifikationsaufgabe.

8. Experiment IIa

8.1 Einleitung

In Experiment I konnte nachgewiesen werden, dass sowohl gedächtnis- als auch erklärungs-basierte Prozesse an der pronominalen Auflösung beteiligt sind und in einer dynamischen Interaktionsrelation zueinander stehen. Diese Aussage basiert zum einen auf der Beobachtung, dass sich die verschiedenen Teileffekte selektiv in unterschiedlichen RT-Komponenten niederschlugen: Die hierarchie-niedrigen Teileffekte (also der *confirmation*-Effekt des Gender-Cues und der Fokus-Effekt des IC-Cues) beeinflussten v.a. den shift-Parameter der Reaktionszeitverteilung, während die hierarchie-hohen Teileffekte (also der *disengagement*-Effekt des Gender-Cues und der Integrations-Effekt des IC-Cues) selektiv den scale-Parameter der Reaktionszeitverteilung beeinflussten und darüber hinaus in einer nachgeschalteten Verifikationsaufgabe noch persistierten. In Experiment Ib konnte auch die postulierte differenzielle Interaktion der beiden IC-Teileffekte mit dem Gender-Cue nachgewiesen werden: Der IC-Fokus-Effekt trat lediglich auf, wenn der Satz bereits disambiguiert war, und führte zu einem schwächeren IC-Integrations-Effekt. Ohne Disambiguierung konnte keine frühe Refokussierung stattfinden und der IC-Integrations-Effekt fiel entsprechend stärker aus.

In der bisherigen Diskussion wurde implizit bereits auf eine weitere empirische Dimension zurückgegriffen, auf der sich die hierarchie-hohen und die hierarchie-niedrigen Effekte der verschieden Auflösungs-Cues voneinander unterscheiden sollten. Gedächtnis-basierte Effekte sollten die Auflösung bereits früh beeinflussen, während erklärungs-basierte Prozesse hauptsächlich satzfinal auftreten sollten. Deswegen kann es überhaupt erst zu der skizzierten Trade-off-Relation des IC-Fokus- mit dem IC-Integrations-Effekt kommen: Wenn bereits ein relevanter Anteil der Erklärungsarbeit von hierarchie-niedrigen Prozessen übernommen wurde, dann können späte hierarchie-hohe Integrations-Prozesse weniger gründlich ausfallen, weil das Verstehensprodukt bereits optimal vorbereitet wurde.

Dadurch ergibt sich aber auch die Möglichkeit, durch die Betrachtung der *Effektdynamik*, d.h. ihrer Verteilung über den Verlauf der Satzverarbeitung hinweg, die einzelnen Effektklassen zu dissoziieren. Die geschilderten hierarchie-niedrigen Teileffekte (samt ihrer differenziellen Interaktionen mit dem Gender-Cue) sollten sich v.a. in der frühen Satzverarbeitung nachweisen lassen, speziell auf oder kurz nach dem Pronomen. Die diskutierten hierarchie-hohen Teilprozesse (samt ihrer differenziellen Interaktion mit dem Gender-Cue) sollten sich hingegen erst vergleichsweise spät nachweisen lassen, insbesondere satzfinal im sen-

tence-wrap-up. In Kombination mit den bereits diskutierten Einflussbeziehungen aus Experiment I ergibt zusammengenommen also folgendes Vorhersagemuster (vgl. Tab. 8.1):

<i>Effekt</i>	<i>Teileffekt</i>	<i>Parameter</i>	<i>Zeitpunkt</i>	<i>Interaktion Gender-Cue</i>
Gender	Confirmation	Shift	Früh	---
	Disengagement	Scale	Spät	---
IC	Fokus	Shift	Früh	Stärker, wenn disambiguiert
	Integration	Scale	Spät	Stärker, wenn nicht disambiguiert

Tab. 8.1: Vorhersagen für die verschiedenen Teileffekte des Gender- und des IC-Cues.

Das bedeutet, in Experiment II wird prinzipiell genau das gleiche Beziehungsgefüge zwischen den einzelnen Teileffekten der beiden Auflösungs-Cues erwartet. Zusätzlich wurde postuliert, dass sie sich auch noch an der richtigen Stelle der Satzverarbeitung nachweisen lassen. Damit ist eine noch strengere Hypothesenprüfung möglich, da zusätzliche Restriktionen eingeführt werden. Sollte sich darüber hinaus die differenzielle Verlaufsdynamik der einzelnen Teileffekte bestätigen, dann wäre dies ein weiterer Hinweis auf die Validität der gewählten Zwei-Prozess-Interpretation der Weibull-Parameter.

Wie bereits am Ende von Kap. 2 diskutiert, ist es jedoch gar nicht so einfach, eine sprachpsychologische Methode zu finden, welche den kontinuierlichen Verlauf der Verarbeitungsdynamik der Satzverarbeitung adäquat abbildet. In Experiment II soll eine *self-paced-reading*-Methode eingesetzt werden, bei der die Vpn die einzelnen Sätze lesen, indem sie nacheinander ein Wort nach dem anderen aufdecken, während das vorhergehende Wort wieder verdeckt wird (sog. nicht-kumulatives *moving-window*-Verfahren). Die zur Rezeption eines einzelnen Wortes benötigte Zeit wird dabei meist als unspezifischer online-Indikator für den Verarbeitungsaufwand interpretiert (eine ausführliche Diskussion findet sich z.B. in Richter, 2003; ein Vergleich verschiedener Lesezeit-Methoden wurde von Just et al., 1982, durchgeführt). Obwohl diese Methode in der Sprachpsychologie ausgesprochen häufig verwendet wird (vgl. Haberlandt, 1994), ist sie sicherlich nicht unproblematisch. Eine häufig geäußerte Kritik ist, dass die resultierenden Lesezeiten sehr viel höher ausfallen, als dies z.B. bei *eye-tracking*-Daten der Fall ist (ca. 400 ms vs. 250 ms, vgl. Just et al., 1982). Diese Diskrepanz wird u.a. darauf zurückgeführt, dass durch das ständige Tastendrücken eine zusätzliche motorische Komponente in die Aufgabe eingeführt wird, die dazu führt, dass die normale Satzverarbeitung sozusagen künstlich verzögert wird. Dadurch kommt es zu einem *lag* zwischen Auge und Hand, bei der die sprachliche Rezeptionskomponente eigentlich schneller fortschreiten könnte, aber auf die Ausführung der motorischen Komponente warten muss. Danks (1986) hat in diesem Zusammenhang darauf hingewiesen, dass dadurch die *immediacy*-Annahme von

Just und Carpenter (1980) nicht mehr gilt. Daraus resultiert u.U., dass bestimmte Verarbeitungsprozesse anders organisiert werden als beim natürlichen Lesen. Insbesondere kann dies dazu führen, dass bestimmte integrationsrelevante Verarbeitungsprozesse durch die ständige Behinderung durch die motorische Komponente ans Satzende verschoben werden, da nur dort die notwendigen Ressourcen zur Verfügung stehen, um sie vorzunehmen. Die *moving-window*-Methode führt also potenziell zu einer artifiziellen Vergrößerung des *sentence-wrap-up* und begünstigt eine propositionsweise im Gegensatz zu einer inkrementellen Verarbeitung (vgl. auch Magliano, Graesser, Eymard, Haberlandt & Gholson, 1993). Im vorliegenden Fall würde dies zu einer Stärkung der satzfinalen, erklärungs-basierten Integrations-Prozesse führen, während frühe Fokussierungs-Effekt schwerer zu detektieren sein müssten.

Mitchell (2004) hat allerdings darauf hingewiesen, dass es zwar grundsätzlich richtig ist, dass *moving-window*-Lesezeiten ein höheres numerisches Grundniveau haben als vergleichbare *eye-tracking*-Daten. Allerdings konnte seiner Meinung nach nie nachgewiesen, dass die Methode zu einer *qualitativ* anderen Verarbeitung führt (vgl. auch Haberlandt & Graesser, 1985; Graesser & Haberlandt, 1986). Im Gegenteil vertritt er die Auffassung, dass zwischen den *Lesezeitprofilen* von *moving-window*- und *eye-tracking*-Daten eine hohe Korrespondenz besteht, wenn man vom unterschiedlichen Ausgangsniveau abstrahiert (Just et al., 1982, ermitteln eine empirische Korrelation von $r = .57$; vgl. auch Aaronson & Ferres, 1984). Entsprechend seien auch alle wesentlichen Erkenntnisse der Sprachpsychologie bis jetzt immer mittels beider Methoden nachweisbar gewesen. Insbesondere finden sich starke *wrap-up*-Prozesse im Gegensatz zur Behauptung Danks auch in *eye-tracking*-Daten (vgl. Just & Carpenter, 1980; Rayner et al., 1989; Rayner, Kambe & Duffy, 2000). *Moving-window*-Methoden verändern also den Rezeptionsprozess nicht, sondern lassen vielmehr die relevanten Verarbeitungsprozesse deutlicher sichtbar werden, weil sie diese künstlich überhöhen. Für den speziellen Fall des IC-Effekts ist ebenfalls von einer weitgehenden methodischen Parallelität auszugehen. So legten Koornneef und van Berkum (2006) sowohl wortweise Lesezeiten als auch *eye-tracking*-Daten vor, mit denen sich konvergierend der IC-Fokus-Effekt nachweisen ließ.

Ohne sich in den Methodenstreit einzumischen (vgl. hierzu Carreiras & Clifton, 2004; Tanenhaus, 2004), sei hier einfach nur festgehalten, dass sich eine wortweise *moving-window*-Methode für die vorliegende Fragestellung noch am ehesten anbietet. Sie besitzt im Vergleich zu den bisher eingesetzten Lesezeit-Methoden zum IC-Effekt (wie satz- oder phrasenweisen Lesezeiten, vgl. Garnham et al., 1992; Stewart, 2000) ein wesentlich höheres Auflösungsvermögen, da sie auch lokale Effekte auf spezifischen Satzelementen nachweisen kann. Eine solch, feinkörnige Auflösung erscheint jedoch gerade für den fragilen IC-Fokus-Effekt zwin-

gend notwendig zu sein (vgl. Koornneef & van Berkum, 2006). Und im Gegensatz zu *eye-tracking*-Daten führen *moving-window*-Lesezeiten zu einer einzigen, homogenen AV. Deswegen steht in diesem Fall ein einheitlicher Verarbeitungsindikator zur Verfügung, während in *eye-tracking*-Studien meist mit einer Vielzahl von Maßen gleichzeitig gearbeitet werden muss (deren Relation nicht immer geklärt ist, vgl. Rayner, 1998; Reichle, Pollatsek, Fisher & Rayner, 1998), die dann meist zu höheren Aggregatmaßen kombiniert werden (wie mittlere Fixationsdauer, Anzahl Regressionen etc.).

Während *eye-tracking*-Studien gewissermaßen additiv vorgehen, indem sie verschiedene molekulare Teilmaße zu Indikatoren verschiedener Verstehensprozesse kombinieren, wird in dieser Studie subtraktiv vorgegangen, indem eine molare AV durch Dekomposition in verschiedene Reaktionszeit-Komponenten zerlegt wird, welche dann wiederum als Indikatoren für verschiedene kognitive Prozesse dienen. Der Schwerpunkt wird also von der Erfassung auf die *Auswertung* der Daten verschoben, indem unterschiedliche Parameter der zugrunde liegenden Reaktionszeitverteilung im Nachhinein ermittelt werden. Obwohl sich das Weibull-Modell von Rouder et al. (2003) bisher ausgesprochen gut bewährt hat, ist fraglich, ob das Verfahren auch für wortweise Lesezeit-Daten gute Ergebnisse liefert. Das Grundproblem besteht darin, dass wortweise Lesezeiten zwar eine höhere Auflösung besitzen, aber auch instabil sind. Da im Gegensatz zu einer globalen Auflösungsaufgabe wie in Experiment I weniger kognitive Prozesse bei jedem einzelnen Wort vorzunehmen sind, haben die resultierenden Reaktionszeitverteilungen bestimmte Eigenschaften, welche eine RT-Verteilungsanalyse erschweren. So sind sie im Normalfall durch den einheitlichen Tastendruck-Rhythmus ausgesprochen varianzarm und die relevanten Unterschiede zwischen den Bedingungen bilden sich nur in einem sehr kleinen Anteil der resultierenden Reaktionszeit ab. Darüber hinaus sind sie sehr anfällig für aufgabenfremde Ausreißer, da es bei der Bearbeitung der Aufgabe leicht zu einem zu schnellen Tastendruck oder einer leichten Verzögerung (z.B. attentionaler Art) kommen kann, die wenig mit dem gerade gelesenen Wort zu tun hat.

Dies alles führt dazu, dass die Schätzung von Reaktionszeitverteilung unter den genannten Bedingungen schwierig ist. Für die Weibull-Verteilung im Speziellen ergeben sich noch zusätzliche Probleme, die mit dem sog. *unbounded likelihood problem* zusammenhängen (vgl. Appendix A.4). Ohne das von Rouder et al. entwickelte Hierarchical Bayes-Verfahren wäre die Schätzung von Verteilungsparametern im vorliegenden Fall auch ziemlich nutzlos, weil dieses durch die Einbeziehung einer hierarchischen Prior-Verteilung stabilisierend wirkt (vgl. Kap. 3). Allerdings ist davon auszugehen, dass die resultierenden Parameterschätzungen nur sehr unreliabel sind, da die Form der Verteilungen von kleinen Schwankungen der empirischen

rischen Verteilung beeinflusst wird, die nicht unbedingt mit den zugrunde liegenden Prozessen korrelieren. Zwar hatte die Simulationsstudie in Kap. 4 deutlich gemacht, dass das Verfahren grundsätzlich auch für Lesezeit-Daten angemessen ist. Allerdings wurde dort eine unrealistisch gute Datenqualität angenommen. Um diese Problematik zumindest ansatzweise zu kompensieren, werden im Folgenden die Werte für die einzelnen Wörter zu Konstituenten aggregiert, um eine Reliabilitätssteigerung der ermittelten Werte zu erreichen. Allerdings kann auch die Konstituentenbildung wahrscheinlich nicht verhindern, dass für die RT-Verteilungsanalyse in Experiment II weniger eindeutige Ergebnisse zu erwarten sind.

8.2 Methode

8.2.1 Stichprobenbeschreibung und -bereinigung

Allgemeine Angaben. An Experiment IIa nahmen insgesamt $N = 25$ Versuchspersonen (21 Frauen, 4 Männer) teil. Keine der Vpn hatte an einem der beiden vorhergehenden Experimente teilgenommen. Das durchschnittliche Alter lag bei $M = 25.3$ Jahren ($SD = 8.3$). Es handelte sich hauptsächlich um Studierende der Psychologie (durchschnittlich im 3. Semester), die für ihre Teilnahme eine Versuchspersonenstunde oder eine Vergütung von 7,-€ erhielten. Alle Teilnehmer/innen berichteten, dass sie normal- oder korrigiert normalsichtig seien, und gaben Deutsch als Muttersprache an.

ELVES. Insgesamt wies die Stichprobe in allen drei eingesetzten hierarchie-niedrigen Subtests von ELVES ein durchschnittliches und unauffälliges Ergebnis auf. Die entsprechenden Mittelwerte lagen für den Subtest *Satzverifikation* bei $M = 15.9$ ($SD = 3.3$), für den Subtest *Sinnhaftigkeit Sätze* bei $M = 16.0$ ($SD = 4.9$) und für den Subtest *Sinnhaftigkeit Satzfolgen* bei $M = 16.3$ ($SD = 3.9$). Die drei Subskalen korrelierten hoch miteinander und wurden deswegen zu einem gemeinsamen Index zusammengefasst (Cronbachs $\alpha = .86$; $M = 16.1$, $SD = 3.6$). Eine Versuchsperson wies für die eingesetzten ELVES-Subskalen einen erhöhten multivariaten Leverage-Wert auf ($h_{ii} = .54$, $p < .01$), da sie in allen Subtests überdurchschnittliche Werte aufwies. Sie gab nach der Untersuchung an, am gleichen Tag in einer parallel laufenden Untersuchung die gleichen ELVES-Subskalen schon einmal bearbeitet zu haben. Da sie sonst ein eher unauffälliges Antwortmuster aufwies, wurde sie nicht von der Analyse ausgeschlossen.

Durchschnittliche Akkuratheit und Antwortlatenz. Die mittleren Reaktionszeiten und die durchschnittlichen Akkuratheitswerte der Verifikationsaufgabe wurden pro Person zu einem Vektor mit insgesamt acht Werten zusammengefasst und ein multivariater Leverage-Wert bestimmt. Zwei Personen wiesen problematische Leverage-Werte auf ($h_{ii} > .70$, $p < .01$). Eine

Analyse ihrer Antwortmuster ergab jedoch, dass sie die Untersuchung *besser* (d.h. korrekter und langsamer) bearbeitet hatten als der multivariate Durchschnitt der Stichprobe. Beide Personen wurden deswegen nicht von der Analyse ausgeschlossen.

8.2.2 Material

Sätze zur pronominalen Auflösung. Das Stimulusmaterial war das gleiche wie in Experiment Ib, d.h. es kam wiederum die Version mit Präpositionalphrase zum Einsatz (vgl. Kap. 5). Es wurden alle 160 Stimuli in jeweils einer der vier Untersuchungsbedingungen bearbeitet, wobei die Zuordnung der Items zu den Untersuchungsbedingungen durch 16 Itemlisten mittels eines Lateinischen Quadrates koordiniert wurde.

Verifikationsaussagen. Das Vorgehen zur Konstruktion der Verifikationsaussagen orientierte sich an Experiment Ia. Das heißt, im Gegensatz zu Experiment Ib kamen wieder nur zwei Arten von Verifikationsaussagen zum Einsatz, nämlich korrekte und inkorrekte (bzw. satzkompatible und -inkompatible). Auf eine zusätzliche Untersuchungsbedingung, bei der die Inhalte des Satzes bereits auf der semantischen Ebene verändert wurden, wurde hingegen verzichtet. Die Koordination der verschiedenen Itemversionen zu den vier Untersuchungsbedingungen sowie die Art ihrer Formulierung erfolgte über die gleichen 16 Itemlisten, die auch für die Auflösungssätze verwendet wurden.

8.2.3 Untersuchungsablauf

Instruktion und Untersuchungsablauf waren weitgehend identisch zu denen in Experiment Ia und Ib. Im Gegensatz zu den beiden letzten Untersuchungen wurde den Vpn jedoch anfänglich nur mitgeteilt, dass sie im Folgenden eine Reihe von Aussagen lesen sollen. Das heißt, eine direkte Auflösung des Pronomens wurde nicht verlangt. Weiterhin wurden den Vpn mitgeteilt, dass ihnen im Anschluss an die Satzrezeption jeweils eine kurze Aussage dargeboten würde. Ihre Aufgabe sei es dann zu entscheiden, ob die Aussage im zuvor gelesenen Satz vorkam oder nicht.

Der Ablauf war dabei für jedes Item so aufgebaut, dass ein Satz zunächst durch einen Druck auf die Leertaste angefordert werden musste. Danach erschien der gesamte Satz, bei dem jedoch alle Buchstaben durch den Buchstaben „X“ ersetzt worden waren (Aufbau und graphische Gestaltung von Seite und Stimulus waren ansonsten vollkommen identisch zu den beiden vorhergehenden Studien). Sowohl Interpunktion als auch Groß- und Kleinschreibung des Satzes blieben dabei vollständig erhalten. Der Satz verteilte sich – je nach Länge – über zwei oder drei Zeilen auf dem Bildschirm, wobei der erste Umbruch meist am Ende der Prä-

positionalphrase lag (nie auf dem Pronomen selbst!) und der zweite Umbruch – wenn überhaupt vorhanden – meist in der Mitte der Verbalphrase.

Die Vpn konnten jeweils das nächste Wort durch Drücken der Leertaste „aufdecken“. Das Aufdecken eines Wortes führte jedoch dazu, dass das jeweils vorhergehende Wort wieder zugedeckt wurde, sodass es den Vpn nicht mehr für eine nachfolgende (Re-)Inspektion zur Verfügung stand. Wenn eine Vp insgesamt länger als 10 s zum Lesen eines Satzes brauchte, wurde sie nach dem letzten Wort durch einen roten Rückmeldungs-Bildschirm dazu aufgefordert, schneller zu lesen. Anschließend wurde dann die zu verifizierende Aussage in dem gleichen Fenster dargeboten wie der Originalsatz. Zur Beantwortung mussten die Vpn entweder die Taste „D“ (*falsch*) oder „K“ (*richtig*) drücken. Auch hier verschwand die Aussage nach Abgabe der Antwort und die Vpn wurden bei Bearbeitungszeiten über 2 s bzw. bei falscher Beantwortung daran erinnert, die Aufgabe schneller und/oder gründlicher zu bearbeiten.

Ansonsten war der Makro-Ablauf der Untersuchung vollkommen identisch zu den beiden vorhergehenden Untersuchungen. Das heißt, die Vpn konnten sich auch hier zunächst durch vier Übungsitems an das Vorgehen gewöhnen, danach folgten 20 Eisbrecher-Items bevor die 160 Stimulussätze in zufälliger Reihenfolge präsentiert wurden. Zum Abschluss der Untersuchung wurden wieder die drei ELVES-Subtests bearbeitet und ein paar soziodemographische Fragen beantwortet. Die Untersuchung dauerte insgesamt ca. 50 Minuten.

8.2.4 Design und Abhängige Variablen

Das Grunddesign der Untersuchung ist das gleiche wie in den beiden vorhergehenden Experimenten, nämlich ein $2(\text{Gender-Cue: vorhanden vs. nicht vorhanden}) \times 2(\text{IC-Cue: vorhanden vs. nicht vorhanden})$ -Design, messwiederholt sowohl über Personen als auch über Items. Als zentrale Abhängige Variable dient v.a. die zur Rezeption eines Wortes benötigte Lesezeit. Die Verifikationsaufgabe im Anschluss an die Satzrezeption war hingegen vergleichsweise sekundär. Ihre Korrektheit kann zur Beurteilung der Qualität der Satzrezeption herangezogen werden sowie zur Feststellung, welche Auflösungsvariante real gewählt wurde. Die Antwortzeit für die Verifikationsaufgabe ist hingegen ein Indikator für die Art des Verarbeitungsprozesses. Beide Variablen sind nicht nur für sich selbst interessant, sondern v.a. im Vergleich zum Vorgehen in Experiment Ia und Ib, bei dem das Pronomen zuvor explizit aufgelöst wurde.

8.3 Ergebnisse

8.3.1 Satzlesezeiten

8.3.1.1 Mittlere Lesezeiten

8.3.1.1.1 Datenbereinigung und Lesezeitprofil

Datenbereinigung. Bei der Datenbereinigung wurde als unterer Cut-off ein Wert von 150 ms angesetzt. Aufgrund dieses Kriteriums wurden 0.1 % aller Datenpunkte durch den bedingungs- und personenspezifischen Mittelwert ersetzt. Weiterhin wurden alle Werte, die 4 Standardabweichungen vom wort- und bedingungsspezifischen Mittelwert einer Person abwichen, durch diesen ersetzt. Diese Maßnahme führte insgesamt zur Ersetzung von ca. 0.7 % aller Datenpunkte.

Lesezeitprofil. Abb. 8.1 zeigt die mittlere Zeit, die zum Lesen der Wörter eines Satzes benötigt wurde. Auf der x-Achse sind die einzelnen Wörter des Satzschemas mit ihrer jeweiligen Funktion angegeben. Die Klammern beziehen sich auf die nachfolgend vorgenommene Konstituentenbildung (s.u.).

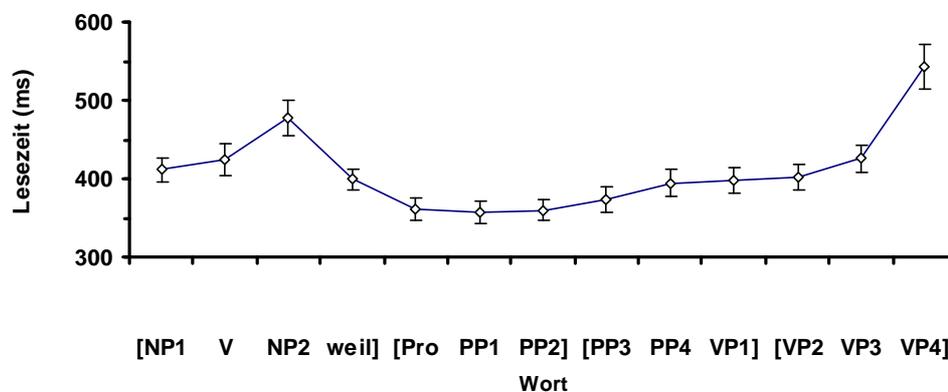


Abb. 8.1: Mittlere Lesezeit über alle vier Untersuchungsbedingungen gemittelt.

Die mittlere Lesezeit pro Wort betrug durchschnittlich ca. 410 ms, wobei der niedrigste Wert bei ca. 350 ms, der höchste bei 550 ms lag. Das sind durchaus typische Werte für Lesezeiten, die mittels einer moving-window-Methode erhoben wurden: Just et al. (1982) geben z.B. bei ihrem Methodenvergleich einen durchschnittlichen Wert 440 ms ($SD = 260$ ms) für die *moving-window*-Lesezeit einzelner Wörter an. Wie in der Abbildung deutlich zu erkennen, variierten die konkreten Lesezeiten erheblich mit der Satzstruktur. Das Profil lässt sich vielleicht am besten als eine zweigipflige Struktur beschreiben, wobei nach einem allmählichen und eher moderaten Anstiegsprozess die Lesezeiten in einem vergleichsweise abrupten Anstieg an

den beiden Teilsatzenden kulminierten. Solche *wrap-up*-Strukturen werden meist als Ort finaler Integrations-, Disambiguierungs- und evtl. Reanalyse-Prozesse interpretiert (z.B. Mitchell & Green, 1978; Just & Carpenter, 1980). Die Tatsache, dass der *wrap-up* des Hauptsatzes ca. 65 ms niedriger ausfiel als für die finale Integration am Ende des Nebensatzes, scheint dabei zunächst für die *connective-integration*-Hypothese von Millis und Just (1994) zu sprechen. Diese geht davon aus, dass die Teilsätze eines Satzes zwar jeweils unabhängig voneinander verarbeitet werden, aber im Falle einer Konjunktion die Teilstrukturen miteinander koordiniert werden müssen, weshalb sie im verbalen Arbeitsgedächtnis zwischengespeichert werden (vgl. Kap. 2.3). Hinzu kommt, dass die Satzstruktur des Deutschen (bei dem – im Gegensatz zur strengen SVO-Struktur des Englischen – Nebensätze Verbendstellung aufweisen) solche *wrap-up*-Prozesse noch begünstigt, da Integrations-, aber auch Prädiktionsprozesse typischerweise bei Verben auftreten (vgl. z.B. Gibson, 1998).

8.3.1.1.2 Ergebnisse auf der Wortebene

Für die vorliegende Fragestellung ist v.a. interessant, *wann* es zu Unterschieden in der (wortweisen) Satzverarbeitung kommt, die sich systematisch auf die beiden manipulierten Auflösungs-Cues für das Pronomen zurückführen lassen. Hierfür wurde in einem ersten Schritt für jedes der 13 Wörter einzeln eine zweifaktorielle, messwiederholte ANOVA mit den beiden Faktoren *Gender-Cue* und *IC-Cue* berechnet. Die entsprechenden Ergebnisse finden sich in Tab. 8.2 Das Ergebnismuster, das sich dabei abzeichnet, ist jedoch einfacher zu verstehen, wenn die Effekte der beiden Cues zunächst einmal voneinander separiert werden. Im Folgenden soll deswegen die Dynamik und die charakteristische Entwicklung der beiden Haupteffekte sowie ihre Interaktion im Laufe der Satzverarbeitung einzeln dargestellt werden.

	[NP1]	V	NP2	weil]	[er	PP1	PP2]	[PP3	PP4	VP1]	[VP2	VP3	VP4]
By-Participants													
<i>Gender-Cue</i>													
<i>F(1,24)</i>	1.01	< 1	< 1	2.47	< 1	2.2	3.17	3.90	8.79	4.64	18.87	12.06	10.72
<i>p</i>	.307	.917	.940	.130	.905	.151	.088	.060	.007**	.045*	.001**	.002**	.003**
η_p^2	.043	< .001	< .001	.093	< .001	.084	.117	.140	.268	.157	.440	.344	.309
<i>IC-Cue</i>													
<i>F(1,24)</i>	< 1	< 1	< 1	< 1	< 1	< 1	1.13	< 1	< 1	< 1	< 1	2.31	25.19
<i>p</i>	.433	.355	.838	.820	.730	.891	.299	.330	.769	.977	.597	.141	.001**
η_p^2	.030	.115	.002	.030	.006	.008	< .001	.006	.006	.018	.092	.293	.512
<i>Gender- x IC-Cue</i>													
<i>F(1,24)</i>	< 1	1.89	1.85	< 1	6.22	1.62	2.98	6.88	< 1	< 1	1.32	< 1	3.12
<i>p</i>	.844	.182	.187	.942	.020*	.215	.097	.015*	.845	.969	.262	.635	.090
η_p^2	.002	.073	.071	<.001	.206	.063	.110	.223	.002	<.001	.052	.010	.115
By-Items													
<i>Gender-Cue</i>													
<i>F(1,159)</i>	< 1	< 1	< 1	4.36	< 1	5.05	7.80	8.41	18.82	11.80	17.62	33.27	152.40
<i>p</i>	.320	.994	.829	.038*	.790	.026*	.006**	.004**	.001**	.001*	.001**	.001**	.001**
η_p^2	.006	< .001	< .001	.027	< .001	.031	.047	.050	.106	.069	.100	.173	.489
<i>IC-Cue</i>													
<i>F(1,159)</i>	< 1	1.51	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	11.27
<i>p</i>	.325	.222	.746	.769	.564	.931	.344	.439	.656	.957	.616	.319	.001**
η_p^2	.003	.016	.000	.006	.001	.002	.000	.000	.001	.001	.013	.030	.057
<i>Gender- x IC-Cue</i>													
<i>F(1,159)</i>	< 1	< 1	1.26	< 1	3.11	< 1	1.81	4.08	< 1	< 1	< 1	< 1	1.72
<i>p</i>	.784	.385	.263	.987	.080	.403	.180	.045*	.786	.932	.626	.824	.192
η_p^2	< .001	.005	.008	< .001	.019	.004	.011	.025	< .001	< .001	.001	< .001	.011

Tab. 8.2: Ergebnisse des varianzanalytischen Vergleichs zwischen den vier Untersuchungsbedingungen auf Wortebene.

Gender-Effekt. Abb. 8.2 a) zeigt die wortweisen Lesezeiten im Satzverlauf in Abhängigkeit von der Ausprägung des Gender-Cues. Die korrespondierende Abb. 8.2 b) stellt hingegen den Effekt des Gender-Cues direkt dar. Das heißt, hier wird vom generellen Profil des Satzverlaufes abstrahiert und nur die Differenz zwischen An- und Abwesenheit des Cues abgetragen. Da diese Differenz durch Abzug des Wertes der Abwesenheits- von dem der Anwesenheitsbedingung gebildet wurde, stehen positive Werte für einen Verarbeitungsvorteil, der sich durch die Realisierung des Cues ergibt, negative hingegen für einen Verarbeitungsnachteil.

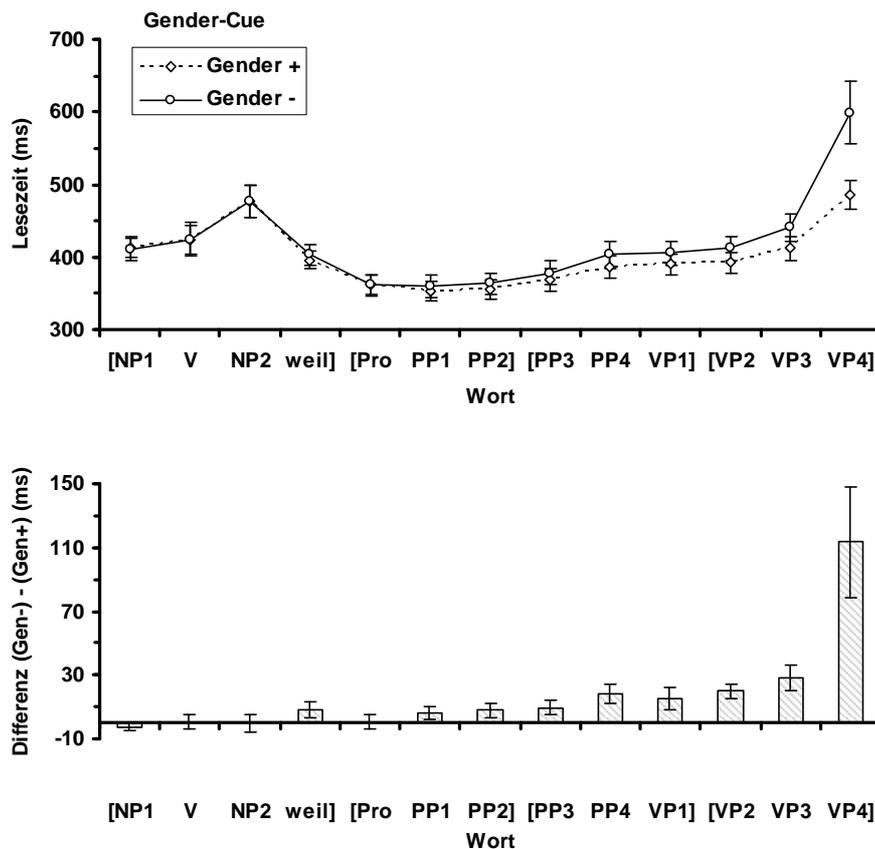


Abb. 8.2: a) Wortweisen Lesezeiten in Abhängigkeit zur Ausprägung des Gender-Cues und b) Effekt des Gender-Cues im Satzverlauf.

In beiden Abbildungen ist deutlich zu erkennen, dass der Effekt des Gender-Cues früh einsetzte: Die ersten substanziellen Effekte fanden sich zwar nicht auf dem Pronomen selbst, spätestens jedoch ab dem ersten Wort der Präpositionalphrase. Ab diesem Zeitpunkt war der Effekt konstant vorhanden, nahm in seiner Stärke im Laufe der Verarbeitung sogar stetig zu bis er dann auf dem letzten Wort des Satzes sein Maximum erreichte. Dieser letzte Effekt war ausgesprochen groß (er ist mit 113 ms ungefähr genauso stark wie die Effekte auf allen anderen Wörtern des Satzes zusammen). Stellt man in Rechnung, dass sich ein gewisses Ausmaß an *spill-over*-Effekten bei *moving-window*-Methoden nicht vermeiden lässt (vgl. Mitchell,

2004), kann man festhalten, dass der Gender-Effekt erwartungsgemäß ziemlich direkt einsetzte und einen stetigen Verarbeitungsvorteil für die disambiguierte Satzversion schuf.

IC-Effekt. Abb. 8.3 a) und b) stellen die wortweisen Lesezeiten und die Einzeleffekte des IC-Cues für jedes Wort des Satzschemas dar und sind parallel zur Darstellung des Gender-Cues aufgebaut.

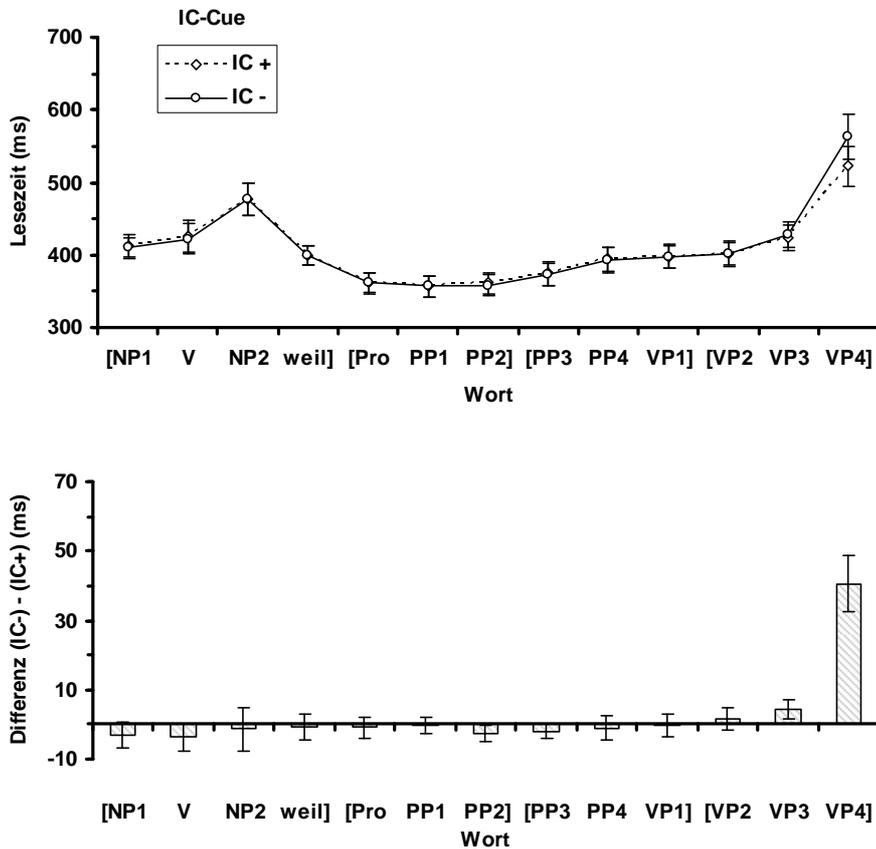


Abb. 8.3: a) Wortweisen Lesezeiten in Abhängigkeit von der Ausprägung des IC-Cues und b) absolute Größe des Kausalitätseffekts im Satzverlauf.

Im Gegensatz zum Gender-Cue fanden sich kaum frühe Effekte für den IC-Cue: Die absoluten Effekte waren vergleichsweise schwach und unsystematisch ausgeprägt. Vielmehr schien sich der IC-Effekt ganz auf das Satzende zu konzentrieren und insbesondere die Größe des *sentence-wrap-ups* zu beeinflussen.

Interaktions-Effekt. Wie Tab. 8.2 zu entnehmen, war die Interaktion der beiden Faktoren vergleichsweise schwach ausgeprägt. Das heißt, die Effekte der beiden Cues schienen weitgehend unabhängig voneinander zu sein. Allerdings gab es in der Mitte des Satzes eine bedeutsame Abweichung von diesem generellen Befund: Sowohl auf dem Pronomen selbst als auch während der ersten Hälfte der Präpositionalphrase fanden sich substanzielle und meist auch (marginal-)signifikante *F*-Werte für den Interaktionsterm, die darauf hinweisen, dass sich der IC-Cue zu diesem (frühen) Zeitpunkt sehr wohl auf die Satzverarbeitung auswirkte, allerdings

in Abhängigkeit von der Ausprägung des Gender-Cues jeweils unterschiedlich. Der differenzielle Verlauf des Kausalitätseffekts ist in Abb. 8.4 dargestellt.

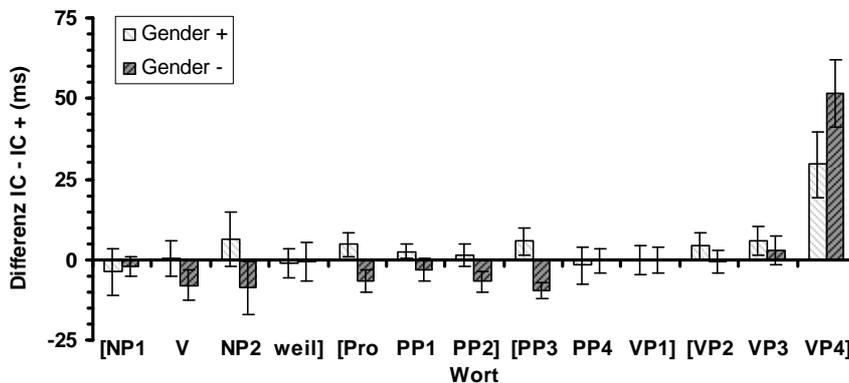


Abb. 8.4: Effekte des IC-Cues in Abhängigkeit von der Ausprägung des Gender-Cues im Satzverlauf.

Auf der Abbildung ist zu sehen, dass der IC-Cue in dem Fall, in dem der Satz nicht bereits durch den Gender-Cue disambiguiert wurde, einen frühen *negativen* Effekt auf die Satzverarbeitung hatte. Im Falle einer gleichzeitigen Disambiguierung ergab sich hingegen ein leichter Verarbeitungsvorteil für die kongruente Bedingung.

8.3.1.1.3 Ergebnisse auf der Konstituentenebene

Konstituentenbildung. Obwohl sich das generelle Befundmuster auch so hinreichend deutlich abzeichnet, bringt es die hohe Auflösung der Wortebene mit sich, dass einzelne Effekte zunächst vergleichsweise schwach und instabil erscheinen. Deswegen wurden in einem zweiten Schritt die einzelnen Wörter des Satzes zu Konstituenten gruppiert und die Lesezeiten zwecks Reliabilitätssteigerung aggregiert. Die gewählte Gruppierung verfolgte dabei die doppelte Zielsetzung, sowohl sinnvoll interpretierbare, als auch ungefähr gleich große Wortgruppen zu erhalten. Die gewählte Einteilung ist in (1) dargestellt:

- (1) [NP1 V NP2 weil] [Pro PP1 PP2] [PP3 PP4 VP1] [VP2 VP3 VP4].

Konstituente 1 (*Hauptsatz*: HS) umfasst alle Wörter, die in den vier Itemversionen identisch sind (natürlich mit Ausnahme der Eigennamen). Diese Konstituente ist damit von eher geringem theoretischem Interesse und es sollten sich eigentlich keine großen Unterschiede für sie finden. Kleinere Differenzen können als leichte Unterschiede im Itemmaterial interpretiert werden. Diese sollten jedoch nicht systematisch auftreten und auch nicht über verschiedene Studien replizierbar sein. Konstituente 2 (*Pronominal-Gruppe*: Pro) umfasst das Pronomen und die beiden nachfolgenden Wörter (also die erste Hälfte der Präpositionalphrase). Der Umfang der Konstituente ergibt sich aus der typischen Reichweite sog. *spill-over*-Effekte, die ungefähr zwei Wörter umfasst (vgl. z.B. Koornneef & van Berkum, 2006). Weiterhin ist aus

der Literatur zur pronominalen Auflösung bekannt, dass sich die meisten Effekte erst ein wenig verzögert auswirken (typischerweise ein oder zwei Worte, vgl. Badecker & Straub, 2002). Die gewählte Konstituentengröße ist also in der Lage auch evtl. zeitverzögerte Effekte des Pronomens zu detektieren. Aus demselben Grund wurde für den Rest des Satzes ebenfalls eine Dreier-Gruppierung gewählt, was zu den beiden Konstituenten 3 (*Präpositionalphrase*: PP) und 4 (*Verbalphrase*: VP) führte. Diese Einteilung steht „quer“ zur syntaktischen Strukturierung des Satzes (von daher sollte die Verwendung des Begriffes „Konstituente“ nicht im linguistischen Sinne verstanden werden). Da jedoch weder das Lesezeitprofil noch die relevanten Effektverläufe eine bemerkbare Beeinflussung durch grammatikalische Konstituentengrenzen aufwiesen (im Gegensatz zur Gliederungsebene „Satz“, die sich ausgesprochen stark im Verarbeitungsprozess niederschlug!), wurde gleich großen Konstituenten der Vorzug gegeben. Auch konnten somit zumindest drei Konstituenten von theoretischem Interesse realisiert werden, was eine zumindest grobe Abbildung der relevanten Effektverläufe ermöglicht.

Durch die Konstituenten-Bildung werden dieselben Daten noch einmal verwendet. Auf den ersten Blick sieht es deshalb so aus, als müsste deswegen eine α -Fehler-Adjustierung vorgenommen werden. Da hier nicht der Platz ist, um diese Frage ausführlich zu diskutieren (vgl. z.B. Cross & Chaffin, 1982), sei schlicht festgestellt, dass im vorliegenden Fall keine α -Fehler-Kumulation stattfindet bzw. nur in einem so geringen Maße, dass eine α -Fehler-Adjustierung nicht notwendig ist. Warum ist das so? Eine α -Fehler-Kumulierung tritt dann auf, wenn ein und dieselbe Hypothese mit mehreren, voneinander unabhängigen inferenzstatistischen Tests überprüft wird. In diesem Fall wird es zunehmend leichter, die Hypothese zu bestätigen. Typische Problemfälle sind z.B. Korrelationsstudien, bei denen alle untersuchten Variablen einfach mit allen anderen korreliert werden, oder Einzelvergleiche bei der Varianzanalyse. In all diesen Fällen wird die gleiche Variable in mehreren empirischen Konstellationen untersucht und in der Tat steigt das α -Fehler-Niveau an, je mehr Kontraste man z.B. prüft. Der vorliegende Fall unterscheidet sich hiervon insofern, dass wirklich *dieselben* Daten und auch keine neuen UVn verwendet werden. Es handelt sich deswegen um *abhängige* Tests, bei denen keine α -Fehler-Kumulierung auftritt (bzw. je nach Abhängigkeitsgrad geringer ausfällt). Ein sehr einfaches Beispiel betrifft den Fall maximaler Abhängigkeit (vgl. Bortz, 1999): Wenn ich ein und denselben inferenzstatistischen Test mehrmals durchführe, dann kommt er immer wieder zum gleichen Ergebnis und wird nicht in einem bestimmten Prozentsatz von Fällen „künstlich“ signifikant werden. Genauso wird eine zweite AV, die perfekt mit der ersten korreliert ist, zum gleichen Ergebnis führen etc. Anders der Fall, wenn für die gleiche Variable mehrere Tests mit verschiedenen *unabhängigen Variablen* durchgeführt werden:

Hier ist der Logik des Signifikanztests entsprechend zu erwarten, dass sich in einem bestimmten Prozentsatz von Fällen ein signifikantes Ergebnis ergibt, obwohl dies auf Populationsebene nicht gilt.

Der Punkt ist nun, dass im vorliegenden Fall durch die Konstituentenbildung hochgradig abhängige Messungen vorliegen (so korrelieren die Werte der einzelnen Wörter mit der Konstituente durchschnittlich im Bereich $r = .95-.99$). Darüber hinaus werden keine neuen UVn in die Hypothesenprüfung eingeführt. Eine α -Fehler-Adjustierung wäre deswegen unangemessen konservativ. Die Wahrscheinlichkeit, unter diesen Umständen einen α -Fehler zu begehen, ist ausgesprochen gering, weshalb von einer Korrektur abgesehen wird. Im Zweifelsfall hat die aggregierte Analyse Vorrang. Ansonsten kann ich mich der Empfehlung von Koch (2001, S. 41) anschließen: „In weiteren Untersuchungen (Replikationen) muss dann überprüft werden, ob man an der getesteten Hypothese festhalten darf oder nicht.“

Tab. 8.3 enthält die Ergebnisse einer zweifaktoriellen messwiederholten ANOVA mit den Faktoren *Gender-Cue* und *IC-Cue*, die für jede Konstituente separat durchgeführt wurde.

	HS	Pro	PP	VP
By-Participants				
<i>Gender-Cue</i>				
<i>F</i> (1,24)	< 1	1.46	6.49	12.76
<i>p</i>	.661	.239	.018*	.002**
η_p^2	.008	.057	.213	.347
<i>IC-Cue</i>				
<i>F</i> (1,24)	< 1	< 1	< 1	17.47
<i>p</i>	.350	.483	.639	.001**
η_p^2	.037	.021	.009	.421
<i>Gender- x IC-Cue</i>				
<i>F</i> (1,24)	1.46	6.91	< 1	< 1
<i>p</i>	.239	.015*	.369	.426
η_p^2	.057	.224	.034	.027
By-Items				
<i>Gender-Cue</i>				
<i>F</i> (1,159)	< 1	4.36	18.06	113.10
<i>p</i>	.780	.038*	.001**	.001**
η_p^2	< .001	.027	.102	.416
<i>IC-Cue</i>				
<i>F</i> (1,159)	< 1	< 1	< 1	10.05
<i>p</i>	.324	.479	.584	.002**
η_p^2	.006	.003	.002	.059
<i>Gender- x IC-Cue</i>				
<i>F</i> (1,159)	< 1	2.29	< 1	< 1
<i>p</i>	.415	.132	.566	.577
η_p^2	.004	.014	.002	.002

Tab. 8.3: Ergebnisse des varianzanalytischen Vergleichs der vier Konstituenten auf Unterschiede zwischen den vier Untersuchungsbedingungen.

Gender-Effekt. Abb. 8.5 zeigt die Effekte des Gender-Cues für die vier Konstituenten.

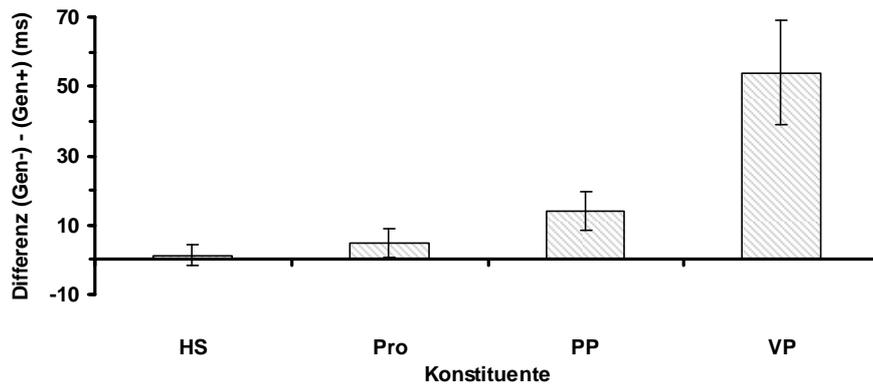


Abb. 8.5: Absolute Größe und Richtung des Gender-Effekts für die vier Konstituenten.

Es ergab sich das gleiche Bild wie für die Wortebene: Der Effekt des Gender-Cues setzt direkt nach dem Pronomen an und vergrößerte sich dann stetig im Verlauf des Satzes. Er kulminierte im *sentence-wrap-up*.

IC-Effekt. Abb. 8.6 zeigt den Haupteffekt des IC-Cues für die vier Konstituenten.

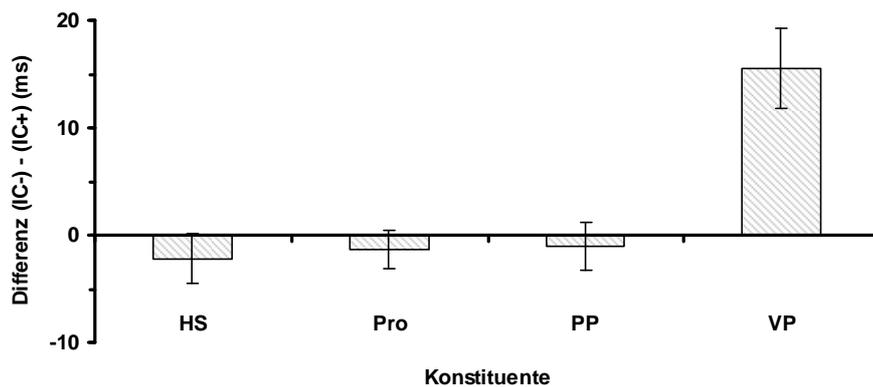


Abb. 8.6: Absolute Größe und Richtung des IC-Effekts für die vier Konstituenten.

Auch hier ergab sich ein sehr ähnliches Effektmuster wie für die Wortebene: Es fanden sich keine frühen Effekte des IC-Cues. Dieser war vielmehr ganz im *sentence-wrap-up* enthalten.

Interaktions-Effekt. Abb. 8.7 zeigt die Einfachen Haupteffekte des IC-Cues für die beiden Ausprägungen des Gender-Cues für die vier Konstituenten.

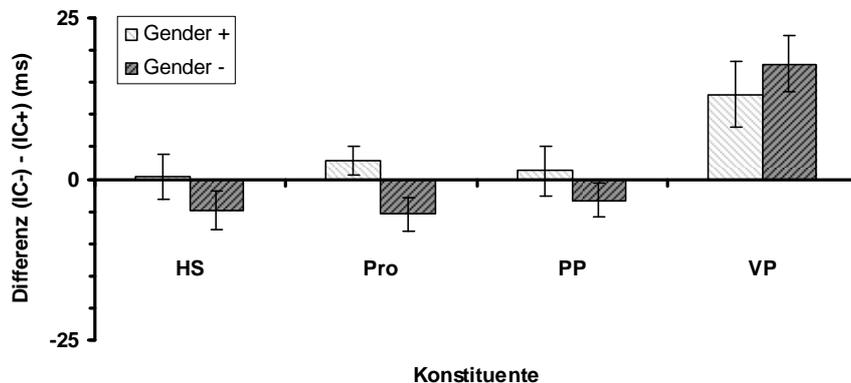


Abb. 8.7: Absolute Größe und Richtung des IC-Effekts in Abhängigkeit von der Ausprägung des Gender-Cues für die vier Konstituenten.

Für die Hauptsatz-Konstituente deutete sich ein leichter Effekt des IC-Cues an, der jedoch nicht signifikant war. Der dissoziierte IC-Effekt auf der Pro-Konstituente war deutlich ausgeprägt: Wenn ein Satz ambig war, dann war der Effekt negativ, wenn aber der Satz bereits eine eindeutige Lesart hatte, dann war der Effekt tendenziell positiv. Diese Interaktion war allerdings nur in der *by-participants*-Analyse signifikant, nicht in der *by-items*-Analyse. Auf der VP-Konstituente deutete sich das gleiche Muster an, war allerdings nicht signifikant.

8.3.1.1.4 Reanalyse der „korrekt“ beantworteten Items

Alle Analysen wurden nochmals nur mit den (bei der nachfolgenden Verifikationsaufgabe) korrekt beantworteten Items durchgeführt. Ein vollständiger Bericht des Ergebnismusters findet sich in Appendix C.3.1. Es blieb vollständig erhalten. Der berichtete Interaktionseffekt war schwächer ausgeprägt und deshalb nur marginal signifikant; trotz der dramatischen Reduktion der Itemmenge bildete er sich in der Mittelwertkonstellation noch deutlich ab.

8.3.1.1.5 Zusammenfassung

Das Ergebnismuster, das sich für die mittleren Lesezeiten ergab, war klar ausgeprägt. Sowohl auf der Wort- als auch auf der Konstituenten-Ebene zeigte sich, dass der Gender-Cue bereits sehr früh die Lesezeit beeinflusste (und in einen großen Effekt am Satzende kulminierte). Der IC-Cue zeigte sich hingegen allein im *sentence-wrap-up*. Zusätzlich gab es einen frühen Interaktionseffekt auf dem Pronomen, der darauf hindeutet, dass der IC-Cue früh die Satzverarbeitung beeinflusste; allerdings je nach Ausprägung des Gender-Cues je unterschiedlich: Wenn der Satz gleichzeitig gender-disambiguiert war, dann hatte der IC-Cue einen positiven Effekt, ansonsten einen negativen.

8.3.1.2 Reaktionszeitverteilung: Weibull-Parameter

8.3.1.2.1 Parameterschätzung, Modellgüte und Parameterprofile

Parameterschätzung. Das Vorgehen bei der Parameterschätzung entsprach prinzipiell dem in Kap. 3.4 skizzierten Vorgehen. An die Verteilung der Leszeiten jedes einzelnen Wortes eines Satzes wurde eine eigene Weibull-Verteilung angepasst, sodass für jeden Satz jeweils drei Parameterprofile vorlagen. Als Startwert für den scale-Parameter wurde ein Wert von $\lambda = \theta^{\beta} = 40$ angesetzt. Die verwendeten Parameter für die beiden Prior-Verteilungen des scale-Parameters waren $a_1 = 2.0$, $b_1 = 0.1$, $a_2 = 2.0$ und $b_2 = 4$.

Modellgüte. Die Passung der Weibull-Verteilung für die einzelnen RT-Verteilungen wurde mittels des in Kap. 3.5 beschriebenen Chi-Quadrat-Verfahrens beurteilt. Der Fit war zwar insgesamt zufrieden stellend (mittleres $\chi^2(3) = 6.32$, $p > .21$), jedoch nicht auf der Ebene der Einzelverteilungen: 161 der insgesamt 1300 (= 25 x 4 x 13) Einzelverteilungen (= 12.7 %) waren auf dem 1 %-Niveau signifikant. Dies ist ein Hinweis darauf, dass die Schätzung von Weibull-Parametern für wortweise Lesezeiten aufgrund der im Vergleich zu üblichen Reaktionszeitverteilungen sehr kurzen und wenig variablen Dauer wesentlich erschwert ist.

Parameterprofile. Abb. 8.8 a) und b) zeigt die durchschnittlichen Weibull-Parameter für die Verteilung der Lesezeiten im Satzverlauf. Der shift- und der scale-Parameter sind dabei in Millisekunden angegeben und gemeinsam in Teilabbildung a) dargestellt. Teilabbildung b) zeigt hingegen den shape-Parameter in seiner üblichen Skalierung.

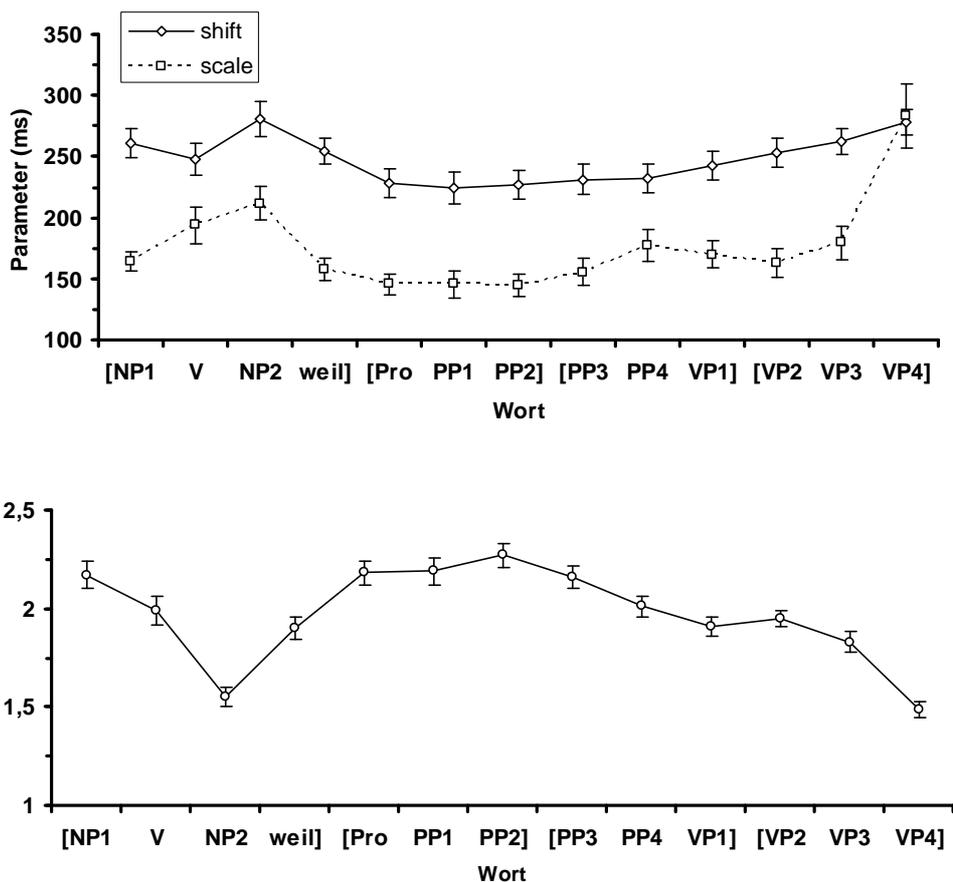


Abb. 8.8: a) Durchschnittliche shift- und scale-Parameter und b) durchschnittliche shape-Parameter für jedes Wort des Satzschemas.

Drei Aspekte sind von besonderer Bedeutung: Erstens, die Lesezeiten zerfielen in zwei unabhängige Komponenten, wobei von der durchschnittlichen Lesezeit pro Wort (ca. 400 ms) ca. 250 ms auf die shift- und ca. 150 ms auf die scale-Komponente entfiel. Zweitens, die shift-Komponente war stationär und lag relativ konstant rund um den Mittelwert von 250 ms (Min: 224 ms, Max: 280 ms, SD = 18.9 ms). Die scale-Komponente war hingegen sehr viel variabler (Min: 145 ms, Max: 283 ms, SD = 37.7 ms) und zeigte ungefähr den gleichen Profilverlauf wie die bereits beschriebenen mittleren Lesezeiten. Diese Beobachtung ist ein weiterer Hinweis auf die Konstruktvalidität der in dieser Arbeit vorgenommenen Interpretation des shift- und scale-Parameters: Die shift-Komponente ist eine minimale Residualkomponente, die vergleichsweise hierarchie-niedrige und periphere Prozesse abdeckt (wie z.B. Stimulus-Enkodierung und lexikalischen Zugriff etc.). Solche Prozesse sind vergleichsweise konstant und variieren nicht wesentlich im Laufe der Satzverarbeitung. Der Wert von 250 ms ist hierbei durchaus realistisch, da er in dem Zeitfenster liegt, der gewöhnlich für basale sprachliche Verarbeitungsprozesse angesetzt wird (vgl. z.B. Balota, 1994; Balota, Yap & Cortese, 2006). Für die scale-Komponente ist hingegen die Annahme wahrscheinlich, dass sie so etwas wie

den kognitiven „overhead“ der Satzverarbeitung abbildet, wenn man sich ihren Verlauf betrachtet: Sie war fast ausschließlich für die Dynamik der Satzverarbeitung verantwortlich und damit sensitiv sowohl für die syntaktische Struktur eines Satzes (syntaktische Komplexität, Arbeitsgedächtnis-Kapazität) als auch für semantische und pragmatische Verarbeitungsprozesse (so wurde z.B. der gesamte *sentence-wrap-up* mit seinen Integrations- und Interpretationskomponenten allein vom scale-Parameter getragen).

Drittens schien es eine komplementäre Beziehung zwischen dem scale- und dem shape-Parameter zu geben, der den Verlauf des scale-Parameters spiegelsymmetrisch fast haargenau nachzeichnete. Für shift- und scale-Parameter lag hingegen ein prinzipiell gleichförmiges Verhältnis vor (wobei der shift-Parameter der schwächere der beiden ist, s.o.). Das Korrelationsmuster der Parameter bestätigte diesen Eindruck: shift- und scale-Parameter korrelierten für die 13 Datenpunkte des Satzes positiv mit $r = .75$ ($p < .003$), während das Profil des shape-Parameters sowohl mit dem des shift- als auch mit dem des scale-Parameters hoch negativ korrelierte (shift: $r = -.86$, $p < .001$; scale: $r = -.86$, $p < .001$). Es muss jedoch darauf hingewiesen werden, dass dieses Beziehungsmuster ausdrücklich nur für die *Parameterprofile* der Sätze galt. Auf der Ebene der Parameterschätzung verhielten sich die Parameter ganz anders zueinander: Hier korrelierte der shape-Parameter negativ mit dem shift- und positiv mit dem scale-Parameter (vgl. Appendix A.5).

8.3.1.2.2 Ergebnisse auf der Wortebene

In einem ersten Schritt wurden die drei für die Verteilung der Lesezeiten eines Wortes geschätzten Weibull-Parameter wieder mittels einer zweifaktoriellen messwiederholten Varianzanalyse daraufhin geprüft, ob sie systematisch mit den beiden manipulierten Auflösungs-Cues variieren. Die entsprechenden Ergebnisse finden sich in Tab. 8.4. Der besseren Übersicht halber sollen jedoch im Folgenden die Effekte der einzelnen Cues wiederum separat besprochen werden.

	[NP1]	V	NP2	weil]	[Pro	PP1	PP2]	[PP3	PP4	VP1]	[VP2	VP3	VP4]
shift													
<i>Gender-Cue</i>													
<i>F(1,24)</i>	< 1	< 1	< 1	2.26	< 1	< 1	< 1	2.56	3.37	1.53	5.69	5.94	4.86
<i>p</i>	.518	.597	.754	.146	.692	.608	.763	.122	.079	.229	.025*	.023*	.037*
η_p^2	.018	.012	.004	.086	.007	.011	.004	.096	.123	.060	.192	.198	.168
<i>IC-Cue</i>													
<i>F(1,24)</i>	1.91	< 1	1.27	< 1	1.58	< 1	1.09	1.38	< 1	< 1	3.60	< 1	4.22
<i>p</i>	.180	.916	.272	.650	.221	.635	.307	.252	.821	.689	.070	.650	.051
η_p^2	.074	< .001	.050	.009	.062	.010	.043	.054	.002	.007	.131	.009	.150
<i>Gender- x IC-Cue</i>													
<i>F(1,24)</i>	3.47	< 1	1.23	2.52	1.31	< 1	< 1	1.45	< 1	< 1	1.75	< 1	1.24
<i>p</i>	.075	.391	.278	.126	.264	.667	.784	.240	.812	.604	.199	.776	.277
η_p^2	.126	.031	.049	.095	.052	.008	.003	.057	.002	.011	.068	.003	.049
scale													
<i>Gender-Cue</i>													
<i>F(1,24)</i>	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	7.19
<i>p</i>	.606	.645	.797	.744	.820	.788	.416	.569	.951	.475	.995	.574	.013*
η_p^2	.011	.009	.003	.005	.002	.003	.028	.014	< .001	.021	< .001	.013	.231
<i>IC-Cue</i>													
<i>F(1,24)</i>	< 1	< 1	1.54	< 1	2.15	< 1	2.98	2.33	< 1	< 1	4.71	< 1	3.82
<i>p</i>	.352	.645	.227	.721	.155	.661	.097	.140	.748	.674	.040*	.899	.062
η_p^2	.036	.009	.060	.005	.082	.008	.111	.088	.004	.007	.164	.001	.137
<i>Gender- x IC-Cue</i>													
<i>F(1,24)</i>	3.21	< 1	< 1	2.56	< 1	< 1	< 1	5.64	< 1	< 1	< 1	< 1	< 1
<i>p</i>	.086	.731	.869	.123	.545	.502	.333	.026*	.890	.540	.330	.940	.636
η_p^2	.118	.005	.001	.096	.016	.019	.039	.190	.001	.016	.040	< .001	.009

	[NP1]	V	NP2	weil]	[er	PP1	PP2]	[PP3	PP4	VP1]	[VP2	VP3	VP4]
shape													
<i>Gender-Cue</i>													
<i>F(1,24)</i>	< 1	< 1	< 1	1.90	< 1	< 1	6.25	< 1	< 1	< 1	4.65	8.73	4.54
<i>p</i>	.374	.864	.762	.181	.492	.889	.020*	.634	.498	.396	.041*	.007**	.044*
η_p^2	.037	.001	.004	.073	.020	.001	.207	.010	.019	.030	.162	.267	.159
<i>IC-Cue</i>													
<i>F(1,24)</i>	1.36	< 1	2.72	1.24	< 1	< 1	< 1	3.35	< 1	< 1	12.63	< 1	2.76
<i>p</i>	.256	.701	.112	.268	.520	.493	.845	.080	.824	.822	.002**	.468	.110
η_p^2	.054	.006	.102	.051	.017	.020	.002	.123	.002	.002	.345	.022	.103
<i>Gender- x IC-Cue</i>													
<i>F(1,24)</i>	< 1	< 1	2.89	4.88	13.27	< 1	< 1	1.10	< 1	< 1	3.83	< 1	< 1
<i>p</i>	.384	.829	.102	.037*	.001**	.999	.501	.304	.497	.434	.062	.540	.565
η_p^2	.037	.002	.107	.169	.356	< .001	.019	.044	.019	.026	.138	.016	.014

Tab. 8.4: Varianzanalytischer Vergleich der drei Weibull-Parameter auf Unterschiede hinsichtlich der vier Untersuchungsbedingungen pro Wort.

Gender-Effekt. Zunächst soll dargestellt werden, wie und vor allen Dingen wann im Laufe der Satzverarbeitung der Gender-Cue die einzelnen Parameter beeinflusst.

shift-Parameter. Abb. 8.9 a) zeigt das Profil des shift-Parameters für die beiden Ausprägungen des Gender-Cues. Die korrespondierende Teilabbildung b) stellt den Effekt des Cues direkt dar (wobei positive Werte wieder für einen Verarbeitungsvorteil stehen).

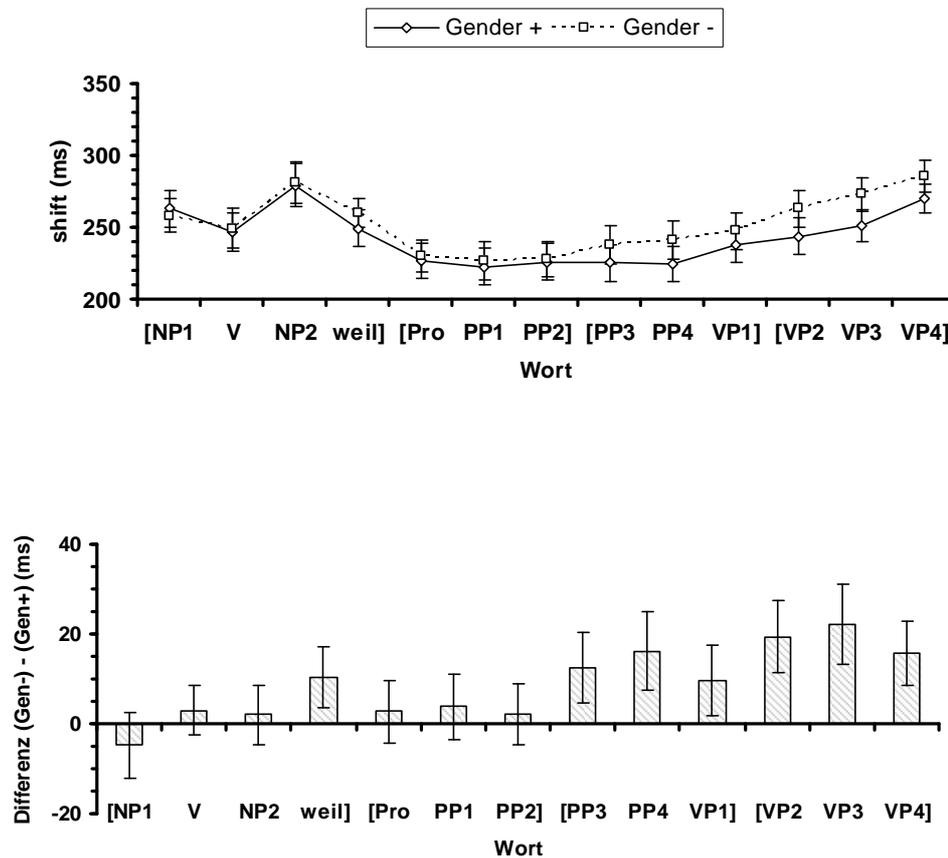


Abb. 8.9: a) Profil des shift-Parameters für die beiden Gender-Bedingungen und b) direkte Darstellung des Gender-Effekts.

Es ist deutlich zu erkennen, dass spätestens ab Mitte der Präpositionalphrase ein deutlicher Effekt für den shift-Parameter ansetzte, der gegen Ende des Satzes an Stärke zunahm.

scale-Parameter. Abb. 8.10 a) und b) zeigen Verlaufsprofile und Effektausprägung des Gender-Cues auf den *scale-Parameter*.

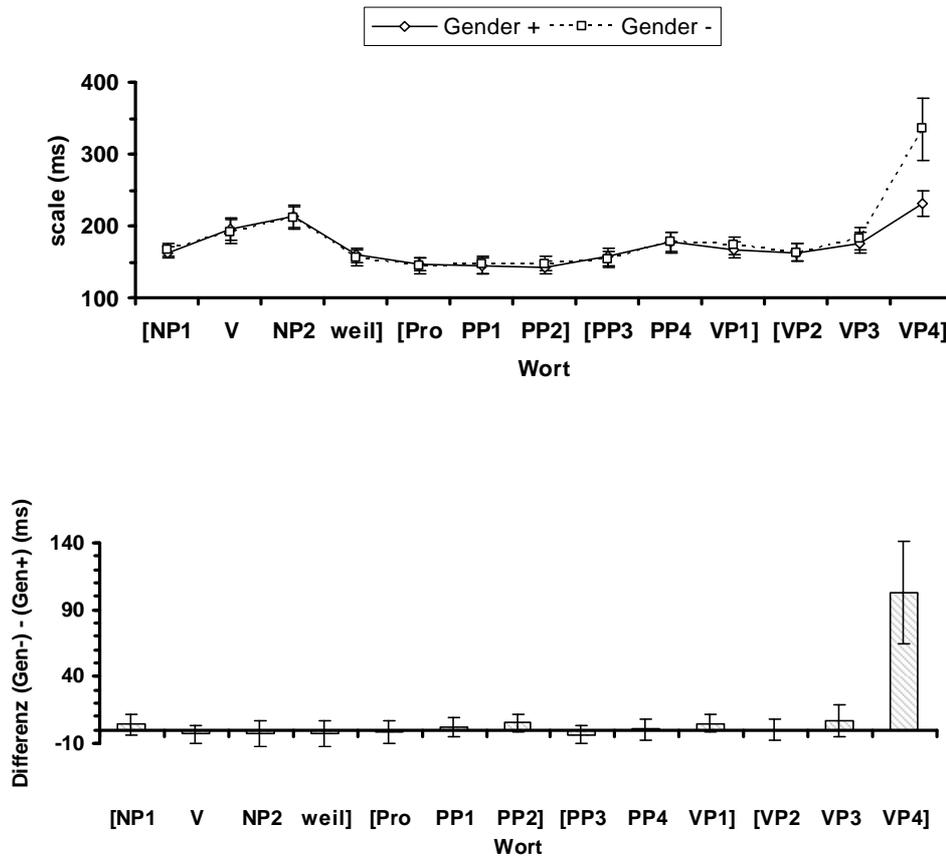


Abb. 8.10: a) Profil des *scale-Parameters* für die beiden Gender-Bedingungen und b) direkte Darstellung des Gender-Effekts

Auch für den *scale-Parameter* war der relevante Effekt des Gender-Cues direkt ersichtlich: Er war ganz im *sentence-wrap-up* enthalten und mit ca. 100 ms ausgesprochen stark ausgeprägt.

shape-Parameter. Abb. 8.11 a) und b) zeigen Verlaufsprofile und Effektausprägung des Gender-Cues auf den *shape*-Parameter.

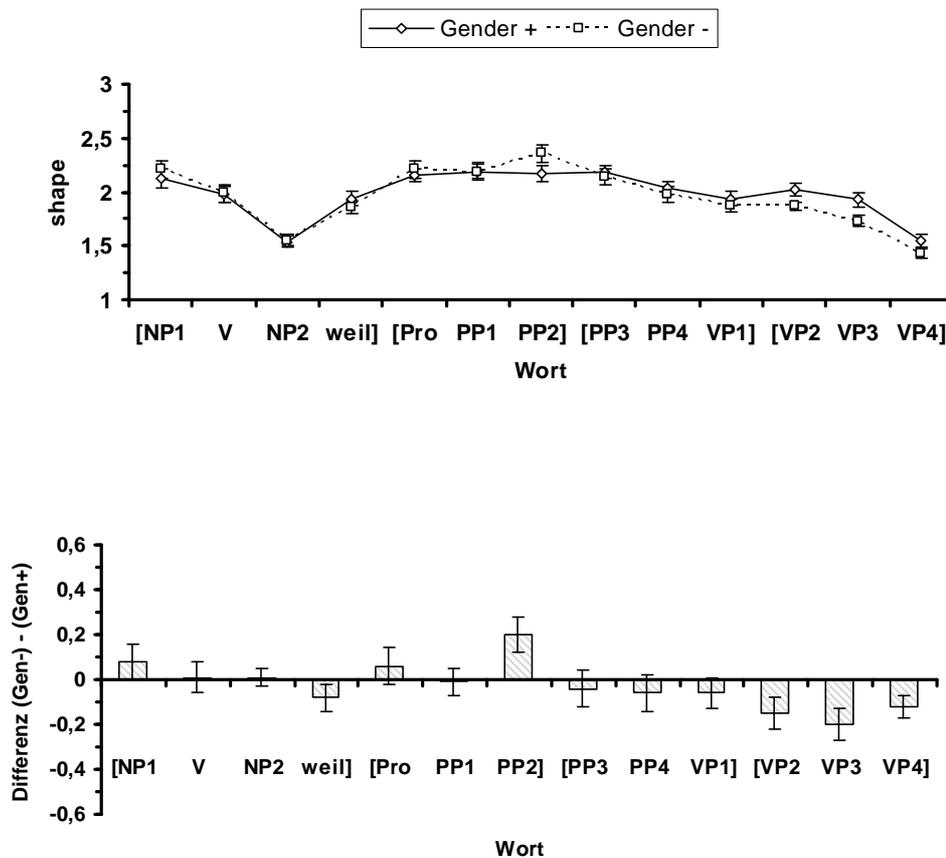


Abb. 8.11: a) Profil des *shape*-Parameters für die beiden Gender-Bedingungen und b) direkte Darstellung des Effekts.

Es gab einen leichten Effekt des Gender-Cues auf den *shape*-Parameter. Er trat gegen Ende des Satzes (genauer: auf der Verbalphrase) auf und führte dazu, dass die Verteilungen in der disambiguierten Bedingung eher normal verteilt war als in der nicht-disambiguierten.

Zusammenfassend kann festgestellt werden, dass die Effektstruktur des Gender-Cues ausgesprochen klar ausgeprägt war: Es gab einen frühen Effekt, der sich allein im *shift*-Parameter niederschlug, und einen (sehr) späten Effekt im *sentence-wrap-up*, der primär vom *scale*-Parameter getragen wurde (und von einem komplementären Effekt des *shape*-Parameters begleitet wurde).

IC-Effekt. Das Bild, das sich für den Kausalitäts-Effekt ergibt, ist komplexer, weshalb seine Darstellung entsprechend differenzierter ausfallen wird. Auch hier soll wieder damit begonnen werden, dass zunächst die direkten Effekte des Cues auf die drei Parameter separat besprochen werden.

shift-Parameter. Abb. 8.12 a) und b) zeigen Verlaufsprofile und Effektausprägung des IC-Cues auf den shift-Parameter.

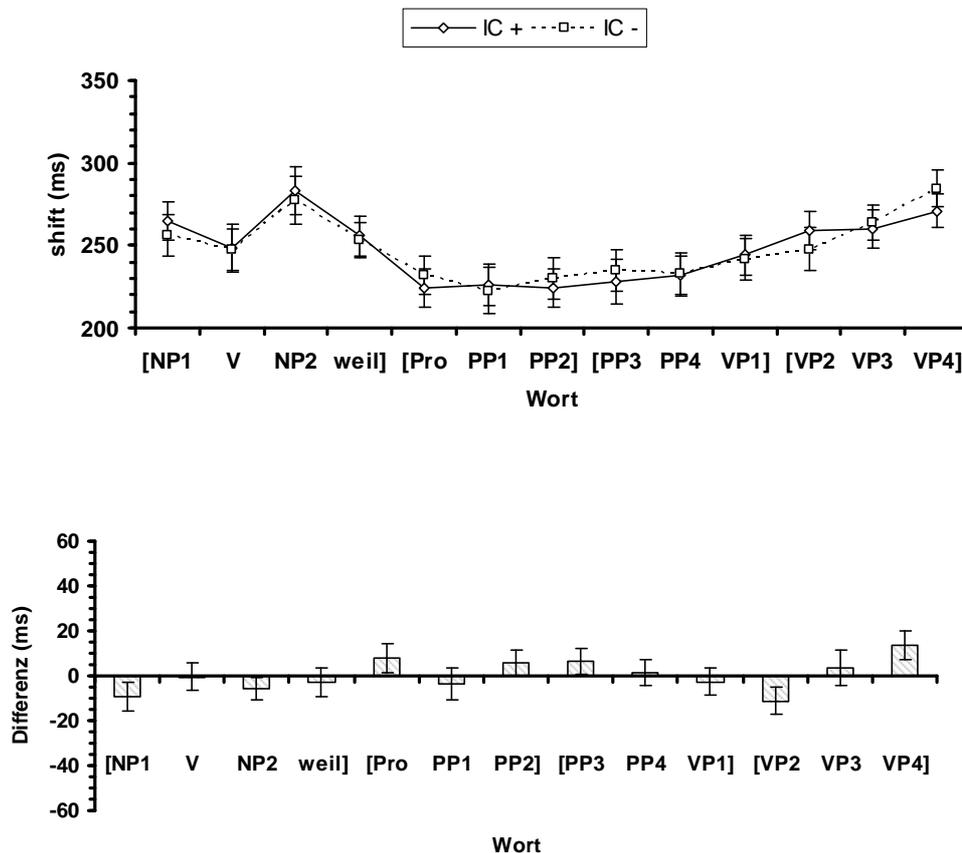


Abb. 8.12: a) Profil des shift-Parameters für die beiden IC-Bedingungen und b) direkte Darstellung des Effekts.

Insgesamt waren die Auswirkungen des Kausalitäts-Cues auf den shift-Parameter vergleichsweise schwach ausgeprägt. An zwei Stellen im Satz waren jedoch Ansätze von „shift-Aktivität“ zu beobachten: 1. Am Ende des Satzes fanden sich marginal signifikante Effekte des IC-Cues, die aber sowohl negativ als auch positiv ausfielen, weshalb sich kein klarer Trend abzeichnete. 2. In der Mitte des Satzes (auf dem Pronomen und der ersten Hälfte der Präpositionalphrase) fanden sich schwache, aber konsistent positive Effekte des IC-Cues, die allerdings statistisch nicht eindeutig ausfielen.

scale-Parameter. Abb. 8.13 a) und b) zeigen Verlaufsprofile und Effektausprägung des IC-Cues auf den scale-Parameter.

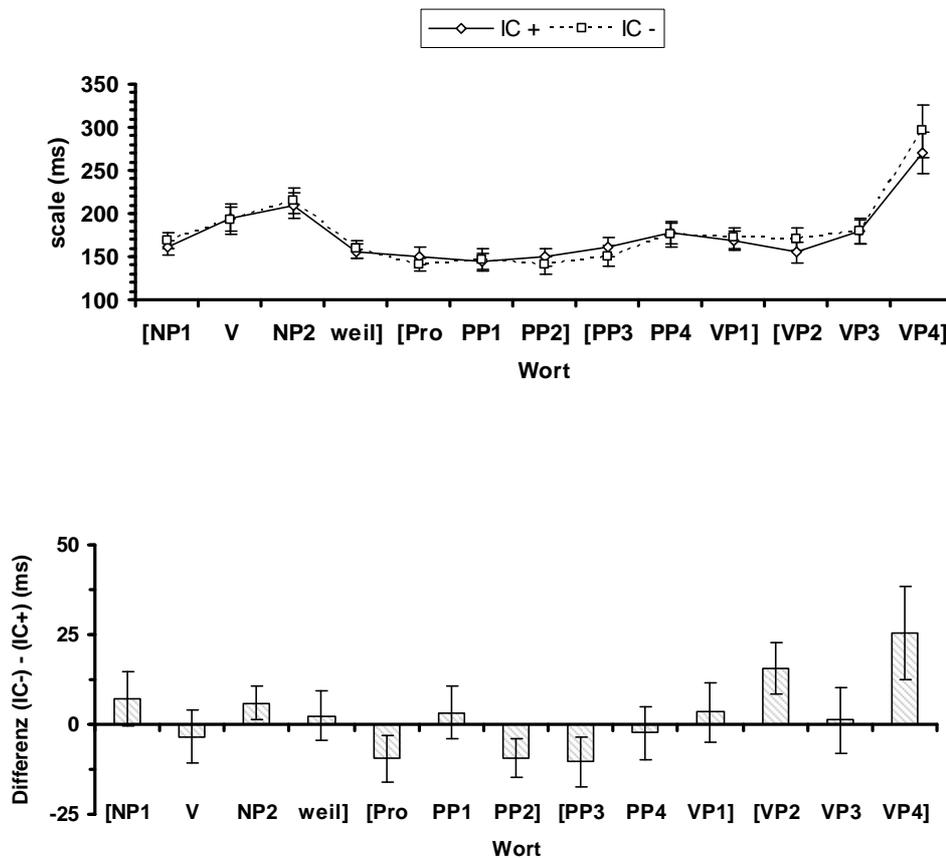


Abb. 8.13: a) Profil des scale-Parameters für die beiden IC-Bedingungen und b) direkte Darstellung des Effekts.

Für den scale-Parameter ergab sich eine zwar grundsätzlich sehr ähnliche, teilweise jedoch komplementäre Effektkonstellation wie für den shift-Parameter, wobei die scale-Effekte prinzipiell stärker ausgeprägt waren als die shift-Effekte. Auch hier gab es zwei „hot spots“: Zunächst einmal gab es einen klaren positiven Effekt am Satzende, auch wenn dieser *wrap-up*-Effekt selbst knapp das Signifikanzniveau von 5 % verfehlte. Darüber hinaus waren jedoch auch für den scale-Parameter zusätzliche Effekte auf dem Pronomen und während der Präpositionalphrase zu erkennen – auch hier waren sie vergleichsweise schwach ausgeprägt und statistisch nicht vollkommen überzeugend. Im Gegensatz zum Positiv-Effekt auf den shift-Parameter an der gleichen Stelle war der Effekt für den scale-Parameter klar *negativ* ausgeprägt, d.h. es gab es einen frühen Verarbeitungsnachteil des IC-Cues.

shape-Parameter. Abb. 8.14 a) und b) zeigen die Verlaufsprofile und Effektausprägungen des IC-Cues auf den *shape-Parameter*.

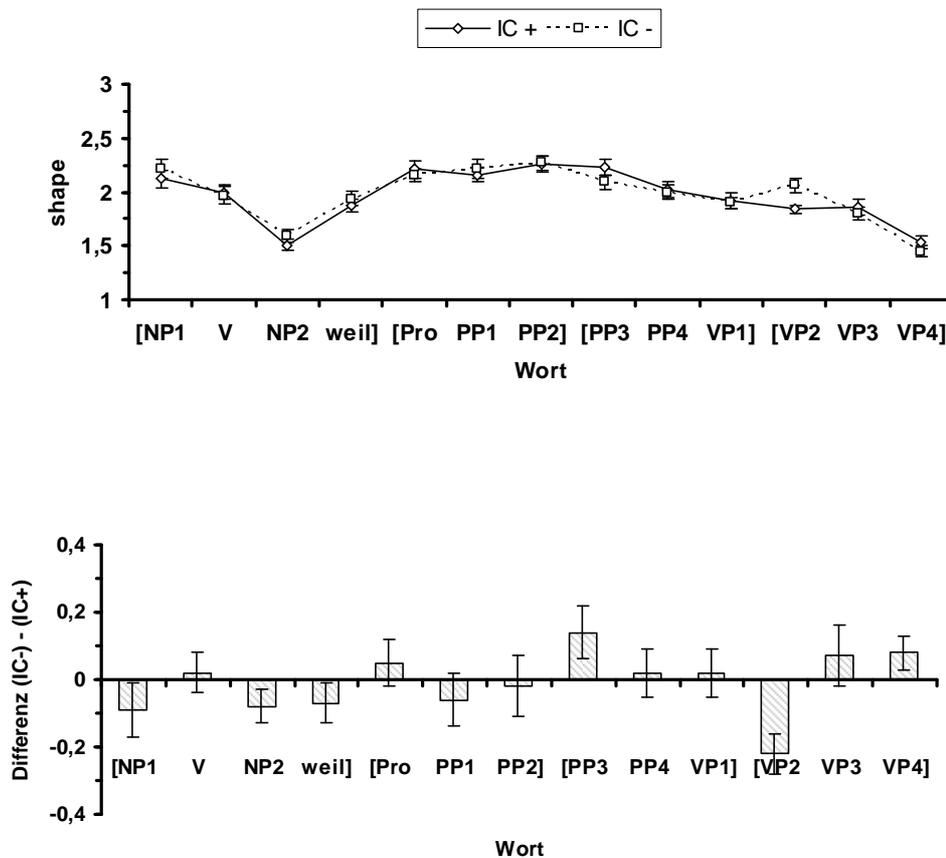


Abb. 8.14: a) Profil des *shape-Parameters* für die beiden IC-Bedingungen und b) direkte Darstellung des Effekts.

Die Auswirkung des IC-Cues auf den *shape-Parameter* war ausgesprochen inkonsistent: Auch wenn die Effekte an zwei Stellen signifikant waren, zeichnete sich kein klarer Trend ab. Grundsätzlich traten jedoch auch hier die Effekt vermehrt am Satzende auf, allerdings gab es auch im Satzinneren ein paar inhomogene Effekte.

Als Zwischenstand lässt sich für den IC-Cue festhalten, dass es zunächst einen klaren Verarbeitungsvorteil für begründungskongruente Nebensätze gab, der sich hauptsächlich im *scale-Parameter* am Satzende niederschlug und von leichten shift-Effekten begleitet wurde. Darüber hinaus gab es jedoch auch vergleichsweise schwache und heterogene Effekte im Satzinneren, speziell auf dem Pronomen selbst: Tendenziell ließ sich hier ein positiver shift-Effekt von einem negativen *scale*-Effekt abgrenzen. Beide Tendenzen waren jedoch nicht so stark ausgeprägt, dass man ihnen an dieser Stelle vollkommen vertrauen sollte.

Interaktions-Effekt. In einem zweiten Schritt ist deswegen besonders die Frage interessant, ob die Effekte des IC-Cues in Abhängigkeit von der Ausprägung des Gender-Cues unterschiedlich ausfallen und sich in unterschiedlichen Reaktionszeit-Komponenten niederschlagen.

shift-Parameter. Abb. 8.15 zeigt die Ausprägungen des IC-Effekts auf den shift-Parameter im Satzverlauf in Abhängigkeit von der Ausprägung des Gender-Cues.

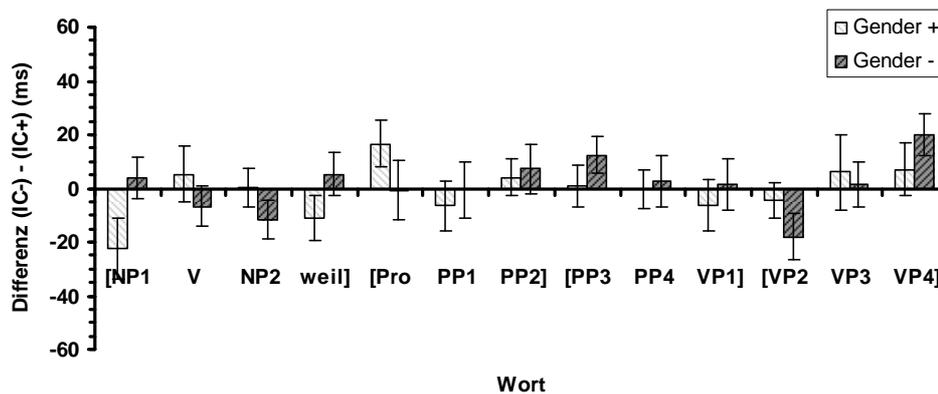


Abb. 8.15: Darstellung des IC-Effekts auf den shift-Parameter in Abhängigkeit vom Gender-Cue.

Auf der Ebene der einfachen Haupteffekte wurde das Effektmuster wesentlich instabiler. Hinzu kommt, dass alle Interaktionseffekte nur sehr schwach ausgeprägt waren. Eine Inspektion der relevanten F -Werte ergab jedoch, dass Interaktionsvarianz zunächst einmal für den Satz-anfang vorhanden war. Wie in Abb. 8.15 zu sehen, sind die Interaktionen meist disordinaler Natur, d.h. die IC-Effekte gingen in den beiden Gender-Bedingungen in jeweils unterschiedliche Richtungen, wobei es keine kohärente Zuordnung zwischen Bedingung und Richtung des Effekts gab. Von größerem Interesse waren hingegen Unterschiede, die sich weiter hinten im Satzverlauf ergaben. Die wichtigste Stelle war das Pronomen selbst: Auch wenn der Interaktionsterm selbst nicht signifikant war, ergibt eine Analyse der einfachen Haupteffekte hier eine deutliche Effektstruktur: Während der IC-Effekt in der Bedingung *Gender+* substanziell und marginal signifikant war ($F(1,24) = 3.64, p = .068, \eta_p^2 = .151$), traf dies für die Bedingung *Gender-* nicht zu ($F(1,24) < 1, p = .949$). Eine weitere Stelle, an der die Effekte in den beiden Bedingungen voneinander abwichen, war der Anfang der Verbalphrase (speziell Wort VP2). Hier gab es einen signifikant negativen einfachen Haupteffekt des IC-Cues in der Bedingung *Gender-* ($F(1,24) = 4.29, p < .049, \eta_p^2 = .152$), jedoch keinen in der Bedingung *Gender+* ($F(1,24) < 1, p = .520$). Und schließlich gab es einen positiven shift-Effekt auf dem letzten Wort des Satzes, allerdings nur in der Bedingung *Gender-* ($F(1,24) = 6.68, p < .016, \eta_p^2 = .218$) nicht aber in der Bedingung *Gender+* ($F(1,24) < 1, p = .472$). Ein ähnlicher Effekt war auch für das Wort PP3 beobachtbar.

scale-Parameter. Abb. 8.16 zeigt die Ausprägungen des IC-Effekts auf den *scale-Parameter* im Satzverlauf in Abhängigkeit von der Ausprägung des Gender-Cues.

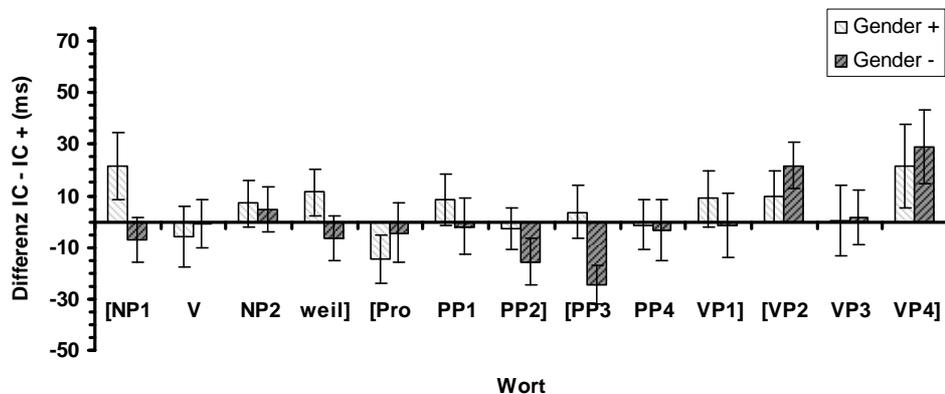


Abb. 8.16: Darstellung des IC-Effekts auf den *scale-Parameter* in Abhängigkeit vom Gender-Cue.

An zwei Stellen war die Interaktion der beiden Faktoren systematisch ausgeprägt: In der Präpositionalphrase gab es einen marginal signifikanten, negativen Effekt in der Bedingung *Gender-* (PP2: $F(1,24) = 3.05$, $p = .094$, $\eta_p^2 = .113$; PP3: $F(1,24) = 10.20$, $p < .004$, $\eta_p^2 = .298$), jedoch nicht in der Bedingung *Gender+*. Des Weiteren war der IC-Effekt in der Verbalphrase (v.a. auf dem Wort VP2) in der Bedingung *Gender-* stärker ausgeprägt als in der Bedingung *Gender+* ($F(1,24) = 5.74$, $p < .025$, $\eta_p^2 = .193$ vs. $F(1,24) < 1$, $p = .347$, $\eta_p^2 = .037$).

shape-Parameter. Abb. 8.17 zeigt die Ausprägungen des IC-Effekts auf den *shape-Parameter* im Satzverlauf in Abhängigkeit von der Ausprägung des Gender-Cues.

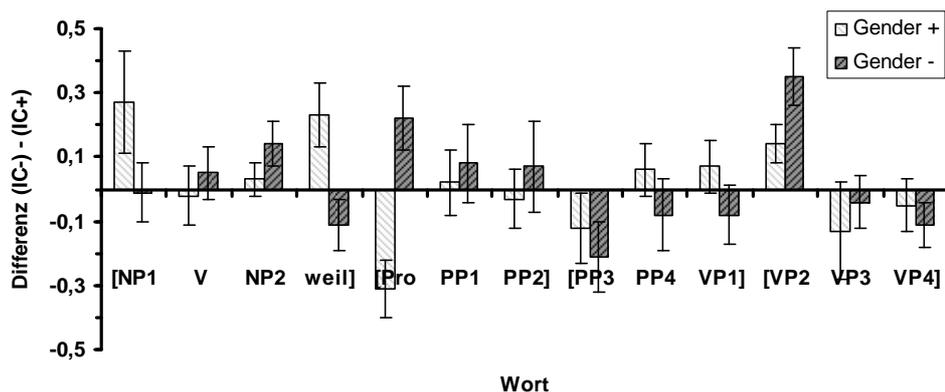


Abb. 8.17: Darstellung des IC-Effekts auf den *shape-Parameter* in Abhängigkeit vom Gender-Cue.

Die Effekte des IC-Cues waren wiederum inkonsistent, auch wenn sie stärker ausfielen als für *shift-* und *scale-Parameter*. An zwei Stellen im Satz ist der Interaktionseffekt signifikant bzw. marginal signifikant. Die wichtigste hiervon war das Pronomen selbst, für das sich eine disordinale Interaktion ergab: Während der Effekt des IC-Cues in der Bedingung *Gender+* negativ war ($F(1,24) = 9.86$, $p < .004$, $\eta_p^2 = .291$), war er in der Bedingung *Gender-* positiv ($F(1,24)$

= 4.72, $p < .040$, $\eta_p^2 = .164$). Diese Konstellation war vollkommen komplementär zum Interaktionseffekt des shift-Parameters. Dies deutet auf Abhängigkeiten in der Parameterschätzung hin (vgl. Appendix A.5).

Ein weitere, diesmal ordinale Interaktionskonstellation ergab sich auf dem Wort V2: Hier war der Effekt in der Bedingung *Gender-* stärker ausgeprägt als in der Bedingung *Gender+* ($F(1,24) = 12.96$, $p < .002$, $\eta_p^2 = .334$ vs. $F(1,24) = 3.33$, $p = .081$, $\eta_p^2 = .122$). Auch hier war der Effekt komplementär zur Interaktion für den shift-Parameter, aber parallel zu der des scale-Parameters. Dies spricht wiederum für Abhängigkeiten in der Parameterschätzung, wobei erwartungsgemäß scale- und shape-Parameter positiv miteinander assoziiert waren, während sich beide zusammen vom shift-Parameter abgrenzen. Auf diese Probleme wird in der Diskussion näher eingegangen.

Zusammenfassend kann zunächst festgehalten werden, dass die Interaktion der beiden Faktoren an insgesamt drei Stellen des Satzes für das Verständnis der Effekte relevant war:

1. Direkt auf dem Pronomen zeigten sich frühe IC-Effekte, allerdings nur, wenn der Satz gleichzeitig gender-disambiguiert war. In diesem Fall ergab sich ein Verarbeitungsvorteil für die ic-kongruente Fortsetzung des Hauptsatzes, der sich v.a. im shift-Parameter abbildete. Komplementär hierzu hatte der IC-Cue einen negativen Effekt auf den shape-Parameter. Beide Effekte waren jedoch sehr schwach und lediglich marginal signifikant.

2. In der Mitte des Satzes, speziell in der Präpositionalphrase, deuteten sich leicht *negative* Effekte des Kausalitäts-Cues auf die Satzverarbeitung an. Diese wurden vollständig vom scale-Parameter getragen. Dieser Effekt war in den Ergebnissen von Experiment I nicht erkennbar. Allerdings ist es aber auch ohne Betrachtung des *Verlaufs* der Effekte nicht möglich, ihn von den finalen scale-Effekten abzugrenzen.

3. Am Satzende, in der Verbalphrase (und insbesondere auf dem Wort VP2) gab es eine ordinale Interaktion der beiden Faktoren. Der IC-Effekt war dabei jeweils in der Bedingung *Gender-* stärker ausgeprägt als in der Bedingung *Gender+*. Die Interaktion der beiden Faktoren war allerdings nicht signifikant.

8.3.1.2.3 Ergebnisse auf Konstituentenebene

In einem zweiten Schritt wurden die Parameter für die einzelnen Wörter zu den bereits in Abschnitt 8.3.1.1.3 beschriebenen Konstituenten zusammengefasst. Dadurch sollten relativ flächig auftretende Effekte stabiler werden, weil durch diese Form der Aggregation Datenvariabilität minimiert wird. Auf der anderen Seite werden so eher lokale, kurzfristige und inkonsistente Effekte evtl. verdeckt, weil sie sich durch diese Maßnahme gegenseitig ausmitteln. Tab. 8.5 stellt die Ergebnisse des varianzanalytischen Vergleichs für die vier Konstituenten und die drei Parameter dar.

Gender-Effekt. Abb. 8.18 zeigt die Effekte des Gender-Cues für die drei Weibull-Parameter auf den vier Konstituenten.

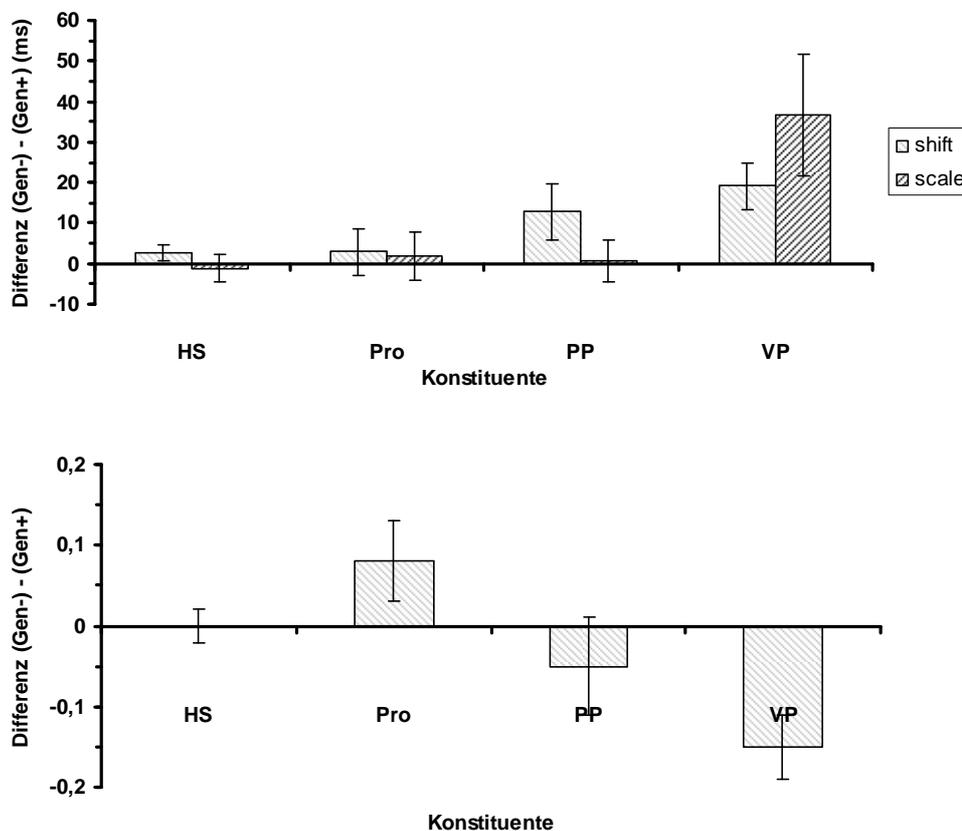


Abb. 8.18: Effekte des Gender-Cues auf a) den shift- und den scale-Parameter sowie b) den shape-Parameter jeder Konstituente des Satzschemas.

Für den Gender-Cue ergab sich das inzwischen bekannte Bild, welches keiner weiteren Erläuterung mehr bedarf: Es gab einen frühen positiven Effekt des Cues auf den shift-Parameter und einen späten positiven für den scale-Parameter, der von einem Negativ-Effekt des shape-Parameters begleitet wurde. Der frühe shift-Effekt war allerdings nur marginal signifikant.

	<i>shift</i>				<i>scale</i>				<i>shape</i>			
	HS	Pro	PP	VP	HS	Pro	PP	VP	HS	Pro	PP	VP
<i>Gender-Cue</i>												
<i>F(1,24)</i>	1.55	< 1	3.31	10.61	< 1	< 1	< 1	5.96	< 1	2.60	< 1	12.39
<i>p</i>	.225	.622	.081	.003**	.754	.768	.908	.022*	.847	.120	.397	.002**
η_p^2	.061	.010	.121	.307	.004	.004	.001	.199	.002	.098	.030	.341
<i>IC-Cue</i>												
<i>F(1,24)</i>	2.15	< 1	< 1	< 1	1.04	1.19	< 1	4.01	3.01	< 1	2.49	< 1
<i>p</i>	.156	.479	.651	.657	.318	.286	.567	.057	.096	.880	.128	.601
η_p^2	.082	.021	.009	.008	.041	.047	.014	.143	.111	.001	.094	.012
<i>Gender- x IC-Cue</i>												
<i>F(1,24)</i>	< 1	< 1	< 1	< 1	2.16	< 1	1.63	< 1	1.66	3.36	1.42	< 1
<i>p</i>	.481	.813	.421	.817	.154	.721	.215	.434	.210	.079	.246	.357
η_p^2	.021	.002	.027	.002	.083	.005	.063	.026	.065	.123	.056	.035

Tab. 8.5: Ergebnisse des varianzanalytischen Vergleichs der vier Konstituenten auf Unterschiede hinsichtlich der vier Untersuchungsbedingungen.

IC-Effekt (Gender+). Abb. 8.19 zeigt die Effekte des IC-Cues für die drei Weibull-Parameter, wenn der Satz gleichzeitig durch den Gender-Cue disambiguiert wurde.

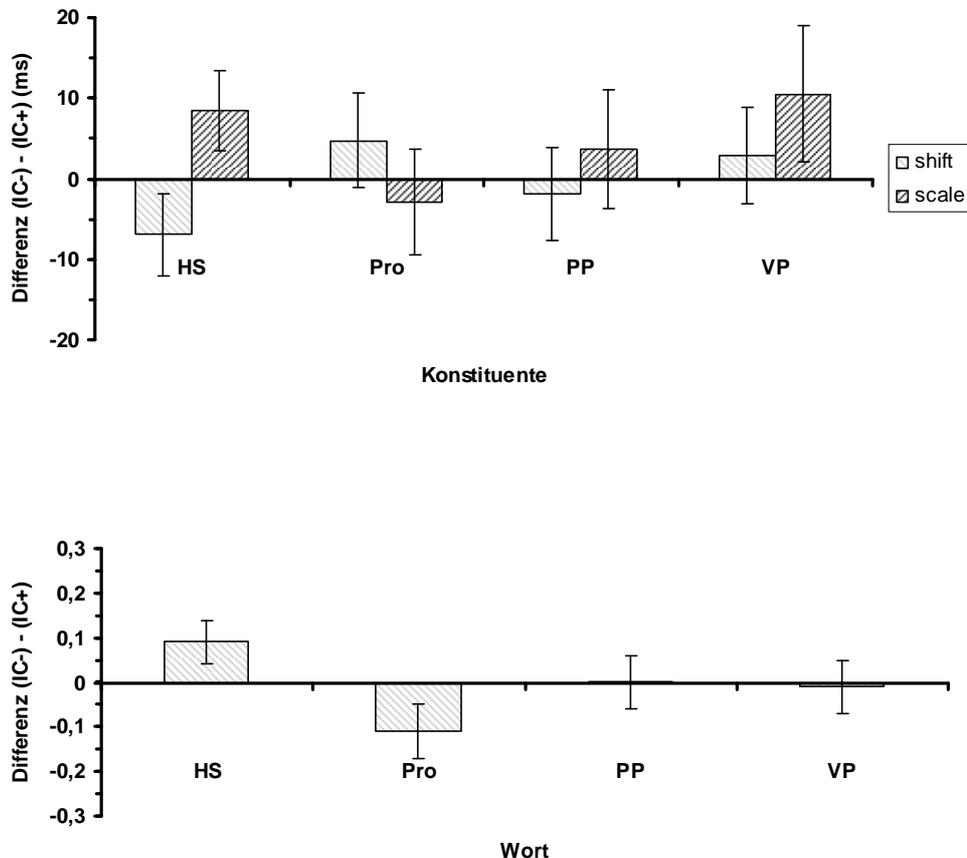


Abb. 8.19: Effekte des IC-Cues in der Bedingung Gender + auf a) den shift- und den scale-Parameter sowie b) den shape-Parameter jeder Konstituente des Satzschemas.

Im Gegensatz zur Wort-Ebene zeigten sich nach der Aggregation zu Konstituenten keine Effekte des shift-Parameters mehr. Zwar gab es auf der Pronomen-Konstituente einen substantiellen Positiv-Ausschlag, der jedoch statistisch nicht reliabel ausfiel ($F < 1$, $p = .431$). Es gab jedoch einen späten Effekt für den scale-Parameter, der auf Haupteffekts-Ebene das angesetzte Signifikanz-Niveau nur knapp verfehlte. Für den shape-Parameter gab es zwei frühe, marginal signifikante Effekte auf der Hauptsatz- und auf der Pro-Komponente (HS: $F(1,24) = 3.65$, $p = .068$, $\eta_p^2 = .132$; Pro: $F(1,24) = 2.94$, $p = .099$, $\eta_p^2 = .109$).

Zusammenfassend kann festgehalten werden, dass es auf der Konstituenten-Ebene im Vergleich zur Wortebene keinen frühen Effekt des IC-Cues auf dem Pronomen gab. Er deutete sich jedoch für den shift-Parameter an und wurde von einem komplementären Effekt des shape-Parameters begleitet. Der bereits bekannte späte scale-Effekt war hingegen vergleichsweise deutlich ausgeprägt, fiel jedoch absolut gesehen auch eher schwach aus.

IC-Effekt (Gender-). Abb. 8.20 zeigt die Effekte des IC-Cues für die drei Weibull-Parameter, wenn der Satz nicht gleichzeitig durch den Gender-Cue disambiguiert wurde.

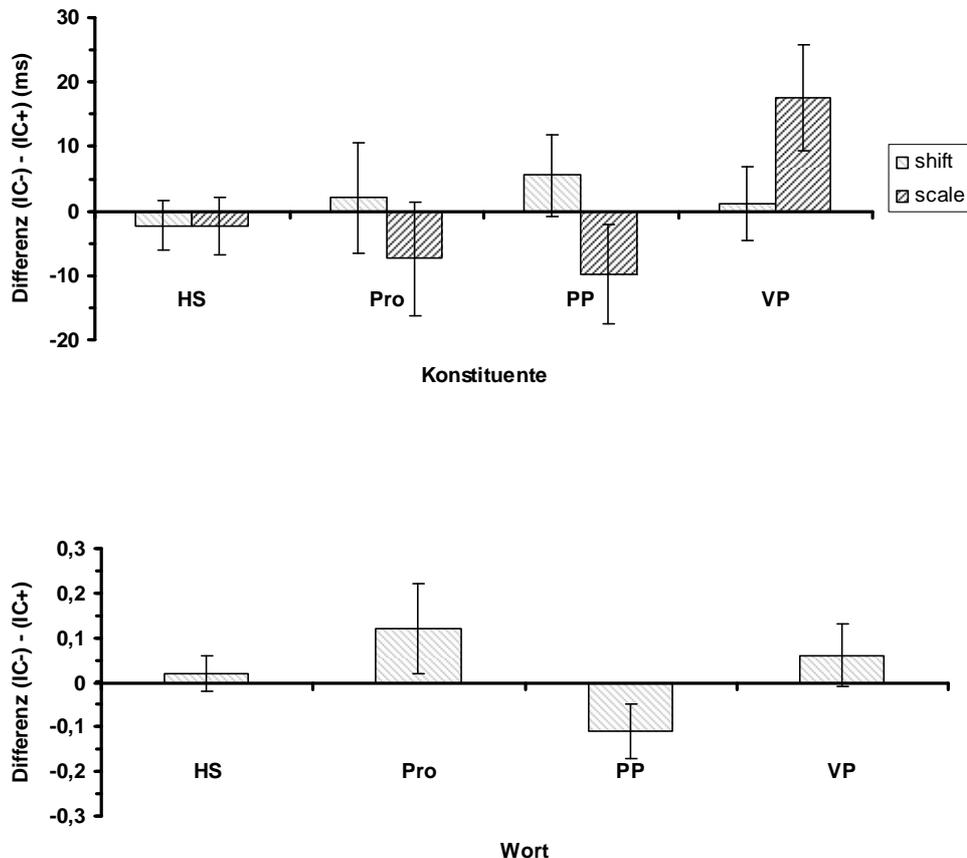


Abb. 8.20: Effekte des IC-Cues in der Bedingung Gender – auf a) den shift- und den scale-Parameter sowie b) den shape-Parameter jeder Konstituente des Satzschemas.

Für den shift-Parameter gab es keine signifikanten Effekte (alle $F < 1$). Der späte scale-Effekt auf der VP-Komponente war jedoch stabil ausgeprägt ($F(1,24) = 4.65, p < .041, \eta_p^2 = .162$). Ein früherer, negativer scale-Effekt deutete sich für die beiden vorhergehenden Konstituenten an, war jedoch ebenfalls zu schwach, um als reliabel bezeichnet werden zu können (Pro: $F < 1, p = .413$; PP: $F(1,24) = 1.65, p = .211$). Auf der Pro- und der PP-Komponente gab es zwei zusätzliche, aber inkonsistente Effekte für den shape-Parameter.

Zusammenfassend lässt sich für die Ergebnisse auf der Konstituenten-Ebene feststellen, dass sich fast alle postulierten Effekte auch nach der Aggregation problemlos wieder finden ließen. Dies galt sowohl für den frühen und den späten Teileffekt des Gender-Cues als auch für den späten des IC-Cues. Weniger eindeutig waren die Ergebnisse für die frühen Effekte des IC-Cues: Hier deutete sich zwar sowohl der frühe shift-Effekt in der disambiguierten Bedingung an als auch der etwas später gelegene scale-Effekt in der Bedingung *Gender+*.

8.3.1.2.4 Re-Analyse der „korrekt“ beantworteten Items

Alle Analysen wurden mit einer Auswahl von 20 Items, die in der nachfolgenden Verifikationsaufgabe korrekt beantwortet wurden, neu berechnet. Ein vollständiger Bericht des Ergebnismusters findet sich in Appendix C.3.2. Alle Effekte waren schwächer ausgeprägt als in der vollständigen Analyse. Insbesondere nach Aggregation auf Konstituentenebene zeigte sich jedoch ein ähnliches Ergebnismuster. Dies galt v.a. für die beiden Effekte des Gender-Cues (früh auf dem shift-Parameter und spät auf dem scale-Parameter) als auch für den späten Effekt des IC-Cues (auf dem scale-Parameter). Auch gab es einen deutlichen Effekt des IC-Cues auf die frühe Satzverarbeitung. Im Gegensatz zur vollständigen Analyse war dieser Effekt jedoch primär im scale-Parameter zu verorten.

8.3.1.2.5 Zusammenfassung

Das Ergebnis der Analyse der Lesezeitverteilungen war in der Lage, das bisherige Befundmuster entscheidend zu ergänzen: Der frühe Gender-Effekt wurde hauptsächlich vom shift-Parameter getragen, während der späte Gender-Effekt und der späte IC-Effekt reine scale-Effekte waren. Des Weiteren gab es einige Hinweise für einen frühen IC-Effekt, der sich je nach Ausprägung des Gender-Cues nicht nur in seiner Richtung (positiv-negativ) unterschied, sondern auch darin, von welchem Parameter er hauptsächlich gespeist wurde: Der frühe IC-Effekt in der gender-disambiguierten Untersuchungsbedingung lag primär im shift-Parameter, während der IC-Effekt in dem Fall, in dem die Referenz des Pronomens noch nicht durch den Gender-Cue festgelegt wurde, v.a. vom scale-Parameter getragen wurde. Allerdings waren beide Effekte vergleichsweise schwach ausgeprägt und konnten sich in der zu Konstituenten aggregierten Analyse inferenzstatistisch nicht absichern lassen.

8.3.1.3 Diskussion Satzlesezeiten

Zusammenfassend ist festzustellen, dass die Evidenz für die verschiedenen Teileffekte der beiden Auflösungs-Cues bei den mittleren Lesezeiten eigentlich recht gut war. Für den Gender-Cue ließen sich bereits früh Effekte ab dem ersten Wort der Präpositionalphrase nachweisen, die in einem starken *wrap-up*-Effekt kulminieren. Für den IC-Cue fand sich hingegen eine frühe Interaktion der beiden Faktoren auf dem Pronomen selbst: Wenn der Satz disambiguiert war, kam es zu einem Verarbeitungsnachteil für den IC-Cue, wenn er jedoch nicht-disambiguiert war, sogar zu einem leichten Verarbeitungsvorteil für den IC-Cue. Für den IC-Cue kam es zusätzliche zu einem starken *wrap-up*-Effekt, der erwartungsgemäß im disambiguierten Fall schwächer ausfiel.

Schwieriger ist die Beurteilung der Ergebnisse der RT-Verteilungsanalyse: Während die satzfinalen Effekte sehr klar dem scale-Parameter zugeordnet werden konnten, waren die Effekte für die beiden frühen Cue-Effekte insgesamt nur eher schwach ausgeprägt und inkonsistent. Der frühe Gender-Effekt konnte vergleichsweise sicher der shift-Komponente zugeordnet werden. Der frühe IC-Effekt war jedoch zu schwach, um sich inferenzstatistisch reliabel absichern zu lassen. Bemerkenswert ist, dass es in Experiment IIa noch einen weiteren frühen Effekt des IC-Cues gab. Bei ihm handelte es sich um einen negativen Effekt für den scale-Parameter. Dieser Effekt trat nur auf, wenn der Satz ambig war und kulminierte in einem extrem stark ausgeprägten *sentence-wrap-up*. Auch dieser Effekt war vergleichsweise schwach und ließ sich nicht generalisieren.

8.3.2 Verifikationsaufgabe

8.3.2.1 Akkuratheit: Anteil „korrekt“ verifizierter Aussagen

Ergebnisse. Abb. 8.21 zeigt die Akkuratheit (Hit-Rate) für die Verifikationsaufgabe in allen vier Untersuchungsbedingungen.

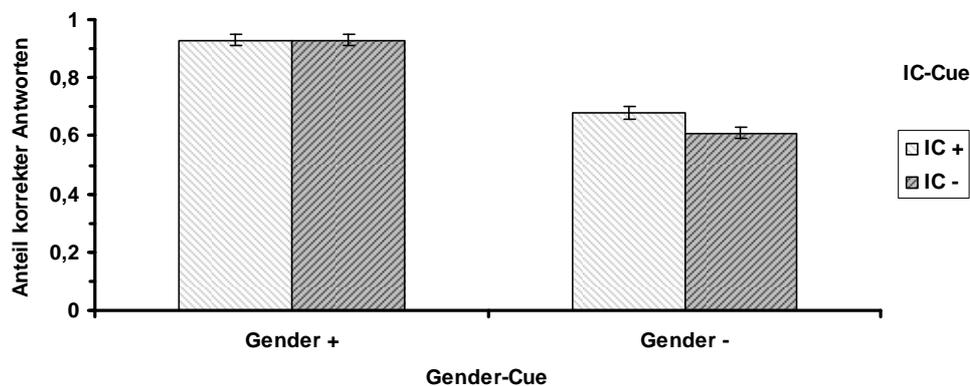


Abb. 8.21: Anteil „korrekt“ verifizierter Aussagen nach Aussagetyp.

Das Antwortmuster ist inzwischen vertraut, und entsprechend ergab eine messwiederholte ANOVA mit den beiden Faktoren *Gender-Cue* (2: vorhanden vs. nicht vorhanden) und *IC-Cue* (2: vorhanden vs. nicht-vorhanden) auch die übliche Effektkonstellation: Einen sehr starken Haupteffekt für den Faktor *Gender-Cue* ($F(1,24) = 304.63, p < .001, \eta_p^2 = .927$; $F(1,159) = 464.07, p < .001, \eta_p^2 = .745$): Wenn ein Satz nicht gender-disambiguiert war, kam es zu einem dramatischen Abfall an korrekten Antworten (ca. 28 %). Darüber hinaus gab es auch einen schwächern Haupteffekt für den Faktor *IC-Cue* ($F(1,24) = 7.77, p < .010, \eta_p^2 = .245$; $F(1,159) = 4.62, p < .033, \eta_p^2 = .028$), der wieder durch eine signifikante Interaktion ($F(1,24) = 9.06, p < .006, \eta_p^2 = .274$; $F(1,159) = 4.69, p < .032, \eta_p^2 = .029$) modifiziert

wurde: Wenn ein Satz bereits gender-disambiguiert war, dann gab es keinen Effekt für den IC-Cue ($F < 1$), wohl aber, wenn der Satz ambig war (ca. 7 %; $F(1,24) = 11.54, p < .002, \eta_p^2 = .325$; $F(1,159) = 5.66, p < .019, \eta_p^2 = .034$).

Arcussinus-Transformation. Um eine Normalverteilung der Residuen zu gewährleisten, wurde die Analyse nochmals mit arcussinus-transformierten Werten durchgeführt. Die vollständigen Ergebnisse sind in Appendix C.3.3 einsehbar. Das Ergebnismuster war jedoch vollkommen identisch zu dem soeben berichteten.

Zusammenfassung. Für die Verifikationsaufgabe ergab sich ein sehr ähnliches Ergebnismuster wie in Experiment I: Auffälliger ist hingegen, dass die Akkuratheitswerte in der Bedingung ohne Gender-Disambiguierung deutlich niedriger ausfiel als bisher (ca. 60 %). Auch ist der Kausalitätseffekt nur sehr schwach ausgeprägt. Die Akkuratheit in der Bedingung mit Gender-Disambiguierung lag hingegen auf dem üblichen hohen Niveau von über 90 %.

8.3.2.2 Mittlere Reaktionszeiten

Datenbereinigung. Für die pronominale Auflösung wurde wiederum ein Wert von 200 ms als unterer Cut-off-Wert angesetzt. Kein Datenpunkt fiel unter diesen kritischen Grenzwert. Weiterhin wurden alle Werte, die 4 Standardabweichungen vom bedingungspezifischen Mittelwert einer Person abwichen, durch diesen ersetzt. Diese Maßnahme führte zu einer Ersetzung von ca. 0.4 % aller Datenpunkte.

Ergebnisse. Abb. 8.22 zeigt die mittlere Zeit, die durchschnittlich zur Verifizierung von Sätzen in den unterschiedlichen Untersuchungsbedingungen benötigt wurde.

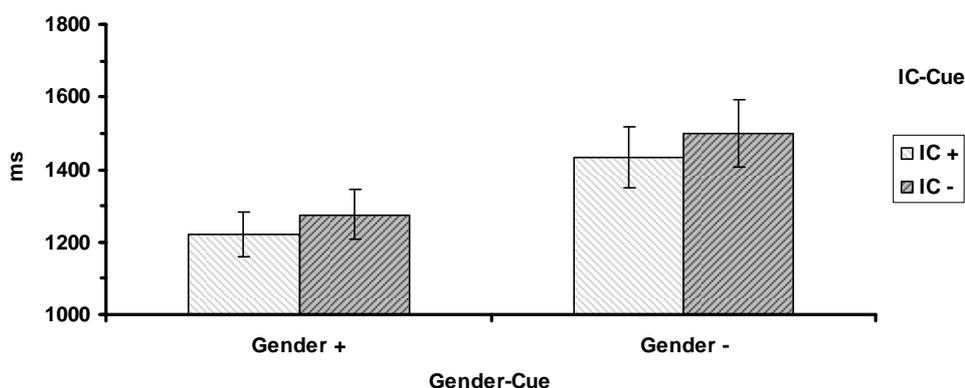


Abb. 8.22: Mittlere RT für die Verifikationsaufgabe in den vier Untersuchungsbedingungen.

Einerseits gab es einen varianzstarken Haupteffekt für den Faktor *Gender-Cue* ($F(1,24) = 53.06, p < .001, \eta_p^2 = .689$; $F(1,159) = 90.81, p < .001, \eta_p^2 = .364$), da gender-disambiguierte Sätze ca. 220 ms schneller verifiziert wurden als nicht-disambiguierte. Ande-

rerseits gab es einen Effekt für den Faktor *IC-Cue* ($F(1,24) = 11.77, p < .002, \eta_p^2 = .329$; $F(1,159) = 8.85, p < .003, \eta_p^2 = .053$), der darauf zurückgeführt werden konnte, dass ikongruente Nebensätze ca. 60 ms schneller verifiziert wurden. Die beiden Faktoren interagierten nicht miteinander (beide $F < 1$).

Re-Analyse der „korrekten“ Antworten. Um zu überprüfen, ob das berichtete Effektmuster evtl. durch Unterschiede zwischen falsch und richtig beantworteten Items moderiert wurde, wurde die eben berichtete Analyse nochmals nur mit den „korrekt“ beantworteten Items gerechnet. Das Ergebnismuster (s. Appendix C.3.4) blieb identisch erhalten; beide Effekte waren sogar noch stärker ausgeprägt.

Zusammenfassung. Die mittleren Antwortlatenzen für die Verifikationsaufgabe zeigten ein ähnliches Muster, wie es für die pronominale Auflösung in Experiment I typisch war: Beide Cues haben einen Effekt auf die pronominale Auflösung, interagierten aber nicht miteinander. Der Gender-Effekt war dabei nach wie vor stärker ausgeprägt als der IC-Effekt, beide Effekte waren jedoch wesentlich kleiner als bei der pronominalen Auflösung in Experiment I. Die Größe der Effekte lag in einem ähnlichen Bereich wie bei der Verifikationsaufgabe in Experiment I. Damit zeigte die Verifikationsaufgabe nun Charakteristika, wie sie sowohl von Vonk (1985a, 1985b) als auch Garnham et al. (1992) berichtet wurden.

8.3.2.3 Reaktionszeitverteilung: Weibull-Parameter

Parameterschätzung. Das Vorgehen bei der Parameterschätzung entsprach dem in Kap. 3.4 skizzierten Vorgehen. Als Startwert für den scale-Parameter wurde wiederum ein Wert von $\lambda = \theta^\beta = 2$ angesetzt. Die Parameter der beiden Prior-Verteilungen des scale-Parameters waren $a_1 = 2.0, b_1 = 0.1, a_2 = 2.0$ und $b_2 = 0.2$.

Modellgüte. Die Passung der Weibull-Verteilung für die einzelnen RT-Verteilungen wurde mittels des in Kap. 3.5 beschriebenen Chi-Quadrat-Verfahrens beurteilt. Der Fit konnte insgesamt als akzeptabel bezeichnet werden: Der mittlere Chi-Quadrat-Wert lag bei $\chi^2(3) = 5.40, p < .27$. Sieben der insgesamt 100 (= 25 x 4) Einzelverteilungen (= 7.0 %) wurden auf dem 1 %-Niveau zurückgewiesen.

Durchschnittliche Parameterausprägung. Abb. 8.23 zeigt die drei Weibull-Parameter der Reaktionszeit-Verteilung und ihre Beziehungen zueinander.

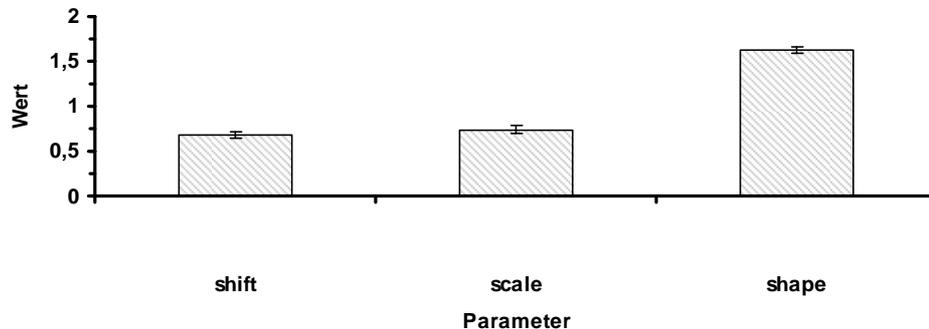


Abb. 8.23: Durchschnittliche Weibull-Parameter der RT-Verteilungen der Verifikationsaufgabe.

Der durchschnittliche shift-Parameter lag bei 0.68 s, der durchschnittliche scale-Parameter bei 0.74 s und der durchschnittliche shape-Parameter bei 1.63. Damit ergaben sich für alle drei Parameter fast identische Ergebnisse wie für die Verifikationsaufgabe in Experiment Ib.

Diese drei Parameter wurden nun jeweils separat mittels einer messwiederholten ANOVA mit den beiden Faktoren *Gender-Cue* (2: *vorhanden* vs. *nicht vorhanden*) und *IC-Cue* (2: *vorhanden* vs. *nicht vorhanden*) auf Unterschiede zwischen den Untersuchungsbedingungen geprüft. Dabei ergaben sich folgende Ergebnisse:

shift-Parameter. Für diesen Parameter gab es keine signifikanten Effekte (alle $F < 1.2$). Lediglich der Interaktionsterm klärte überhaupt ein wenig Varianz auf, verfehlte jedoch das angesetzte Signifikanzniveau ($F(1,24) = 2.74, p = .111$).

scale-Parameter. Es gab einen varianzstarken Haupteffekt für den Faktor *Gender-Cue* ($F(1,24) = 52.47, p < .001, \eta_p^2 = .686$): Reaktionszeitverteilungen für die Sätze, bei denen das Pronomen nicht gender-disambiguiert war, wiesen einen um ca. 240 ms höheren scale-Parameter auf als Sätze mit disambiguierten Pronomen (vgl. Abb. 8.24). Gleichzeitig gab es einen Haupteffekt für den Faktor *IC-Cue* ($F(1,24) = 15.85, p < .001, \eta_p^2 = .398$): In den Untersuchungsbedingungen, in denen die IC und die explizite Auflösung übereinstimmten, fiel der scale-Parameter ca. 100 ms niedriger aus. Die Interaktion der beiden Faktoren war marginal signifikant ($F(1,24) = 2.95, p = .099$) und in der Mittelwertkonstellation klar ausgeprägt: Der IC-Cue wies nur dann einen Effekt auf, wenn der Satz nicht bereits gender-disambiguiert war ($F(1,24) = 13.01, p < .001, \eta_p^2 = .352$); in der Bedingung *Gender+* war der Effekt hingegen weniger stark ausgeprägt und nicht signifikant ($F(1,24) = 2.06, p = .164, \eta_p^2 = .079$).

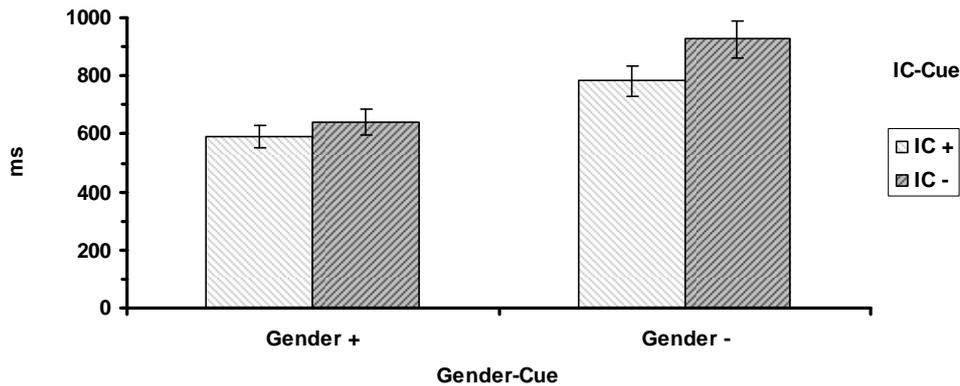


Abb. 8.24: Ausprägung des scale-Parameters der RT-Verteilungen der Verifikationsaufgabe.

shape-Parameter. Der Haupteffekt des Gender-Cues war nicht signifikant ($F < 1$) und auch für den IC-Cue gab es nur einen sehr schwachen Effekt ($F(1,24) = 4.87, p < .037, \eta_p^2 = .169$). Die Interaktion der beiden Faktoren war marginal signifikant ($F(1,24) = 3.36, p = .079, \eta_p^2 = .123$): Komplementär zum Ergebnismuster für den scale-Parameter gab es nur dann einen Effekt des IC-Cues, wenn der Satz bereits gender-disambiguiert war ($F(1,24) = 8.54, p < .007, \eta_p^2 = .262$), sonst nicht ($F < 1$).

Re-Analyse der „korrekten“ Antworten. Für die Analyse der „korrekt“ beantworteten Items ergab sich ein leicht unterschiedliches Ergebnismuster (vgl. Appendix C.3.5): Der Gender-Cue beeinflusste weiterhin nur den scale-Parameter, der IC-Cue hingegen unterschiedliche Parameter in Abhängigkeit davon, ob der Satz bereits durch den Gender-Cue disambiguiert war: In der Bedingung *Gender+* bildete er sich im shift-Parameter ab, in der Bedingung *Gender-* jedoch im scale-Parameter. Damit verhielt sich der IC-Effekt ähnlich wie bei der pronominalen Auflösung in Experiment Ib.

Zusammenfassung. Ähnlich wie die mittlere Reaktionszeit zeigten auch die Weibull-Parameter der Verifikationsaufgabe Charakteristika, wie sie in Experiment I für die pronominalen Auflösung typisch waren: Die Effekte wurden ausschließlich vom scale-Parameter getragen und interagierten nicht miteinander. Allerdings ist interessant, dass sich beim IC-Cue das in Experiment Ib beschriebene Trade-off-Muster tendenziell auch bei der Verifikationsaufgabe zeigt: Der scale-Effekt war in der Bedingung *Gender-* stärker ausgeprägt, während es bei den korrekt beantworteten Items einen zusätzlichen shift-Effekt in der Bedingung *Gender+* gab.

8.3.2.4 Diskussion Verifikationsaufgabe

Vielleicht lässt sich das Ergebnismuster, wie es sich für die Verifikationsaufgabe des vorliegenden Experiments zeigte, am besten als Mischung der Ergebnismuster von pronomi-

nalere Auflösung und Verifikationsaufgabe in Experiment I beschreiben. Einerseits war die Interaktion der beiden Cues vergleichsweise schwach ausgeprägt bzw. überhaupt nicht detektierbar. Dies galt sowohl für die mittleren Reaktionszeiten als auch für die Weibull-Parameter. Das ist ein Befund, wie er für die pronominale Auflösung in Experiment I typisch war. Vonk (1985a, 1985b) berichtete jedoch ein ähnliches Muster für ihre Verifikations-Studien, in denen ebenfalls keine vorherige Auflösung des Pronomens erforderlich war. Allerdings deutete sich bei der Analyse der korrekt beantworteten Items die Trade-off-Relation der beiden IC-Effekte an: In der disambiguierten Bedingung war der IC-Effekt im shift-Parameter zu verorten, während er in der nicht-disambiguierten Bedingung im scale-Parameter lag.

Die Tatsache, dass die Effekte beider Cues insgesamt etwas stärker ausfielen als für die Verifikationsaufgabe in Experiment I (und auch stärker als bei Vonk, 1985a) sprach jedoch gleichzeitig dafür, dass das Verstehensprodukt bei Weitem nicht so stabil war, als wenn man den ganzen Satz wie in Experiment I auf einmal liest. Dafür, dass sich die Art der Satzdarbietung insgesamt negativ auf das Satzverständnis und speziell die Auflösung des Pronomens ausgewirkt hatte, sprachen übrigens auch die im Vergleich zu Experiment I wesentlich erhöhten Fehlerraten. Dies galt speziell für die Bedingungen, in denen die Referenz des Pronomens allein durch das semantisch-pragmatische Weltwissen etabliert werden musste.

8.4 Diskussion Experiment IIa

Insgesamt hat Experiment IIa eher gemischte Ergebnisse erbracht. In einigen Punkten hat es die Ergebnisse aus Experiment I bestätigt und konnte sie sogar noch weiter ausdifferenzieren. In Bezug auf andere Aspekte ist die Befundlage jedoch uneindeutig und nur schwer zu interpretieren.

Sehr eindeutig sind die Ergebnisse in Bezug auf den Gender-Effekt. Hier konnte wie bereits in Experiment I ein hierarchie-hoher *disengagement*- und ein hierarchie-niedriger *confirmation*-Effekt unterschieden werden, wobei beide unabhängig voneinander zum Gender-Effekt beitragen. Die wesentliche Erkenntnis, die in Experiment IIa hinzutrat, ist, dass sich die beiden Komponenten in der Tat auch zeitlich voneinander dissoziieren lassen. Der *confirmation*-Effekt bildete sich in der Analyse der mittleren Reaktionszeiten schon ziemlich früh ab: Er fing bereits auf dem ersten Wort nach dem disambiguierenden Pronomen (PP1) an und steigerte sich dann im Laufe der Satzverarbeitung allmählich in seiner Größe. Die differenzierende Analyse der einzelnen Reaktionszeit-Parameter bestätigte, dass an diesem frühen Effekt vorwiegend hierarchie-niedrige Prozesse beteiligt sind, da er sich ausschließlich im shift-Parameter der Weibull-Verteilung niederschlug. Insgesamt waren jedoch die Ergebnisse der

RT-Verteilungsanalyse schwächer ausgeprägt als für die mittleren Reaktionszeiten; der Effekt konnte auch erst später, nämlich am Ende der Präpositionalphrase, reliabel nachgewiesen werden. Nicht nur der Ansatzpunkt des Effekts (früh) und die RT-Verteilungskomponente (shift) sprechen dafür, dass es sich um einen primär gedächtnis-basierten Prozess handelt. Auch seine Verlaufsdynamik macht dies sehr wahrscheinlich, da er sich im Laufe der Satzverarbeitung steigert. Ein plausibles Erklärungsmodell für diese Dynamik bietet der Resonanz-Mechanismus der gedächtnis-basierten Sprachverarbeitungstheorien. Im Rahmen dieses Modells kommt es zu einem sehr ähnlichen Effekt, nämlich dem sog. Distraktor- oder Fan-Effekt (vgl. Kap. 1.2). Dieser entsteht, wenn für ein anaphorisch anzubindendes Textelement mehrere Antezedenten zur Verfügung stehen, die Merkmale mit ihm teilen. Die Merkmalsüberlappung von Referent und Nicht-Referent führt dazu, dass das Abrufelement das gewünschte Antezedens nicht spezifisch ansprechen kann. Daraus ergibt sich eine langsamere Konvergierdynamik des Resonanzprozesses.

Allerdings liegt auch vergleichsweise gute Evidenz dafür vor, dass ein Gender-Cue zu einer erleichterten Integration des Pronomens am Satzende führte. Bei den mittleren Lesezeiten machte sich dieser Sachverhalt v.a. durch einen starken Effekt des Cues im *sentence-wrap-up* bemerkbar. Dieser fiel durch die methodenspezifische Betonung finaler Integrationsprozesse sogar besonders groß aus (vgl. Abschnitt 8.1). Die RT-Verteilungsanalyse ergab weiterhin, dass der satzfinale Integrationseffekt tatsächlich hauptsächlich vom scale-Parameter getragen wurde. Auch hier konzentrierte er sich ausschließlich auf den *wrap-up*. Er wurde von einem negativen Effekt des shape-Parameters begleitet. Wie in Experiment I persistierte der späte Integrationseffekt des Gender-Cues auch noch in der nachfolgenden Verifikationsaufgabe.

Die Befundlage für die IC-Effekte ist hingegen sehr viel schwieriger einzuschätzen. Der IC-Integrations-Effekt war in Experiment IIa noch vergleichsweise deutlich ausgeprägt. Bei den mittleren Lesezeiten fanden sich stark erhöhte Lesezeiten im *sentence-wrap-up*, wenn inkongruente Fortsetzungen gelesen wurden. Erwartungsgemäß fiel der Effekt stärker aus, wenn der Satz ambig war, allerdings ist die Interaktion nur marginal signifikant. Weniger eindeutig sind hingegen die Ergebnisse der RT-Verteilungsanalyse. Diese ergab für den finalen Effekt des IC-Cues neben klaren scale-Effekten auch leichte shift-Effekte. Beide Effekte waren jedoch nur äußerst schwach ausgeprägt und lediglich marginal signifikant. Der scale-Anteil war erwartungsgemäß nur in der ambigen Bedingung signifikant, während in der disambiguierten Bedingung weder für den shift- noch den scale-Parameter ein signifikanter Effekt verzeichnet werden konnte. Insgesamt ist die RT-Verteilungsanalyse damit relativ in-

konklusiv. Tendenziell spricht sie jedoch eher dafür, dass es im vorliegenden Experiment zu einem Integrationseffekt kam, der differenziell mit dem Gender-Cue interagierte. Ziemlich eindeutig sind wiederum die Ergebnisse für die Verifikationsaufgabe: Wie in Experiment I fanden sich hier persistierende Integrationsschwierigkeiten für ic-inkongruente Fortsetzungen, wenn die Sätze gender-ambig waren.

Am uneindeutigsten ist jedoch die Evidenz für den IC-Fokus-Effekt. Einerseits gibt es starken Grund zur Annahme, dass es in Experiment IIa zu einer frühen Refokussierung in der disambiguierten Bedingung kam. Dafür sprechen u.a. signifikante Interaktionsterme auf dem Pronomen und in der Präpositionalphrase, die indizieren, dass inkongruente Fortsetzungen *sofort* langsamer verarbeitet wurden, während sich in der ambigen Bedingung die Lesegeschwindigkeit sogar erhöhte. Dieser Schereneffekt hielt prinzipiell die gesamte Präpositionalphrase über an, war jedoch insgesamt eher lokal begrenzt. Insbesondere zeigte er nicht die kumulierende Dynamik des frühen Gender-Effekts. Von daher scheint er eher auf einer punktuellen Verarbeitungsschwierigkeit der inkonsistenten Fortsetzung zu beruhen, die wahrscheinlich auf eine direkte Refokussierung bzw. Szenario-Umstrukturierung zurückzuführen ist. Dafür spricht zumindest, dass der satzfinale Integrationseffekt in der disambiguierten Bedingung kleiner ausfiel. Hier scheinen durch die vorbereitende Refokussierung weniger Integrationsprozesse notwendig zu werden. Obwohl das berichtete Effektmuster nur sehr schwach ist, stellt es eine direkte Replikation der Ergebnisse von Koornneef und van Berkum (2006) dar. Diese konnten ebenfalls den frühen Integrationseffekt mit einer *moving-window*-Methode nachweisen. Allerdings geht die vorliegende Untersuchung noch einen Schritt weiter und belegt, dass dieser Effekt nur auftritt, wenn der Satz bereits gender-disambiguiert ist. Da Koornneef und van Berkum keine Daten zur satzfinalen Integration berichten, kann nicht beurteilt werden, ob es bei ihnen auch zu einem finalen Integrations-Effekt kam.

Die Ergebnisse der RT-Verteilungsanalyse sind wiederum unklar. Insbesondere lassen sich keine vernünftigen Aussagen zur Zusammensetzung des IC-Fokus-Effekts machen, da die einzelnen Teileffekte insgesamt nur sehr schwach und inhomogen waren. Deswegen ließen sich auch keine signifikanten Interaktionseffekte auf den Pronomen nachweisen. – weder für den shift- noch für den scale-Parameter. Eine Analyse der einfachen Haupteffekte ergab allerdings, dass der frühe Fokus-Effekt in der Tat auf eine Erhöhung des shift-Parameters zurückgeführt werden konnte. Dieser Effekt war allerdings nur marginal signifikant. Auffällig war auf dem Pronomen hingegen ein starker Interaktionseffekt für den shape-Parameter: Während in der disambiguierten Bedingung inkongruente Fortsetzungen höhere shape-Werte

aufwiesen als in der kongruenten (d.h. eher normal verteilt waren), galt in der ambigen Bedingung genau der umgekehrte Fall.

Insgesamt ist die Befundlage zum IC-Fokus-Effekt uneindeutig: Während die mittleren Lesezeiten klar auf einen frühen Fokus-Effekt hinwiesen, ließ er sich in der RT-Verteilungsanalyse nicht uneindeutig lokalisieren. Zwar sprechen einige Teilbefunde (leicht erhöhter shift-Parameter, stark erniedrigter shape-Parameter) dafür, dass es sich in der Tat um einen shift-Effekt handelt, aber zu diesem Zeitpunkt muss die Sachlage als unklar bezeichnet werden. Tab. 8.6 zeigt die in Experiment IIa detektierten Effekte noch einmal im Überblick:

<i>Effekt</i>	<i>Teileffekt</i>	<i>Satzlesezeiten</i>	<i>Verifikation</i>
Gender	Confirmation	✓	---
	Disengagement	✓	✓
IC	Fokus	?	---
	Integration	(✓)	✓

Tab. 8.6: Effekte in Experiment IIa für die Satzlesezeiten und die Verifikationsaufgabe.

Es ist auffällig, dass die berichteten Ergebnisse der Satzlesezeiten sowie der Verifikationsaufgabe eigentlich recht eindeutig sind, während alle Interpretationsprobleme mit der RT-Verteilungsanalyse verbunden sind. Wie bereits angedeutet, ist dies wahrscheinlich darauf zurückzuführen, dass die Ausgangsbedingungen für eine RT-Verteilungsanalyse bei wortweisen Lesezeiten sehr viel schlechter sind, als bei einer globalen, aber stabilen pronominalen Auflösungsaufgabe wie in Experiment I. Wortweise Lesezeiten haben äußerst varianzarme Verteilungen und sind gleichzeitig ausgesprochen anfällig für Ausreißer, die nicht immer detektierbar sind. Deswegen führen manchmal schon sehr kleine Veränderungen der empirischen Verteilungskurve zu großen Unterschieden in der Parameterschätzung. Konkret macht sich dies zunächst einmal in den wesentlich schlechteren Fit-Eigenschaften der Verteilung bemerkbar: So war in Experiment IIa der Anteil signifikanter Modelle wesentlich erhöht. Hinzu kommt, dass die Schätzer der einzelnen Verteilungen instabil und unreliabel werden. Dies führt dann in einem nächsten Schritt zu inhomogenen Effektmustern und dramatischen Effektstärke-Verlusten.

Bedeutet das, dass der Einsatz von Verteilungsanalysen bei wortweisen Lesezeiten nicht sinnvoll ist? Vielleicht. Allerdings sollte man hier nicht zu schnell aufgeben. Einige Aspekte des Ergebnismusters deuten durchaus darauf hin, dass sich ihr Einsatz sehr wohl lohnen und zu validen Ergebnissen führen kann. So ist es ausgesprochen erstaunlich, dass die verschiedenen Verteilungsparameter differentiell unterschiedliche Aspekte der Satzverarbeitung erfassen. Am deutlichsten zeigte sich dies bei der Betrachtung der Lesezeitprofile. Die shift-

Komponente war im Satzverlauf vergleichsweise stabil und scheint damit eher Basisprozesse bei der Verarbeitung einzelner Wörter abzubilden (wie den lexikalischen Zugriff etc.). Die scale-Komponente war hingegen variabel und bildete den dynamischen Verlauf der Satzverarbeitung ab. Sie scheint deswegen eher für die höheren kognitiven Prozesse, v.a. an den Teilsatzgrenzen verantwortlich zu sein. Nach Just und Carpenter (1980) werden an den Satzgrenzen u.a. alle Prozesse notwendig, die alte Informationen mit neuen verbinden und zu einer gemeinsamen Diskursrepräsentation integrieren. Im Zweifelsfall wird eine Reanalyse vorgenommen und Inkonsistenzen werden aufgelöst. Danach wird dann der Buffer des verbalen Arbeitsgedächtnisses geleert und der nächste (Teil-)Satz kann verarbeitet werden (vgl. auch Millis & Just, 1994). Der scale-Parameter scheint ausgesprochen sensitiv für diese Art von Integrationsprozessen zu sein. Der shift-Parameter reagiert hingegen fast gar nicht auf sie.

Hingegen hat sich die RT-Verteilungsanalyse bei der Dissoziation der beiden Gender-Effekte ausgesprochen gut bewährt. Hier konnte vergleichsweise sicher eine Einteilung in eine frühe, gedächtnis-basierte Effekt-Komponente (die sich im shift-Parameter abbildet) und eine späte, erklärungs-basierte Effekt-Komponente (die sich allein im scale-Parameter abbildet) vorgenommen werden. Warum ist das Verfahren dann bei den IC-Effekten problematisch? Eine mögliche Antwort auf diese Frage ist, dass der (frühe) Gender-Effekt sehr viel stabiler ausgeprägt war und auch eine viel größere Effektstärke hatte als der IC-Effekt. Während der Gender-Effekt Unterschiede in der mittleren Reaktionszeit zwischen 10-20 ms generierte, waren die IC-Fokus-Effekte sehr viel schwächer ausgeprägt (ca. 6-7 ms). Es ist möglich, dass die RT-Verteilungsanalyse für vergleichsweise große Unterschiede noch gut anwendbar ist. Bei kleinen Effekten ist ihr Einsatz hingegen problematisch, weil bei ihnen die relevanten Unterschiede in den Verteilungskurven nicht eindeutig genug ausfallen. Sie wird dann zwar immer noch eine Aufteilung in verschiedene RT-Komponenten (shift vs. scale) vornehmen. Diese werden aber von vergleichsweise geringen Unterschieden in der Verteilungskurve bestimmt, weshalb sich Probleme bei der Parameterschätzung ergeben (wie z.B. Abhängigkeiten bei der Parameterschätzung, vgl. Appendix A.5). Dies führt dann wiederum zu inhomogenen und unreliaiblen Schätzungen, welche die Analyse erschweren und sich bei Aggregation gegenseitig ausmitteln. Auf der Grundlage der vorliegenden Daten würde ich den Grenzwert, ab dem eine RT-Verteilungsanalyse reliable Ergebnisse liefert, bei einem Mittelwerts-Unterschied von ca. 10 ms ansetzen. In Kap. 10 wird die Anwendbarkeit des Verfahrens noch ausführlich diskutiert. Zunächst soll aus dieser Problemanalyse die Konsequenz abgeleitet werden, dass es eine sinnvolle Strategie der Effizienzmaximierung ist, die Effekte der beiden Cues zu vergrößern. Dies wird in Experiment IIb versucht.

9. Experiment IIb

9.1 Einleitung

Bei Experiment IIb handelte es sich in Prinzip um eine Replikation von Experiment IIa, da ebenfalls wortweise Lesezeiten mittels eines *moving-window*-Verfahrens erhoben wurden. Im Gegensatz zu Experiment IIa wurde jedoch diesmal eine kumulative Variante eingesetzt. Der einzige Unterschied zwischen den beiden Verfahren besteht darin, dass in der kumulativen Variante die bereits gelesenen Wörter nicht wieder zugedeckt werden, sondern auf dem Bildschirm stehen bleiben. Die Entscheidung für eine kumulative *moving-window*-Präsentation ist eher ungewöhnlich, weil in der Literatur eigentlich von der Verwendung dieses Verfahrens abgeraten wird (vgl. z.B. Just et al., 1982; Mitchell, 2004): Die offensichtliche Gefahr besteht darin, dass die Vpn zunächst sehr schnell alle Wörter aufdecken und erst dann den Satz als Ganzes verarbeiten. Entsprechend findet man für gewöhnlich, dass bei dieser Methode die satzfinalen *wrap-up*-Effekte, die schon für die nicht-kumulative Variante typisch sind, noch extremer ausfallen, während alle satzinternen Effekte nivelliert werden (vgl. z.B. Kennedy & Murray, 1984). Bei dieser Methode sind also die zum Lesen eines Wortes benötigte Zeit und die hierfür relevanten Verarbeitungsprozesse noch weitaus stärker entkoppelt als bei der nicht-kumulativen Version, weshalb sich auch wesentlich geringere Korrelationen mit der „wirklichen“ Lesezeit (gemessen mittels Blickbewegung) ergeben (Just et al., 1982).

Der Grund, warum in diesem Fall dennoch zu dieser eher ungewöhnlichen Methode gegriffen wurde, ist, dass sie Eigenschaften besitzt, die für die pronominale Auflösung wichtig sind. Bei dieser geht es ja darum, für einen referentiellen Ausdruck ein Antezedens im vorherigen Satzkontext zu finden. Bei der nicht-kumulativen *moving-window*-Methode steht der vorhergehende Kontext jedoch nicht (mehr) zur Verfügung. Deshalb ist es möglich, dass eine andere Verarbeitungsstrategie angewendet wird. Man versucht dann nicht mehr, möglichst direkt einen passenden Referenten für ein Pronomen zu finden (indem man z.B. zurückschaut und sich überlegt, welcher der bisher genannten Aktanten wohl der wahrscheinlichere Referent ist), sondern man wartet vielmehr (notgedrungen), bis einem im weiteren Satzverlauf noch weitere Informationen gegeben werden, die einem bei der Entscheidung helfen. So werden evtl. relevante Verarbeitungsprozesse bis ans Satzende verlagert, die bei verfügbarem Kontext eigentlich sofort vorgenommen würden. Dies führt dann wiederum einer Schwächung der frühen Prozesse, was für die diesbezüglich eher schwache Befundlage in Experiment IIa bezüglich der Reaktionszeitverteilungsparameter verantwortlich sein könnte.

Darüber hinaus ist davon auszugehen, dass die nicht-kumulative Präsentationsmethode den Rezeptionsprozess insgesamt erschwert. Dafür spricht zumindest die Beobachtung aus Experiment IIa, dass die Fehlerquote bei der abschließenden Verifikationsaufgabe deutlich höher ausfiel als bei der vergleichbaren Auflösungsaufgabe in Experiment I. Es ist zu erwarten, dass eine kumulative Version zu einem deutlich besseren Verständnis des Satzes führt, was sich in einem höheren Akkuratheitsniveau bei der Verifikationsaufgabe niederschlagen sollte.

Damit befindet man sich aber in einem methodischen Zwiespalt: Einerseits könnte gerade diese Methode eine in Bezug auf die pronominale Auflösung realistischere Untersuchungssituation bereitstellen, die eine frühe Auflösung des Pronomens begünstigt und zu einem besserem Satzverständnis führt. Andererseits ist jedoch bekannt, dass gerade die kumulative *moving-window*-Methode zu einer Betonung lediglich der satzfinalen Integrationsprozesse führt und frühe Effekte schwächt. Das vorliegende Experiment hatte deswegen die primär explorative Zielstellung zu überprüfen, wie sich die einzelnen Teileffekte der beiden Cues (Gender-Cue: *confirmation* vs. *disengagement*; IC-Cue: Fokus- vs. Integrationseffekt) unter dieser anderen Präsentationsmethode verhalten. Speziell ist in diesem Kontext interessant, ob sich die frühen *bonding*-Prozesse nun besser abzeichnen.

Die theoretischen Vorhersagen sind deswegen auch prinzipiell die gleichen wie für Experiment IIa: Es wird erwartet, dass sich ein früher Gender-*confirmation*-Effekt ergibt, der sich bereits kurz nach dem Pronomen in der Satzverarbeitung finden lässt. Er sollte primär vom shift-Parameter getragen werden. Komplementär hierzu sollte es einen späten Gender-*disengagement*-Effekt geben, der allein den scale-Parameter beeinflusst. Bezüglich des IC-Cues wird erwartet, dass es in der disambiguierten Bedingung zu einem frühen IC-Fokus-Effekt kommt, der mit dem shift-Parameter assoziiert ist. Parallel hierzu sollte es auch einen satzfinalen IC-Integrations-Effekt geben, der lediglich vom scale-Parameter getragen wird. Dieser Effekt sollte in der Bedingung stärker ausfallen, in der der Satz nicht bereits früh durch den Gender-Cue disambiguiert werden konnte (vgl. Tab. 9.1).

<i>Effekt</i>	<i>Teileffekt</i>	<i>Parameter</i>	<i>Zeitpunkt</i>	<i>Interaktion Gender-Cue</i>
Gender	Confirmation	Shift	Früh	---
	Disengagement	Scale	Spät	---
IC	Fokus	Shift	Früh	Stärker, wenn disambiguiert
	Integration	Scale	Spät	Stärker, wenn nicht disambiguiert

Tab. 9.1: Vorhersagen für die verschiedenen Teileffekte des Gender- und des IC-Cues.

9.2 Methode

9.2.1 Stichprobenbeschreibung und -bereinigung

Allgemeine Angaben. An Experiment IIb nahmen insgesamt $N = 23$ Versuchspersonen (22 Frauen, 1 Mann) teil. Keine der Vpn hatte an einer der vorherigen Studien teilgenommen. Das durchschnittliche Alter lag bei $M = 26.6$ Jahren ($SD = 8.0$). Es handelte sich hauptsächlich um Studierende der Psychologie (meist im 3. Semester), die für ihre Teilnahme eine Versuchspersonenstunde oder eine Vergütung von 7,-€ erhielten. Eine Versuchsperson hatte an der Untersuchung teilgenommen, obwohl sie Deutsch nicht als Erstsprache erlernt hatte. Sie wurde von allen weiteren Analysen ausgeschlossen. Ansonsten berichteten alle Teilnehmer/innen, dass sie normal- oder korrigiert normalsichtig seien, und gaben Deutsch als Muttersprache an.

ELVES. Auch die Stichprobe dieses Experimentes wies in allen drei eingesetzten hierarchieniedrigen Subtests von ELVES ein durchschnittliches und unauffälliges Profil auf. Die entsprechenden Mittelwerte lagen für den Subtest *Satzverifikation* bei $M = 16.0$ ($SD = 3.7$), für den Subtest *Sinnhaftigkeit Sätze* bei $M = 15.8$ ($SD = 3.9$) und für den Subtest *Sinnhaftigkeit Satzfolgen* bei $M = 15.5$ ($SD = 4.4$). Die drei Subskalen korrelierten hoch miteinander und wurden deswegen zu einem gemeinsamen Index zusammengefasst (Cronbachs $\alpha = .92$; $M = 15.8$, $SD = 3.7$). Zwei Versuchspersonen wiesen für die eingesetzten ELVES-Subskalen einen erhöhten multivariaten Leverage-Wert auf ($h_{ii} = .39$, $p < .01$). Eine von ihnen schnitt in allen Subtests überdurchschnittlich ab, die andere in einem über- und in einem anderen unterdurchschnittlich. Da ihr Antwortverhalten sonst unauffällig war, wurden sie nicht von der Analyse ausgeschlossen.

Durchschnittliche Akkuratheit und Antwortlatenz. Die mittlere Reaktionszeit und durchschnittliche Akkuratheit für die Verifikationsaufgabe wurden wiederum pro Person zusammengefasst. Zwei Versuchspersonen hatten einen erhöhten multivariaten Leverage-Wert ($h_{ii} > .69$, $p < .01$). Weil beide sehr niedrige Akkuratheitswerte aufwiesen, wurde sie von allen nachfolgenden Analysen ausgeschlossen.

9.2.2 Material, Untersuchungsablauf, Design und Abhängige Variablen

Material. Sowohl die Stimulussätze selbst als auch die Verifikationsaussagen waren exakt die gleichen wie in Experiment IIa. Das gleiche gilt für das Procedere der Zuordnung der einzelnen Items zu den vier Untersuchungsbedingungen und der wahren vs. falschen Aussageform bei der Verifikationsaufgabe

Untersuchungsablauf. Instruktion, Gestaltung und Ablauf der Untersuchung waren weitestgehend identisch zum Vorgehen in Experiment IIa. Der einzige relevante Unterschied zwischen den beiden Untersuchungen bestand darin, dass in Experiment IIb die einzelnen Wörter des Stimulussatzes nach dem Drücken der Leertaste nicht wieder „zugedeckt“ wurden, sondern vielmehr sichtbar blieben.

Design und Abhängige Variablen. Entsprechend waren auch das Grunddesign und die abhängigen Variablen der Untersuchung die gleichen wie in Experiment IIa.

9.3 Ergebnisse

9.3.1 Satzlesezeiten

9.3.1.1 Mittlere Lesezeiten

9.3.1.1.1 Datenbereinigung und Lesezeitprofil

Datenbereinigung. Für die Datenbereinigung wurde ein unterer Cut-off-Wert von 100 ms angesetzt. Aufgrund dieses Kriteriums wurden 0.1% aller Datenpunkte durch den bedingungs- und personenspezifischen Mittelwert ersetzt. Weiterhin wurden alle Werte, die 4 Standardabweichungen vom dem wort- und bedingungsspezifischen Mittelwert einer Person abwichen, durch diesen ersetzt. Diese Maßnahme führte zur Ersetzung von ca. 1.0 % aller Datenpunkte.

Lesezeitprofil. Abb. 9.1 zeigt die mittlere Zeit, die zum Lesen der einzelnen Wörter des Satzes benötigt wurde.

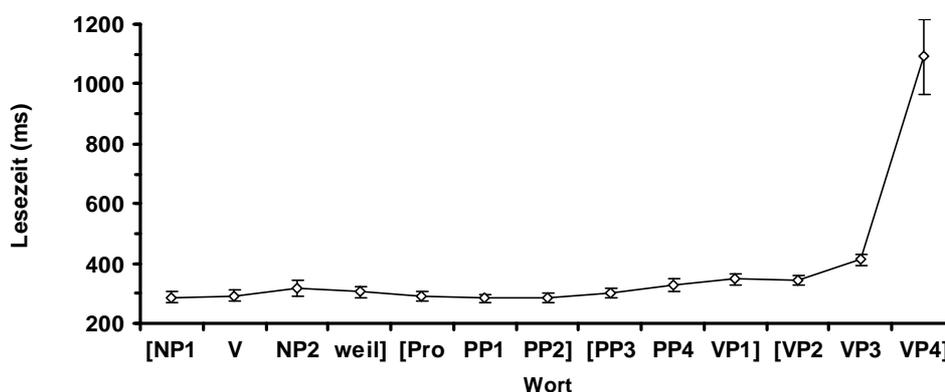


Abb. 9.1: Mittlere Lesezeit über alle vier Untersuchungsbedingungen gemittelt.

Die mittlere Lesezeit pro Wort lag bei 375 ms und ist damit im Vergleich zu Experiment IIa um ca. 10 % gesunken. Trotz dieses eher geringen absoluten Abfalls zeigte eine Inspektion des Lesezeit-Profiles, dass er sich sehr ungleich über die einzelnen Wörter des Satzes verteilte:

Die ersten 12 Wörter wiesen extrem geringe Lesezeiten auf ($M = \text{ca. } 320 \text{ ms}$; $SD = \text{ca. } 38 \text{ ms}$), die ca. 80 ms unter dem entsprechenden Mittelwerten von Experiment IIa lagen. Die Lesezeit des letzten Wortes war hingegen dramatisch angestiegen (von ca. 550 ms auf ca. 1100 ms). Damit scheint zunächst einmal eingetreten zu sein, wovor bei der kumulativen *moving-window*-Methode immer gewarnt wird: eine Nivellierung der satzinternen Lesezeiten durch eine allgemeine Aufblähung des *sentence-wrap-ups* (vgl. Mitchell, 2004).

Allerdings ist es nicht so, dass die Vpn einfach alle Wörter aufdeckten und dann den Satz als Ganzes gelesen haben: Zum einen war bereits die Lesezeit des letzten Wortes hierfür (mit ca. 1000 ms) viel zu gering ausgeprägt (die durchschnittliche Lese- und Auflösungszeit für die strukturidentischen Sätze in Experiment Ib lag bei ca. 4300 ms). Zumindest einige verarbeitungsrelevante Prozesse mussten also bereits vorher stattgefunden haben. Zum anderen kovariierten die durchschnittlichen Lesezeiten zumindest halbwegs mit der zugrunde liegenden Satzstruktur (vgl. Abb. 9.2, bei der die Lesezeit des letzten Wortes nicht mit abgebildet ist, damit sich die Struktur des Lesezeit-Profiles deutlicher abzeichnet).

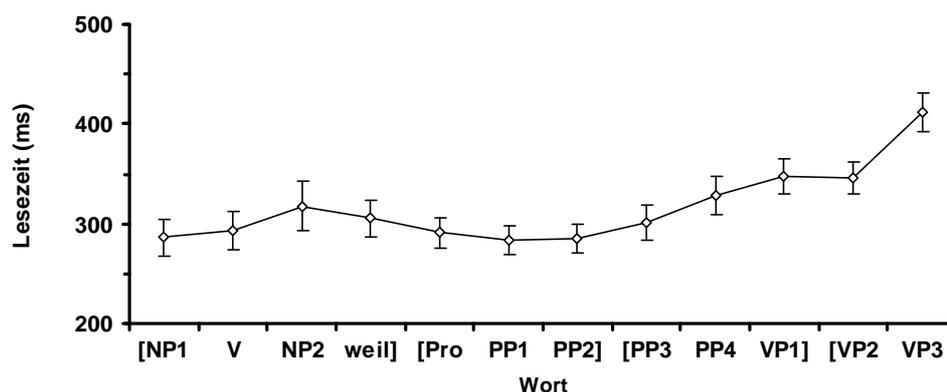


Abb. 9.2: Mittlere Lesezeit über alle vier Untersuchungsbedingungen gemittelt (ohne VP4).

Der Verlauf der Lesezeiten ähnelt stark dem von Experiment IIa (und entsprechen korrelieren die beiden Profile auch zu $r = .80$, $p < .001$ miteinander): Wieder gab es einen deutlich ansteigenden globalen Trend in den Lesezeiten und eine klare Betonung der Konstituenten- bzw. (Neben-)Satzstruktur, an deren Ende die Lesezeiten jeweils erhöht waren. Dieser Befund spricht für die *connective-integration*-Hypothese von Millis und Just (1994).

9.3.1.1.2 Ergebnisse auf der Wortebene

In einem ersten Schritt wurde wiederum für jedes der 13 Wörter eine eigene zweifaktorielle, messwiederholte ANOVA mit den beiden Faktoren *Gender-Cue* und *IC-Cue* berechnet. Die entsprechenden Ergebnisse finden sich in Tab. 9.2. Die Effekte werden nachfolgend für jeden der beiden Cues (und ihre Interaktion) einzeln besprochen.

Gender-Effekt. Abb. 9.3 a) und b) zeigen die wortweisen Lesezeiten für die beiden Bedingungen des Gender-Cues bzw. dessen direkten Effekt. Wieder stehen positive Werte für einen Verarbeitungsvorteil, der sich durch die Realisierung des Cues ergibt, negative hingegen für einen Verarbeitungsnachteil.

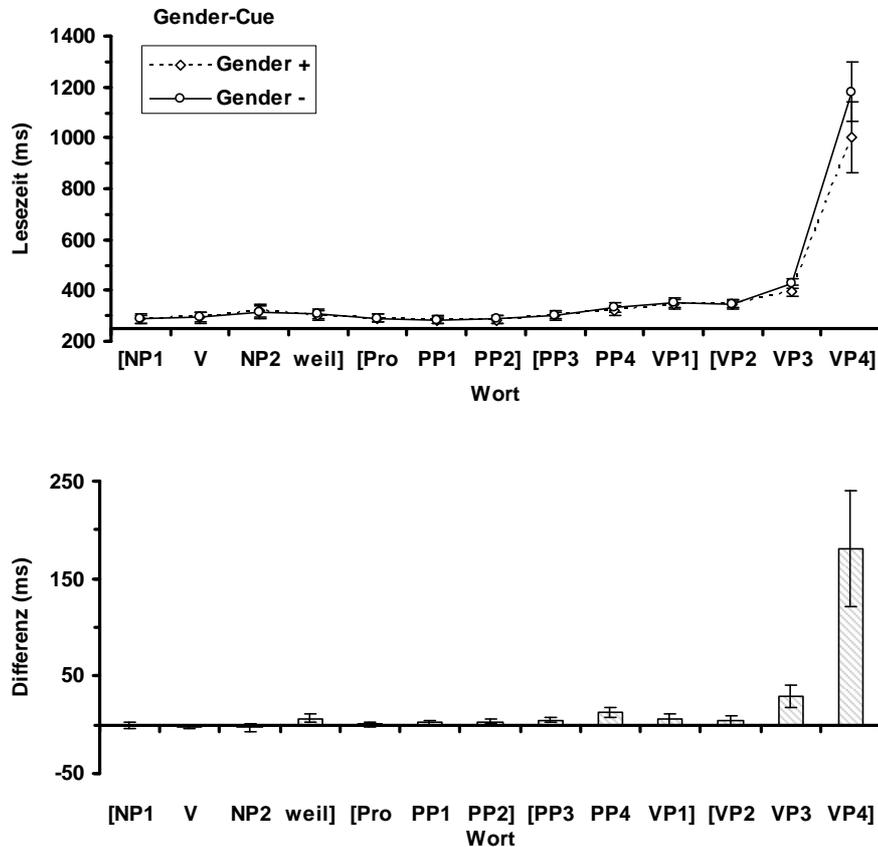


Abb. 9.3: a) Wortweise Lesezeiten in Abhängigkeit von der Ausprägung des Gender-Cues und b) Effekt des Gender-Cues im Satzverlauf.

Wie schon in Experiment IIa dominierte ein später Effekt des Gender-Cues im *sentence-wrap-up* das Bild. Mit einer Differenz von ca. 180 ms hatte er im Vergleich sogar noch deutlich an Stärke zugelegt (ca. 60 %), was daran liegt, dass durch die kumulative *moving-window*-Methode prinzipiell alle *closure*-Effekte wesentlich gestärkt werden (s.o.). Alle anderen Effekte schienen im Vergleich zu diesem späten Effekt vernachlässigbar zu sein. Abb. 9.4, in welcher der letzte Effekt nicht mit abgebildet und die y-Achse reskaliert wurde, macht jedoch deutlich, dass dieser erste Eindruck trägt.

	[NP1]	V	NP2	weil]	[er	PP1	PP2]	[PP3	PP4	VP1]	[VP2	VP3	VP4]
By-Participants													
<i>Gender-Cue</i>													
<i>F(1,19)</i>	< 1	1.48	< 1	3.77	< 1	< 1	3.43	4.56	6.48	1.05	< 1	5.64	9.26
<i>p</i>	.985	.239	.461	.067	.914	.397	.080	.046*	.020*	.318	.346	.028*	.007**
η_p^2	< .001	.072	.029	.166	.001	.038	.153	.194	.254	.052	.043	.229	.328
<i>IC-Cue</i>													
<i>F(1,19)</i>	4.52	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	2.56	2.10	25.55
<i>p</i>	.047*	.781	.390	.934	.536	.869	.501	.686	.459	.862	.126	.164	.001**
η_p^2	.192	.004	.039	.001	.020	.001	.024	.009	.029	.002	.119	.099	.573
<i>Gender- x IC-Cue</i>													
<i>F(1,19)</i>	< 1	1.57	< 1	< 1	< 1	< 1	< 1	< 1	1.32	6.76	6.00	< 1	1.12
<i>p</i>	.596	.225	.508	.596	.642	.330	.573	.528	.266	.018*	.024*	.694	.303
η_p^2	.015	.076	.023	.015	.012	.050	.017	.021	.065	.262	.240	.008	.056
By-Items													
<i>Gender-Cue</i>													
<i>F(1,159)</i>	< 1	< 1	< 1	3.07	< 1	< 1	1.41	2.72	9.00	1.87	1.21	14.44	16.87
<i>p</i>	.966	.424	.395	.082	.945	.530	.237	.101	.003**	.174	.273	.001**	.001**
η_p^2	< .001	.004	.005	.019	< .001	.002	.009	.017	.054	.012	.008	.083	.096
<i>IC-Cue</i>													
<i>F(1,159)</i>	4.57	< 1	1.44	< 1	< 1	< 1	< 1	< 1	< 1	< 1	2.65	2.76	12.17
<i>p</i>	.034*	.730	.231	.959	.486	.724	.621	.716	.455	.677	.105	.099	.001**
η_p^2	.028	.001	.009	< .001	.003	.001	.002	.001	.004	.001	.016	.017	.071
<i>Gender- x IC-Cue</i>													
<i>F(1,159)</i>	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	6.71	4.10	< 1	< 1
<i>p</i>	.685	.660	.674	.445	.319	.323	.453	.329	.418	.010**	.045*	.718	.715
η_p^2	.001	.001	.001	.004	.006	.006	.004	.006	.004	.040	.025	.001	.001

Tab. 9.2: Ergebnisse des varianzanalytischen Vergleichs zwischen den vier Untersuchungsbedingungen auf der Wortebene.

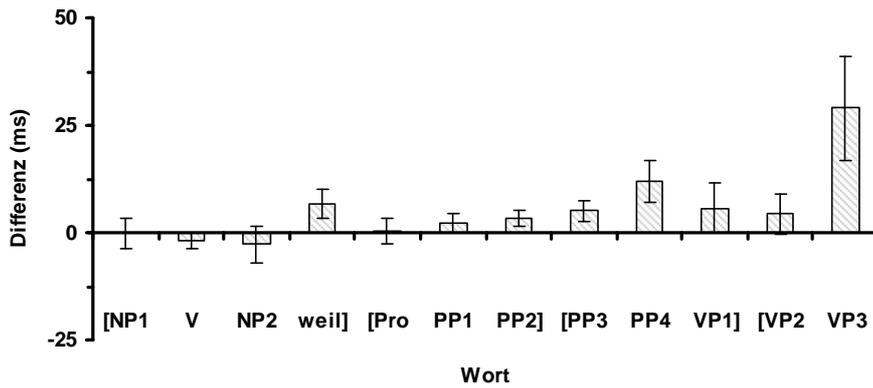


Abb. 9.4: Effekt des Gender-Cues im Satzverlauf (ohne VP4)

Wie erwartet fanden sich frühe Effekte des Gender-Cues, und zwar signifikant ab dem Wort PP3. Im Gegensatz zu Experiment IIa gab es dann auf der Strecke VP1-2 einen leichten Abfall in der Effektstärke, ab dem Wort VP3 war er jedoch wieder voll ausgeprägt. Zusammenfassend ist festzuhalten, dass es zwar frühe Effekte des Gender-Cues gab, diese aber im Vergleich zu Experiment IIa später einsetzten. Auch hatte der Effekt nun eine andere Verlaufsdynamik

IC-Effekt. Abb. 9.5 a) und b) stellen die wortweisen Lesezeiten und die Einzeleffekte des IC-Cues für jedes Wort des Satzschemas dar.

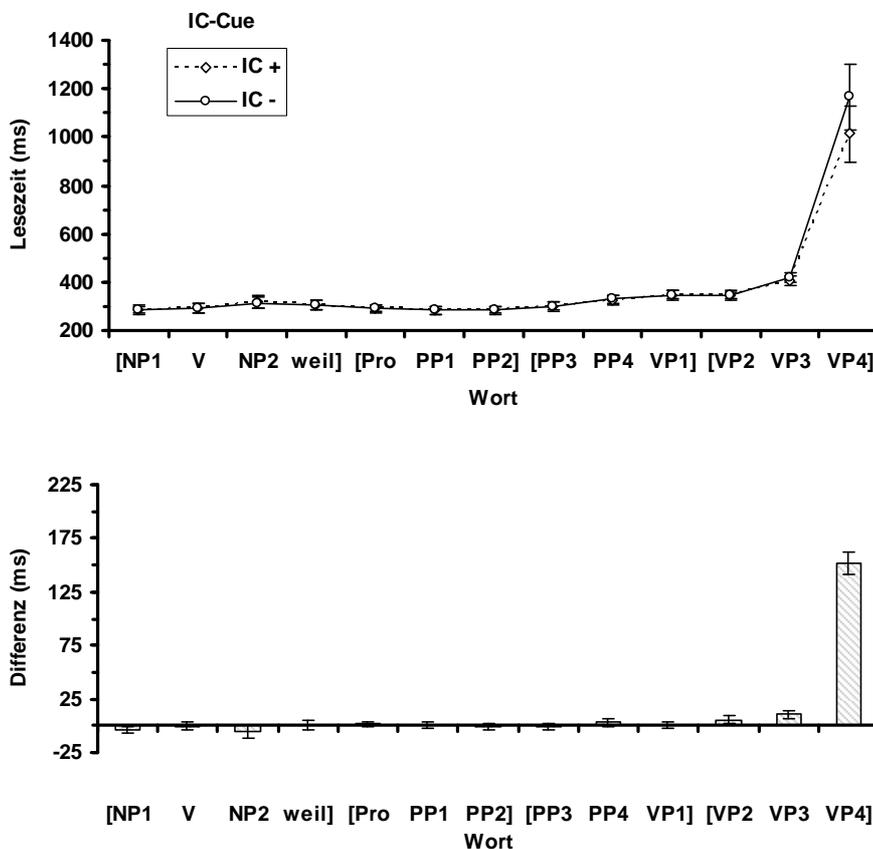


Abb. 9.5: a) Wortweise Lesezeiten in Abhängigkeit von der Ausprägung des IC-Cues und b) absolute Größe des Kausalitätseffekts im Satzverlauf.

Auch hier war der Effekt des IC-Cues auf dem letzten Wort des Satzes sicherlich am prägnantesten ausgeprägt, da er in seiner Stärke dramatisch zugenommen hat (ca. 40 ms in Experiment IIa vs. ca. 150 ms in Experiment IIb).

Der Abbildung kann man jedoch ebenfalls entnehmen, dass es keine frühen Effekte des IC-Cues gab. Dieser Eindruck bleibt auch erhalten, wenn man den Effekt des letzten Wortes wieder von der Abbildung ausschließt und diese reskaliert (vgl. Abb. 9.6): Signifikante Effekte des IC-Cues fanden sich erst in der Verbalphrase (genauer: ab dem Wort VP2).

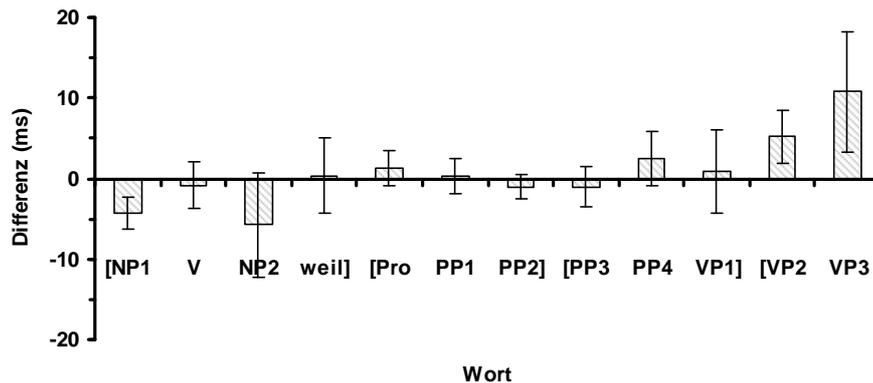


Abb. 9.6: Haupteffekte des IC-Cues auf die Lesezeit (ohne VP4).

Interaktions-Effekt. Wie in Tab. 9.3 zu sehen, gab es im Gegensatz zu Experiment IIa auch keine frühen Interaktionseffekte. Die Interaktion der beiden Faktoren war allgemein vergleichsweise irrelevant zur Beurteilung der Effekte, allerdings fanden sich schwache Effekte am Anfang der Verbalphrase, insbesondere auf den Wörtern VP1 und VP2. Abb. 9.7 zeigt die einfachen Haupteffekte des IC-Cues in Abhängigkeit von der Ausprägung des Gender-Cues (und zwar mit und ohne *wrap-up*-Effekt).

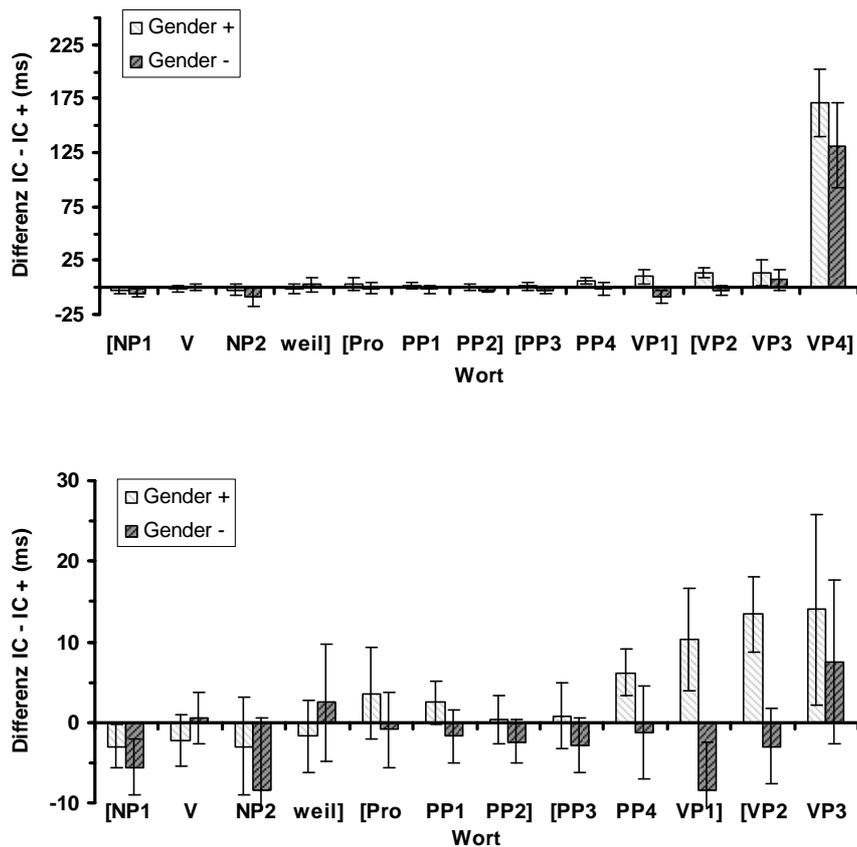


Abb. 9.7: Einfache Haupteffekte des IC-Cues in den zwei Gender-Bedingungen: a) mit und b) ohne VP4.

Wenn die Effekte des IC-Cues danach differenziert wurden, ob der Satz bereits durch den Gender-Cue disambiguiert wurde oder nicht, ergab sich ein deutlicheres Bild: Denn nun zeigte sich, dass der Effekt des IC-Cues bereits vor der Verbalphrase ansetzte (nämlich auf dem Wort PP4, also zum gleichen Zeitpunkt, an dem auch der Gender-Effekt zu wirken begann), allerdings tritt dieser Effekt nur in der Bedingung *Gender+* auf. In der Bedingung *Gender-* fanden sich hingegen am Anfang der Verbalphrase tendenziell sogar negative Effekte des IC-Cues. Allerdings waren sie zu schwach, um signifikant zu werden. Eine ähnliche Effekt-Dissoziation (positive Effekte in der Bedingung *Gender+*, negative in der Bedingung *Gender-*) fanden sich auch in der Satzmitte. Man beachte, dass der satzfinale Effekt in der Bedingung *Gender+* größer ausfiel als in der Bedingung *Gender-*, allerdings nicht signifikant.

9.3.1.1.3 Ergebnisse auf der Konstituentenebene

Um zu überprüfen, ob sich das Ergebnismuster dadurch stabilisierte, wurden die einzelnen Wörter des Satzes in einem zweiten Schritt wieder zu Konstituenten gruppiert. Die Gruppierung erfolgte dabei nach dem in Kap. 8.2 ausführlich besprochenen Schema. Tab. 9.3 enthält die Ergebnisse einer zweifaktoriellen messwiederholten ANOVA, die nachfolgend wieder für jeden Cue einzeln besprochen werden sollen.

	HS	Pro	PP	VP
By-Participants				
<i>Gender-Cue</i>				
<i>F(1,19)</i>	< 1	1.53	4.82	10.16
<i>p</i>	.781	.231	.041*	.005**
η_p^2	.004	.075	.202	.347
<i>IC-Cue</i>				
<i>F(1,19)</i>	< 1	< 1	< 1	21.90
<i>p</i>	.331	.863	.715	.001**
η_p^2	.050	.002	.007	.535
<i>Gender- x IC-Cue</i>				
<i>F(1,19)</i>	< 1	< 1	7.13	2.17
<i>p</i>	.946	.391	.015*	.157
η_p^2	< .001	.039	.273	.102
By-Items				
<i>Gender-Cue</i>				
<i>F(1,159)</i>	< 1	< 1	8.59	19.58
<i>p</i>	.952	.470	.004**	.001**
η_p^2	< .001	.003	.051	.110
<i>IC-Cue</i>				
<i>F(1,159)</i>	1.48	< 1	< 1	13.48
<i>p</i>	.226	.778	.586	.001**
η_p^2	.009	.001	.002	.078
<i>Gender- x IC-Cue</i>				
<i>F(1,159)</i>	< 1	1.19	4.06	< 1
<i>p</i>	.887	.267	.046*	.585
η_p^2	< .001	.007	.025	.002

Tab. 9.3: Ergebnisse des varianzanalytischen Vergleichs der vier Konstituenten auf Unterschiede zwischen den vier Untersuchungsbedingungen.

Gender-Effekt. Abb. 9.8 stellt die Effekte des Gender-Cues für die vier Konstituenten dar.

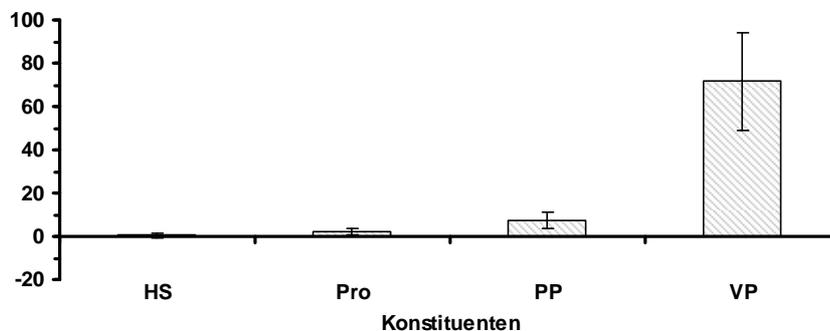


Abb. 9.8: Absolute Größe und Richtung des Gender-Effekts für die vier Konstituenten.

Wie bereits in der Analyse der wortweisen Lesezeiten gab es einen starken Effekt des Cues auf der Verbalphrase und einen etwas kleineren auf der Präpositionalphrase.

IC-Effekt. Abb. 9.9 zeigt den Haupteffekt des IC-Cues für die vier Konstituenten.

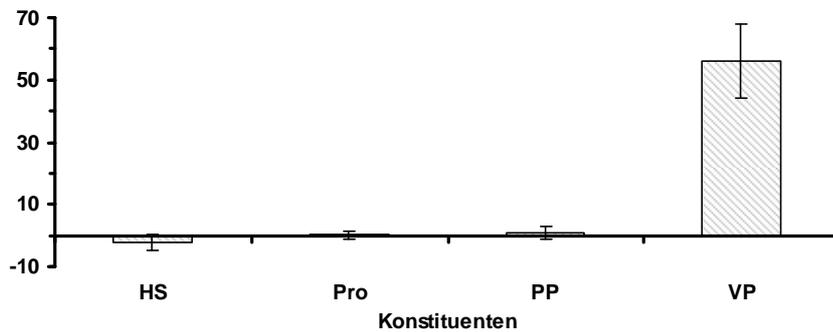


Abb. 9.9: Absolute Größe und Richtung des IC-Effekts für die vier Konstituenten.

Auch hier wurde das Ergebnis der Wort-Ebene vollkommen bestätigt: Es gab nur einen Haupteffekt des IC-Cues auf der Verbalphrase. Dieser fiel allerdings sehr stark aus.

Interaktions-Effekt. Abb. 9.10 zeigt die Einfachen Haupteffekte des IC-Cues für die beiden Ausprägungen des Gender-Cues für die vier Konstituenten.

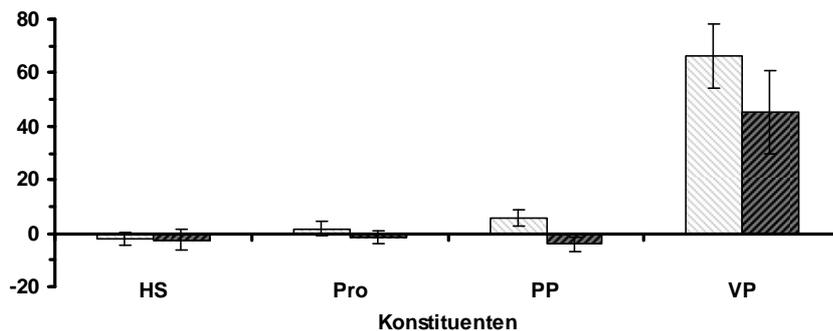


Abb. 9.10: Absolute Größe und Richtung des IC-Effekts in Abhängigkeit von der Ausprägung des Gender-Cues für die vier Konstituenten.

In der zu Konstituenten aggregierten Analyse war der Interaktions-Term der Präpositionalphrase signifikant: Während der IC-Cue in der Bedingung *Gender+* einen (marginal) signifikanten, positiven Effekt auf die mittlere Lesezeit hatte ($F(1,19) = 3.41, p = .081, \eta_p^2 = .152$; $F(1,159) = 4.40, p < .038, \eta_p^2 = .027$), war er in der Bedingung *Gender-* negativ ausgeprägt (aber nicht signifikant: $F(1,19) = 2.74, p = .114, \eta_p^2 = .126$; $F(1,159) = 1.18, p = .279, \eta_p^2 = .007$). Auffällig ist weiterhin, dass der satzfinale Effekt erwartungskonträr in der disambiguierten Bedingung sogar größer ausfiel als in der nicht-disambiguierten Bedingung; allerdings ist die Interaktion der beiden Faktoren nicht signifikant.

9.3.1.1.4 Re-Analyse der „korrekt“ beantworteten Items

Alle Analysen wurden nochmals nur mit den (bei der nachfolgenden Verifikationsaufgabe) korrekt beantworteten Items berechnet. Ein vollständiger Bericht des Ergebnismusters findet sich in Appendix C.4.1. Es blieb vollständig erhalten; allerdings waren alle Effekte ein wenig schwächer ausgeprägt.

9.3.1.1.5 Zusammenfassung

Das Ergebnismuster für die mittlere Lesezeit war das gleiche wie in Experiment IIa: Für den Gender-Cue gab es einen starken satzfinalen Effekt, der jedoch bereits in der Präpositionalphrase ansetzte. Für den IC-Cue gab es auf der Ebene der Haupteffekte nur einen *wrap-up*-Effekt, der ebenfalls groß ausfiel. Auf der Ebene der einfachen Haupteffekte zeigte sich, dass der Effekt des IC-Cues schon früh ansetzte, allerdings nur, wenn der Satz bereits genderdisambiguiert war. In der ambigen Version gab es sogar einen leicht negativen Effekt des IC-Cues, der allerdings nicht signifikant war.

Zwei Abweichungen zum letzten Experiment sind hingegen von Bedeutung: Erstens waren – wie bereits ausführlich besprochen – alle satzfinalen Effekte (wesentlich) stärker ausgeprägt, was direkt auf die veränderte Art der Lesezeiterfassung zurückgeführt werden kann. Zweitens waren im Vergleich zu Experiment IIa alle Effekte „verschoben“, d.h. sie setzten erst später im Verlauf der Satzverarbeitung ein. Insbesondere fand sich der Interaktionseffekt des IC-Cues erst später in der Präpositional- und nicht mehr in der Pronominalphrase.

9.3.1.2 Weibull-Parameter

9.3.1.2.1 Parameterschätzung, Modellgüte und Parameterprofile

Parameterschätzung. Das Vorgehen bei der Parameterschätzung entsprach prinzipiell dem in Kap. 3.4 skizzierten Vorgehen. An die Verteilung der Leszeiten jedes einzelnen Wortes eines Satzes wurde eine eigene Weibull-Verteilung angepasst, sodass für jeden Satz jeweils drei Parameterprofile vorlagen. Als Startwert für den scale-Parameter wurde ein Wert von $\lambda = \theta^b = 50$ angesetzt. Die verwendeten Parameter für die beiden Prior-Verteilungen des scale-Parameters waren $a_1 = 2.0$, $b_1 = 0.1$, $a_2 = 2.0$ und $b_2 = 5$.

Modellgüte. Die Passung der Weibull-Verteilung für die einzelnen RT-Verteilungen wurde mittels des in Kap. 3.5 beschriebenen Chi-Quadrat-Verfahrens beurteilt. Der mittlere Fit kann als zufrieden stellend bezeichnet werden ($\chi^2(3) = 6.39$, $p < .18$). Auf der Ebene der einzelnen Verteilungen war er nicht optimal: 209 der insgesamt 1040 (=20x4x13) Einzelver-

teilungen (= 20.1 %) waren auf dem 1 %-Niveau signifikant. Aufgrund der eingeschränkten Varianz der Verteilung ist das beschriebene Chi-Quadrat-Verfahren jedoch nur schlecht zur Beurteilung des Modellfits geeignet: In einem substantiellen Anteil der Fälle können lediglich zwei *bins* extrahiert werden (vgl. Kap. 3.5). Der erhöhte Anteil an signifikanten Modellen ist wiederum ein Zeichen dafür, dass die geschätzten Verteilungsparameter vergleichsweise instabil und unreliabel sind.

Parameterprofile. Abb. 9.11 zeigt die durchschnittlichen shift- und scale-Parameter für die Verteilung der Lesezeiten im Satzverlauf.

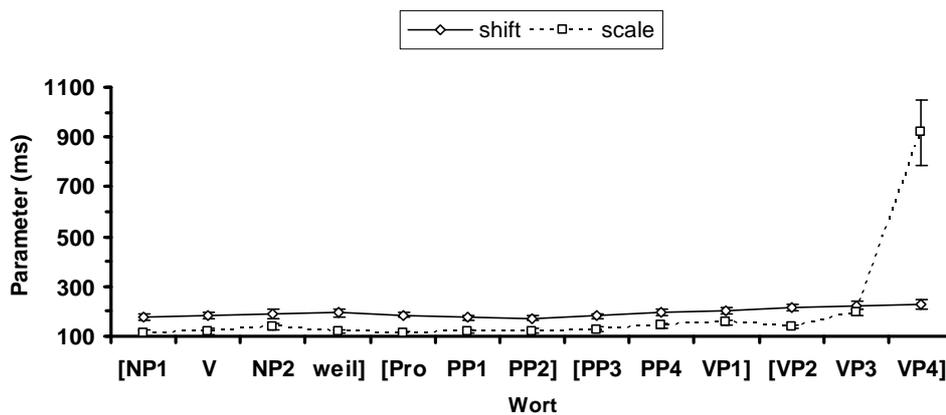


Abb. 9.11: Durchschnittliche shift- und scale-Parameter für jedes Wort des Satzschemas.

Wieder ist es so, dass der shift-Parameter sehr stabil um einen Wert von ca. 200 ms lag, während der scale-Parameter Werte zwischen ca. 100-150 ms annahm. Am auffälligsten war jedoch sicherlich, dass der satzfinale *wrap-up*-Effekt wieder allein vom scale-Parameter getragen wurde – ein Hinweis auf die Konstruktvalidität der vorgeschlagenen Interpretation der beiden Parameter. Um den Verlauf der beiden Parameter besser nachvollziehen zu können, wurde in Abb. 9.12 a) wieder das letzte Wort weggelassen und die Abbildung reskaliert. Teilabbildung b) zeigt die korrespondierenden Werte des shape-Parameters.

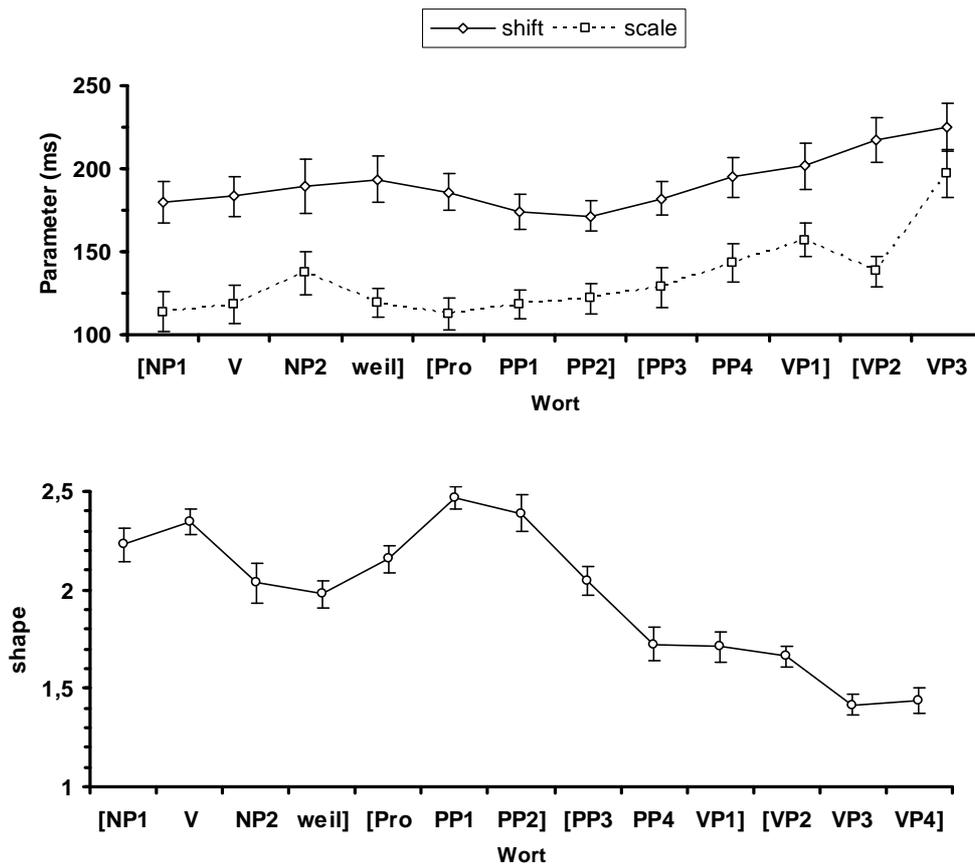


Abb. 9.12: a) Durchschnittliche shift- und scale-Parameter (ohne VP4) und b) durchschnittliche shape-Parameter für jedes Wort des Satzschemas.

Die Beziehungen drei Parameter untereinander lassen sich ähnlich beschreiben wie für Experiment IIa: Der shift-Parameter war vergleichsweise konstant ausgeprägt (*Min*: 170 ms, *Max*: 226 ms; $M = 194$ ms, $SD = 18$ ms). Im Laufe der Satzverarbeitung stieg er kontinuierlich an. Im Vergleich hierzu war der scale-Parameter eher variabel (*Min*: 112 ms, *Max*: 919 ms, $M = 194$ ms, $SD = 219$ ms; ohne letztes Wort: *Max* = 197 ms, $M = 134$, $SD = 24$ ms), da er die Satz- bzw. Konstituentenstruktur mit seinem Verlauf nachzeichnete. Der Verlauf des shape-Parameters war wieder spiegelverkehrt zu dem des scale-Parameters. Allgemein waren die Beziehungen zwischen den einzelnen Parametern ähnlich ausgeprägt wie in der vorhergehenden Untersuchung (shift-scale: $r = .81$, $p < .001$; shift-shape: $r = -.93$, $p < .001$; scale-shape: $r = -.84$, $p < .001$; jeweils ohne letztes Wort berechnet).

9.3.1.2.2 Ergebnisse Wortebene

Die drei Weibull-Parameter wurden in einem ersten Schritt wieder mittels einer zweifaktoriellen messwiederholten ANOVA für jedes Wort einzeln daraufhin geprüft, ob sie systematisch mit den beiden manipulierten Auflösungs-Cues variieren. Die entsprechenden Ergebnisse finden sich in Tab. 9.4, sie werden nachfolgend für jeden Cue einzeln besprochen.

	[NP1]	V	NP2	weil]	[Pro	PP1	PP2]	[PP3	PP4	VP1]	[VP2	VP3	VP4]
shift													
<i>Gender-Cue</i>													
<i>F(1,19)</i>	< 1	1.35	5.20	3.45	1.68	< 1	< 1	3.62	2.62	2.79	< 1	< 1	2.19
<i>p</i>	.588	.260	.034*	.079	.211	.691	.923	.072	.122	.111	.977	.580	.155
η_p^2	.016	.066	.215	.154	.081	.009	.001	.160	.121	.128	< .001	.016	.104
<i>IC-Cue</i>													
<i>F(1,19)</i>	< 1	< 1	< 1	< 1	1.36	5.00	< 1	< 1	< 1	2.25	2.35	1.65	< 1
<i>p</i>	.965	.595	.458	.934	.258	.038*	.711	.850	.645	.150	.141	.214	.366
η_p^2	< .001	.015	.029	< .001	.067	.208	.007	.002	.011	.106	.110	.080	.043
<i>Gender- x IC-Cue</i>													
<i>F(1,19)</i>	1.72	1.21	< 1	< 1	< 1	5.66	3.54	< 1	2.78	2.12	1.98	< 1	< 1
<i>p</i>	.206	.285	.394	.963	.564	.028*	.075	.461	.112	.161	.176	.507	.355
η_p^2	.083	.060	.038	< .001	.018	.229	.157	.029	.128	.101	.094	.023	.045
scale													
<i>Gender-Cue</i>													
<i>F(1,19)</i>	< 1	3.92	2.85	8.68	1.28	< 1	< 1	1.64	< 1	< 1	< 1	5.88	9.89
<i>p</i>	.569	.062	.108	.008	.272	.875	.681	.216	.769	.425	.416	.025*	.005**
η_p^2	.017	.171	.130	.313	.063	.001	.009	.079	.005	.034	.035	.236	.342
<i>IC-Cue</i>													
<i>F(1,19)</i>	< 1	< 1	< 1	< 1	< 1	4.42	< 1	< 1	< 1	1.18	< 1	< 1	28.70
<i>p</i>	.635	.785	.921	.908	.388	.049*	.465	.678	.863	.292	.791	.685	.001**
η_p^2	.012	.004	.001	.001	.040	.189	.028	.009	.002	.058	.004	.009	.602
<i>Gender- x IC-Cue</i>													
<i>F(1,19)</i>	1.25	< 1	< 1	< 1	< 1	4.00	2.14	< 1	1.40	< 1	< 1	< 1	< 1
<i>p</i>	.277	.426	.795	.854	.669	.060	.160	.723	.251	.600	.942	.956	.830
η_p^2	.062	.034	.004	.002	.010	.174	.101	.007	.069	.015	< .001	< .001	.002

	[NP1]	V	NP2	weil]	[er	PP1	PP2]	[PP3	PP4	VP1]	[VP2	VP3	VP4]
shape													
<i>Gender-Cue</i>													
<i>F(1,19)</i>	< 1	2.21	11.11	< 1	< 1	< 1	8.81	< 1	< 1	< 1	5.39	< 1	7.43
<i>p</i>	.944	.153	.003**	.821	.463	.967	.008**	.830	.551	.359	.032*	.372	.013*
η_p^2	.037	.001	.004	.073	.020	.001	.207	.010	.019	.030	.162	.267	.159
<i>IC-Cue</i>													
<i>F(1,19)</i>	2.18	< 1	3.26	1.21	12.95	7.20	6.78	< 1	< 1	1.57	1.60	1.97	< 1
<i>p</i>	.156	.763	.087	.286	.002**	.015*	.017*	.906	.874	.225	.222	.177	.444
η_p^2	.103	.005	.147	.060	.405	.275	.263	.001	.001	.076	.077	.094	.031
<i>Gender- x IC-Cue</i>													
<i>F(1,19)</i>	1.05	19.90	< 1	< 1	< 1	4.65	11.20	< 1	4.16	< 1	3.98	< 1	< 1
<i>p</i>	.319	.001**	.372	.528	.414	.044*	.003**	.644	.056	.673	.061	.342	.456
η_p^2	.052	.512	.042	.021	.035	.197	.371	.011	.180	.010	.173	.048	.030

Tab. 9.4: Varianzanalytischer Vergleich der drei Weibull-Parameter auf Unterschiede hinsichtlich der vier Untersuchungsbedingungen pro Wort.

Gender-Effekt. Zunächst sollen wieder die Effekte des Gender-Cues auf die drei Parameter einzeln besprochen werden.

shift-Parameter. Abb. 9.13 a) zeigt das Profil des shift-Parameters für die beiden Ausprägungen des Gender-Cues. Die korrespondierende Teilabbildung b) stellt den Effekt des Cues direkt dar (wobei positive Werte wieder für einen Verarbeitungsvorteil stehen).

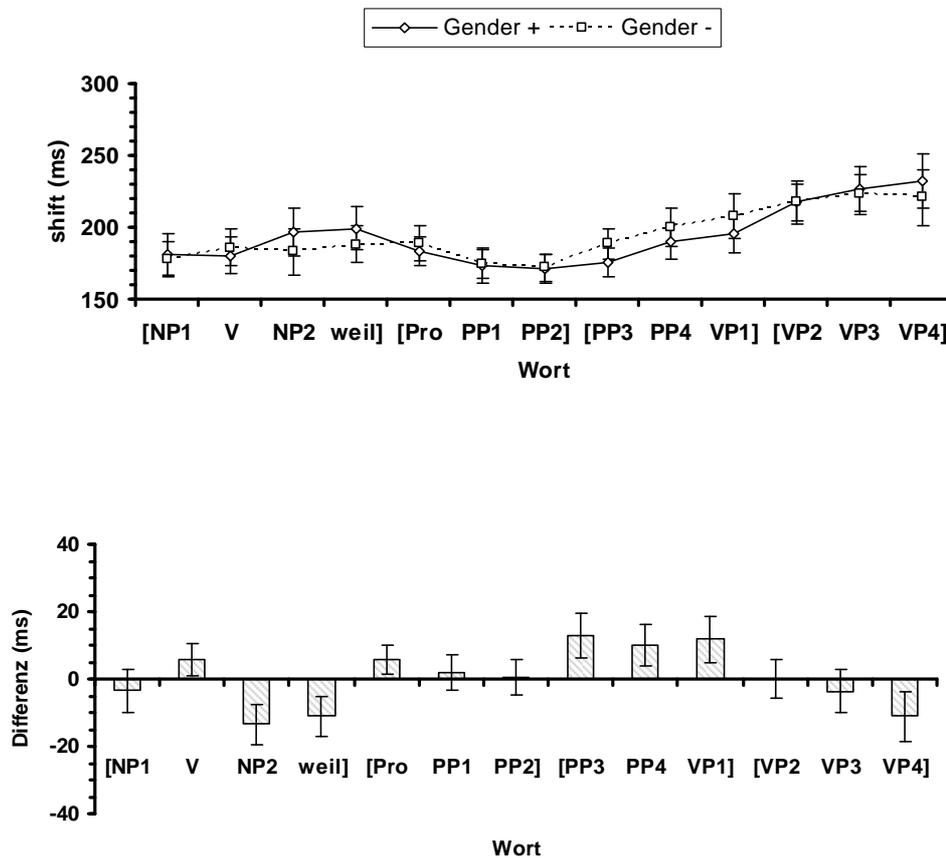


Abb. 9.13: a) Profil des shift-Parameters für die beiden Gender-Bedingungen und b) direkte Darstellung des Gender-Effekts.

Der Gender-Cue beeinflusste den shift-Parameter v.a an zwei Stellen im Satz: Einmal gab es einen negativen Effekt am Ende des Hauptsatzes (evtl. als Reflex auf erhöhte scale-Werte aufgrund der Hauptsatz-Integration: s.u.), und zum anderen gab es positive shift-Effekte am Ende der Präpositionalphrase (PP3-VP1), die jedoch nur marginal signifikant waren.

scale-Parameter. Abb. 9.14 a) und b) zeigen Verlaufsprofile und Effektausprägung des Gender-Cues auf den scale-Parameter.

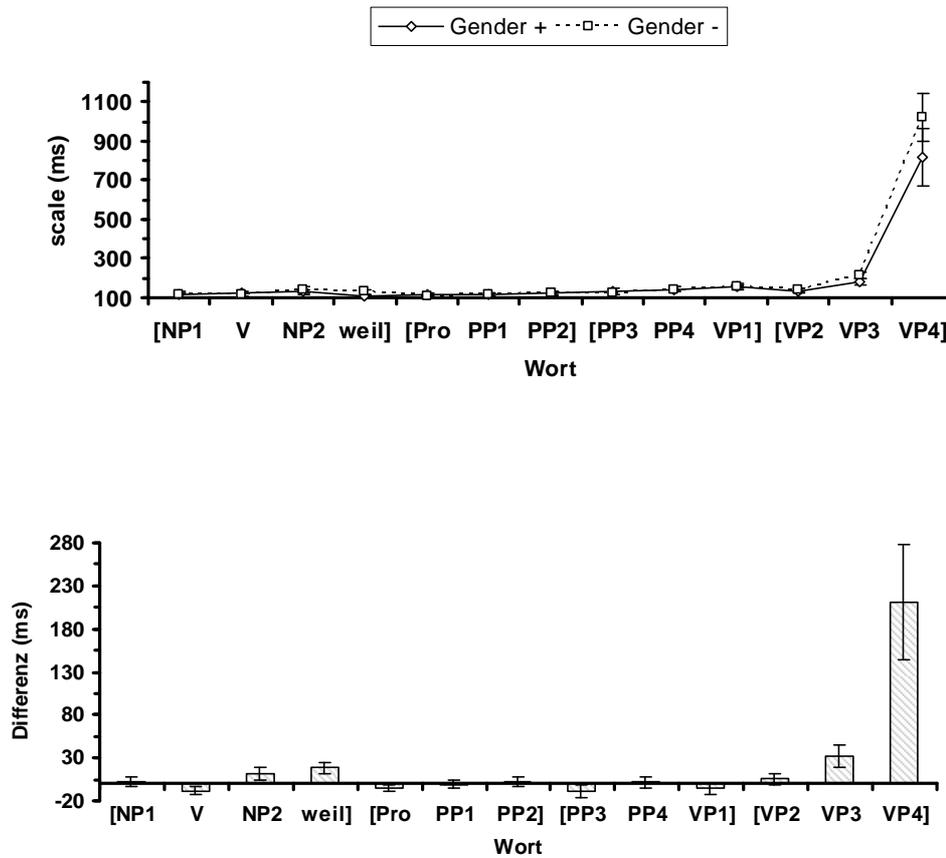


Abb. 9.14: a) Profil des scale-Parameters für die beiden Gender-Bedingungen und b) direkte Darstellung des Gender-Effekts.

An zwei Stellen des Satzes gab es einen Einfluss des Gender-Cues auf den scale-Parameter. Zum einen gab es einen kleinen Effekt am Ende des Hauptsatzes und (besonders prominent) einen sehr starken Effekt am Ende des Nebensatzes (VP3-4).

shape-Parameter. Abb. 9.15 a) und b) zeigen Verlaufprofile und Effektausprägung des Gender-Cues auf dem *shape-Parameter*.

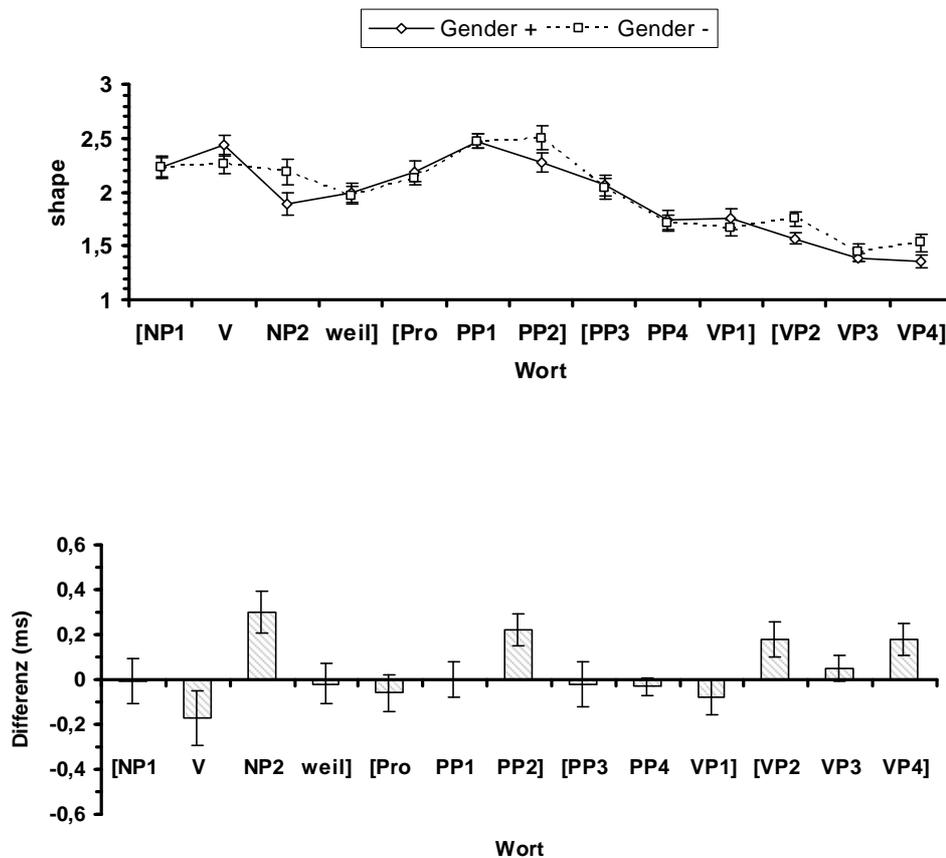


Abb. 9.15: a) Profil des *shape-Parameters* für die beiden Gender-Bedingungen und b) direkte Darstellung des Effekts.

Die Effekte des Gender-Cues auf den *shape-Parameter* fielen wieder vergleichsweise unsystematisch aus. An drei Stellen gab es signifikante Effekte: 1. Zunächst am Ende des Hauptsatzes einen positiven Effekt des Cues, der mit dem Negativ-Effekt des *shift-Parameters* und dem Positiv-Effekt des *scale-Parameters* korrespondierte. 2. Dann gab es auf dem Wort PP2 in der Mitte des Satzes einen signifikanten Positiv-Effekt (der theoretisch nicht begründet werden kann). Und 3. gab es eine Positiv-Strecke am Ende des Satzes (VP2-4), die wahrscheinlich primär vom starken Ausschlag des *scale-Parameters* verursacht wurde (vgl. Appendix A.5).

Zusammenfassend kann also festgestellt werden, dass alle bereits beschriebenen Effekte des Gender-Cues wieder erkennbar waren: sowohl der späte Effekt des *scale-Parameters* als auch der frühe Effekt, der hauptsächlich vom *shift-Parameter* getragen wird.

IC-Effekt. Da Anlass zur Annahme besteht, dass sich die Effekte des IC-Cues unterscheiden, je nachdem ob gleichzeitig auch der Gender-Cue realisiert ist, sollen im Folgenden die Effekte des IC-Cues für die beiden Gender-Bedingungen separat dargestellt werden. Dieses Vorgehen ist auch dadurch gerechtfertigt, dass sich für alle drei Parameter signifikante Interaktions-Terme ergaben, insbesondere am Anfang der Präpositionalphrase (PP1-2). Eine weitere Stelle, an der die Interaktion der beiden Cues relevant zu sein scheint, ist am Anfang der Verbalphrase (PP4-VP2), auch wenn die Effekte hier etwas schwächer ausgeprägt waren.

IC-Effekt (Gender+). Zunächst soll der Effekt des IC-Cues für den Fall dargestellt werden, in dem das Pronomen durch den Gender-Bezug eindeutig interpretierbar ist.

shift-Parameter. Abb. 9.16 zeigt den Einfachen Haupteffekt des IC-Cues auf den shift-Parameter in der Bedingung *Gender+*.

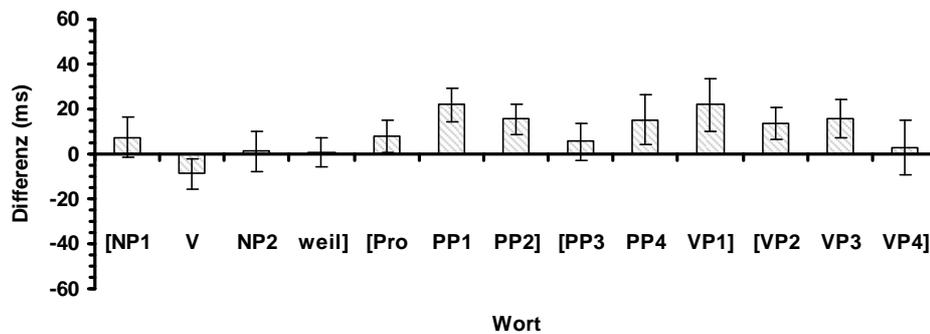


Abb. 9.16: Einfacher Haupteffekt des IC-Cues in der Bedingung *Gender+* auf den shift-Parameter.

Es ist klar erkennbar, dass es einen sehr frühen Effekt des IC-Cues auf den shift-Parameter gab, wenn der Satz gleichzeitig gender-disambiguiert war. Dieser Effekt setzte direkt am Anfang der Präpositionalphrase ein (PP1: $F(1,19) = 8.61, p < .009, \eta_p^2 = .312$; PP2: $F(1,19) = 4.69, p < .043, \eta_p^2 = .198$) und trat an deren Ende erneut zu Tage, auch wenn die einfachen Haupteffekte nur marginal signifikant waren (VP1: $F(1,19) = 3.45, p = .079, \eta_p^2 = .154$; VP2: $F(1,19) = 3.86, p < .064, \eta_p^2 = .169$; VP3: $F(1,19) = 3.38, p < .082, \eta_p^2 = .151$).

scale-Parameter. Abb. 9.17 a) und b) zeigen das Verlaufsprofil des IC-Effekts auf den *scale-Parameter* in der Bedingung *Gender+* mit und ohne den satzfinalen Effekt.

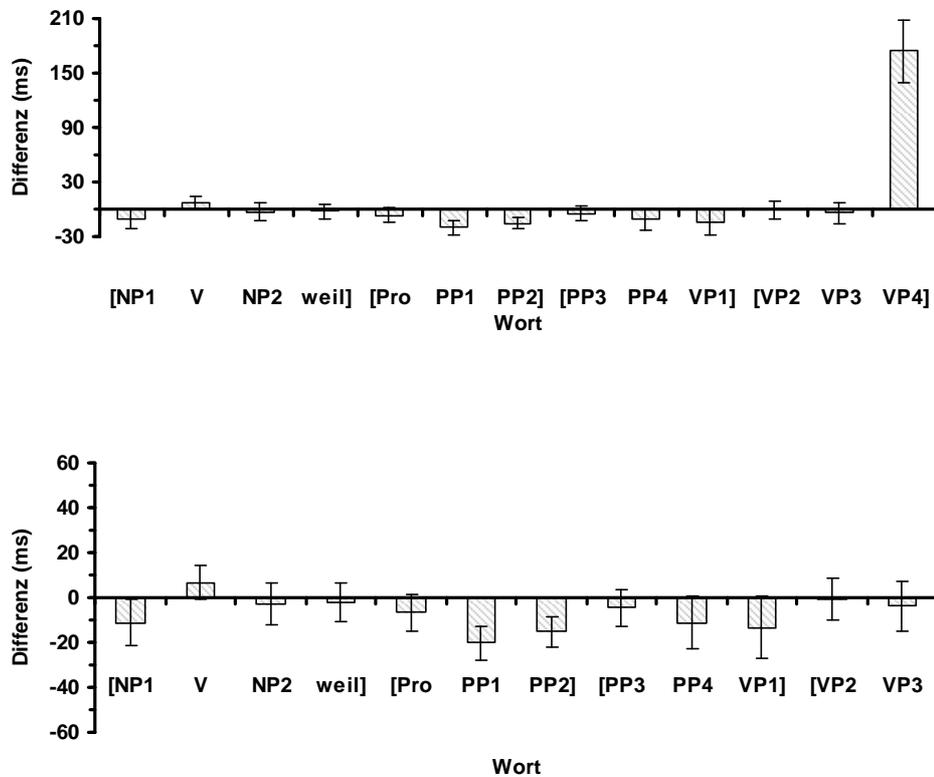


Abb. 9.17: Einfacher Haupteffekt des IC-Cues in der Bedingung *Gender+* auf den *scale-Parameter* a) mit und b) ohne satzfinalen Effekt.

Neben der kaum überraschenden Beobachtung, dass der satzfinale Effekt allein vom *scale-Parameter* getragen wurde, ist für die vorliegende Fragestellung vor allem die Tatsache wichtig, dass es – parallel zum positiven Effekt des Cues auf den *shift-Parameter* – einen komplementären Negativ-Effekt für den *scale-Parameter* gab. Er setzt an der gleichen Stelle ein (PP1: $F(1,19) = 7.04, p < .016, \eta_p^2 = .270$; PP2: $F(1,19) = 5.39, p < .032, \eta_p^2 = .221$) und hatte ungefähr den gleichen Verlauf, auch wenn die späten Effekte nicht signifikant wurden. Dieser Effekt war neu, er in Experiment IIa war nicht beobachtet worden.

shape-Parameter. Abb. 9.18 zeigt die IC-Effekte für den *shape-Parameter*.

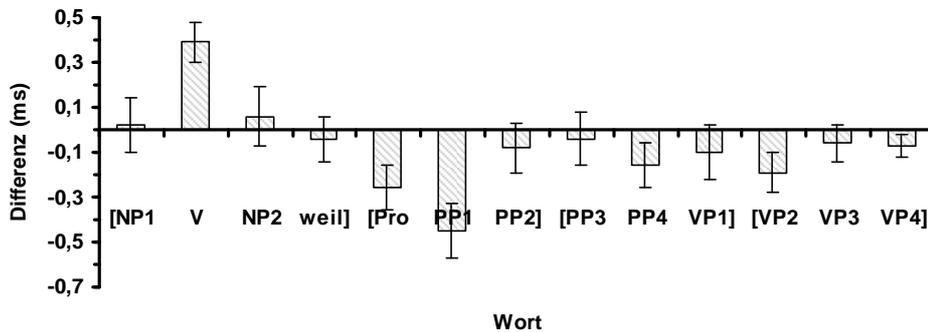


Abb. 9.18: Einfacher Haupteffekt des IC-Cues auf den *shape-Parameter* in der Bedingung *Gender+*.

Die Effekte des IC-Cues auf den *shape-Parameter* schienen ebenfalls mit den Effekten auf die anderen beiden Parameter in Verbindung zu stehen, da sie einen sehr ähnlichen Verlauf aufwiesen. Es gab zwei sehr frühe Effekte des Cues, einer davon auf dem Pronomen selbst (*Pro*: $F(1,19) = 7.39, p < .014, \eta_p^2 = .280$) und einer auf dem ersten Wort der Präpositionalphrase (*PP1*: $F(1,19) = 14.03, p < .001, \eta_p^2 = .425$). Darüber hinaus existierten noch weitere Effekte des Cues satzabwärts, die jedoch schwächer ausgeprägt waren (weshalb auch nur der Effekt auf dem Wort *VP2* die Signifikanzgrenze überschritt).

Als Zwischenstand lässt sich für den IC-Effekt bei gleichzeitiger *Gender-Disambiguierung* festhalten, dass es – neben dem obligatorischen späten *scale-Effekt* – im frühen Bereich der Satzverarbeitung sowohl einen positiven *shift-* als auch einen negativen *scale-Effekt* gab, wobei beide von weiteren *shape-Effekten* begleitet wurden.

IC-Effekt (Gender-). Im Folgenden wird dargestellt, wie die Effekte des IC-Cues auf die drei Parameter der Reaktionszeitverteilung in der ambigen Bedingung ausfallen.

shift-Parameter. Abb. 9.19 zeigt die einfachen Haupteffekte des IC-Cues auf den *shift-Parameter* in der Bedingung *Gender-*.

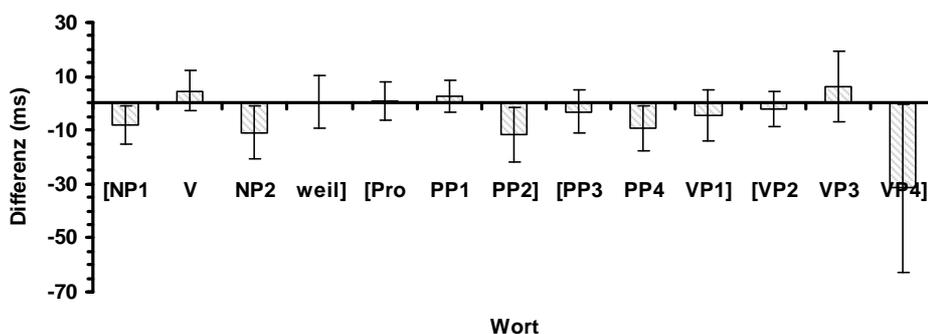


Abb. 9.19: Einfacher Haupteffekt des IC-Cues in der Bedingung *Gender-* auf den *shift-Parameter*.

Ohne Gender-Disambiguierung gab es keine signifikanten Effekte des IC-Cues auf den shift-Parameter (alle $F < 1.4$). Zwar deuteten sich gegen Ende der Präpositionalphrase leichte Effekte des Parameters an, diese waren jedoch – wenn überhaupt vorhanden – negativ ausgeprägt und schon deswegen nicht mit den parallelen Effekten in der Bedingung *Gender+* zu vergleichen.

scale-Parameter. Abb. 9.20 a) und b) zeigen das Verlaufprofil des IC-Effekts auf den *scale-Parameter* in der Bedingung *Gender-* mit und ohne den satzfinalen Effekt.

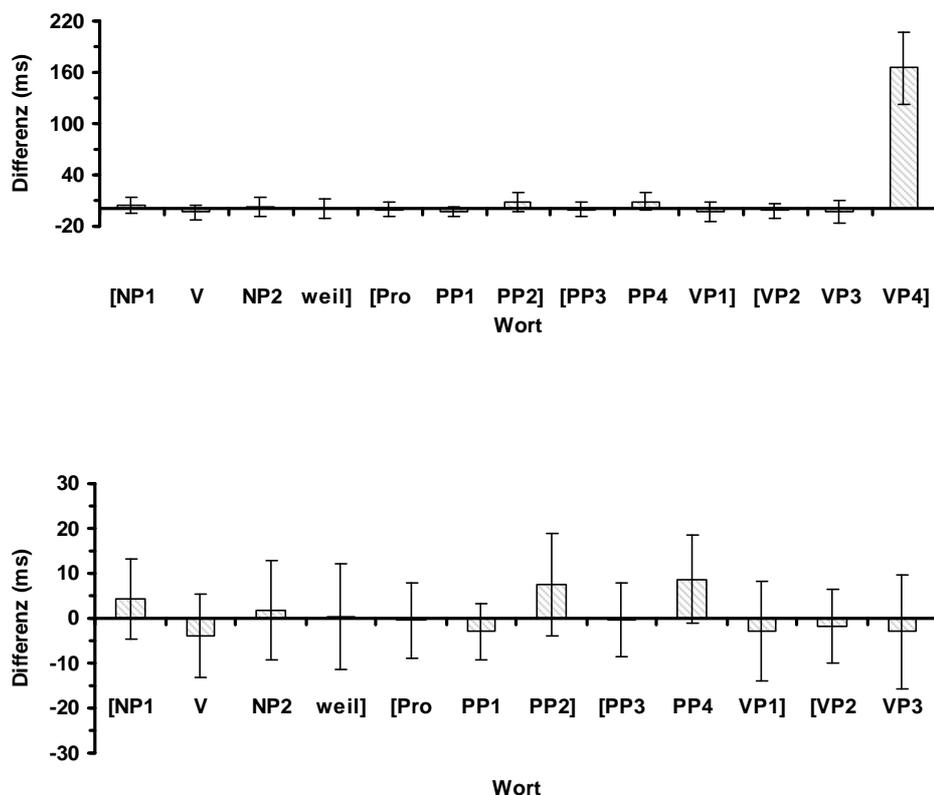


Abb. 9.20: Einfacher Haupteffekt des IC-Cues in der Bedingung *Gender-* auf den *scale-Parameter* a) mit und b) ohne satzfinalen Effekt.

Auch in der Bedingung *Gender-* war der satzfinale *scale*-Effekt besonders prominent. Dieser war sogar vergleichsweise stärker ausgeprägt als in der Bedingung *Gender+*. Es gab keine frühen Effekte für diesen Parameter (alle $F < 1$).

shape-Parameter. Abb. 9.21 zeigt die IC-Effekte für den *shape-Parameter*.

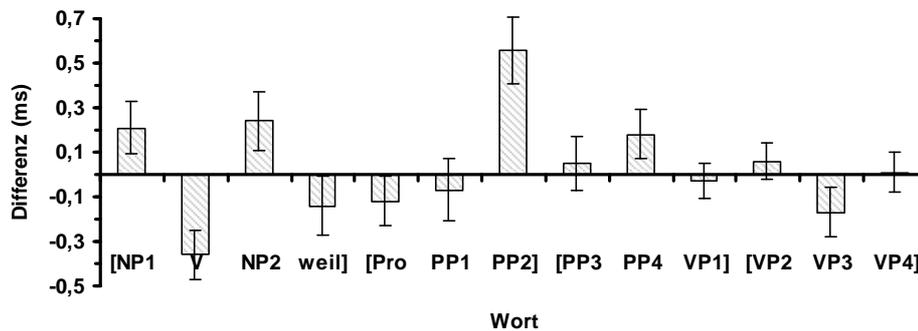


Abb. 9.21: Einfacher Haupteffekt des IC-Cues auf den *shape-Parameter* in der Bedingung Gender-.

Die Effekte des IC-Cues auf den *shape-Parameter* waren unsystematisch und meist nicht signifikant. Lediglich auf dem Wort PP2 und dem Wort PP4 gab es einen positiven Effekt des Cues (PP2: $F(1,19) = 13.07, p < .002, \eta_p^2 = .407$; PP4: $F(1,19) = 2.70, p = .117, \eta_p^2 = .124$).

Zusammenfassend kann festgehalten werden, dass es – wieder mit Ausnahme des obligatorischen *scale*-Effekts im *sentence-wrap-up* – keine systematischen (frühen) Effekte des IC-Cues gab, wenn der Satz nicht gleichzeitig durch den Gender-Cue disambiguiert war, und zwar für keinen der drei Parameter. Hinzu kommt, dass die Effekte für den *shift*- und den *scale-Parameter* – wo tendenziell vorhanden – im Vergleich zu Experiment IIa umgekehrt gepolt waren; d.h. der *shift*-Effekt war tendenziell negativ und der *scale*-Effekt positiv.

9.3.1.2.3 Ergebnisse auf Konstituentenebene

In einem zweiten Schritt wurden die Parameter für die einzelnen Wörter wieder zu Konstituenten zusammengefasst, um die Stabilität der Effekte zu überprüfen (vgl. Tab. 9.5).

Gender-Effekt. Abb. 9.22 zeigt die Effekte des Gender-Cues für die drei Weibull-Parameter.

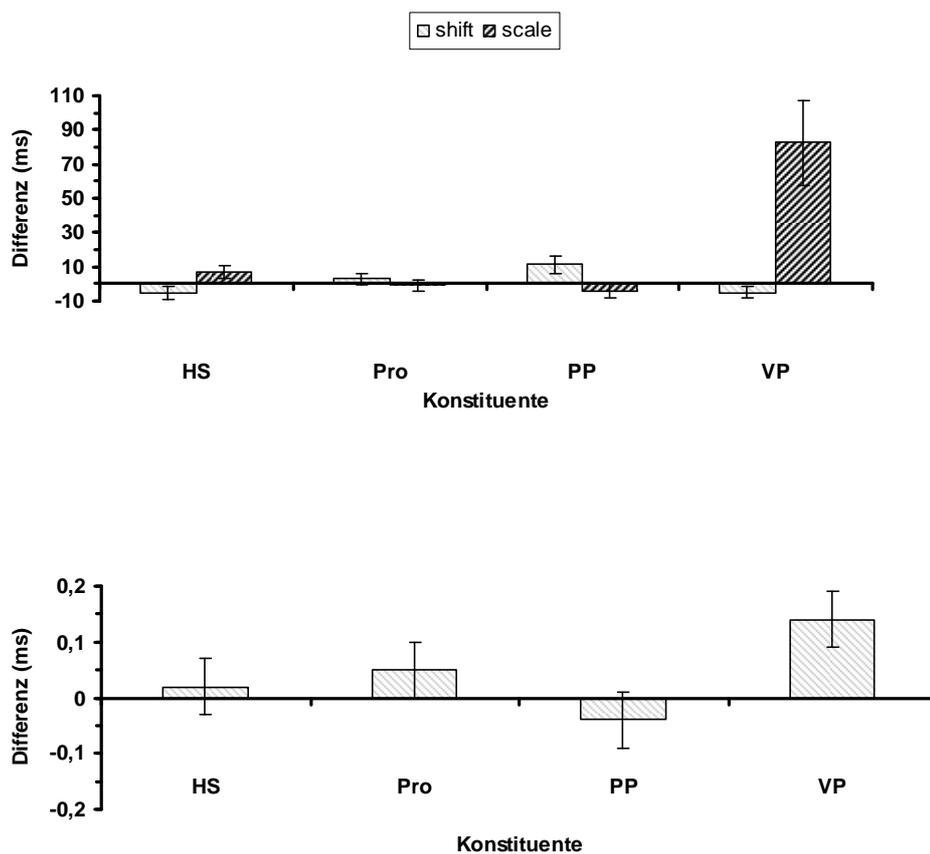


Abb. 9.22: Effekte des Gender-Cues auf a) den shift- und den scale-Parameter sowie b) den shape-Parameter jeder Konstituente des Satzschemas.

Die Aggregation der Einzeleffekte der einzelnen Wörter bestätigte voll das bereits skizzierte Effektmuster: Es gab einen frühen Effekt des Cues im shift-Parameter der Präpositionalphrase und einen späten für den scale- und den shift-Parameter in der Verbalphrase.

	<i>shift</i>				<i>scale</i>				<i>shape</i>			
	HS	Pro	PP	VP	HS	Pro	PP	VP	HS	Pro	PP	VP
<i>Gender-Cue</i>												
<i>F(1,19)</i>	2.29	< 1	5.09	1.78	2.93	< 1	1.06	11.01	< 1	1.29	< 1	7.70
<i>p</i>	.147	.395	.036*	.198	.103	.762	.317	.004**	.622	.271	.389	.012*
η_p^2	.107	.038	.211	.086	.134	.005	.053	.367	.013	.063	.039	.288
<i>IC-Cue</i>												
<i>F(1,19)</i>	< 1	4.99	< 1	< 1	< 1	4.40	< 1	19.43	1.15	3.07	< 1	3.57
<i>p</i>	.642	.038*	.338	.913	.814	.050*	.390	.001**	.298	.096	.741	.074
η_p^2	.012	.208	.048	.001	.003	.188	.039	.506	.057	.139	.006	.158
<i>Gender- x IC-Cue</i>												
<i>F(1,19)</i>	< 1	4.68	4.69	3.18	< 1	3.19	1.44	< 1	< 1	9.08	2.49	1.82
<i>p</i>	.677	.043*	.043*	.090	.751	.090	.244	.833	.334	.007**	.131	.193
η_p^2	.009	.198	.198	.143	.005	.144	.071	.002	.049	.323	.116	.087

Tab. 9.5: Ergebnisse des varianzanalytischen Vergleichs der vier Konstituenten auf Unterschiede hinsichtlich der vier Untersuchungsbedingungen.

IC-Effekt (*Gender+*). Abb. 9.23 zeigt die Effekte des IC-Cues für die drei Weibull-Parameter, wenn der Satz gleichzeitig durch den Gender-Cue disambiguiert wurde.

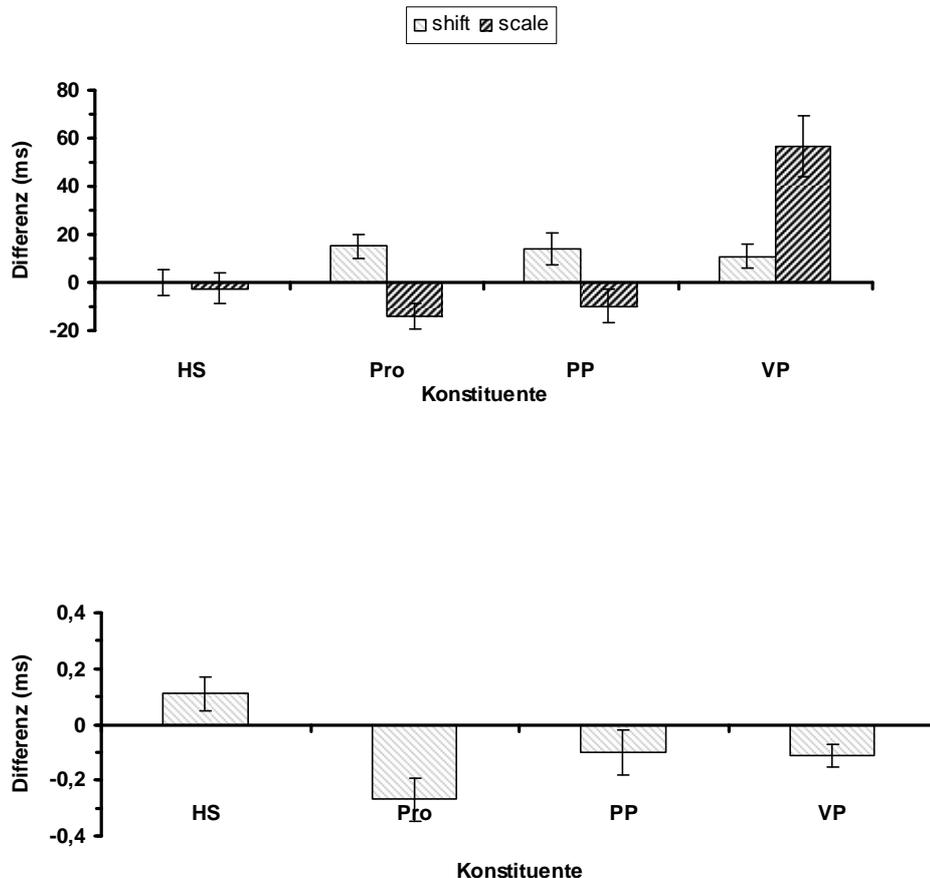


Abb. 9.23: Effekte des IC-Cues in der Bedingung *Gender+* auf a) den shift- und den scale-Parameter sowie b) den shape-Parameter jeder Konstituente des Satzschemas.

Es gab einen sehr frühen positiven Effekt des Cues für den shift-Parameter direkt auf der Pro-Konstituente, der sich in den beiden folgenden Konstituenten noch fortsetzte (Pro: $F(1,19) = 8.38, p < .009, \eta_p^2 = .306$; PP: $F(1,19) = 4.69, p < .043, \eta_p^2 = .198$; VP: $F(1,19) = 4.67, p < .044, \eta_p^2 = .197$). Er wurde begleitet von komplementären Ausschlägen des scale- (Pro: $F(1,19) = 7.42, p < .014, \eta_p^2 = .281$) und des shape-Parameters (Pro: $F(1,19) = 10.84, p < .004, \eta_p^2 = .363$; VP: $F(1,19) = 6.00, p < .024, \eta_p^2 = .240$). Diese frühen negativen Effekte waren neu – sie wurden in Experiment IIa noch nicht beobachtet. Hinzu kam der späte scale-Effekt des Cues, der den *sentence-wrap-up* beschleunigte (VP: $F(1,19) = 19.79, p < .001, \eta_p^2 = .510$).

IC-Effekt (*Gender-*). Abb. 9.24 zeigt die Effekte des IC-Cues für die drei Weibull-Parameter, wenn der Satz nicht gleichzeitig durch den Gender-Cue disambiguiert wurde.

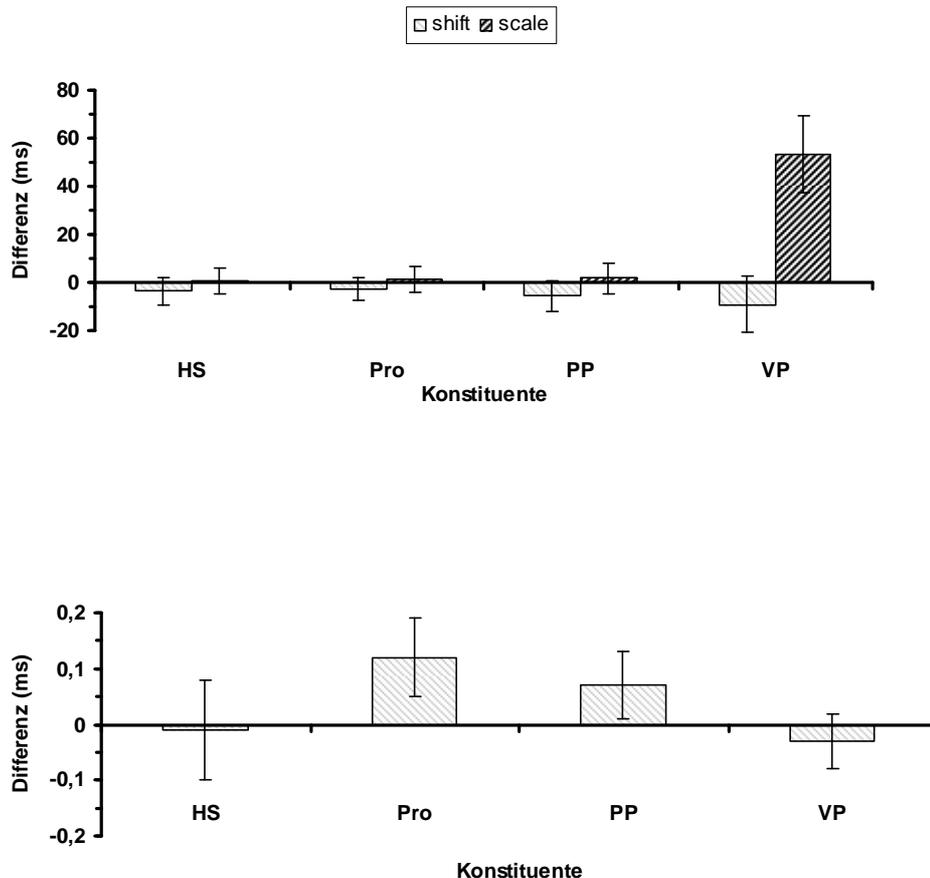


Abb. 9.24: Effekte des IC-Cues in der Bedingung *Gender-* auf a) den shift- und den scale-Parameter sowie b) den shape-Parameter jeder Konstituente des Satzschemas.

Abgesehen vom späten scale-Effekt (VP: $F(1,19) = 11.08$, $p < .004$, $\eta_p^2 = .368$) hatte der IC-Cue keinen systematischen Einfluss auf die drei Parameter. Eine Ausnahme war ein früher Effekt für den shape-Parameter, der auf der Pro-Konstituente marginal signifikant war ($F(1,19) = 2.99$, $p = .100$, $\eta_p^2 = .136$).

Zusammenfassend kann festgestellt werden, dass die Aggregation das berichtete Ergebnismuster stabilisierte: Unsystematische Effekte (vor allem im Hauptsatz) mitteln sich erwartungsgemäß gegenseitig aus, die berichteten, theoretisch relevanten Effekte treten hingegen nun deutlicher zutage.

9.3.1.1.4 Re-Analyse der „korrekt“ beantworteten Items

Alle Analysen wurden mit nur 20 Items, die in der nachfolgenden Verifikationsaufgabe korrekt beantwortet wurden, neu berechnet. Ein vollständiger Bericht des Ergebnismusters findet sich in Appendix C.4.2. Auch wenn alle Effekte etwas schwächer ausgeprägt waren als

in der vollständigen Analyse, ergab sich im Großen und Ganzen das gleiche Ergebnismuster – allerdings mit zwei Ausnahmen: 1. Der frühe Gender-Effekt war zu schwach, um signifikant zu werden. 2. Die Parameterschätzungen für den shape-Parameter waren bei einer Datenbasis von nur 20 Items offensichtlich zu instabil, um sinnvolle Ergebnisse zu liefern. Es finden sich sehr viele Ausschläge für diesen Parameter, die teilweise abnorm groß ausfallen und darüber hinaus sehr inkonsistent sind. Das heißt, für diesen Parameter etablierte sich kein klares oder gar mit dem der Hauptanalyse korrespondierendes Muster.

9.3.1.1.5 Zusammenfassung

Die Analyse der Parameter der einzelnen Reaktionszeitverteilungen erbrachte in Bezug auf die theoretisch relevanten Effekte wiederum ein klares Bild, welches später noch erläutert werden soll. Interessanter ist hingegen ein früher, zusätzlicher scale-Effekt, der in Experiment IIa noch nicht nachweisbar war. Da er vollkommen parallel zum frühen shift-Effekt verlief, ist zunächst unklar, ob es sich bei ihm evtl. um ein methodisches Artefakt handelt. Dieses Problem soll in der abschließenden Diskussion besprochen werden.

9.3.1.3 Diskussion Satzlesezeiten

Im Prinzip konvergieren die Ergebnisse der einzelnen Analysen: Bei den mittleren Leseseiten zeigten sich Effekte sowohl für den Gender-Cue als auch für den IC-Cue spätestens ab dem Ende der Präpositionalphrase. Beide Effekte waren also erst etwas später nachzuweisen als in Experiment IIa. Sie kulminierten beide in einen extremen *wrap-up*-Effekt, der ungefähr doppelt so groß war wie alle anderen Effekte zusammen. Die Analyse der Reaktionszeitverteilungen konnte dieses Bild komplettieren, indem sie sicherstellte, dass die beiden *wrap-up*-Effekte erwartungsgemäß reine scale-Effekte sind. Auch die Interpretation der beiden frühen Effekte ist vergleichsweise eindeutig. Der frühe Gender-Effekt wurde dem shift-Parameter zugeordnet. Dieser ließ sich allerdings erst in der Präpositionalphrase nachweisen. Beim frühen IC-Effekt (der erwartungsgemäß nur in der disambiguierten Bedingung auftrat) ergab sich hingegen ein Doppeleffekt: Einerseits gab es einen deutlichen Verarbeitungsnachteil in der inkongruenten Version, der sich im shift-Parameter niederschlug. Dieser wurde jedoch gleichzeitig von einem frühen scale-Effekt ausgeglichen. Dieser Sachverhalt wurde in Experiment IIa noch nicht beobachtet. Da zu vermuten ist, dass es sich bei diesem Effekt um ein Artefakt handelt, soll er in der abschließenden Diskussion noch eingehend besprochen werden.

9.3.2 Verifikationsaufgabe

9.3.2.1 Akkuratheit: Anteil „korrekt“ verifizierter Aussagen

Ergebnisse. Abb. 9.25 zeigt die Akkuratheit (Hit-Rate) für die Verifikationsaufgabe in allen vier Untersuchungsbedingungen.

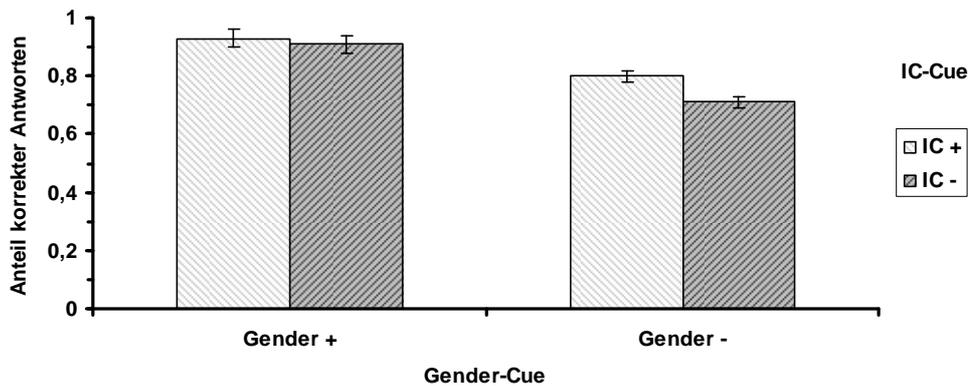


Abb. 9.25: Anteil „korrekt“ verifizierter Aussagen nach Aussagetyt.

Eine messwiederholte ANOVA mit den beiden Faktoren *Gender-Cue* (2: vorhanden vs. nicht vorhanden) und *IC-Cue* (2: vorhanden vs. nicht-vorhanden) ergab einen sehr starken Haupteffekt für den Faktor *Gender-Cue* ($F(1,19) = 94.70, p < .001, \eta_p^2 = .833; F(1,159) = 114.56, p < .001, \eta_p^2 = .419$): Wenn ein Satz nicht gender-disambiguiert war, kam es zu einem dramatischen Abfall an korrekten Antworten (ca. 17 %). Darüber hinaus gab es auch einen Haupteffekt für den Faktor *IC-Cue* ($F(1,19) = 16.59, p < .001, \eta_p^2 = .466; F(1,159) = 16.15, p < .001, \eta_p^2 = .092$), der allerdings durch einen signifikanten Interaktionsterm ($F(1,24) = 7.01, p < .016, \eta_p^2 = .270; F(1,159) = 4.98, p < .027, \eta_p^2 = .030$) weiter qualifiziert wurde: Wenn ein Satz bereits gender-disambiguiert war, fiel der Effekt des IC-Cues wesentlich schwächer aus als wenn das Pronomen ambig war (Gender+: ca. 2 %; $F(1,19) = 5.90, p < .025, \eta_p^2 = .237; F(1,159) = 1.85, p = .175, \eta_p^2 = .012$; Gender-: ca. 9 %; $F(1,19) = 12.61, p < .002, \eta_p^2 = .399; F(1,159) = 12.50, p < .001, \eta_p^2 = .073$).

Arcussinus-Transformation. Die Analyse wurde nochmals mit arcussinus-transformierten Werten durchgeführt. Die vollständigen Ergebnisse sind in Appendix C.4.3 einsehbar. Das Ergebnismuster war jedoch vollkommen identisch zu dem soeben berichteten.

Zusammenfassung. Die Effektstruktur ist hinlänglich bekannt und soll nicht mehr besprochen werden. Wichtiger ist hingegen, dass das grundsätzliche Akkuratheitsniveau im Vergleich zu Experiment IIa wieder angestiegen ist und wie in Experiment I bei 85 % liegt.

9.3.2.2 Mittlere Reaktionszeiten

Datenbereinigung. Für die pronominale Auflösung wurde wiederum ein Wert von 200 ms als unterer Cut-off-Wert angesetzt. Lediglich ein Datenpunkt fiel unter diesen kritischen Grenzwert. Weiterhin wurden alle Werte, die 4 Standardabweichungen vom dem bedingungs-spezifischen Mittelwert einer Person abwichen, durch diesen ersetzt. Diese Maßnahme führte zu einer Ersetzung von ca. 0.2 % aller Datenpunkte.

Ergebnisse. Abb. 9.26 zeigt die mittlere Zeit, die durchschnittlich zur Verifizierung von Sätzen in den unterschiedlichen Untersuchungsbedingungen gebraucht wurde.

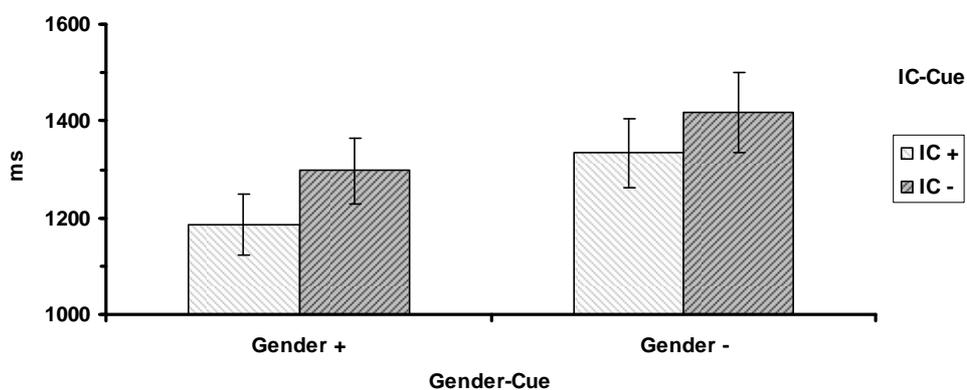


Abb. 9.26: Mittlere RT für die Verifikationsaufgabe in den vier Untersuchungsbedingungen.

Einerseits gab es einen varianzstarken Haupteffekt für den Faktor *Gender-Cue* ($F(1,19) = 33.46$, $p < .001$, $\eta_p^2 = .638$; $F(1,159) = 43.71$, $p < .001$, $\eta_p^2 = .216$), da gender-disambiguierte Sätze ca. 130 ms schneller verifiziert wurden als nicht-disambiguierte. Andererseits gab es auch einen Effekt für den Faktor *IC-Cue* ($F(1,19) = 24.11$, $p < .001$, $\eta_p^2 = .559$; $F(1,159) = 26.66$, $p < .001$, $\eta_p^2 = .144$), der darauf zurückgeführt werden kann, dass verbkongruente Nebensätze ca. 100 ms schneller verifiziert wurden. Allerdings interagieren die beiden Faktoren nicht miteinander (beide $F < 1$).

Re-Analyse der „korrekten“ Antworten. Die Analyse wurde nochmals nur mit den „korrekt“ beantworteten Items durchgeführt. Das Ergebnismuster (s. Appendix C.4.4) war ähnlich, allerdings deutete sich in der *by-participants*-Analyse eine Interaktion der beiden Faktoren an.

Zusammenfassung. Die mittleren Antwortlatenzen für die Verifikationsaufgabe zeigten ein ähnliches Muster wie in Experiment IIa. In charakteristischem Gegensatz zur Verifikationsaufgabe in den Experimenten Ia und Ib, interagierten die beiden Cues nicht miteinander. Dies weist darauf hin, dass die Verifikationsaufgabe nun Eigenschaften hat, wie sie dort für die pronominale Auflösung typisch waren. Das generelle Niveau der Antwortlatenzen war mit ca. 1300 ms ebenfalls ungefähr vergleichbar zu Experiment IIa.

9.3.2.3 Reaktionszeitverteilung: Weibull-Parameter

Parameterschätzung. Das Vorgehen bei der Parameterschätzung entsprach dem in Kap. 3.4 skizzierten. Als Startwert für den scale-Parameter wurde wiederum ein Wert von $\lambda = \theta^{\beta} = 2$ angesetzt. Die Parameter der beiden Prior-Verteilungen des scale-Parameters waren $a_1 = 2.0$, $b_1 = 0.1$, $a_2 = 2.0$ und $b_2 = 0.2$.

Modellgüte. Die Passung der Weibull-Verteilung für die einzelnen RT-Verteilungen wurde mittels des in Kap. 3.5 beschriebenen Chi-Quadrat-Verfahrens beurteilt. Der Fit kann insgesamt als akzeptabel bezeichnet werden: Der mittlere Chi-Quadrat-Wert lag bei $\chi^2(3) = 5.47$, $p < .25$. Sieben der insgesamt 80 (= 20 x 4) Einzelverteilungen (= 8.8 %) wurden auf dem 1 %-Niveau zurückgewiesen.

Durchschnittliche Parameterausprägung. Abb. 9.27 zeigt die drei Weibull-Parameter der Reaktionszeit-Verteilung und ihre Beziehungen zueinander.

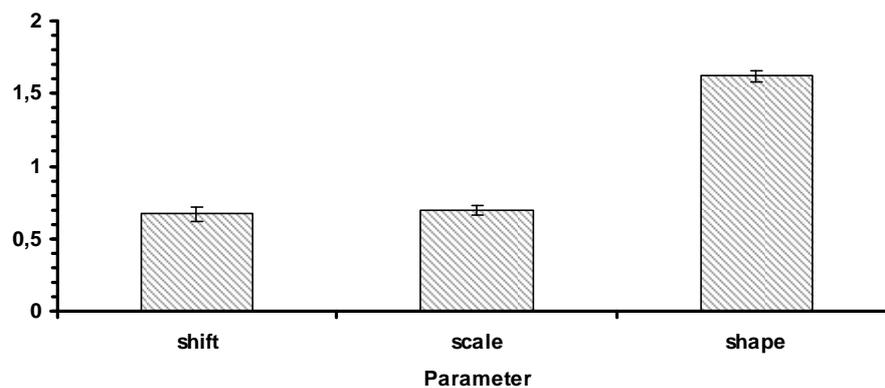


Abb. 9.27: Durchschnittliche Weibull-Parameter der RT-Verteilungen der Verifikationsaufgabe.

Der durchschnittliche shift-Parameter lag bei 0.67 s, der durchschnittliche scale-Parameter bei 0.70 s und der durchschnittliche shape-Parameter bei 1.62. Damit ergaben sich für alle drei Parameter fast identische Ergebnisse wie für die Verifikationsaufgabe in den vorangegangenen Experimenten.

Diese drei Parameter wurden nun jeweils separat mittels einer messwiederholten ANOVA mit den beiden Faktoren *Gender-Cue* (2: *vorhanden* vs. *nicht vorhanden*) und *IC-Cue* (2: *vorhanden* vs. *nicht vorhanden*) auf Unterschiede zwischen den Untersuchungsbedingungen geprüft. Dabei ergaben sich folgende Ergebnisse:

shift-Parameter. Für diesen Parameter gab es einen signifikanten Haupteffekt für den IC-Cue ($F(1,19) = 7.68$, $p < .012$, $\eta_p^2 = .288$): Der shift-Parameter für ic-kongruente Sätze war ca. 60 ms niedriger als für ic-inkongruente (vgl. Abb. 9.28). Alle anderen Effekte waren nicht signifikant (alle $F < 1.4$).

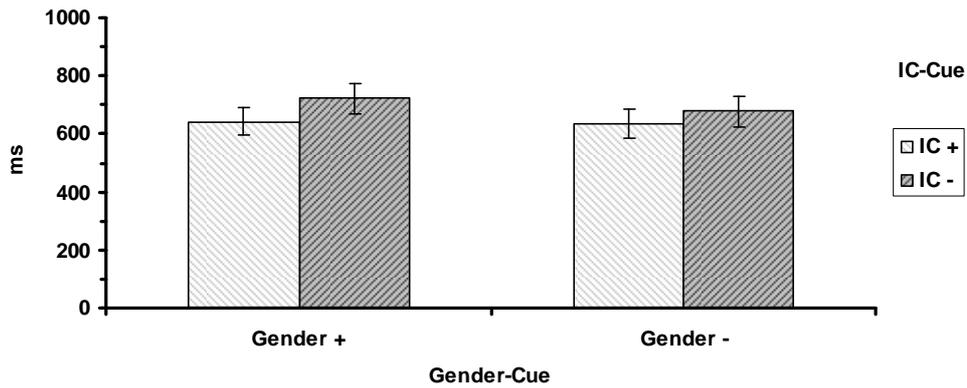


Abb. 9.28: Ausprägung des shift-Parameters der RT-Verteilungen der Verifikationsaufgabe.

scale-Parameter. Es gab einen varianzstarken Haupteffekt für den Faktor *Gender-Cue* ($F(1,19) = 35.56, p < .001, \eta_p^2 = .652$): Reaktionszeitverteilungen für die Sätze, bei denen das Pronomen nicht gender-disambiguiert war, wiesen einen um ca. 170 ms höheren scale-Parameter auf als Sätze mit disambiguierten Pronomen (vgl. Abb. 9.29). Der Haupteffekt für den Faktor *IC-Cue* war hingegen nicht signifikant ($F(1,19) = 2.62, p = .122, \eta_p^2 = .121$), obwohl sich die beiden Bedingungen um ca. 40 ms unterschieden. Die Interaktion der beiden Faktoren war ebenfalls nicht signifikant ($F < 1$).

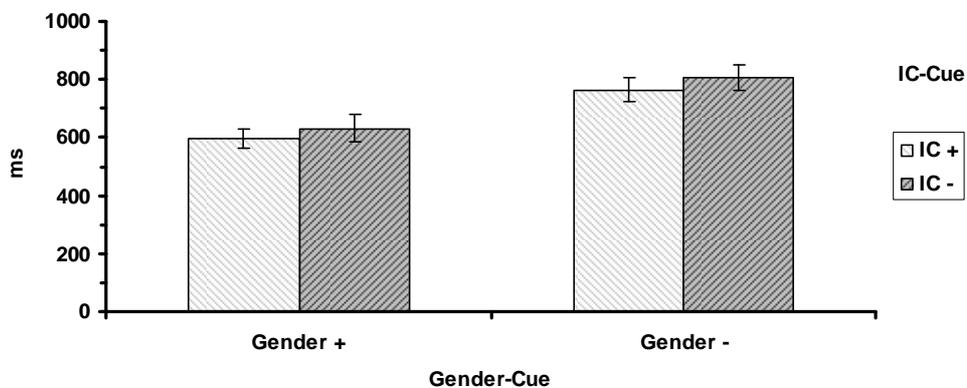


Abb. 9.29: Ausprägung des scale-Parameters der RT-Verteilungen der Verifikationsaufgabe.

shape-Parameter. Für diesen Parameter gab es keine signifikanten Effekte (alle $F < 1$), nur einen sehr schwachen und lediglich marginal signifikanten Effekt des IC-Cues ($F(1,19) = 3.90, p = .063, \eta_p^2 = .170$): IC-kongruente Sätze wiesen einen um ca. .14 Einheiten niedrigeren shape-Parameter auf.

Re-Analyse der „korrekten“ Antworten. Für die Analyse der „korrekt“ beantworteten Items ergab sich ein unterschiedliches Ergebnismuster (vgl. Appendix C.4.5): Die Effekte der beiden Cues lagen – wie in Experiment Ila – allein im scale-Parameter und waren ähnlich stark ausgeprägt.

Zusammenfassung. Auch die Weibull-Parameter wiesen Eigenschaften auf, wie sie in Experiment I für die pronominale Auflösung typisch waren. Insbesondere interagierten die beiden Cues nicht miteinander. Bezüglich der Lokalisierung der Effekte war das Bild jedoch uneinheitlich: Zum einen legte die Analyse aller Antworten nahe, dass die Effekte der beiden Cues von unterschiedlichen Parametern getragen werden. Der IC-Effekt befand sich dabei im shift-Parameter. Die Analyse nur der korrekt beantworteten Items sprach hingegen eher dafür, dass die beiden Effekte allein im scale-Parameter zu verorten sind.

9.3.2.3 Diskussion Verifikationsaufgabe

Die vielleicht wichtigste Erkenntnis aus der Betrachtung der Verifikationsaufgabe in Experiment IIb ist, dass die generellen Akkuratheitswerte – wie intendiert – durch die Verwendung der kumulativen *moving-window*-Methode wieder auf ein Niveau gebracht werden konnten, wie es aus Experiment I bekannt ist. Hier ist es offensichtlich besser möglich, auf das bereits gelesene Material zu rekurrieren und so direkt den Antezedenten des Pronomens zu identifizieren. Dies scheint sich förderlich auf die Verstehensleistung auszuwirken.

Hinsichtlich der Reaktionszeiten ergaben sich allerdings einige Abweichungen zum Ergebnismuster der bisherigen Untersuchungen: Bei der Analyse der mittleren Reaktionszeiten interagierten die beiden Faktoren nicht miteinander (da der IC-Cue auch in der disambiguierten Bedingung die Verifikationszeit beeinflusste). Die RT-Verteilungsanalyse ergab, dass der IC-Effekt im shift-Parameter lag. Allerdings galt das nur für die Gesamtauswertung. Wenn hingegen nur die korrekt beantworteten Items betrachtet wurden, dann war der Effekt wie erwartet im scale-Parameter angesiedelt.

9.4 Diskussion Experiment IIb

Das Ergebnismuster in Experiment IIb ist ebenfalls nicht vollkommen eindeutig. Es zeichnet sich jedoch ein wesentlich eindeutigeres Bild ab als in Experiment IIa.

Besonders klar sind die Ergebnisse bezüglich des Gender-Cues. Hier konnte eindeutig ein früher *confirmation*- von einem späten *disengagement*-Effekt abgegrenzt werden. Der *confirmation*-Effekt setzt dabei früher ein als der *disengagement*-Effekt. Da jedoch generell der Ansatzpunkt des Effekts in Experiment IIb im Vergleich zur nicht-kumulativen *moving-window*-Methode verzögert war, ließ er sich erst am Ende der Präpositionalphrase nachweisen. Die RT-Verteilungsanalyse bestätigte, dass es sich bei diesem Effekt um einen shift-Effekt handelt, was auf die Beteiligung vergleichsweise hierarchie-niedriger Verarbeitungsprozesse hindeutet. Der *disengagement*-Effekt ist hingegen ausschließlich im *sentence-wrap-up* zu verorten. Er fiel im Vergleich zu Experiment IIa ausgesprochen massiv aus. Die Analy-

se der Verteilungscharakteristika bestätigte, dass es sich bei diesem Effekt ausschließlich um einen scale-Effekt handelt, der v.a. hierarchie-hohe Verarbeitungsprozesse abbildet. Erwartungsgemäß persistierte dieser Effekt auch noch in der nachfolgenden Verifikationsaufgabe.

Für die beiden IC-Effekte ist in Experiment IIb die Befundlage ebenfalls wesentlich eindeutiger. Im Gegensatz zu den Effekten des Gender-Cues ist das Bild allerdings auch nicht vollkommen unproblematisch. Der IC-Fokus-Effekt zeichnete sich auf der Ebene der mittleren Lesezeiten bereits in der Präpositionalphrase ab, wo er erwartungsgemäß nur in der disambiguierten Bedingung auftrat. Er wies damit die gleiche Ansatzcharakteristik auf wie der Gender-*confirmation*-Effekt. Dies spricht dafür, dass inkongruente Fortsetzungen bereits früh als solche erkannt wurden und direkt zu einer erschwerten Verarbeitung führten. Die Analyse der korrespondierenden Reaktionszeitverteilungen ergab, dass sich dieser Effekt erwartungsgemäß auf eine Erhöhung des shift-Parameters zurückführen ließ, die bereits auf der Pronominal-Konstituente einsetzte. Allerdings führte diese frühe Elaboration inkonsistenter Informationen nicht zu einer Reduktion satzfinaler Integrationsprozesse, was evtl. an den Aufgabencharakteristika der kumulativen *moving-window*-Methode liegt. Diese können ebenfalls dafür verantwortlich gemacht werden, dass der IC-Integrations-Effekt in Experiment IIb sogar besonders deutlich ausfiel: Er war ungefähr doppelt so stark ausgeprägt wie alle frühen IC-Effekte zusammen. Auf der Ebene der mittleren Reaktionszeiten zeigte sich, dass er ausschließlich den satzfinalen *wrap-up* betraf und die RT-Verteilungsanalyse bestätigte, dass es sich bei ihm um einen reinen scale-Effekt handelte. Allerdings fiel er erwartungskonträr nicht schwächer aus, wenn das Pronomen durch den Gender-Cue disambiguiert wurde.

Dies könnte auch die Ursache dafür sein, dass sich bezüglich des IC-Cues in der anschließenden Verifikationsaufgabe uneinheitliche Ergebnisse zeigten: Grundsätzlich wurde auch hier ein persistierender IC-Effekt gefunden. Allerdings lag er bei der Analyse aller Items hauptsächlich im shift-Parameter. Erst die Betrachtung nur der korrekt beantworteten Items zeigte, dass es sich um einen scale-Effekt handelt. Dieses Ergebnismuster kann evtl. darauf zurückgeführt werden, dass die satzfinale Integration in dieser Bedingung sogar so gründlich vorgenommen wurde, dass die Verifikationsaufgabe lediglich durch einen oberflächlichen Check-Prozess zu lösen war, der sich im shift-Parameter niederschlägt. Es würde allerdings zusätzlicher Untersuchungen bedürfen, um zu überprüfen, ob es durch die veränderte Darbietungsmethode wirklich zu einem anderen Antwortprozess kommt oder dieser Befund auf Zufälligkeiten der vorliegenden Untersuchung zurückführbar ist. Tab. 9.6 zeigt die in der vorliegenden Untersuchung beobachteten Effekte noch einmal im Überblick:

<i>Effekt</i>	<i>Teileffekt</i>	<i>Satzlesezeiten</i>	<i>Verifikation</i>
Gender	Confirmation	✓	---
	Disengagement	✓	✓
IC	Fokus	✓	---
	Integration	✓	(---)

Tab. 9.6: Effekte in Experiment IIb für die Satzlesezeiten und die Verifikationsaufgabe.

Der entscheidende Grund, warum in Experiment IIb auch die frühen Effektkomponenten mit der RT-Verteilungsanalyse reliabel nachgewiesen werden konnten, ist ihre im Vergleich zu Experiment IIa größere Effektstärke. Sowohl der *Gender-confirmation*- als auch der IC-Fokus-Effekt überschreiten bereits auf der Ebene der mittleren Lesezeiten die „magische“ Grenze von 10 ms, was sich positiv auf die Schätzung der Verteilungsparameter auswirkt. Dies kann ziemlich klar auf die veränderte Art der Satzpräsentation zurückgeführt werden. Offensichtlich werden frühe Verarbeitungsprozesse gestärkt, wenn ein direkter Rückgriff auf den relevanten Diskurskontext möglich ist. Denn wie in der Einleitung dieses Kapitels ausführlich diskutiert, ist es bei der kumulativen im Gegensatz zur nicht-kumulativen *moving-window*-Methode möglich, direkt nach dem Pronomen den vorhergehenden Satzkontext noch einmal zu inspizieren und zu versuchen, die Inkonsistenz durch einen Rückgriff auf bereits vorhandene Szenario-Informationen zu reanalysieren. Dies ist theoretisch v.a. deswegen interessant, weil sich damit eine ähnliche Konstellation ergibt, wie beim Übergang von Experiment Ia zu Experiment Ib. Dort konnte ebenfalls durch die zusätzliche Einführung diskursiver Hintergrundinformationen die frühe Auflösungskomponente gestärkt werden. Dies ist ein weiterer Hinweis darauf, dass die frühen, gedächtnis-basierten Prozesse auf Informationsbestände zurückgreifen, die bereits vorher (evtl. durch erklärungs-basierte Prozesse) zur Verfügung gestellt wurden und für eine frühe Reanalyse herangezogen werden können. Daraus resultiert offensichtlich eine derartige Stärkung der frühen Prozesskomponenten, dass sie sich sogar gegen die starke Tendenz der kumulativen *moving-window*-Methode, Integrationsprozesse v.a. am Satzende vorzunehmen, noch durchsetzen können. Für eine verbesserte Verstehens- und damit Reanalyseleistung durch Rückgriff auf bisherige Satzinformationen spricht ebenfalls die Tatsache, dass die Akkuratheitsrate bei der Verifikationsaufgabe deutlich gestiegen ist. Bei der nicht-kumulativen Präsentation war diese auf Werte um die 60 % abgesunken, insbesondere in den Untersuchungsbedingungen, in denen eine Auflösung ausschließlich aufgrund pragmatischer (Verb-)Informationen vorgenommen werden konnte. Wenn also ein diskursiver Kontext für die pronominale Auflösung notwendig ist, dieser aber nicht mehr zur Verfügung steht, kommt es zu Verständnisschwierigkeiten.

Besonderer Diskussion bedarf ein anderer Effekt, der in Experiment IIb erstmalig beobachtet werden konnte: Ein Verarbeitungsnachteil für ic-inkongruente Fortsetzungen, der bereits sehr früh auftrat. Im Gegensatz zum IC-Fokus-Effekt betraf dieser Effekt jedoch allein den scale-Parameter. In Experiment IIa gab es zwar einen sehr ähnlichen Effekt, dort betraf er allerdings die ambige Bedingung. Dieser Effekt ist theoretisch v.a. deshalb interessant, weil er darauf hindeutet, dass Integrationsprozesse nicht allein satzfinal stattfinden (wie z.B. vom *bonding-resolution*-Ansatz von Garrod & Sanford, 1989, angenommen wird). Stattdessen könnte es auch sein, dass die pragmatisch-diskursive Interpretation eines Satzes streng inkrementell vorgenommen wird. Diese Auffassung wird speziell für den IC-Effekt von Koornneef und van Berkum (2006) vertreten. Für sie ist der IC-Fokus-Effekt ein Indikator dafür, dass die abweichenden Diskursinformationen direkt integriert werden und die Gesamtinterpretation des Satzes kontinuierlich aufgebaut wird. Es wäre durchaus möglich, dass ein ähnlicher Mechanismus dem frühen scale-Effekt zugrunde liegt: Dies würde erklären, warum er lediglich dann auftritt, wenn das Pronomen disambiguiert ist. Denn es darf nicht vergessen werden, dass der Effekt negativ ist, d.h. kongruente Fortsetzungen brauchen *länger*, um verarbeitet zu werden. Dies ist ja zunächst einmal kontra-intuitiv. Unter der Prämisse, dass frühe Integrationsprozesse Verarbeitungskapazität beanspruchen, wäre es theoretisch allerdings durchaus plausibel.

Eine solche Interpretation des frühen scale-Effekts wäre sehr reizvoll. Sie passt insbesondere sehr gut in das in Kap. 1 skizzierte integrative Rahmenmodell der pronominalen Auflösung, welches von einer dynamischen Wechselwirkung der verschiedenen Prozesskomponenten im Auflösungsprozess ausgeht. Allerdings muss man hier sehr vorsichtig sein: Der Effekt wurde in Experiment IIa und IIb in unterschiedlichen Bedingungen beobachtet und konnte jeweils einzeln nicht repliziert werden. So traurig das auch ist: Im vorliegenden Fall ist eher davon auszugehen, dass es sich bei diesem Effekt um ein reines Methoden-Artefakt handelt, das auf Abhängigkeitsbeziehungen bei der Parameterschätzung zurückzuführen ist. Wahrscheinlich ist der negative Ausschlag des scale-Parameters an dieser Stelle nur ein Reflex der Tatsache, dass der korrespondierende shift-Parameter an der gleichen Stelle einen starken Positiv-Ausschlag aufwies. Da die beiden Parameter bei ihrer Schätzung negativ miteinander korreliert sind, kann es u.U. komplementär zum „realen“ shift-Effekt zu einem artifiziellen scale-Effekt kommen. Auf die Problematik abhängiger Parameterschätzungen und ihre Konsequenzen für die vorgestellten Studien soll jedoch noch ausführlich in der abschließenden Diskussion eingegangen werden.

10. Zusammenfassung der Ergebnisse und Diskussion

In dieser Arbeit wurde die Auffassung vertreten, dass es sinnvoll ist, die Beziehung zwischen den gedächtnis-basierten und der erklärungs-basierten Modellen der pronominalen Auflösung nicht in einer Ausschließlichkeitsrelation zu konstruieren (vgl. Kap. 1). Beide Prozessarten sind nicht nur gleichzeitig, aber unabhängig voneinander am pronominalen Auflösungsprozess beteiligt, sondern interagieren dynamisch miteinander. Entsprechend sollten sich gedächtnis- und erklärungs-basierte Auflösungsprozesse auch nicht oder nur sehr schwer voneinander trennen lassen: Beide Prozessarten sind notwendige und essenzielle Bestandteile der pronominalen Auflösung, und das Verstehensprodukt ergibt sich gerade aus ihrer Interaktion.

In Kap. 2 wurde dargelegt, inwieweit dieses theoretische Rahmenmodell auch zur Erklärung der komplexen Effekte morpho-syntaktischer Gender- und semantisch-pragmatischer IC-Informationen auf die pronominalen Auflösung herangezogen werden kann. Hierfür wurde zunächst aufgearbeitet, dass in der relevanten Literatur sowohl für den Gender- als auch für den IC-Cue bislang Ansätze formuliert wurden, welche ihre Wirkung jeweils nur auf einen der beiden Prozesse zurückführen wollen. Beim IC-Cue führte dies zur Kontrastierung einer Fokussierungs-Hypothese, welche von einer Voraktivierung des kausal salienten Aktanten durch den IC-Cue ausgeht, und einer Integrations-Hypothese, welche davon ausgeht, dass ic-kongruente Fortsetzungen im Zuge der satzfinalen Kohärenzbildung leichter in das Diskursmodell zu integrieren sind. In Bezug auf den Gender-Cue wurden parallel zwei ähnliche Erklärungsansätze formuliert: Der eine (sog. *confirmation*-Hypothese) geht davon aus, dass die morpho-syntaktische Übereinstimmung von Pronomen und Antezedens zu einer Stabilisierung des aktuellen Diskursfokus führt. Die andere (sog. *disengagement*-Hypothese) behauptet hingegen, dass Gender-Informationen zu einer distinkteren Repräsentation der Aktanten führen, weshalb die Diskursmodellbildung erleichtert ist. Es wurde unter Rückgriff auf den *bonding-resolution*-Ansatz von Garrod und Sanford (1989) festgestellt, dass sich beide Erklärungsansätze nicht wechselseitig ausschließen (müssen), sondern vielmehr gegenseitig ergänzen (können): ic-kongruente Fortsetzungen sind leichter in das Diskursmodell integrierbar, *weil* sie bereits durch die Struktur der IC-Verben präsupponiert und damit diskursiv hervorgehoben sind. Und Aktanten, die durch eine eindeutige Gender-Zuordnung ausgewiesen sind, lassen sich eindeutiger mittels gedächtnis-basierter Prozesse identifizieren, *weshalb* sie sich später besser in das Diskursmodell integrieren lassen.

In Kap. 3 wurde die Methode der RT-Verteilungsanalyse vorgestellt, mit der sich Hypothesen über die komplexe Interdependenz-Beziehung gedächtnis- und erklärungs-basierter

Prozesse bei der pronominalen Auflösung empirisch überprüfen lassen. Denn diese ermöglicht u.U. durch die separate Betrachtung verschiedener Aspekte der *Verteilung* von Reaktionszeiten eine unabhängige Erfassung der beiden theoretisch relevanten Prozessklassen sowie ihrer dynamischen Interaktion während der Satzverarbeitung. Speziell wurde dafür argumentiert, dass die Weibull-Verteilung ein adäquates RT-Modell ist, das sich zudem auch unter Bedingungen schätzen lässt, die für eine sprachpsychologische Untersuchung auf der Satzebene realistischerweise anzusetzen sind. Dies wurde in zwei Simulationsstudien überprüft (vgl. Kap. 4). Ein weiterer Vorteil der Weibull-Verteilung ist, dass sich ihre Parameter heuristisch als Indikatoren verschiedener kognitiver Prozesse interpretieren lassen: Der shift-Parameter bezieht sich auf Zeit, die minimal zur Ausführung eines Prozesses benötigt wird, und ist damit ein Maß für die hierarchie-niedrige, gedächtnis-basierte kognitive Verarbeitung. Der scale- und der shape-Parameter sind hingegen Maße der Form und Skalierung der RT-Verteilung und damit indikativ für die Beteiligung hierarchie-hoher, erklärungs-basierter Verarbeitungsprozesse.

Im empirischen Teil der Arbeit wurden zwei Experimentserien vorgestellt, die differenzierte Hypothesen über den interaktiv-parallelen Einfluss von IC- und Gender-Informationen auf die pronominalen Auflösung überprüften. In beiden Experimentserien lasen die Vpn kausale Nebensatzgefüge, die ein ambiges Pronomen enthielten, welches sich entweder auf das Subjekt oder das Objekt des Satzes beziehen konnte. Variiert wurde einerseits, ob der Referent des Pronomens eindeutig durch morpho-syntaktische Gender-Informationen identifizierbar war (Gender-Cue), und andererseits, ob die pronominalen Auflösung mit dem impliziten kausalen Bias des Verbs übereinstimmte (IC-Cue). Beide Manipulationen wurden vorher mittels zweier Voruntersuchungen auf ihre Validität überprüft (Kap. 5). Die Experimente Ia (Kap. 6) und Ib (Kap. 7) bedienten sich einer pronominalen Auflösungsaufgabe, die als Indikator für den globalen Verarbeitungsaufwand interpretiert werden kann. Die Vpn lasen die Sätze als Ganzes und sollten so schnell wie möglich per Tastendruck angeben, welcher Hauptsatz-Aktant mit dem Pronomen im Nebensatz gemeint war. Anschließend bearbeiteten sie noch eine zusätzliche Verifikationsaufgabe, in der eine pronominal aufgelöste Variante des Nebensatzes vorgegeben wurde und auf ihre Richtigkeit beurteilt werden musste. In den Experimenten IIa (Kap. 8) und IIb (Kap. 9) wurde hingegen eine *self-paced-reading*-Aufgabe verwendet. Die Vpn lasen die Sätze Wort für Wort und bearbeiteten abschließend eine Verifikationsaufgabe. Durch die Erfassung des lokalen Verarbeitungsaufwandes über den ganzen Satzverlauf hinweg ließen sich differenziertere Hypothesen über die temporale Dynamik des Auflösungsprozesses und die Interaktion der beteiligten Teilprozesse überprüfen.

Die Ergebnismuster der vier Experimente waren nicht immer einheitlich, aber eindeutig genug, um erste Rückschlüsse auf die zugrunde liegende Prozessarchitektur zuzulassen. Die Ergebnisse sollen im nächsten Abschnitt noch einmal kurz zusammengefasst und integrativ über die verschiedenen Experimente hinweg dargestellt werden.

10.1 Zusammenfassung und Integration der Ergebnisse aus Experiment I und II

10.1.1 Gender-Effekte

Ausgesprochen eindeutig fielen die Ergebnisse bezüglich der Effekte des Gender-Cues aus. Diese sind allerdings für die Fragestellung der Interaktion gedächtnis- und erklärungs-basierter Prozesse bei der pronominalen Auflösung theoretisch weniger interessant. In Experiment I konnte zunächst auf der deskriptiven Ebene der mittleren Auflösungszeiten klar der Befund bestätigt werden, dass gender-disambiguierte Pronomen einen klaren Auflösungsvorteil haben: Sie wurden sehr viel verlässlicher und schneller aufgelöst als Pronomen, für die zwei gleichgeschlechtliche Referenten zur Verfügung standen. Ebenso verlässlich konnte in Experiment I durch die RT-Verteilungsanalyse nachgewiesen werden, dass dieser generelle Verarbeitungsvorteil sowohl auf eine Veränderung der shift- als auch der scale-Komponente zurückzuführen war. Unter Rückgriff auf die in Kap. 3 vorgestellte Interpretation dieser beiden Reaktionszeitkomponenten als Indikatoren hierarchie-hoher und -niedriger Verarbeitungsprozesse, lässt sich also feststellen, dass für den Gender-Erleichterungseffekt sowohl eine bessere Identifizierbarkeit durch gedächtnis-basierte Fokus- und *feature-matching*-Prozesse als auch eine schnellere Integrierbarkeit in das aktuelle Diskursmodell verantwortlich sind. Für eine prinzipiell bessere Diskursverankerung gender-disambiguierten Aktanten spricht ebenfalls, dass sich der Gender-Effekt auch in der auf die eigentliche Auflösung folgende Verifikationsaufgabe zeigte, wo er ausschließlich im scale-Parameter verortet war. Da die eigentliche Auflösung zu diesem Zeitpunkt bereits stattgefunden hatte, ist anzunehmen, dass die Verifikationsaufgabe primär Prozesse der Verstehenskonsolidierung abbildete, bei der die bestehende Diskursrepräsentation noch einmal auf ihre Kohärenz geprüft wurde bzw. – im Falle einer Fehlaufklärung – restrukturiert werden musste. Solche Prozesse sind immanent erklärungs-basiert und sollten sich deswegen in den Verteilungskomponenten widerspiegeln, welche hierarchie-hohe Prozesse abbilden.

In Experiment II konnten die beiden unterschiedlichen Effektkomponenten zusätzlich noch zeitlich voneinander dissoziiert werden: Der Gender-Cue beeinflusste die Satzverarbeitung bereits sehr früh, kurz nach der Verarbeitung des Pronomens in der Präpositionalphrase, sein Einfluss steigerte sich kontinuierlich. Gleichzeitig fand sich ein starker Effekt des Cues

im *sentence-wrap-up*. Die nachfolgende RT-Verteilungsanalyse bestätigte, dass die frühen Effekte primär vom *shift*-, der späte Effekt jedoch allein vom *scale*-Parameter getragen wurde. Dieses Muster war in Experiment IIa noch schwach ausgeprägt, in Experiment IIb jedoch deutlich, wenn auch zeitlich ein wenig verzögert (s.u.). Der späte *scale*-Effekt persistierte in beiden Experimenten ebenfalls in der abschließenden Verifikationsaufgabe.

Zusammengenommen liegt damit vergleichsweise starke Evidenz dafür vor, dass der Effekt des Gender-Cues zwei unabhängige Teilkomponenten umfasst, nämlich eine frühe *confirmation*- und eine späte *disengagement*-Komponente (vgl. Rigalleau et al., 2004). Der *confirmation*-Effektanteil beeinflusst die Satzverarbeitung früh und wird hauptsächlich von hierarchie-niedrigen Prozessen getragen, die sich im *shift*-Parameter der Reaktionszeitverteilung niederschlagen. Insbesondere seine Verlaufsdynamik mit ihrer kontinuierlichen Steigerung weist offensichtliche Parallelen zu dem von den gedächtnis-basierten Sprachverarbeitungstheorien beschriebenen Distraktor- oder Fan-Effekt auf (vgl. z.B. Myers & O'Brien, 1998): Durch die Tatsache, dass beide Hauptsatz-Aktanten das gleiche Gender-Merkmal haben, kann der korrespondierende Gender-Cue im Resonanz-Prozess keinen der beiden Referenten spezifisch ansprechen und reaktivieren. Die Abruf- und Stabilisierungsdynamik verzögert sich dadurch, was sich in einem kontinuierlich steigenden Effektverlauf niederschlägt. Dieser frühe Reaktivierungsprozess des Antezedens durch einen gedächtnis-basierten Resonanzmechanismus lässt sich als Teil des initialen *bondings* im Rahmen des *bonding-resolution*-Ansatzes von Garrod und Sanford rekonstruieren (vgl. McKoon & Ratcliff, 1989, für eine sehr ähnliche Einteilung in frühe *recovery*- und späte Integrationsprozesse).

Der *disengagement*-Effektanteil findet sich hingegen ausschließlich im späten *sentence-wrap-up* und wird allein vom *scale*-Parameter getragen. Beides zusammen spricht dafür, dass es sich bei diesem Teilprozess primär um einen Integrations-Effekt handelt. Nach Just und Carpenter (1980) sind Satzgrenzen der Ort, an dem (teil-)satzübergreifende Integrationsprozesse stattfinden. Unter anderem werden dort noch nicht zugewiesene anaphorische Relationen aufgelöst und eine kohärente Satzrepräsentation hergestellt. Mit Garnham et al. (1992) kann davon ausgegangen werden, dass der Gender-Cue solche Zuweisungsprozesse beschleunigt, weil Gender-Informationen die beteiligten Aktanten distinkter machen. Es lässt sich deshalb leichter entscheiden, wer was in dem zugrunde liegenden Szenario getan hat. Die kausale Ereignisstruktur, die durch das Kausalkonnektiv verlangt wird, lässt sich in diesem Fall leichter rekonstruieren. Im Rahmen des *bonding-resolution*-Ansatzes von Garrod und Sanford bedeutet dies, dass Gender-Informationen die satzfinale Auflösung des Pronomens erleichtern,

weil sie die Zuordnung der verschiedenen Aktanten zu den *slots* des zugrunde liegenden Szenarios erleichtern und dessen abschließende Evaluation beschleunigen.

Durch die Kombination dieser Befunde kann m.E. recht sicher davon ausgegangen werden, dass Gender-Informationen sowohl die frühe *bonding*- als auch die finale *resolution*-Phase beschleunigen. Allerdings lässt sich auf der Grundlage der vorliegenden Daten nur wenig über die genaue Relation der beiden Teilprozesse zueinander sagen, außer dass sie beide gleichzeitig, aber unabhängig voneinander den Auflösungsprozess beeinflussen. Um die weitergehende Hypothese zu prüfen, dass beide Teilprozesse in einer dynamischen Interaktionsbeziehung stehen, wäre ein anderes Design notwendig. In diesem müsste nicht nur manipuliert werden, ob das Pronomen gender-ambig ist oder nicht, sondern auch, ob überhaupt ein gender-kongruenter Referent zur Auflösung zur Verfügung steht (vgl. z.B. Rigalleau et al., 2004). Das heißt, es müssten Sätze wie

(1) Peter beschuldigte Hans, weil sie die Situation nicht mehr aushielt.

untersucht werden, bei deren das Pronomen mit keinem der beiden genannten möglichen Antezedenten den Gender-Cue teilt. Bei solchen Sätzen ist es notwendig, einen weiteren Aktanten (z.B. *Maria*) in das Diskursuniversum einzuführen, was jedoch ohne zusätzliche referenzielle Aktivität kaum möglich ist. In Analogie zur noch zu besprechenden Trade-off-Relation der beiden Teilprozesse des IC-Cues ist zu vermuten, dass dieser Satz immer noch leichter zu verarbeiten wäre als ein Satz wie

(2) Peter_i beschuldigte Hans_j, weil er_k die Situation nicht mehr aushielt.

Bei diesem bezieht sich das Pronomen ebenfalls auf keinen der beiden Hauptsatzaktanten (sondern z.B. auf eine nicht spezifizierte, dritte Person). Dies wird in Satz (2) jedoch nicht bereits früh durch den Gender-Cue signalisiert. Deshalb können auch keine frühen Reanalyse-Prozesse stattfinden und kein alternatives Diskursmodell bereits online aufgebaut werden. Die satzfinale Integration ist deshalb ganz erheblich erschwert.

Genau aus diesem Grund wäre ein solches Vorgehen auch problematisch: Schon der gender-disambiguierte Satz (1) ist ohne weiteren Diskurskontext pragmatisch äußerst fragwürdig. Der korrespondierende Satz (2) ist sogar schlicht inakzeptabel: Wie in Kap. 1 dargestellt, werden Pronomen dazu verwendet, um sich auf nahe und gut verfügbare Diskursentitäten zu beziehen. Ein Pronomen zu verwenden, welches sich auf eine nicht verfügbare Person bezieht, obwohl zwei plausible Referenten im aktuellen Diskurskontext vorhanden sind, ist eine Verletzung der Kooperationsmaxime der sprachlichen Kommunikation (vgl. Grice, 1975). Sie kann ohne eine plausibilisierende Diskurseinbettung nicht normal verarbeitet werden (mit entsprechender Einbettung allerdings sehr wohl, vgl. McKoon et al., 1996).

10.1.2 IC-Effekte

Die Effekte des IC-Cues bedürfen einer etwas ausführlicheren Erläuterung. Einerseits, weil bei ihnen die Effektstruktur von der Ausprägung des Gender-Cues abhängig ist, andererseits aber auch, weil bei ihnen die Befundlage nicht ganz so eindeutig ist wie für den Gender-Cue. Zunächst einmal konnte in allen vier Experimenten deskriptiv nachgewiesen werden, dass IC-Informationen einen bedeutsamen Einfluss auf die pronominale Auflösung haben: Experiment I ergab zum einen, dass in ic-inkongruenten Bedingungen mehr Auflösungs-„fehler“ begangen wurden. Das heißt, es wurden mehr Auflösungen vorgenommen, die mit dem impliziten Bias des Verbs übereinstimmten und nicht mit der pragmatischen Begründung des Nebensatzes. Anscheinend wurde der Nebensatz in diesen Fällen einfach in Richtung des Verb-Bias „uminterpretiert“. Zum anderen brauchten ic-inkongruente Sätze länger, um aufgelöst zu werden. Dieser Effekt war ca. 250 ms groß, was ungefähr 30 % des Gender-Effekts entspricht. Sowohl seine absolute Größe als auch die Relation zwischen Gender- und IC-Effekt entsprachen vollkommen den in der Literatur für Auflösungsaufgaben berichteten Werten (Caramazza et al., 1977, Vonk, 1984).

Ähnlich wie für den komplexen Effekt des Gender-Cues scheint auch für den IC-Effekt eine weitergehende Differenzierung in verschiedene Teileffekte sinnvoll zu sein, nämlich in einen frühen, hierarchie-niedrigen IC-Fokus-Effekt und einen späten, hierarchie-hohen IC-Integrations-Effekt. Sehr eindeutig waren die Ergebnisse in den Experimenten I und II für den IC-Integrations-Effekt. Dafür, dass ic-inkompatible Informationen schlechter in das aktuelle Diskursmodell integrierbar sind, sprach v.a. die RT-Verteilungsanalyse, die für fast alle Experimente einen IC-Effekt im scale-Parameter der Reaktionszeitverteilung nachwies. Lediglich Experiment Ib stellte eine Ausnahme dar, auf die weiter unten noch eingegangen wird. Darüber hinaus konnte in Experiment II nachgewiesen werden, dass die scale-Effekte v.a. den *sentence-wrap-up* betrafen. Dieser Befund steht im Einklang mit der Integrations-Hypothese von Garnham et al. (1996) und Stewart et al. (2000), da satzfinale Effekte auf eine erschwerte Integration von Haupt- und Nebensatz hinweisen. Unter Rückgriff auf die ereignisstrukturelle Rekonstruktion des IC-Effekts in Kap. 2.1.2 kann der Effekt dahingehend interpretiert werden, dass zur Auflösung des Pronomens Informationen über Abfolge und Relation der verschiedenen Teilereignisse der Ereignisstruktur notwendig sind. Der Kausalkonnektor spezifiziert, dass der im Nebensatz genannte Aktant an einem dem Fokalereignis vorangegangenen Ereignis (ursächlich) beteiligt war. Wenn die Auflösung mit dem IC-Bias des Verbs übereinstimmt, dann können die im Satz genannten Aktanten und ihre Eigenschaften bzw. Aktivitäten direkt den *slots* der durch das Hauptsatzverb aufgespannten Ereignisstruktur zugewiesen

werden. Wenn dieses Ereignis hingegen nicht in der Ereignisstruktur enthalten bzw. nicht bereits voraktiviert ist, dann wird eine komplexe Zusatzinferenz notwendig, bei der das zusätzliche Ereignis in die Ereignisstruktur eingefügt wird. Dieser Prozess ist inhärent erklärungs-basiert und erfordert höhere kognitive Prozesse, z.B. (quasi-)syllogistische Schlüsse, wie sie von Hobbs (1978) angenommen wurden. Im Rahmen des SMF-Ansatzes von Sanford und Garrod entspricht dies dem sog. *secondary processing*, bei dem die Aktanten explizit den verschiedenen Diskursrollen des Skripts oder Szenarios zugeordnet werden (*role-to-role-mapping* sensu Garnham et al., 1992). Weiterhin ist anzunehmen, dass solche Zuordnungsprozesse bei der finalen Auflösung des Pronomens vorgenommen werden, also in der *resolution*-Phase, in der die vorher etablierten *bonds* hinsichtlich ihrer Passung zum aktuellen Diskursmodell evaluiert werden.

Ein weiterer Hinweis darauf, dass die IC-Kompatibilität der Satzfortführung Aufbau und Aufrechterhaltung des Diskursmodells beeinflusst, ist, dass sich in allen Experimenten persistierende Integrations-Effekte auch in den Verifikationsaufgaben fanden. Da diese auf die eigentliche Auflösung folgten, kann davon ausgegangen werden, dass sie primär Konsolidierungsprozesse abbildeten (s.o.). Entsprechend wurden auch hier (wie beim *Gender-disengagement*-Effekt) alle Effekte allein vom scale-Parameter getragen. Nur für Experiment IIb ergab sich ein abweichendes Muster, hier lag der Effekt im shift-Parameter. Dieses Ergebnis war unerwartet und lässt sich theoretisch nur schwer begründen. Allerdings ist eine nahe liegende, primär methodisch motivierte Alternativerklärung, dass der Effekt auf Spezifika der kumulativen *moving-window*-Präsentation zurückgeführt werden kann. Diese tendiert zur Betonung satzfinaler Integrationsprozesse (vgl. Kap. 9), die in diesem Fall gerade bei den ic-inkompatiblen Fortsetzungen ganz besonders gründlich durchgeführt wurden. Evtl. waren deswegen hierarchie-niedrige Prüfprozesse ausreichend, um die Aussage zu verifizieren.

In Bezug auf den IC-Integrations-Effekt muss jedoch noch eine weitere methodische Einschränkung angesprochen werden. Obwohl bei der Konstruktion der Stimulussätze große Sorgfalt darauf verwendet wurde, dass sich die ic-kongruenten und -inkongruenten Fortsetzungen in ihrer Plausibilität und Inferenzweite nicht voneinander unterscheiden, hatte eine Voruntersuchung ergeben, dass inkompatible Fortsetzungen unplausibler eingeschätzt wurden als kompatible (vgl. Kap. 5). Die vorgelegten Experimente sind deswegen angreifbar, u.a. durch die Argumentation von Garnham (2001), dass sich der IC-Effekt allein auf Unterschiede im Stimulusmaterial zurückführen lässt. Solche Vorwürfe können an dieser Stelle nicht vollkommen ausgeräumt werden, allerdings kann darauf hingewiesen werden, dass eine vollständige Rückführung des IC-Integrations-Effekts auf Plausibilitätsunterschiede im Stimu-

lusmaterial im vorliegenden Fall eher unwahrscheinlich ist. Zum einen war der Plausibilitätsunterschied ausgesprochen klein und kaum hinreichend, um den sehr stabilen IC-Effekt allein verursachen zu können. Darüber hinaus wurden die ic-inkongruenten Fortsetzungen zwar geringfügig unplausibler eingeschätzt als die kongruenten Fortsetzungen, aber nicht als *unplausibel* im absoluten Sinne. Eine Zusatzanalyse hatte ergeben, dass sich die Fortsetzungen hinsichtlich dieses Kriteriums nicht voneinander unterschieden. Die vorliegende Studie leidet hier sicherlich darunter, dass sie im Vergleich zu anderen Untersuchungen eine wesentlich größere Itemmenge untersuchte. Es konnten deswegen nicht nur selektiv IC-Verben ausgesucht werden, zu denen sich vollkommen gleich plausible Fortsetzungen konstruieren lassen. Zum anderen – und das ist der gewichtigere Grund – sind gewisse Plausibilitätsunterschiede zwischen den ic-kongruenten und -inkongruenten Fortsetzungen gewissermaßen sachimmanent: ic-inkongruente Fortsetzungen *sind* nicht ähnlich plausibel wie kongruente. Denn wie oben und in Kap. 2 skizziert, ist bei ic-inkongruenten Fortsetzungen eine zusätzliche Ereignisinferenz notwendig, die unabhängig motiviert sein muss. Ohne unterstützenden Diskurskontext ist es jedoch ausgesprochen schwer, eine solche unabhängige Motivation für das Zusatzereignis zu gewährleisten ohne gleichzeitig die kongruente Auflösungsrichtung zu schwächen. Bei einem ic-inkongruenten Satz wie

(3) Peter verdächtigte Hans, weil er misstrauisch geworden war.

sind Handlungen von *Hans* präsupponiert, welche *Peters* Misstrauen rechtfertigen bzw. begründen. Diese Handlungen konstituieren das zusätzliche, distale Ereignis, welches noch in die Ereignisstruktur integriert werden muss. Bei dem komplementären, ic-kongruenten Satz

(4) Peter verdächtigte Hans, weil er etwas zu verbergen versuchte.

wird hingegen der Grund für *Peters* Misstrauen direkt benannt. Da damit alle relevanten Ereignisse im expliziten Satzkontext selbst genannt werden, ist keine unabhängige Inferenz (samt ihrer Motivierung) notwendig. Ein solcher Satz erscheint zwangsläufig plausibler. Werden *Hans* Misstrauen evozierende Handlungen ebenfalls explizit im Diskurskontext erwähnt wie in (5)

(5) Seit einiger Zeit fehlte immer wieder Geld in der Kasse. Peter verdächtigte Hans, weil er ...,

dann ergeben sich kaum Unterschiede in der Plausibilität der beiden Sätze. Hieran lässt sich ersehen, dass die ereignisstrukturelle Ausformung der mit IC-Verben verbundenen Diskursrepräsentation durch Kontextinformationen moduliert werden kann (vgl. Kap. 2).

Im Gegensatz zum IC-Integrationseffekt ist der IC-Fokus-Effekt labil. Das liegt auch daran, dass er erwartungsgemäß nur in den gender-disambiguierten Bedingung auftrat und

damit lediglich auf der Ebene der einfachen Haupteffekte nachgewiesen werden kann. Darüber hinaus zeigte er sich nicht in allen Experimenten (was jedoch theoretisch begründbar ist, s.u.). Insgesamt kann bei der systematischen Zusammenschau der verschiedenen Experimente mit einiger Sicherheit davon ausgegangen werden, dass ic-inkompatible Fortsetzungen zu einem frühen IC-Fokus-Effekt führen. Eine Voraussetzung für sein Auftreten ist allerdings, dass der Satz gleichzeitig gender-disambiguiert ist und bestimmte diskursive Voraussetzungen erfüllt sind (vgl. Abschnitt 10.1.3). Evidenz für den Fokus-Effekt ergab die RT-Verteilungsanalyse in Experiment Ib, weil der IC-Effekt in der disambiguierten Bedingung nicht im scale-, sondern im shift-Parameter verortet war. Aber auch in Experiment II gab es Hinweise auf den IC-Fokus-Effekt: Zunächst einmal dadurch, dass sich auf der Ebene der mittleren Lesezeiten konsistent IC-Effekte in der Nähe des Pronomens finden ließen, wenn der Satz gender-disambiguiert war. In Experiment IIa betraf dies die Pronomen-Region selbst, in Experiment IIb war der Effekt zeitlich ein wenig verschoben und fand sich erst in der nachfolgenden Präpositionalphrase. Es muss hervorgehoben werden, dass dieser Effekt (im Gegensatz zum IC-Integrationseffekt, bei dem das nicht ganz ausgeschlossen ist) nicht auf Unterschiede im Stimulusmaterial zurückgeführt werden kann, weil dieses bis zur finalen Verbalphrase in allen Itemversionen identisch war. Dieser Befund bestätigt ähnliche Ergebnisse von Koornneef und van Berkum (2006), die ebenfalls frühe Effekte des IC-Cues in der Pronomen-Region finden konnten, wenn der Satz gender-disambiguiert war. Im Vergleich zur Untersuchung von Koornneef und van Berkum waren die Effekte in den mittleren Lesezeiten allerdings sehr viel schwächer. Während Koornneef und van Berkum mittlere Effekte von ca. 20 ms berichten, lagen sie in der vorliegenden Untersuchung im Bereich zwischen 5 und 15 ms. Dieser Umstand kann evtl. allein schon durch die Tatsache erklärt werden, dass in den vorgelegten Experimenten wesentlich mehr Stimulussätze verwendet wurden, wodurch der Effekt durch sprachliche Abweichungen zwischen den Items „verwaschen“ wurde. Eine plausible und v.a. theoretisch interessantere Erklärung ist jedoch, dass Koornneef und van Berkum mit einer Diskurseinbettung arbeiteten. Wie oben demonstriert, kann die Einführung zusätzlicher Diskursinformationen den Effekt u.U. entscheidend modulieren (s.u.). Das Fehlen eines einbettenden Diskurskontextes ist deswegen auch als einer der größten Nachteile der vorgelegten Experimente zu betrachten: Zum einen ist durch die Verarbeitung lediglich isolierter Sätze die externe Validität des Stimulusmaterials sehr eingeschränkt. Wir verarbeiten im Alltag Sätze nie für sich alleine, sondern rezipieren sie immer in einem größeren kontextuellen bzw. ko-textuellen Zusammenhang. Darüber hinaus führt die fehlende Diskurseinbettung offensichtlich auch zu einer Schwächung der empirisch beobachtbaren Effekte und

schränkt damit die Präzision der Untersuchung (als einen Aspekt ihrer internen Validität) ein. Dieser Punkt sollte deswegen unbedingt in nachfolgenden Untersuchungen aufgegriffen werden, evtl. in Verbindung mit einer systematischen Manipulation der Kontextinformationen.

Die zunächst eher kleinen Effektstärken für den frühen IC-Fokus-Effekt sind auch dafür verantwortlich, dass die RT-Verteilungsanalyse in Experiment IIa kaum reliable Unterschiede zwischen den einzelnen Reaktionszeitkomponenten aufdeckte. Erwartungsgemäß wurde der Fokus-Effekt hauptsächlich vom shift-Parameter getragen (und von einem starken shape-Effekt begleitet). Dies bestätigt den Befund aus Experiment Ib, dass für den IC-Fokus-Effekt hierarchie-niedrige Verarbeitungsprozesse relevant sind. Allerdings waren diese Unterschiede zu schwach, um sich inferenzstatistisch absichern zu lassen. Erst nachdem in Experiment IIb die frühen Effektkomponenten durch die direkte Verfügbarkeit des zurückliegenden Diskurskontextes gestärkt wurden, konnten frühe shift-Effekte reliabel nachgewiesen werden, und zwar sowohl in der Pronomen-Region als auch in der Präpositionalphrase.

Die prozessuale Fundierung des IC-Fokus-Effekts ist allerdings unklar. Die traditionelle Vorstellung zur Fokussierungs-Hypothese ist, dass IC-Verben den kausal verantwortlichen Aktanten salienter werden lassen. Er ist deswegen fokussiert und dadurch leichter auflösbar (vgl. z.B. McKoon et al., 1993). Wie aber bereits bei der Diskussion der korrespondierenden Salienz-Hypothese zum sozialpsychologischen IC-Effekt angesprochen (Edwards & Potters, 1993), ist eher davon auszugehen, dass die Salienz des kausalen Aktanten nicht Ursache, sondern vielmehr Folge der Rekonstruktion des (attributionalen) Kausalgefüges ist. In Anschluss an Koornneef und van Berkum neige ich zu der Auffassung, dass der kausal relevante Aktant nicht nur fokussiert, sondern aufgrund der verbspezifischen Ereignisstruktur sogar *erwartet* wird. Im Rahmen der ereignisstrukturellen Rekonstruktion des IC-Effekts bedeutet das, dass mit einem NP2-Verb wie *beschuldigen* ein implizites Ereignis präsupponiert ist, das eine Handlung des NP2-Aktanten spezifiziert, die kausal zur Beschuldigung führt. Der explizite Kausalkonnektor signalisiert, dass im Satz Informationen über die kausale Ursache des finalen Ereignisses folgen werden. Spätestens an dieser Stelle ist die Vorhersage möglich, dass sich das nachfolgende Pronomen auf den NP2-Aktanten beziehen wird, weil er der kausal verantwortliche Aktant ist. Im Rahmen des SMF-Ansatzes von Sanford und Garrod entspricht eine derartige, schnelle Zuweisung dem sog. *primary processing*, bei dem ein Aktant des expliziten Textfokus auf die entsprechende Rolle des Verb-Szenarios im impliziten Fokus gemappt wird. Dies geschieht im Rahmen des initialen *bondings* der pronominalen Auflösung. Der betreffende Aktant ist deswegen zwangsläufig auch salienter, gerade *weil* er eine prominente Rolle im zugrunde liegenden Kausalgefüge spielt, das durch das Verb vorgegeben wird.

10.1.3 Relation und Interaktion der verschiedenen IC-Teileffekte

Ähnlich wie beim Gender-Cue kann damit mit einiger Sicherheit festgehalten werden, dass am IC-Effekt sowohl frühe, gedächtnis-basierte als auch späte, erklärungs-basierte Verarbeitungsprozesse beteiligt sind. Diese Erkenntnis ist bereits ein wichtiger Schritt hin zu einer nicht-dichotomisierenden Sichtweise der Relation gedächtnis- und erklärungs-basierter Sprachverarbeitung. Allerdings macht das integrative Rahmenmodell der pronominalen Auflösung, wie es am Ende von Kap. 1 skizziert wurde, noch eine weitere, darüber hinausgehende Vorhersage: Die beiden Prozessklassen sollten nicht nur gleichzeitig, aber unabhängig voneinander die pronominale Auflösung beeinflussen, sondern im Laufe des Auflösungsprozesses dynamisch interagieren. Bei der theoretischen Erläuterung des IC-Fokus-Effekts ist deutlich geworden, dass gedächtnis- und erklärungs-basierte Verarbeitung gar nicht so unterschiedlich sind, wenn man einmal auf die Ebene der zugrunde liegenden Repräsentationen geht: Aktanten sind fokussierter, *weil* sie kausal relevanter sind.

Entsprechend wurde auch postuliert, dass die verschiedenen Teilprozesse des IC-Effekts dynamisch miteinander interagieren. Speziell sollte sich der späte, erklärungs-basierte IC-Integrations-Effekt reduzieren, wenn die Integration ic-inkompatibler Informationen mittels gedächtnis-basierter Prozesse bereits vorbereitet werden kann. Die relevante empirische Situation, in der sich diese Vorhersage überprüfen lässt, ist der Vergleich der verschiedenen IC-Effekte in Abhängigkeit von der Ausprägung des Gender-Cues.

In gender-ambigen Sätzen wird den obigen Ausführungen entsprechend der kausal saliente Aktant als potenzieller Referent für das Pronomen voraktiviert. Wenn dann eine Fortsetzung folgt, die mit dem Verb-Szenario übereinstimmt, dann ist die abschließende Integration vergleichsweise einfach. Wenn hingegen in der abschließenden Verbalphrase eine inkonsistente Fortsetzung folgt, dann hat gewissermaßen die falsche Voraktivierung stattgefunden. Der Auflösungsmechanismus muss hier also die ereignisstrukturelle Integration „aus dem Stand“ vornehmen, was entsprechend größere Verarbeitungsressourcen erfordert. Eventuell finden ähnlich wie bei einem *garden-path*-Satz zusätzlich noch weitere Reanalyse-Prozesse statt.

In gender-disambiguierten Sätzen verändert sich die Sachlage jedoch entscheidend: Die Inkongruenz der frühen, durch das Verb verursachten Voraktivierung wird direkt mit dem Pronomen sichtbar. Auch hier findet also eine frühe prädiktive Fokussierung auf den kausal salienten Aktanten statt, allerdings wird sofort bemerkt, dass die Vorhersage falsch war und die Diskursstruktur kann entsprechend modifiziert werden. Diese Umstrukturierung führt dann zu dem oben ausführlich beschriebenen IC-Fokus-Effekt.

An dieser Stelle wird allerdings klar, dass die Bezeichnung „IC-Fokus-Effekt“ irreführend oder zumindest missverständlich ist. Denn die eigentliche Fokussierung hat ja bereits vorher stattgefunden (und zwar in allen Bedingungen). Die Prozesse, die nach der Detektion der Inkonsistenz wirksam sind, sind hingegen nur ein Reflex auf diese frühe Fokussierung. Ich werde sie im Folgenden *IC-Refokussierungs-Prozesse* nennen und begrifflich vom eigentlichen, mit jedem IC-Verb verbundenen Fokus-Effekt abgrenzen. Man bemerke, dass der Fokus-Effekt selbst mit dem vorliegenden Design nicht direkt erfasst werden kann. Dies wäre nur dann möglich, wenn man IC-Verben mit normalen Verben kontrastieren würde. Dies hat jedoch meinem Wissen nach noch keine Studien realisiert. Stattdessen wird in den meisten Studien der IC-Refokussierungs-Effekt als ein *Indikator* für das Vorliegen des frühen IC-Fokus-Effekts benutzt (Stewart et al., 2000; Koornneef & van Berkum, 2006). Tab. 10.1 macht die Beziehungen der verschiedenen Teileffekte der beiden Cues in den einzelnen Untersuchungsbedingungen noch einmal deutlich.

	Cue-Effekte	Hauptsatz + weil	Früher Nebensatz	Später Nebensatz
Gender+IC+	Gender	---	---	---
	IC	Fokus (+)	---	---
Gender+IC-	Gender	---	---	---
	IC	Fokus(-)	Refokussierung (-)	Integration (-)
Gender-IC+	Gender	---	Confirmation (-)	Disengagement (-)
	IC	Fokus (+)	---	---
Gender-IC-	Gender	---	Confirmation (-)	Disengagement (-)
	IC	Fokus (-)	---	Integration (-)

Tab. 10.1: Verschiedene Teileffekte der beiden Cues in den vier Untersuchungsbedingungen.

Der entscheidende Punkt ist, dass man sich den frühen Refokussierungsprozess als einen gedächtnis-basierten Verarbeitungsprozess vorstellen kann, der zu einer Umfokussierung in der aktuellen Ereignisstruktur führt. Im Rahmen des resonanztheoretischen Modells wäre es sogar äußerst plausibel, dass es gerade das Gender-Merkmal ist, welches die Refokussierung steuert. Denn wenn das Pronomen verarbeitet wird, dann wirkt dieses Merkmal als Abruf-Cue und wird den einzigen gender-kongruenten Referenten im Diskurskontext entsprechend reaktivieren. Da der zwar nicht gender-, aber dafür ic-kongruente Referent ebenfalls durch die Verbkausalität voraktiviert ist, kommt es hier zu einer ähnlichen Interferenz wie beim *gender-confirmation*-Effekt, nämlich einer verlangsamten Abrufdynamik und einer schlechteren Konvergenz des Resonanz-Algorithmus. Deswegen sind hier mehr Verarbeitungsressourcen notwendig, die sich v.a. im shift-Parameter abbilden.

Eine bedeutende theoretische Konsequenz der frühen, gedächtnis-basierten Refokussierung ist nun, dass sie sich zwar zunächst einmal negativ auf die Satzverarbeitung auswirkt, sich jedoch später bei der satzfinalen Integration „auszahlen“ kann. Denn dadurch, dass der kausal verantwortliche Aktant bereits voraktiviert werden konnte, lässt sich die satzfinale Zusatzinferenz leichter etablieren und evaluieren. Entsprechend wurde vorhergesagt, dass der IC-Integrations-Effekt in der gender-disambiguierten Bedingung weniger stark ausfallen sollte. In den vier vorgelegten Experimenten fand sich durchaus Evidenz hierfür, allerdings sind auch hier die Befunde nicht vollkommen eindeutig. Für die geschilderte Trade-off-Relation zwischen dem IC-Refokussierungs- und dem IC-Integrations-Effekt spricht zunächst einmal, dass sich in Experiment Ib der IC-Effekt in der relevanten gender-disambiguierten Bedingung allein im shift-Parameter fand und nicht wie in der ambigen Bedingung im scale-Parameter. Dieser Befund ist deswegen besonders interessant, weil er darauf aufmerksam macht, dass sich zwar in beiden Bedingungen Effekte des IC-Cues finden ließen, diese aber evtl. auf sehr unterschiedlichen Verarbeitungsprozessen basierten: Während in der ambigen Bedingung der Integrations-Anteil überwog, war es in der disambiguierten Bedingung der Refokussierungs-Anteil. Dies fällt allerdings nicht auf, wenn man lediglich globale Verarbeitungsmaße erhebt und die prozessuale Fundierung der Effekte unberücksichtigt lässt. Dadurch lässt sich das alte „Rätsel“ der IC-Forschung erklären, warum es überhaupt zu einem IC-Effekt kommt, wenn das Pronomen bereits disambiguiert ist (vgl. Caramazza et al., 1977): Denn zumindest die ursprüngliche Intuition war ja, dass sich die satzfinale Integration wesentlich erleichtert, wenn man schon durch den Gender-Cue weiß, wer mit dem Pronomen gemeint ist. Diese Vermutung war zwar durchaus richtig, aber leider wurde übersehen, dass sich im Fall einer frühen Gender-Disambiguierung die mit ic-inkonsistenten Fortsetzungen verbundenen Verarbeitungsprobleme einfach an den Anfang der Satzverarbeitung verschoben werden: Statt einer aufwändigen, satzfinalen Integration sind also frühe Refokussierungsprozesse notwendig. Dieses Modell kann auch die Heterogenität der Befunde zum frühen IC-(Re)Fokussierungseffekt erklären: Alle Studien, welche eine frühe Verarbeitung von IC-Informationen nachweisen konnten, haben mit einer disambiguierten Bedingung gearbeitet.

Ein weiterer Hinweis auf die Trade-off-Relation des IC-Fokus- und des IC-Integrations-Effekts ist, dass es in allen Experimenten persistierende Integrationseffekte bei der Verifikationsaufgabe gab. Eine Ausnahme war jedoch Experiment IIb, auf dessen Besonderheit jedoch schon oben eingegangen wurde. Charakteristischer Weise traten die persistierenden Integrationseffekte nur in den ambigen Untersuchungsbedingungen auf, also dann, wenn die satzfinale Integration besonders schwierig ausfallen musste. In der bereits disambiguierten Bedingung

fanden sich hingegen keine oder nur sehr kleine Effekte. Dies spricht dafür, dass die Integration bei inkongruenten, aber disambiguierten Sätzen nicht nur leichter bzw. schneller vorgenommen werden kann, sondern auch dafür, dass die resultierende Diskursstruktur stabiler ist.

Weitere Hinweise auf eine Interaktion der beiden Prozesse gab es in Experiment II. In Experiment IIa war der scale-Effekt im *sentence-wrap-up* in der disambiguierten Bedingung nur sehr klein und wurde im Einzelvergleich auch nicht signifikant. In der ambigen Bedingung war er jedoch sehr stabil ausgeprägt. Der entsprechende Interaktionsterm überschritt allerdings nicht das geforderte Signifikanzniveau, und zwar weder bei den mittleren Lesezeiten, noch bei der Analyse der Verteilungsparameter. Ähnliches gilt für Experiment IIb, bei dem sich ebenfalls keine Unterschiede in der Größe der beiden Effekte absichern ließen. Über die Gründe für die fehlenden Interaktionseffekte lässt sich nur spekulieren. Eine eher theoretisch motivierte Begründung wäre z.B., dass in der disambiguierten Bedingung sogar eine besonders gründliche Integration und Evaluation der Informationen vorgenommen wurde, die Struktur also tiefer verarbeitet wurde als in der ambigen. Hierfür spricht, dass sich keine weiteren Effekte mehr in der nachfolgenden Verifikationsaufgabe finden ließen. Eine eher methodisch motivierte Erklärung ist hingegen, dass solche diffizilen Interaktionen mit der *moving-window*-Methode nicht nachweisbar sind, weil diese zu einer allgemeinen Aufblähung des *wrap-ups* neigt. Diese Erklärung ist insbesondere für Experiment IIb wahrscheinlich, bei dem die satzfinalen Effekte in der Tat sehr extrem ausfielen. Eine Replikation in einem anderen methodischen Rahmen wäre deshalb ausgesprochen wünschenswert.

Allerdings gibt es ja auch noch die umgekehrte Einflussrichtung, nämlich die von der erklärungs-basierten auf die gedächtnis-basierte Komponente. Diese ließ sich in den vorliegenden Untersuchungen nicht direkt prüfen, weil die Logik des IC-Paradigmas dies zunächst einmal nicht zulässt: Die oben skizzierte Abfolge der Effekte ist jeweils so, dass die hierarchie-niedrigen Teilprozesse den hierarchie-hohen vorangehen, weshalb sich Abhängigkeitsbeziehungen immer nur in eine Richtung prüfen lassen. Allerdings lässt sich eine vorläufige Beurteilung aufgrund der Beziehungen *zwischen* den verschiedenen Experimenten vornehmen.

Denn es ist auffällig, dass der IC-Refokussierungs-Effekt nicht in allen Experimenten nachgewiesen werden konnte. So trat er z.B. in Experiment Ia zunächst nicht auf, sondern erst, nachdem in Experiment Ib eine zusätzliche Präpositionalphrase in die Stimulussätze eingefügt worden war. Dies geschah mit der Absicht, satzintern noch weitere Kontextinformationen zur Verfügung zu stellen, welche die Refokussierung leiten können. Parallel hierzu konnte der IC-Refokussierungs-Effekt in Experiment IIa mit einer nicht-kumulativen Präsentati-

onsmethode nicht reliabel abgesichert werden, bei einer kumulativen Präsentation der Sätze wie in Experiment IIb jedoch schon. Bei dieser steht den Rezipienten/innen der vorhergehende Satzkontext direkt für eine Reanalyse zur Verfügung. Beide Befunde zusammen lassen die Schlussfolgerung zu, dass frühe gedächtnis-basierte Refokussierungsprozesse nur dann möglich sind (oder zumindest stärker ausfallen), wenn überhaupt verstehensrelevante Informationen vorhanden und zugänglich sind, die solche Prozesse initiieren und unterstützen. Auf der theoretischen Ebene gäbe hierfür es mehrere Begründungsmöglichkeiten. In einem eher szenario-orientierten Ansatz wie dem von Sanford und Garrod dienen zusätzliche Diskursinformationen dazu, das Szenario im impliziten Fokus weiter auszudifferenzieren. Dadurch sind dann u.U. auch andere *mappings* möglich, die bereits im Rahmen des *primary processing* etabliert werden können. Konkret kann durch das von der Präpositionalphrase zusätzlich zur Verfügung gestellte Material direkt eine alternative Lesart entwickelt werden, welche die inkongruente Zuordnung des Pronomens plausibilisiert und damit seine nachfolgende Integration erleichtert. Eine andere theoretische Begründung ist eher gedächtnis-basiert, kommt aber zu einem sehr ähnlichen Ergebnis: Hier stellt das zusätzliche Material der Präpositionalphrase weitere Abruf-Cues zur Verfügung, welche eine Reaktivierung des nicht-fokussierten Referenten erleichtern.

Allerdings ist die oben vertretene Auffassung, dass der IC-Refokussierungs-Effekt nur auftritt, wenn weitere, verarbeitungsleitende Informationen vorhanden sind, bei weitem nicht zwingend. Eine plausible Alternativerklärung, warum der Refokussierungs-Effekt nur mit der zusätzlichen Präpositionalphrase auftrat, ist z.B., dass sie sonst direkt von den erklärungs-basierten Prozessen überlagert bzw. maskiert wurde. In solchen Sätzen gab es also gar nicht die Notwendigkeit für eine frühe Refokussierung, weil die Auflösung ja direkt erfolgte. Darüber hinaus standen hier nicht die notwendigen zeitlichen und kapazitätsbezogenen Ressourcen zur Verfügung. Erst wenn mit der Präpositionalphrase eine gewisse Zone der interpretatorischen Unsicherheit etabliert wurde, wurde die frühe Refokussierung aktiv. In ähnlicher Weise lässt sich auch eine Alternativerklärung dafür finden, warum Refokussierungsprozesse bei der kumulativen Präsentationsweise stärker ausfallen: Da man hier den vorherigen Diskurskontext noch direkt „vor Augen“ hat, wirkt sich die Asymmetrie zwischen den schnellen sprachlichen Verstehensprozessen und der langsamen motorischen Komponente zum Aufdecken des nächsten Wortes noch extremer aus als bei der nicht-kumulativen Version des *moving-window*-Paradigmas. Die Vpn werden dadurch zu einer verstärkt inkrementellen Verarbeitung „gezwungen“, weil ihre Augen immer schon weiter sind als ihre Hand (vgl. Koornneef & van Berkum, 2006).

Zusammengefasst bedeutet das, dass sich die Annahme einer starken, dynamischen Wechselwirkungsbeziehung zwischen den beiden Prozessen theoretisch plausibel begründen lässt und empirisch durchaus gerechtfertigt erscheint. Allerdings sollte man bei der Interpretation der Ergebnisse auch nicht zu weit gehen, weil sie ausschließlich auf Zwischen-Experiment-Manipulationen basieren. Auch wenn sich die Veränderung des Ergebnismusters theoretisch schlüssig auf die Manipulation der experimentellen Anordnung zurückführen lässt, kann man nicht vollständig ausschließen, dass sie lediglich Zufälligkeiten bei der Stichprobenziehung geschuldet ist. Darüber hinaus stellen die vorgenommenen Veränderungen der Untersuchungsprozedur zwar aktive Manipulationen dar, allerdings kann nicht sichergestellt werden, dass sie nicht mit anderen relevanten Variablen konfundiert sind.

Es gäbe allerdings auch experimentelle Möglichkeiten, die oben skizzierte Einflusshypothese des vorangehenden Diskurskontextes auf die erklärungs-basierte Verarbeitung zu prüfen. So liegt es z.B. nahe, die in dieser Studie isoliert dargebotenen Sätze in einen weiteren Diskurskontext einzubinden, z.B. in Drei-Satz-Diskurse wie bei McKoon et al. (1993). Zusätzlich zur IC-Kongruenz könnte dann nämlich manipuliert werden, ob der bisherige Diskurskontext die ic-kongruente oder die ic-inkongruente Lesart des Satzes unterstützt (oder sich diesbezüglich neutral verhält). Hier wäre zu erwarten, dass Refokussierungsprozesse stärker ausfallen, wenn die Fortsetzung zwar inkongruent ist, der vorherige Diskurskontext die neue Lesart aber nahelegt.

10.2 Methodische Reflexion: Analyse von Reaktionszeitverteilungen

Mit der RT-Verteilungsanalyse wurde in der vorliegenden Arbeit ein zumindest für die sprachpsychologische Forschung auf der Satzebene eher ungewöhnliches Verfahren eingesetzt. Dies geschah vor dem Hintergrund, dass ein solches Verfahren in Bezug auf die spezielle Fragestellung der Interaktion gedächtnis- und erklärungs-basierter Verarbeitungsprozesse bei der pronominalen Auflösung die Möglichkeit verspricht, unabhängige Indikatoren für die beiden relevanten Prozessklassen zur Verfügung zu stellen. In Kap. 3 wurde die Weibull-Verteilung als ein theoretisch plausibles Reaktionszeitmodell eingeführt und das Hierarchical Bayes-Verfahren von Rouder et al. (2003) als ein für die vorliegende Untersuchung adäquates Schätzverfahren vorgestellt. Dabei wurde für eine Zwei-Prozess-Interpretation der resultierenden Verteilungsparameter plädiert, in deren Rahmen der shift-Parameter primär hierarchie-niedrige (gedächtnis-basierte) und der scale- bzw. der shape-Parameter hierarchie-hohe (erklärungs-basierte) Verarbeitungsprozesse abbildet. In Kap. 4 wurde das Verfahren im Hinblick auf die zu erwartende Datensituation in zwei Simulationsstudien auf seine Effizienz

geprüft und entsprechende Adaptionen vorgenommen. Jetzt, nachdem das Verfahren in vier Experimenten eingesetzt wurde, ist es an der Zeit, zurückzublicken, um die Möglichkeiten und Grenzen des Verfahrens zu beurteilen. Darüber hinaus ist es notwendig zu diskutieren, ob die berichteten Ergebnisse von bestimmten Spezifika des Verfahrens abhängig sind.

1.2.1 Stabilität des Verfahrens und Abhängigkeiten in den Parameterschätzungen

Selbstverständlich ist es an dieser Stelle nicht möglich, Aussagen über die Reliabilität des Verfahrens im engeren Sinne zu machen, da – im Gegensatz zur Simulationsstudie in Kap. 4 – nicht die „wahren“ Werte als Vergleichsmaßstab vorliegen. Allerdings lassen sich einige allgemeine Aussagen über das Verfahren der Parameterschätzung sowie die Stabilität der Parameter formulieren. Dabei ist zunächst einmal festzuhalten, dass sich das hierarchische Schätzverfahren von Rouder et al. (2003) (wie schon in der Simulationsstudie) durchaus bewährt hat: Alle MCMC-Ketten konvergierten nach einer angemessenen *burn-in*-Periode und wiesen ein moderates Maß an Autokorrelation auf. Auch die Aussage von Rouder et al., dass die Prior-Spezifizierung keinen übermächtigen Einfluss auf die Parameterschätzung hat, kann bestätigt werden: Viele der berichteten Ergebnisse wurden sicherheitshalber mit alternativen Prior-Werten wiederholt. Es ergaben sich weitgehend identische Ergebnisse.

Bei der Beurteilung der Modellgüte der resultierenden Weibull-Verteilung muss nach den verschiedenen Experimenten bzw. Aufgaben differenziert werden: Für die „langen“ Auflösungs- und Verifikationsaufgaben, die RT-Werte über einer Sekunde generieren, ergaben sich recht zufrieden stellende Fit-Werte ($p > .25$, Missfit-Quote von ca. 3.5 % bei den Auflösungsaufgaben, von ca. 7 % bei den Verifikationsaufgaben). Die Weibull-Verteilung scheint für diese Daten ein adäquates, wenn auch nicht überragendes Verteilungsmodell darzustellen. Der Fit für die wortweisen Lesezeiten fiel hingegen sehr viel schlechter aus ($p > .2$, Missfit-Quote von ca. 13 % bei Experiment IIa und ca. 20 % bei Experiment IIb). Allerdings muss man in Rechnung stellen, dass Lesezeiten durch ihre Varianzarmut im verwendeten χ^2 -Verfahren von Heathcote (2002) benachteiligt werden, weil sich bei ihnen weniger *bins* extrahieren lassen. Nichtsdestotrotz muss festgehalten werden, dass sich wortweise Lesezeiten von ihren Verteilungseigenschaften anscheinend sehr viel schlechter für eine RT-Verteilungsanalyse eignen. Dies macht sich nicht nur in einer verschlechterten Gesamtpassung der Verteilungskurve bemerkbar, sondern auch darin, dass die Parameterschätzungen in diesem Fall wesentlich inkonsistenter werden, weil schon kleine Abweichungen von der Verteilung zu unverhältnismäßigen Ausschlägen der entsprechenden Parameterschätzer führen.

Dies manifestiert sich u.a. in instabilen und inkonsistenten Schätzwerten, welche die Fehlervarianz dramatisch erhöhen.

Für die Reliabilität der Parameterschätzungen lassen sich wie erläutert keine genauen Angaben machen. Allerdings kann die Stabilität der Parameterschätzungen zwischen den einzelnen Untersuchungsbedingungen als eine Form der Retest-Reliabilität oder internen Homogenität approximativ beurteilt werden. Typischerweise gilt für (mittlere) Reaktionszeitdaten, dass sie enorm hoch (d.h. über $r = .90$) zwischen den Untersuchungsbedingungen korrelieren. Das heißt, die Niveauunterschiede zwischen den Personen sind ausgesprochen stabil. Das gilt auch für die vorliegenden Daten: So sind z.B. bei den gepoolten Daten für Experiment I ($N = 45$) die beiden aggregierten Gender-Bedingungen (*Gender+* vs. *Gender-*) zu $r = .90$ miteinander korreliert. Die korrespondierenden shift-Parameter sind mit $r = .90$ ähnlich reliabel, die scale-Parameter mit $r = .63$ schon wesentlich weniger. Die shape-Parameter korrelieren mit $r = .55$ auf einem ähnlich niedrigen Niveau. Für Experiment II gelten tendenziell vergleichbare Werte, allerdings ist der shape-Parameter mit Korrelationen um $r = .30$ noch wesentlich unrelabler. Insgesamt kann damit festgehalten werden, dass der shift-Parameter am reliabelsten (im Sinne von stabilsten) geschätzt wird, während die Schätzung des scale-Parameters weniger reliabel ist und der shape-Parameter sogar zur Unreliabilität tendiert. Dies erklärt auch, warum eigentlich in allen Experimenten der shift- und der scale-Parameter die deutlichsten und konsistentesten Effekte zeigten und sich auch theoretisch mit dem meisten Gehalt interpretieren ließen.

Die wahrscheinlich größten Probleme ergeben sich allerdings daraus, dass die Parameterschätzungen nicht unabhängig voneinander sind, sondern vielmehr stark miteinander korrelieren. Perfiderweise zeigt sich dieses Problem noch nicht auf der Ebene der Parameterschätzungen selbst (wo die Interkorrelation der Parameter vergleichsweise moderat ausfällt). Das Problem wird erst virulent, wenn man die Parameter der einzelnen Bedingungen voneinander abzieht und miteinander vergleicht. Dies ist aber genau das, was bei der varianzanalytischen Kontrastbildung getan wird. Wenn man also den shift-Parameter in Bedingung 1 von dem in Bedingung 2 abzieht und das gleiche mit den scale-Parametern der beiden Bedingungen macht, dann finden sich teilweise hohe Korrelationen zwischen den beiden Kontrasten. Die genaue Höhe variiert je nach gewähltem Kontrast und fällt generell für die Haupteffekte geringer aus als für die einfachen Haupteffekte. Tendenziell liegen jedoch die Korrelationen zwischen dem shift- und dem scale-Kontrast in Experiment I im Bereich um $r = -.50$, in Experiment II sogar um $r = -.80$. In Experiment II kommen ähnlich hohe Abhängigkeitsbeziehungen zwischen dem shift- und dem shape-Parameter (negativ) sowie dem scale- und dem sha-

pe-Parameter (positiv) hinzu. Dieses Muster ist kein Spezifikum der angefallenen empirischen Daten, sondern tritt tendenziell (wenn auch nicht so stark) auch in den Daten der in Kap. 4 beschriebenen Simulationsstudie auf. Bei den Lesezeiten aus Experiment II ist es jedoch ganz besonders stark ausgeprägt.

Das Problem abhängiger Parameterschätzungen wird zwar in der relevanten RT-Literatur nicht intensiv behandelt, ist aber prinzipiell sowohl für die Weibull- als auch für die Ex-Gauss-Verteilung bekannt (vgl. z.B. Cousineau et al., 2004). Dass es allerdings so massiv auftritt, ist überraschend. Das liegt auch daran, dass in der Literatur meist nur die Schätzwerte selbst miteinander verglichen werden und keine Kontraste. Die Abhängigkeitsbeziehung resultiert daraus, dass (wie in Kap. 3 dargestellt) sowohl bei der Weibull- als auch bei der Ex-Gauss-Verteilung eine additive Zerlegung des Erwartungswertes der Verteilung in zwei unabhängige Teilkomponenten vorgenommen wird. Für die Weibull-Verteilung gilt speziell:

$$(6) \quad E(X) = \Gamma\left(1 + \frac{1}{c}\right) \times \sigma + \theta.$$

Das heißt, der Erwartungswert besteht aus der Summe des shift-Parameters und des mit einer Funktion des shape-Parameters gewichteten scale-Parameters. (Diese Gewichtung ist im vorliegenden Fall so geringfügig, dass sie der Anschaulichkeit halber im Folgenden vernachlässigt werden kann). Mit einem kleinen Beispiel kann verdeutlicht werden, wie es nun zu den abhängigen Kontrasten kommt:

Bedingung 1: AM = 1000 → 500 (shift) + 500 (scale)

Bedingung 2: AM = 1000 → 400 (shift) + 600 (scale)

Differenz: 0 = 100 + -100

Nehmen wir an, wir hätten zwei Untersuchungsbedingungen mit dem gleichen Mittelwert von 1000. Bedingung 1 wird in eine shift-Komponente von 500 und eine scale-Komponente von ebenfalls 500 unterteilt. Bei Bedingung 2 beträgt hingegen der shift-Wert 400 und der korrespondierende scale-Wert 600. Wenn man nun die Differenz der einzelnen Parameter in den einzelnen Bedingungen bildet, dann beträgt sie für den Mittelwert 0, für den shift-Wert 100 und für den scale-Wert -100. Shift- und scale-Differenz sind also genau gegenläufig, weil sie der Restriktion unterliegen, zusammen die Mittelwertsdifferenz zu ergeben. Egal wie man die Zahlen im obigen Beispiel dreht, es wird immer eine gegenläufige Beziehung zwischen den Parametern bestehen, die sich in einer massiven negativen Korrelation der Parameter niederschlägt. Diese hat jedoch nichts mit der Struktur der Daten an sich zu tun, sondern ist einzig

und allein ein methodisches Artefakt, welches sich aus der Tatsache ergibt, dass die beiden Komponenten additiv miteinander verknüpft sind.

Diese Erkenntnis ist deshalb deprimierend, weil sie klarmacht, dass die RT-Verteilungsanalyse wohl nicht immer zu vollkommen unabhängigen Parametern für verschiedene kognitive Prozesse gelangt. Allerdings muss es nicht immer und auch nicht zwangsläufig zu einer solch starken Korrelation der Kontraste kommen. Die Bedingungen, unter denen sie auftritt, sind jedoch leider bislang noch unerforscht, zumindest ist mir keine Literatur dazu bekannt. Speziell für die Weibull-Verteilung ist davon auszugehen, dass die Größe der Kontrast-Korrelation einerseits von der Größe der Differenz der Bedingungen abhängig ist, andererseits aber auch von sehr speziellen Parameterkonstellationen. Es ist zu vermuten, dass das Problem der abhängigen Parameterschätzungen eine Konsequenz des sog. *unbounded likelihood problem* ist (vgl. Appendix A.4), dass gerade für die Schätzung der Weibull-Verteilung mit einem shift-Parameter seit Langem diskutiert wird (vgl. Cheng & Illes, 1990; Cheng & Traylor, 1995; vgl. auch Rouder et al., 2005).

Wichtiger als eine derart generelle Fragestellung ist jedoch, dass die gegenläufige Parameter-Interkorrelation zur Beurteilung der berichteten Ergebnisse relevant ist. Dies wird sofort deutlich, wenn man sich z.B. die Effekte für den shift- und den scale-Parameter betrachtet, wie sie sich in Experiment IIa in der gender-disambiguierten Bedingung ergeben haben (vgl. Abb. 10.1).

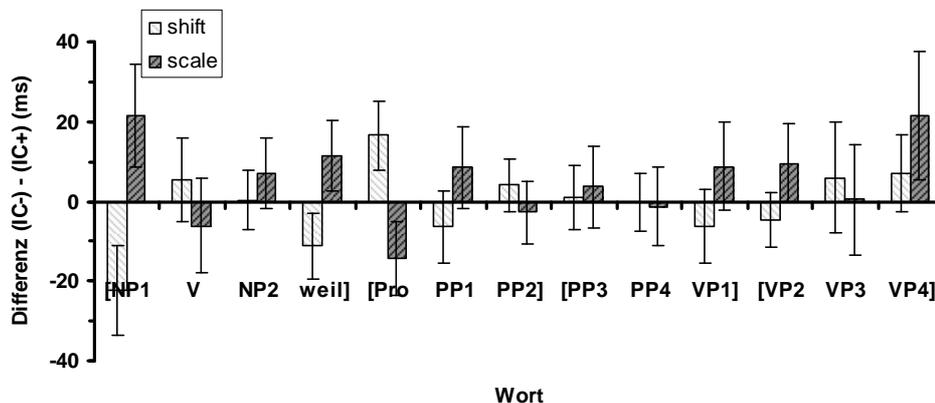


Abb. 10.1: Effekte des IC-Cues auf den shift- und den scale-Parameter in der Bedingung Gender +.

Es ist klar zu erkennen, dass die Effekte des IC-Cues auf die beiden Parameter klar gegenläufig ausgeprägt sind (in der Tat korrelieren sie zu $r = -.70$ miteinander). Das heißt, immer dort, wo es einen Positiv-Effekt für den shift-Parameter gibt, wird es zwangsläufig auch einen komplementären Negativ-Effekt für den scale-Parameter geben (und umgekehrt). Beide zusammen addieren sich zum (meist sehr viel geringeren) Unterschied in der mittleren Lesezeit

auf. Man steht nun vor der Aufgabe herauszubekommen, welche der Effekte „real“ und welche ein methodisches Artefakt sind.

Bei dieser Aufgabe hilft, sich zunächst einmal vor Augen zu führen, welche Effekte überhaupt von dem Problem der abhängigen Parameterschätzung betroffen sein können. Denn das sind lediglich Effekte, die in ein und derselben übergeordneten Untersuchungsbedingung erhoben wurden und nur auf einem Kontrast beruhen (also einfache Haupteffekte). Da die einzelnen Bedingungen komplett unabhängig voneinander geschätzt wurden, kann es zu keinen Abhängigkeitsbeziehungen auf der Ebene der Haupteffekte und auch nicht auf der Ebene der Interaktionen kommen (in der Tat korrelieren die Kontraste in der disambiguierten und der ambigen Bedingung nur marginal miteinander). Damit sind insbesondere alle Gender-Effekte und die differenziellen Effektverläufe des IC-Cues in den beiden Gender-Bedingungen vor jeder Verzerrung sicher. Wenn überhaupt, dann ist nur der IC-Refokussierungs-Effekt von dieser Problematik betroffen, weil er auf dem einfachen Kontrast des IC-Cues in der disambiguierten Gender-Bedingung basiert. Zweitens ist beruhigend festzustellen, dass die Effekte auch nur lokal abhängig sind, d.h. es gibt keine Möglichkeit, dass z.B. die Parameterschätzung des Pronomens die Schätzung des satzfinalen Wortes beeinflusst haben kann. Drittens sagt die Tatsache, dass der shift- und der scale-Kontrast negativ miteinander korreliert sind, noch nichts darüber aus, in welche Richtung die Kontraste gehen. Es ist dadurch also noch nicht determiniert, ob der shift-Effekt den Positiv-Ausschlag zeigt oder aber der scale-Effekt. Dies ergibt sich nach wie vor allein aus der Schätzung der relevanten Verteilungskomponenten, die damit weiter theoretisch gehaltvoll sind. Dies gilt insbesondere dann, wenn die Effekte konsistent auftreten, d.h. nicht auf einzelne Wörter beschränkt sind, sondern sich auch in den aggregierten Konstituenten finden lassen.

Wenn man alle diese Punkte zusammennimmt, dann erweist sich überhaupt nur ein einziger Effekt als inspektionswürdig, nämlich der frühe shift-Effekt, der auf der Pronomen-Region in Experiment IIB gefunden wurde (der vergleichbare Effekt auf der nachfolgenden Präpositionalkonstituente ist hingegen nicht betroffen). Bei diesem kommt es zu einem Positiv-Ausschlag des shift-Parameters von ca. 15 ms, während der scale-Parameter einen komplementären Negativ-Ausschlag von ca. 14 ms aufweist. Für diesen Effekt ist deswegen zu entscheiden, ob es sich bei ihm nun um einen frühen gedächtnis-basierten Verarbeitungsvorteil oder um einen erklärungs-basierten Verarbeitungsnachteil für ic-kongruente Fortsetzungen handelt. Meines Erachtens ist hier die Wahl sowohl aus theoretischen (größere Plausibilität eines IC-Refokussierungs-Prozesses) als auch statistischen Gründen (höhere Reliabilität des shift-Parameters) ziemlich eindeutig. M.E. handelt es sich um einen frühen shift-Effekt

und der komplementäre scale-Effekt ist ein methodischer Reflex der Parameter-Interkorrelation. Aber selbst wenn man diesen problematischen Effekt ganz aus der abschließenden Betrachtung ausklammert, kann die Argumentation vollständig aufrechterhalten werden.

Zusammen genommen bedeutet das, dass das Verfahren der RT-Verteilungsanalyse bei Weitem nicht so unproblematisch ist, wie dies nach den durchweg positiven Ergebnissen der in Kap. 4 dargestellten Simulationsstudie zu erwarten war. Einerseits gibt es methodeninhärente Probleme des Verfahrens, die sich insbesondere daraus ergeben, dass die Schätzung der Parameter der geshifteten Weibull-Verteilung ausgesprochen schwer ist (vgl. Appendix A.4). Andererseits liegt das aber auch an den Besonderheiten sprachpsychologischer Daten auf der Satzebene. Dies gilt bedingt zwar schon für die varianzreichen und durch komplexe Verarbeitungsprozesse verrauschten Daten der pronominalen Auflösungsaufgabe in Experiment I, macht sich jedoch am stärksten bei der Analyse der wortweisen Lesezeiten in Experiment II bemerkbar. Beide Probleme zusammen machen darauf aufmerksam, dass die Analyse von RT-Verteilung für komplexe sprachpsychologische Daten mit großer Vorsicht angewendet werden sollte, da sie u.U. zu verzerrten und ineffizienten Parameterschätzungen führt.

Speziell für die vorliegende Studie muss festgehalten werden, dass insbesondere die Ergebnisse der RT-Verteilungsanalyse der wortweisen Lesezeiten aus Experiment II nur sehr vorsichtig interpretiert werden sollten. Allerdings ist das Fehlerrisiko nur sehr gering, weil alle relevanten Ergebnisse nicht vom Problem abhängiger Parameterschätzungen betroffen sind. Darüber hinaus sind die Effekte meist durch korrespondierende und eindeutig interpretierbare Unterschiede auf der Ebene der mittleren Lesezeiten gedeckt. Die RT-Verteilungsanalyse führt hier also zu einer weiteren Qualifizierung, auf welche kognitiven Prozesse ein Effekt in der mittleren Lesezeit zurückgeführt werden kann. Im vorliegenden Fall ergab sich dabei immer das (theoretisch ausgesprochen plausible) Ergebnis, dass die frühen Effekte allein vom shift-Parameter getragen wurden, während sich die späten Effekte allein im scale-Parameter fanden. Die Argumentation könnte damit auch vollständig ohne Rückgriff auf die beiden Verteilungskomponenten allein über die früh-spät-Unterscheidung erfolgen, wie dies ja auch bei anderen Studien üblich ist (z.B. Koornneef & van Berkum, 2006). Dennoch wäre es wünschenswert, dass die shift-früh- bzw. scale-spät-Zuordnung noch einmal unabhängig mit einer Methodik repliziert würde, die weniger anfällig für die oben beschriebene Problematik ist. Anbieten würde sich hier v.a. ein *probe-recognition*-Verfahren, welches ähnlich wie wortweise Lesezeiten eine zeitliche Stratifizierung der Effekte erlaubt, aber weniger problematische Reaktionszeitverteilungen generieren dürfte.

1.2.2 Konstruktvalidität der Parameter

Dass es bei der Anwendung des Verfahrens u.U. zu schätztechnischen Problemen kommt, muss noch nicht zwangsläufig heißen, dass die mit ihm getroffenen Aussagen weniger valide sind. Im Gegenteil könnte man sogar behaupten, dass die Ergebnisse, die sich *trotz* der verfahrensinternen Probleme konsistent und reliabel abzeichnen, sogar ganz besonders stabil und damit aussagekräftig sind. In Kap. 3 wurde für die Parameter der Weibull-Verteilung in Anlehnung an Balota und Spieler (1996) sowie Rouder et al. (2003) eine Zwei-Prozess-Interpretation vorgeschlagen, in deren Rahmen die shift-Komponente der Verteilung primär hierarchie-niedrige kognitive Prozesse, die scale-/shape-Komponente der Verteilung hingegen primär hierarchie-hohe kognitive Prozesse abbildet. Eine sehr ähnliche Interpretation wird häufig für die Parameter der Ex-Gauss-Verteilung vorgeschlagen (vgl. z.B. Schwarz, 2001). Hinter dieser Zuordnung steht die allgemeine Beobachtung, dass periphere, stimulus-getriebene kognitive Prozesse meist zu kurzen, wenig varianzreichen und fast normal verteilten Reaktionszeitverteilungen führt, während zentrale kognitive Prozesse meist varianzreiche und stark rechtsschiefe Verteilungen generieren. Es wurde dafür argumentiert, dass die beiden Reaktionszeitkomponenten in der vorliegenden Untersuchung als Indikatoren für die Anteile an gedächtnis- und erklärungs-basierter Verarbeitung bei der pronominalen Auflösung fungieren können.

Wie hat sich die generelle und eher unspezifische Zwei-Prozess-Interpretation nun bewährt? Selbstverständlich lassen sich an dieser Stelle keine Aussagen über die Validität der Parameter formulieren, weil kein relevanter externer Vergleichsmaßstab für eine Beurteilung der Abbildungsqualität zur Verfügung steht. Allerdings ist eine Beurteilung der Konstruktvalidität der Parameter möglich, d.h. eine Einschätzung darüber, ob sie sich theoretisch plausibel in den bisherigen Kenntnisstand zur Satzverarbeitung und zur pronominalen Auflösung einbetten lassen. Hier muss man sagen, dass sich die Zwei-Prozess-Interpretation ausgesprochen gut bewährt hat. Erstaunlich gut sogar, wenn man bedenkt, wie schwierig die Parameterschätzung selbst ist. Mehrere Aspekte der Ergebnisse sprechen m.E. für die Validität der Zwei-Prozess-Interpretation zumindest in Hinblick auf die vorliegende Problem- und Fragestellung. So ist schon auf der Ebene der rein deskriptiven Betrachtung des Komponentenverlaufs auffällig, dass die beiden Parameter differenzielle Profile hatten, welche sich zudem schlüssig aus allgemeinen Eigenschaften des Sprachverarbeitungsmechanismus herleiten ließen. Die shift-Komponente war im Satzverlauf vergleichsweise stabil ausgeprägt, wies aber einen steigenden Trend auf. Ein solches Verhalten ist von vergleichsweise hierarchie-niedrigen Prozessen zu erwarten, weil die basalen Verarbeitungsprozesse (lexikalischer Zugriff, syntaktisches

Parsing, motorische Komponente innerhalb des *moving-window*-Paradigmas etc.) im Satzverlauf vergleichsweise konstant bleiben. Der leicht positive Trend lässt sich wiederum auf die steigende Ressourcenbelastung des verbalen Arbeitsgedächtnisses im Laufe der Satzverarbeitung zurückführen, weil zunehmend mehr Aktanten und offene syntaktische Prädiktionsbeziehungen im Buffer gehalten werden müssen (vgl. Just & Carpenter, 1992; Gibson, 1998). Der scale-Parameter hatte hingegen ein diskontinuierliches, zweigipfliges Profil mit starken Spitzen an den (Teil-)Satzgrenzen und weniger starken an den Phrasengrenzen. Da an diesen Stellen die zentralen interpretativen Integrationsprozesse stattfinden (vgl. Just & Carpenter, 1980; Millis & Just, 1994), spricht dies dafür, dass der scale- im Gegensatz zum shift-Parameter sensitiv für diese Prozessklassen ist.

Auch auf der Ebene der beschriebenen Effekte gab es eine erstaunlich kohärente Zuordnung der beiden Komponenten zu den einzelnen Teileffekten, welche durchweg dem theoretisch spezifizierten Muster folgten. Wie bereits im letzten Abschnitt diskutiert, gilt dies bereits für die zeitliche Dissoziation der beiden Effekte: shift-Effekte traten vornehmlich satzintern auf, während scale-Effekte satzfinal positioniert waren. Speziell fanden sich shift-Effekte in der Pronomen-Region sowie innerhalb der Präpositionalphrase, wo sie jeweils systematisch und theoretisch plausibel mit einem der Cues kovariierten. Desgleichen waren scale-Effekte v.a. im *sentence-wrap-up* dominant und wurden auch hier selektiv und spezifisch von den beiden Auflösungs-Cues angesprochen. Es ist schwer vorstellbar, dass sich eine solche Systematik zeigen könnte, wenn mit den beiden Komponenten kein weiterführender Gehalt verbunden wäre als eine rein deskriptive Beschreibung der Reaktionszeitverteilung.

Ein weiterer Punkt, der für die Zwei-Prozess-Interpretation der Parameter spricht, ist, dass sich in den nachfolgenden Verifikationsaufgaben fast ausschließlich Effekte des scale-Parameters finden ließen, jedoch keine im shift-Parameter (mit der Ausnahme von Experiment IIb, doch s.o.). Dies ist außerordentlich plausibel, weil nicht davon auszugehen ist, dass die schnelllebigen, hierarchie-niedrigen Gedächtnisprozesse zu diesem Zeitpunkt noch wirksam sind. Da die Verifikationsaufgabe auf die eigentliche Auflösung des Pronomens folgt, ist davon auszugehen, dass das Verstehensprodukt zu diesem Zeitpunkt schon etabliert war und in der Aussage v.a. die bestehende Diskursrepräsentation noch einmal abgesucht wurde, um die Aufgabe zu verifizieren. Solche Prozesse sind klar erklärungs-basiert und sollten entsprechend auch nur den scale-Parameter betreffen.

Wenn es also durchaus Anhaltspunkte für eine systematische Zuordnung der beiden Reaktionszeitparameter zu den beiden relevanten Prozessklassen gibt, so ist diese Zuordnung vor dem Hintergrund der vorliegenden Daten auch nicht absolut zwingend. Insbesondere kann

man bezweifeln, dass der shift-Parameter wirklich hierarchie-niedrige Prozesse erfasst, wie sie das Resonanz-Modell impliziert. Der Grund für diese Zweifel besteht einerseits darin, dass sich die meisten hierarchie-niedrigen Effekte lediglich etwas verzögert zeigten und nicht direkt mit dem Pronomen selbst einsetzten. Dies gilt sowohl für den Gender-*confirmation*- als auch für den IC-Refokussierungs-Effekt. Zwar sind solche *spill-over*-Effekte in der Literatur hinlänglich bekannt (vgl. Mitchell, 2004; Koornneef & van Berkum, 2006), allerdings fiel die Distanz zwischen Cue und Pronomen in der vorliegenden Untersuchung mit 3-4 Wörtern vergleichsweise groß aus. Man könnte deswegen durchaus die Auffassung vertreten, dass es sich hier nicht um eine automatische, gedächtnis-basierte Verarbeitung handelt, sondern um verzögerte, strategische Prozesse.

Dieser Einwand wird durch die Beobachtung verstärkt, dass einige Aspekte des realisierten Untersuchungssettings eine (frühe) strategische Verarbeitung auch noch unterstützt haben. So wurde den Vpn explizit mitgeteilt, dass es in der Untersuchung um die Auflösung von Pronomen in Sätzen geht, und sie haben fortlaufend Rückmeldung über die Korrektheit ihrer Antworten erhalten. Zudem erforderten alle nachfolgenden Verifikationsfragen eine explizite Auflösung des Pronomens (sog. *deep questions* sensu Stewart et al., 2000). Weiterhin gab es in den Untersuchungen keine Filler-Sätze, sondern alle Stimuli enthielten Informationen, die zur Auflösung genutzt werden konnten. Insgesamt liegt damit eine Situation vor, bei der die Vpn ausreichend motiviert waren und die notwendigen zeitlichen Ressourcen zur Verfügung hatten, um eine Auflösung strategisch vorzunehmen. Darüber hinaus standen ihnen in den Stimulussätzen die notwendigen Informationen zur Verfügung, um solche Antwortstrategien bereits früh bei der Satzverarbeitung einzusetzen. Es ist wohlbekannt, dass Rezipienten/innen auf solche Hinweisreize sensitiv reagieren und ihre Verarbeitung entsprechend anpassen (vgl. z.B. Greene et al., 1992; Garnham et al., 1992; Stewart et al., 2000).

Es ist nicht möglich, diese Einwände vollkommen zu entkräften. Das liegt v.a. daran, dass die vorliegende Untersuchung explizit auf eine strategische Verarbeitung der pronominalen Auflösung abzielte. Die Vpn *sollten* zumindest am Satzende zu einem sicheren Verstehensprodukt gelangen und sie *sollten* hierzu die notwendige Motivation und Informationen zur Verfügung haben. Dies liegt an der spezifischen Fragestellung dieser Arbeit. Denn eine zentrale Implikation des am Ende von Kap. 1 vorgestellten Modells der pronominalen Auflösung ist, dass hierarchie-niedrige, gedächtnis-basierte Prozesse dabei helfen können, eine satzfinale Auflösung mittels erklärungs-basierter Verarbeitungsprozesse vorzunehmen. Diese Annahme lässt sich nur dann adäquat prüfen, wenn es überhaupt zu einer satzfinalen Auflösung des Pronomens kommt. Spätestens seit der Untersuchung von Greene et al. (1992) ist

bekannt, dass Pronomen nicht automatisch aufgelöst werden, sondern nur, wenn Vpn die notwendigen Informationen und Ressourcen zur Verfügung stehen, um eine Auflösung vorzunehmen. Ebenfalls ist bekannt, dass Pronomen im alltäglichen Sprachverstehen nicht aufgelöst werden, wenn dies nicht zwingend notwendig ist (vgl. Poesio et al., 2006). Deswegen musste zunächst eine Untersuchungssituation etabliert werden, die eine stabile satzfinale Auflösung gewährleistet. Dies führt aber in der Tat dazu, dass eine frühe strategische Verarbeitung nicht vollkommen ausgeschlossen werden kann. Allerdings müsste man dann eine Erklärung dafür finden, warum die Vpn überhaupt auf eine frühe strategische Verarbeitung zurückgreifen, wenn ihnen später noch genügend Zeit zur Auflösung zur Verfügung steht. Unter Rückgriff auf die Ergebnisse von Experiment II könnte man nämlich sogar behauptet, dass es sogar den gegenläufigen Trend gab, alle strategischen Verarbeitungsprozesse alleine im *wrap-up* zu konzentrieren. Insbesondere in Experiment Iib, bei dem am Ende der Satzrezeption noch der gesamte zurückliegende Diskurskontext vorhanden war, wäre dies eine ausgesprochen sinnvolle Verarbeitungsstrategie. Wie die stark erhöhten satzfinalen Effekte belegen, wurde sie auch offensichtlich angewendet.

Die Frage, ob es in den vorgelegten Experimenten zu einer frühen strategischen Verarbeitung kam, lässt sich allerdings nur empirisch entscheiden. Es wäre ausgesprochen spannend zu erfahren, ob sich die Effekte in einem Design replizieren lassen, welches die Gefahr einer frühen strategischen Verarbeitung minimiert. Allerdings ist es nicht möglich, hierfür eine experimentelle Anordnung zu wählen, bei der die strategische Komponente ganz unterdrückt wird (wie dies z.B. bei Greene et al., 1992, der Fall ist). Alle pronominalen Auflösungsprozesse sind auf eine ressourcenintensive satzfinale Integration angewiesen. Ohne diese lässt sich die beschriebene Abhängigkeitsbeziehung der beiden Verarbeitungsprozesse nicht überprüfen. Die Wahrscheinlichkeit einer frühen strategischen Verarbeitung ließe sich jedoch auch auf moderatere Weise manipulieren. Eine viable Möglichkeit wäre z.B. konstruktionsfremde Filler-Items zu verwenden, die keine pronominale Auflösung verlangen, und Fragen zu stellen, die sich auf auflösungsfremde Sachverhalte des Satzes beziehen (Garnham et al., 1992).

10.3 Ausblick: Interaktion gedächtnis- und erklärungs-basierter Verarbeitungsprozesse bei der pronominalen Auflösung

Für die allgemeine Perspektive der generellen Relation gedächtnis- und erklärungs-basierter Verarbeitungsprozesse bei der pronominalen Auflösung erweist sich eine Betrachtung des IC-Effekts bzw. der verschiedenen IC-Teileffekte am fruchtbarsten. Ihre Relation ist zusammenfassend noch einmal in Abb. 10.2 dargestellt. Wie in Kap. 2 ausführlich erläutert, wird in der vorliegenden Arbeit davon ausgegangen, dass in der konzeptuellen Struktur der IC-Verben vergleichsweise ausführliche Informationen über die zugrunde liegende Ereignisstruktur enthalten sind. Diese spezifiziert u.a. in welcher kausalen Relation die verschiedenen Teilereignisse zueinander stehen und damit auch ihre zeitliche Abfolge. Wenn ich also höre, dass eine Person eine andere *fürchtet*, dann weiß ich, dass sich diese Person in einem Zustand befindet (Ereignis 1), bei dem das Eintreten eines zweiten Ereignisses (z.B. ein Treffen mit der anderen Person) zu einem Zustand der Angst führen würde. Es wird davon ausgegangen, dass solche Informationen als verb-inhärente Mini-Szenarien direkt bei ihrer Verarbeitung abgerufen werden und in einem Prozess ähnlich dem des *primary processing* innerhalb des SMF-Ansatzes verfügbar sind. In Kombination mit einem Kausalkonnektor (wie *weil*) kommt es zu zwei Effekten: Zum einen wird der kausal verantwortliche Aktant des Szenarios fokussiert, evtl. weil der Kausalkonnektor als Abruf-Cue in einem gedächtnis-basierten Abrufprozess wirkt und so den mit dem CAUSE-Operator assoziierten Aktanten reaktiviert. Dies ist der IC-Fokus-Effekt. Es ist wichtig zu betonen, dass dieser nicht einfach auf einem unspezifischen Merkmal des Verbs basiert, wie dies noch Caramazza et al. (1977) annahmen, sondern vielmehr direkt aus der kausalen, ereignisstrukturellen Repräsentation der vom Verb implizierten Ereignisse resultiert, also einem komplexen, primär erklärungs-basierten Prozess.

Zum anderen werden aber an dieser Stelle u.U. umfangreiche semantische Prädiktionsprozesse möglich, weil das Verb-Szenario mit seiner Verursachungsstruktur einen leicht verfügbaren Grund für die nachfolgende Begründung zur Verfügung stellt: Wenn in dem Verb *fürchten* eine inhärente Kausalrelation enthalten ist und durch einen Kausalkonnektor indiziert wird, dass Informationen über ein kausal vorgängiges Ereignis folgen, dann ist es vergleichsweise nahe liegend, dass diese Informationen das im Verb *fürchten* enthaltene Verursachungsereignis (bzw. den mit ihm assoziierten Aktanten) thematisieren werden.

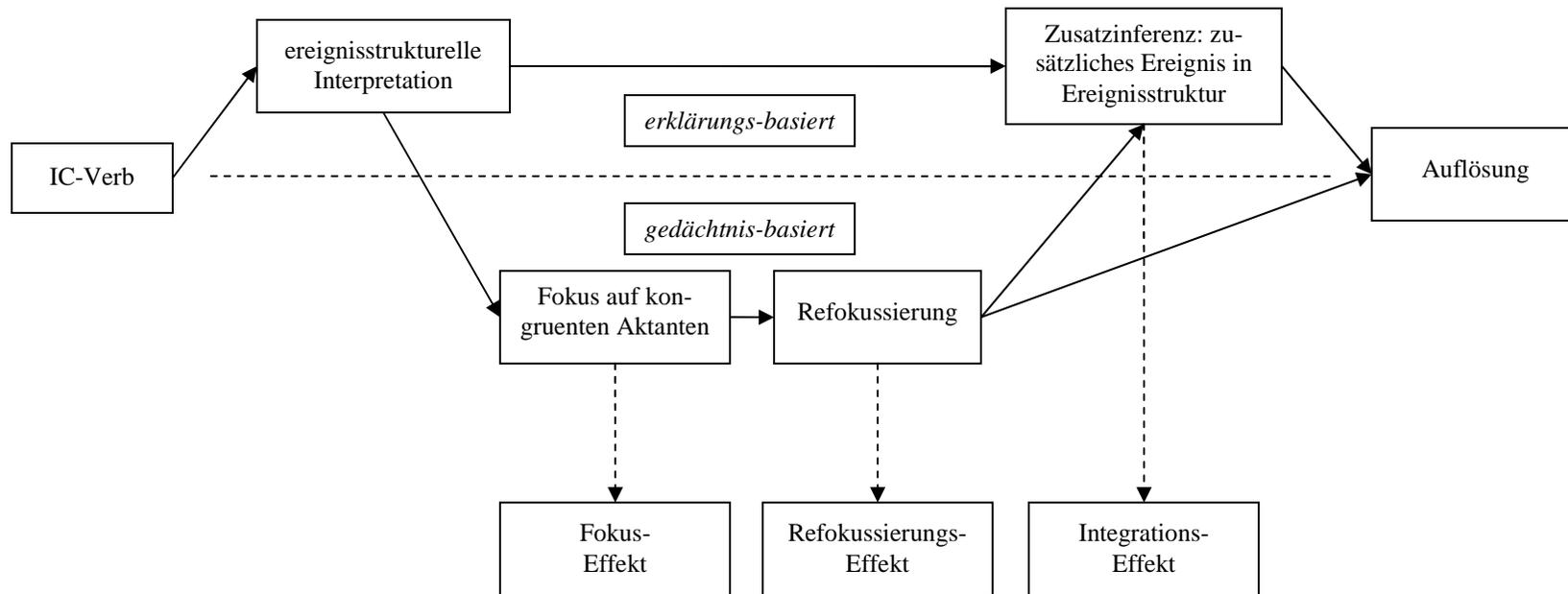


Abb. 10.2: Teileffekte des IC-Cues auf die pronominale Auflösung.

Bis hierhin ist die Verarbeitung für alle Untersuchungsbedingungen gleich: Die im Verb enthaltenen Kausalinformationen werden abgerufen und führen zu einer Voraktivierung des entsprechenden Aktanten. Wenn nun tatsächlich eine zum Verb-Bias kongruente Fortsetzung folgt, dann ist die nachfolgende Auflösung des Pronomens vergleichsweise einfach: Der passende Aktant ist bereits voraktiviert und die Integration des vom Verb präsupponierten Ereignisses erfolgt problemlos. Interessanter ist der Fall, bei dem impliziter Verb-Bias und explizite Begründung nicht übereinstimmen. In Kap. 2 wurde dafür argumentiert, dass in diesem Fall eine ereignisstrukturelle Zusatzinferenz notwendig ist: Die bereits im Verb enthaltene Kausalrelation kann nicht genutzt werden. Stattdessen muss eine neue, zusätzliche Kausalrelation in die Ereignisstruktur eingefügt werden. Das heißt, in einem Satz wie *Peter fürchtet Klaus, weil er...* ist bereits eine implizite Kausalrelation enthalten (irgendetwas an *Klaus* führt dazu, dass *Peter* in einen Zustand der Angst versetzt wird). Wenn nun eine Fortsetzung wie *weil er menschen-scheu ist.* folgt, dann ist diese Fortsetzung nicht nur unerwartet, sondern man muss eine neue, zweite Erklärung in die Ereignisstruktur einfügen. Es wird also nicht mehr begründet, warum Klaus Peter in einen Zustand der Angst versetzt, sondern warum sich Peter in einem Zustand befindet, in dem Klaus ihn in Angst versetzen kann. Allerdings ist nicht nur eine Zusatzinferenz notwendig, sondern die resultierende Ereignisstruktur ist auch komplexer (zwei vs. eine Kausalrelation; drei vs. zwei Ereignisse), weshalb sie auch länger bei der Integration bzw. Evaluation braucht. Dies ist der IC-Integrations-Effekt.

Es ist davon auszugehen, dass in dem Fall, in dem keine Gender-Informationen zur Verfügung stehen, die finale Integration noch schwieriger ist, weil der auflösungsrelevante Aktant nicht voraktiviert ist. Wenn jedoch zusätzliche Gender-Informationen vorhanden sind, dann kann die Inkongruenz von impliziter und expliziter Begründung bereits früh erkannt werden und es kommt zu dem, was weiter oben als IC-Refokussierungs-Effekt beschrieben wurde. Wie erläutert, gibt es (mindestens) zwei theoretisch unterschiedliche Möglichkeiten, wie man sich ihn vorstellen kann: Einmal als eher gedächtnis-basierten Reaktivierungsprozess, bei dem der nicht-fokussierte Aktant wieder reaktiviert wird. Zum anderen steht aber auch die eher erklärungs-basierte Interpretation offen, dass durch die frühe Entdeckung der Inkongruenz eine Umstrukturierung des aktuellen Diskurskontextes vorgenommen werden kann, evtl. unter Rückgriff auf zusätzliche Szenario-Informationen aus dem Diskurskontext. Wie im letzten Abschnitt festgestellt, kann auf der Grundlage der vorliegenden Daten nicht eindeutig zwischen beiden Mechanismen unterschieden werden.

Weiterhin ist interessant, dass die frühen Refokussierungsprozesse nur dann beobachtet werden konnten, wenn ein unterstützender diskursiver Kontext vorhanden war, der den Vpn

unmittelbar zugänglich war. Dieser Befund ist theoretisch v.a. deshalb von Bedeutung, weil er darauf aufmerksam macht, dass die Effizienz der frühen Refokussierung von externen Moderatorbedingungen abhängig ist. In Bezug auf die beiden Erklärungsmöglichkeiten des Refokussierungseffekts verhält sie sich jedoch zunächst einmal neutral: Die Verfügbarkeit weiterer Diskursinformationen kann sowohl eine eher gedächtnis-basierte Verarbeitung durch die Bereitstellung weiterer Abruf-Cues als auch eine eher erklärungs-basierte durch Verarbeitung die Bereitstellung weiterer Szenario-Informationen fördern. Wichtig ist allerdings, dass es in beiden Fällen zu einer Erleichterung bei der satzfinalen Integration kommt, weil der inkongruente Referent voraktiviert wurde.

Ich bin der Auffassung, dass diese Modellvorstellung durch die vorgelegten Experimente eine zumindest anfängliche, tentative Stützung erfahren hat. Was lässt sich jetzt aber daraus für die allgemeine Relation gedächtnis- und erklärungs-basierter Prozesse bei der pronominalen Auflösung lernen? Zunächst einmal sicherlich, dass es theoretisch fruchtbar ist, gedächtnis- und erklärungs-basierte Prozesse nicht als einander ausschließend zu betrachten, sondern vielmehr in einem gemeinsamen Rahmenmodell zu integrieren. Man stelle sich z.B. Abb. 10.2 vor, in der entweder nur die erklärungs-basierten (oberhalb der gestrichelten Linie) oder nur die gedächtnis-basierten Prozesse (unterhalb der gestrichelten Linie) vorhanden sind. Keine dieser beiden Teilabbildungen könnte das in der vorliegenden Untersuchung gefundene Ergebnismuster hinreichend erklären.

Wie in Kap. 1 und Kap. 2 dargestellt, sind solche impliziten oder sogar expliziten Reduktionstendenzen jedoch ausgesprochen häufig anzutreffen. So stellten z.B. Gerrig und McKoon (2001) bezüglich der pronominalen Auflösung allgemein fest:

In the current experiments resonance made the referent of a pronoun available [...] – without readers engaging in a special process of “pronoun resolution” (Gerrig & McKoon, 2001, p. 84).

Wie sonst lässt sich dieser Satz verstehen, als dass keine speziellen erklärungs-basierten Prozesse zur Auflösung eines Pronomens benötigt werden, weil hierzu bereits gedächtnis-basierte Verarbeitungsprozesse ausreichend sind? Es ist jedoch unklar, welche Konsequenzen dies für die in Abb. 10.2 dargestellte Modellierung hätte: Sollen alle erklärungs-basierten Prozesse aus dem Modell gestrichen werden, weil sie überflüssig sind?

Auch für den IC-Cue wurden in der bisherigen Diskussion zwei einander ausschließende Erklärungsmodelle vertreten. Eine (allein?) gedächtnis-basierte, primär mit Aufmerksamkeits-Prozessen argumentierende Fokussierungs-Hypothese wurde einer (ausschließlich!) erklärungs-basierten Integrations-Hypothese gegenübergestellt (vgl. Garnham et al., 1996; Stewart et al., 2000). In Bezug auf Abb. 10.2 unterscheiden sie sich also v.a. darin, ob sie le-

diglich die erklärungs-basierte, obere oder aber nur die gedächtnis-basierte, untere „Route“ zulassen und die jeweils andere aus ihrer Betrachtung ausklammern. Im Hinblick auf die vorliegenden Ergebnisse muss festgestellt werden, dass beide Hypothesen die Komplexität des Gegenstandsbereiches enorm unterschätzen: IC-Informationen wirken sich sowohl auf die erklärungs- als auch auf die gedächtnis-basierte Verarbeitung aus.

Darüber hinaus – das ist die zweite wichtige Schlussfolgerung dieser Arbeit – sind die beiden „Routen“ nicht unabhängig voneinander. Vielmehr existieren mannigfaltige Querbeziehungen zwischen ihnen. Das heißt, gedächtnis- und erklärungs-basierte Prozesse sind nicht einfach nur zwei unterschiedliche, aber voneinander unabhängige Arten, wie IC-Informationen verarbeitet werden, sondern sie setzen sich gegenseitig voraus. Die Interdependenz der beiden Prozessklassen hat dabei ein so großes Ausmaß, dass man sich fragen kann, ob es überhaupt Sinn macht, zwischen ihnen zu unterscheiden: Gedächtnis-basierte Prozesse operieren auf dem Produkt der erklärungs-basierten Verarbeitung, transformieren dieses u.U. und stellen es wiederum nachfolgend der erklärungs-basierten Verarbeitung zur Verfügung. Der Informationsfluss zwischen den beiden Prozessebenen ist also keine informationelle Einbahnstraße, sondern kann explizit in beide Richtungen verlaufen: von den gedächtnis- zu den erklärungs-basierten, aber auch umgekehrt von den erklärungs- zu den gedächtnis-basierten Prozessen. Am deutlichsten wird dies, wenn man die komplexe Wirkung der in IC-Verben enthaltenen ereignisstrukturellen Informationen betrachtet. Diese führen einerseits zu weit reichenden Prädiktionsprozessen, welche die erklärungs-basierte Integration beeinflussen. Darüber hinaus sind sie aber auch für die gedächtnis-basierte Fokussierung entscheidend, da sie überhaupt erst den relevanten Bedeutungsraum aufspannen, in dem Aufmerksamkeitsressourcen allokiert und evtl. auch umdisponiert werden.

Abschließend muss jedoch noch auf eine dritte Konsequenz hingewiesen werden, die sich aus der in Abb. 10.2 dargestellten Modellierung des IC-Effekts bei der pronominalen Auflösung ergibt. Dort ist klar zu sehen, dass die erklärungs-basierte Verarbeitung der gedächtnis-basierten zwar zeitlich nachgeordnet, logisch aber vorgeordnet ist. Das heißt, das ultimative Ziel der Sprachverarbeitung ist prinzipiell, ein unter erklärungs-basierter Perspektive möglichst optimales Verstehensprodukt aufzubauen. Gedächtnis-basierte Prozesse sind eine Ebene tiefer angesiedelt und können bei dieser Etablierung u.U. helfen. Manchmal werden sie aber auch eher hinderlich sein. Aparenterweise ist beim IC-Effekt beides gleichzeitig der Fall: Wenn verb-gesteuerte Voraktivierung und explizite Begründung übereinstimmen, dann ergänzen sie sich synergetisch. Andererseits können die beiden Prozesse aber auch gegeneinander arbeiten, wodurch sich die Verarbeitung zusätzlich erschwert. Wichtig ist jedoch v.a. zu

sehen, dass die beiden Prozesse in einer Ergänzungsrelation stehen: Wenn eine Aufgabe bereits durch gedächtnis-basierte Prozesse hinreichend gut erledigt werden kann, dann sind nur wenige ressourcen-intensive Erklärungsprozesse notwendig, um das finale Verstehensprodukt zu evaluieren und zu etablieren. Von daher kann das obige Zitat von Gerrig und McKoon (2001) auch so verstanden werden, dass hier die resonanz-basierte Reaktivierung so erfolgreich war, dass der zweite erklärungs-basierte Auflösungsschritt so minimal ausfiel, dass er sich nicht mehr nachweisen ließ. Daraus jedoch zu folgern, dass es ihn prinzipiell nicht gibt, wäre ein fataler Fehlschluss. Denn wenn rein gedächtnis-basierte Prozesse keinen eindeutigen Antezedens reaktivieren oder sogar das „falsche“ Antezedens aktivieren, dann werden erklärungs-basierte Prozesse wieder relevant. Sie müssen sogar verstärkt eingesetzt werden, um die fehlerhafte Bearbeitung auf der gedächtnis-basierten Ebene auszugleichen.

An dieser Stelle zeigt sich, dass die erklärungs-basierten Verarbeitungsprozesse gewissermaßen als Schnittstelle zum kognitiv-motivationalen Gesamtsystem fungieren. Denn selbstverständlich ist davon auszugehen – wie dies insbesondere die konstruktivistische Inferenztheorie betont –, dass Sprachverarbeitung nicht in einem kognitiven, emotionalen und motivationalen Vakuum stattfindet, sondern auch von den verfügbaren individuellen Verarbeitungsressourcen und motivationalen Zielperspektiven entscheidend mitbestimmt wird. Speziell ist beim IC-Effekt davon auszugehen, dass insbesondere die satzfinale Integration und die ressourcen-intensive Zusatzinferenz, die für die inkongruenten Fortsetzungen notwendig ist, nur dann vorgenommen werden, wenn beide Voraussetzungen gegeben sind, also wenn die Leser/innen hinreichend motiviert sind, den Inhalt des Satzes zu verstehen und die zur Durchführung notwendigen Informations- und Ressourcenbestände.

Wenn man alle diese Punkte zusammen nimmt und mit den in Kap. 1 dargestellten Integrationsperspektiven für die gedächtnis- und die erklärungs-basierten Sprachverarbeitungstheorien vergleicht, dann fällt auf, dass das in Abb. 10.2 dargestellte Erklärungsmodell des IC-Effekts eigentlich alle Varianten gleichzeitig realisiert, wenn auch mit unterschiedlicher Schwerpunktsetzung. Denn zunächst einmal wird klar konzediert, dass nicht unter allen Umständen eine erklärungs-basierte Verarbeitung zur Etablierung eines adäquaten Verstehensproduktes notwendig ist. Wenn gedächtnis-basierte Verarbeitungsprozesse ein ähnlich adäquates Verstehensprodukt gewährleisten, dann kann sich die erklärungs-basierte Komponente auf eine Evaluation des Produktes beschränken. Gleichzeitig wird aber auch festgestellt, dass diese gedächtnis-basierten Prozesse nicht unabhängig und allein für sich operieren, sondern vielmehr nur dann wirksam sind, wenn sie am Produkt einer erklärungs-basierten Verarbeitung ansetzen, die u.U. auch höhere Wissensstrukturen wie z.B. implizite Ereignisstrukturen

von Verben bzw. komplexe Szenarien umfassen können. Diese „lenken“ die gedächtnis-basierte Verarbeitung, indem sie diese vorstrukturieren und bestimmte Abrufrichtungen wahrscheinlicher machen. Wenn sich im Laufe der Satzverarbeitung herausstellt, dass die gedächtnis-basierte Verarbeitung zu einem unplausiblen bzw. inzwischen vom neuen Diskurskontext überholten Ergebnis kommt, dann fungiert die erklärungs-basierte Verarbeitungskomponente als Kontrollinstanz, welche remediale Reanalyse-Prozesse solange ausführt bis ein individuell variabler Kohärenzstandard erreicht ist. Damit aber sind gedächtnis- und erklärungs-basierte Prozesse gerade als prinzipiell gleichberechtigte und parallel-interagierende Verarbeitungsmechanismen anzusetzen.

Literatur

- Aaronson, D. & Ferres, S. (1984). The word-by-word reading paradigm: An experimental and theoretical approach. In D.E. Kieras & M.A. Just (Eds.), *New methods in reading comprehension research* (pp. 31-68). Hillsdale, NJ: Erlbaum.
- Aaronson, D. & Ferres, S. (1983a). Lexical categories and reading tasks. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 675-699.
- Aaronson, D. & Ferres, S. (1983b). A model for coding lexical categories during reading. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 700-725.
- Aaronson, D. & Scarborough, H.S. (1976). Performance theories for sentence coding: Some quantitative evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 56-70.
- Aaronson, D. & Scarborough, H.S. (1977). Performance theories for sentence coding: Some quantitative models. *Journal of Verbal Learning and Verbal Behavior*, 16, 277-303.
- Abelson, R.P. & Kanouse, D.E. (1972). Subjective acceptance of verbal generalization. In S. Feldman (Ed.), *Cognitive consistency: Motivational antecedents and behavioral consequents* (pp. 171-197). New York: Academic Press.
- Abney, S.P. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research*, 18, 129-144.
- Albrecht, J.E. & Myers, J.L. (1995). Role of context in accessing distant information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1459-1468.
- Albrecht, J.E. & Myers, J.L. (1998). Accessing distant text information during reading: effects of contextual cues. *Discourse Processes*, 26, 87-107.
- Alisch, L.-M. (Hrsg.). (2001). Nichtlineare Zeitreihenanalyse [Themenheft]. *Empirische Pädagogik*, 15.
- Albrecht, J.E. & O'Brien, E.J. (1991). Effects of centrality on retrieval of text-based concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 932-939.
- Albrecht, J.E. & O'Brien, E.J. (1993). Updating a mental model: maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1061-1070.
- Albrecht, J.E. & Myers, J.L. (1995). Role of context in accessing distant information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1459-1468.

- Altman, N.S. (1992). An introduction to Kernel and nearest-neighbor nonparametric regression. *The American Statistician*, *46*, 175-185.
- Altmann, G.T.M. & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*, 247-264.
- Altmann, G.T.M. & Steedman, M.J. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191-238.
- Anderson, A., Garrod, S.C. & Sanford, J. (1983). The accessibility of pronominal antecedents as a function of episode in narrative text. *Quarterly Journal of Experimental Psychology*, *35A*, 427-440.
- Anderson, J.R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.
- Andrews, S. & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 514-544.
- Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.
- Ariel, M. (2004). Accessibility marking: discourse functions, discourse profiles, and processing cues. *Discourse Processes*, *37*, 91-116.
- Arnold, J.E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, *31*, 137-162.
- Arnold, J.E., Eisenband, J.G., Brown-Schmidt, S. & Trueswell, J.C. (2000). The rapid use of gender information: evidence of the time course of pronoun resolution from eye tracking. *Cognition*, *76*, B13-B26.
- Ashby, F.G. & Townsend, J.T. (1980). Decomposing the reaction time distribution: Pure insertion and selective influence revisited. *Journal of Mathematical Psychology*, *21*, 93-123.
- Au, T.K. (1986). A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language*, *25*, 104-122.
- Badecker, W. & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 748-769.
- Balota, D.A. (1994). Visual word recognition. the journey from features to meaning. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 303-358). San Diego: Academic Press.

- Balota, D.A., Yap, M.L. & Cortese, M.J. (2006). Visual word recognition. the journey from features to meaning (A travel update). In M.J. Traxler & M.A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 285-376). San Diego: Academic Press.
- Balota, D.A., Cortese, M.J., Sergent-Marshall, S.D., Spieler, D.H. & Yap, M.J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, 133, 283-318.
- Balota, D.A. & Spieler, D.H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, 128, 32-55.
- Bartlett, F.C. (1932/1995). *Remembering. A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- Bargh, J.A. (1992). The ecology of automaticity: Toward establishing the conditions needed to produce automatic processing effects. *American Journal of Psychology*, 105, 181-199.
- Basu, A.P. (1964). Estimates of reliability for some distributions useful in life testing. *Technometrics*, 6, 215-219.
- Battacharya, S.R. (1967). Bayesian approach to life testing and reliability estimation. *Journal of the American Statistical Association*, 62, 48-62.
- Beckermann, A. (2000). *Analytische Einführung in die Philosophie des Geistes* (2., überarb. Aufl.). Berlin: de Gruyter.
- Belletti, A. & Rizzi, L. (1988). PSYCH-verbs and θ -theory. *Natural Language and Linguistic Theory*, 6, 291-352.
- Bieri, P. (Hrsg.). (1981). *Analytische Philosophie des Geistes*. Weinheim: Beltz.
- Bransford, J.D., Barclay, J.R. & Franks, J.J. (1972). Sentence memory: A constructive versus interpretative approach. *Cognitive Psychology*, 3, 193-209.
- Bransford, J.D. & Franks, J.J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, 2, 331-350.
- Berger, J.O. & Sun, D. (1993). Bayesian analysis for the poly-Weibull distribution. *Journal of the American Statistical Association*, 88, 1412-1418.
- Bhatt, C. (1991). *Einführung in die Morphologie*. Hürth: Gabel.
- Blanco, M.J. & Alvarez, A.A. (1994). Psychometric intelligence and visual focused attention: relationships in nonsearch tasks. *Intelligence*, 18, 77-106.
- Blumenberg, H. (1979). *Über die Lesbarkeit der Welt*. Frankfurt a.M.: suhrkamp.
- Bolstad, W.M. (2004). *Introduction to Bayesian statistics*. Hoboken: Wiley.

- Bornkessel, I. & Schlesewsky, M. (2006). The extended Argument Dependency Model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113*, 787-821.
- Brennan, S.E. (1995). Centering attention in discourse. *Language and Cognitive Processes*, *10*, 137-167.
- Brown, R. & Fish, D. (1983a). The psychological causality implicit in language. *Cognition*, *14*, 237-273.
- Brown, R. & Fish, D. (1983b). Are there universal schemas of psychological causality. *Archives de Psychologie*, *51*, 145-153.
- Brown, R. & Heathcote, A. (2003). QMLE: Fast, robust, and efficient estimation of distribution functions based on quantils. *Behavior Research Methods, Instruments, & Computers*, *35*, 485-492.
- Budiu, R. (2001). *The role of background knowledge in sentence processing*. Unpublished PhD thesis. Carnegie Mellon University.
- Budiu, R. & Anderson, J.R. (2002). Comprehending anaphoric metaphors. *Memory & Cognition*, *30*, 158-165.
- Budiu, R. & Anderson, J.R. (2004). Interpretation-based processing: a unified theory of semantic sentence comprehension. *Cognitive Science*, *28*, 1-44.
- Büring, D. (2005). *Binding theory*. Cambridge: Cambridge University Press.
- Burgess, C., Livesay, K. & Lund, K. (1997). Explorations in context space: words, sentences, discourse. *Discourse Processes*, *25*, 211-257.
- Bußmann, H. (2002). *Lexikon der Sprachwissenschaft* (3., akt. u. erw. Aufl.). Stuttgart: Kröner.
- Butt, M. (2006). *Theories of case*. Cambridge: Cambridge University Press.
- Cacciari, C., Carreiras, M. & Cionini, C.B. (1997). When words have two genders: Anaphor resolution for Italian functionally ambiguous words. *Journal of Memory and Language*, *37*, 517-532.
- Canavos, G.C. & Tsokos, C.P. (1973). Bayesian estimation of life parameters in the Weibull distribution. *Journal of Operational Research*, *21*, 755-763.
- Caramazza, A. & Gupta, S. (1979). The roles of topicalization, parallel function and verb semantics in the interpretation of pronouns. *Linguistics*, *17*, 497-518.
- Caramazza, A., Grober, E., Garvey, C. & Yates, J. (1977). Comprehension of anaphoric pronouns. *Journal of Verbal Learning and Verbal Behavior*, *16*, 601-609.

- Carpenter, P.A., Miyake, A. & Just, M.A. (1994). Working memory constraints in comprehension. Evidence from individual differences, aphasia, and aging. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1075-1122). San Diego: Academic Press.
- Carpenter, P.A., Miyake, A. & Just, M.A. (1995). Language comprehension: Sentence and discourse processing. *Annual Review of Psychology*, 46, 91-120.
- Carrieras, M. (1993). Minimalist misconceptions of mental models. *Psychology*, 4 (<http://journals.ecs.soton.ac.uk/resource/psychology>, Datum letzter Zugriff: 17.7.2007).
- Carreiras, M. & Clifton, C. (2004). On the on-line study of language comprehension. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension. Eyetracking, ERPs and beyond* (pp. 1-32). New York: Psychology Press.
- Carreiras, M., Garnham, A., Oakhill, J. & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *Quarterly Journal of Experimental Psychology*, 49, 639-663.
- Carrithers, C. & Bever, T.G. (1984). Eye-fixation patterns during reading confirm theories of language comprehension. *Cognitive Science*, 8, 157-172.
- Casella, G. & George, E.I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46, 167-174.
- Chafe, W.L. (1970). *Meaning and the structure of language*. Chicago: Chicago University Press.
- Charniak, E. (1972). *Toward a model of children's story comprehension*. Technical Report 266 MIT Artificial Intelligence Laboratory.
- Chen, L. (1995). Testing the mean of skewed distributions. *Journal of the American Statistical Association: Theory and Methods*, 90, 767-772.
- Cheng, R.C.H. & Amin, N.A.K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society B*, 45, 394-403.
- Cheng, R.C.H. & Iles, T.C. (1987). Corrected maximum likelihood in non-regular problems. *Journal of the Royal Statistical Society B*, 49, 95-101.
- Cheng, R.C.H. & Iles, T.C. (1990). Embedded models in three-parameter distributions and their estimation. *Journal of Royal Statistical Society B*, 52, 135-149.
- Cheng, R.C.H. & Traylor, L. (1995). Non-regular maximum likelihood problems. *Journal of the Royal Statistical Society B*, 57, 3-44.

- Chib, S. & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49, 327-335.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Christmann, U. & Groeben, N. (1999). Psychologie des Lesens. In B. Franzmann, K. Hasemann, D. Löffler & E. Schön (Hrsg.), *Handbuch Lesen* (S. 145-223). München: Sauer.
- Clark, H.H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335-359.
- Clark, H.H. & Sengul, C.J. (1979). In search of referents for nouns and pronouns. *Memory & Cognition*, 7, 35-41.
- Clifton, C. & Duffy, S.A. (2001). Sentence and text comprehension: Roles of linguistic structure. *Annual Review of Psychology*, 52, 167-196.
- Clifton, C. & Kennison, S.M. & Albrecht, J.E. (1997). Reading the words *her, his, him*: Implication for parsing principles based on frequency and on structure. *Journal of Memory and Language*, 36, 276-292.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Erlbaum.
- Colonus, H. (1995). The instance theory of automaticity: Why the Weibull? *Psychological Review*, 102, 744-750.
- Comrie, B. (1999). Grammatical gender systems: A linguist's assessment. *Journal of Psycholinguistic Research*, 28, 457-466.
- Congdon, P. (2001). *Bayesian statistical modelling*. Chichester: Wiley.
- Cook, A.E. & Guéraud, S. (2005). What have we been missing? The role of general world knowledge in discourse processing. *Discourse Processes*, 39, 265-278.
- Cook, A.E. & Myers, J.L. (2004). Processing discourse roles in scripted narratives: The influences of context and knowledge. *Journal of Memory and Language*, 50, 268-288.
- Corbett, A.T. (1984). Prenominal adjectives and the disambiguation of anaphoric nouns. *Journal of Verbal Learning and Verbal Behavior*, 23, 683-695.
- Corbett, A.T. & Doshier, B.A. (1978). Instrument inferences in sentence encoding. *Journal of Verbal Learning and Verbal Behavior*, 17, 283-294.
- Cornish, F. (1989). Anaphoric pronouns: Under linguistic control or signalling particular discourse representations? *Journal of Semantics*, 5, 233-260.

- Corrigan, R. (1992). The relationship between causal attributions and judgements of the typicality of events described by sentences. *British Journal of Social Psychology*, *31*, 351-368.
- Corrigan, R. (1993). Causal attributions to the states and events encoded by different types of verbs. *British Journal of Social Psychology*, *32*, 335-348.
- Corrigan, R. & Stevenson, C. (1994). Children's causal attribution to states and events described by different classes of verbs. *Cognitive Development*, *9*, 235-256.
- Cousineau, D., Brown, S. & Heathcote, A. (2004). Fitting distributions using maximum likelihood: Method and packages. *Behavior Research Methods, Instruments & Computers*, *36*, 742-756.
- Cousineau, D., Goodman, V.W. & Shiffrin, R.M. (2002). Extending statistics of extremes to distributions varying in position and scale and the implications for race models. *Journal of Mathematical Psychology*, *46*, 431-454.
- Cousineau, D. & LaRoche, S. (1997). PASTIS: A program for curve and distribution analysis. *Behavior Research Methods, Instruments, & Computers*, *29*, 542-548.
- Crawley, R.A., Stevenson, R.J. & Kleinman, D. (1990). The use of heuristics strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, *19*, 245-264.
- Crinean, M. & Garnham, A. (2006). Implicit causality, implicit consequentiality and semantic roles. *Language and Cognitive Processes*, *21*, 636-648.
- Croft, William (1991). *Syntactic categories and grammatical relations. The cognitive organisation of information*. Chicago: University of Chicago Press.
- Croft, W. (1993). Case marking and the semantics of mental verbs. In J. Pustejovsky (Ed.), *Semantics and the lexicon* (pp. 55-72). Dordrecht: Kluwer.
- Croft, W. & Cruse, D.A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Cross, E.M. & Chaffin, W.W. (1982). Use of the binomial theorem in interpreting results of multiple tests of significance. *Educational and Psychological Measurement*, *42*, 25-34.
- Danks, J.H. (1986). Identifying component processes in test [sic!] comprehension: Comment on Haberlandt and Graesser. *Journal of Experimental Psychology: General*, *115*, 193-197.
- Davison, A.C. & Louzada-Neto, F. (2000). Inference for the poly-Weibull model. *The Statistician*, *49*, 189-198.
- Dawson, M.R.W. (1988). Fitting the ex-Gaussian equation to reaction time distributions. *Behavior Research Methods, Instruments & Computers*, *20*, 54-57.

- DeCarlo, L.T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, 2, 292-307.
- DeCharms, R. (1968). *Personal causation: The internal affective determinants of behavior*. New York: Academic Press.
- de Vincenzi, M. (1999). Differences between the morphology of gender and number: Evidence from establishing coreference. *Journal of Psycholinguistic Research*, 28, 537-553.
- Dell, G.S., McKoon, G. & Ratcliff, R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. *Journal of Memory and Language*, 22, 121-132.
- Dolan, C.V., van der Maas, H.L.J., Molenaar, P.C.M. (2002). A framework for ML estimation of parameters of (mixtures of) common reaction time distributions given optional truncation or censoring. *Behavior Research Methods, Instruments & Computers*, 34, 304-323.
- Donders, F.J. (1868/1969). On the speed of mental processes. *Acta Psychologica*, 30, 412-431.
- Dowty, D. (1979). *Word meaning and Montague grammar*. Dordrecht: Reidel.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67, 547-619.
- Dubey, S.D. (1965). Asymptotic properties of several estimators of Weibull parameters. *Technometrics*, 7, 423-434.
- Dubey, S.D. (1967). Monte Carlo Study of the moment and maximum likelihood estimators of Weibull parameters. *Trabajos de estadística*, 18, 156-162.
- Edwards, D. & Potter, J. (1993). Language and causation: a discursive action model of description and attribution. *Psychological Review*, 100, 23-41.
- Ehrich, V. (1992). *Hier und Jetzt. Studien zur lokalen und temporalen Deixis im Deutschen*. Tübingen: Niemeyer.
- Ehrlich, K. (1980). Comprehension of pronouns. *Quarterly Journal of Experimental Psychology*, 32, 247-255.
- Ehrlich, K. & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior*, 22, 75-87.
- Erb, E. (1997). Gegenstand- und Problemkonstituierung: Subjekt-Modelle (in) der Psychologie. In N. Groeben (Hrsg.), *Zur Grammatik einer sozialwissenschaftlichen*

- Psychologie, Bd. I Metatheoretische Perspektiven, 2. Halbbd. Gegenstandsverständnis, Menschenbilder, Methodologie und Ethik* (S. 139-239). Münster: Aschendorff.
- Ericsson, K.A. & Simon, H.A. (1980). Verbal reports as data. *Psychological Review*, 87, 215-251.
- Ericsson, K.A., & Simon, H.A. (1993). *Protocol analysis; Verbal reports as data* (2nd ed.). Cambridge: MIT Press.
- Fanselow, G. & Felix, S.W. (1993a). *Sprachtheorie. Bd. 1: Grundlagen und Zielsetzungen*. Tübingen: Francke.
- Fanselow, G. & Felix, S.W. (1993b). *Sprachtheorie. Bd. 2: Die Rektions- und Bindungstheorie*. Tübingen: Francke.
- Faul, F., Erdfelder, E., Lang, A.G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Faust, M.E., Balota, D.A., Spieler, D.H. & Ferraro, F.R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin*, 125, 777-799.
- Fernandez, A. & Carriedo, N. (1993). Reading comprehension: In defence of a mental model approach. *Psychology*, 4 (<http://journals.ecs.soton.ac.uk/resource/psychology>, Datum letzter Zugriff: 17.7.2007).
- Ferretti, T.R., McRae, K. & Hatherall, A. (1997). Integrating verbs, situation schemas and thematic role concepts. *Journal of Memory and Language*, 44, 516-547.
- Fiedler, K. & Semin, G.R. (1988). On the causal information conveyed by different interpersonal verbs: The role of implicit sentence context. *Social Cognition*, 6, 21-39.
- Fillenbaum, S. & Rapaport, A. (1971). *Structures in the subjective lexicon*. London: Academic Press.
- Fillmore, C.J. (1968). The case for case. In E. Bach & J. Harms (Eds.), *Universals in linguistic theory* (pp. 1-88). New York: Holt, Rinehart & Winston.
- Fillmore, C.J. (1970). Types of lexical information. In F. Kiefer (Ed.), *Studies in syntax and semantics* (pp. 109-132). New York: Humanities Press.
- Fillmore, C.J. (1971). Verbs of judging: A exercise in semantic description. In C.J. Fillmore & D.T. Langendoen (Eds.), *Studies in linguistic semantics* (pp. 273-296). New York: Holt, Rinehart & Winston.
- Fillmore, C.J. (1977). Scenes-and-frames semantics. In A. Zampolli (Ed.), *Linguistic structures processing* (pp. 55-81). Amsterdam: North Holland Publishing.

- Fisher, R.A. and Tippett, L.H.C. (1928), Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24, 180-190.
- Fodor, J.A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Colthart (Ed.), *Attention and performance XII* (pp. 559-586). Hillsdale: Erlbaum.
- Frazier, L. & Clifton, C., (1996). *Construal*. Cambridge: MIT Press.
- Frazier, L. & Clifton, C. (2000). On bound variable interpretations: the LF-only hypothesis. *Journal of Psycholinguistic Research*, 29, 125-139.
- Friederici, A.D. & Jacobson, T. (1999). Processing grammatical gender during language comprehension. *Journal of Psycholinguistic Research*, 28, 467-484.
- Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.
- Garnham, A. (1992). Minimalism versus constructionism: A false dichotomy in theories of inference during reading. *Psychology*, 3 (<http://journals.ecs.soton.ac.uk/resource/psychology>, letzter Zugriff: 17.7.2007).
- Garnham, A.. (1993a). Dichotomy or not dichotomy?: That is the question. *Psychology*, 4. (<http://journals.ecs.soton.ac.uk/resource/psychology>, letzter Zugriff: 17.7.2007).
- Garnham, A. (1993b). An impartial view of inference making. *Psychology*, 4. (<http://journals.ecs.soton.ac.uk/resource/psychology>, letzter Zugriff: 17.7.2007).
- Garnham, A. (1999). Reference and anaphora In S. Garrod & M.J. Pickering (Eds.), *Language processing* (pp. 335-362). Hove: Psychology Press.
- Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Hove: Psychology Press.
- Garnham, A. & Oakhill, J. (1985). On-line resolution of anaphoric pronouns: Effects of inference making and verb semantics. *British Journal of Psychology*, 76, 385-393.
- Garnham, A. & Oakhill, J. (1990). Mental models as context for interpreting texts: implications from studies of anaphora. *Journal of Semantics*, 7, 379-393.
- Garnham, A., Oakhill, J. & Cruttenden, H. (1992). The role of implicit causality and gender cue in the interpretation of pronouns. *Language and Cognitive Processes*, 7, 231-255.
- Garnham, A., Oakhill, J., Ehrlich, M.-F. & Carreiras, M. (1995). Representations and processes in the interpretation of pronouns: New evidence from Spanish and French. *Journal of Memory and Language*, 34, 41-62.

- Garnham, A., Traxler, M., Oakhill, J. & Gernsbacher, M.A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language*, 35, 517-543.
- Garrod, S., Freudenthal, D. & Boyle, E. (1994). The role of different types of anaphor in the on-line resolution of sentences in a discourse. *Journal of Memory and Language*, 33, 39-68.
- Garrod, S., O'Brien, E.J., Morris, R.K. & Rayner, K. (1990). Elaborative inferencing as an active or passive process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 250-257.
- Garrod, S. & Sanford, A. (1977). Interpreting anaphoric relations: The integration of semantic information while reading. *Journal of Verbal Learning and Verbal Behavior*, 16, 77-90.
- Garrod, S. & Sanford, A.J. (1983). The mental representation of discourse in a focused system: Implications for the interpretation of anaphoric noun-phrases. *Journal of Semantics*, 1, 21-41.
- Garrod, S. & Sanford, A.J. (1985). On the real-time character of interpretation during reading. *Language and Cognitive Processes*, 1, 43-59.
- Garrod, S. & Sanford, A.J. (1989). On the real-time character of interpretation during reading. *Language and Cognitive Processes*, 1, 43-59.
- Garrod, S.C. & Sanford, A.J. (1994). Resolving sentences in a discourse context. How discourse representation affects language understanding. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 675-698). San Diego: Academic Press.
- Garrod, S. & Sanford, A. (1999). Incrementality in discourse understanding. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 3-27). Mahwah: Erlbaum.
- Garrod, S.C. & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: bonding and resolution. *Journal of Memory and Language*, 42, 526-544.
- Garvey, C. & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, 5, 459-464.
- Garvey, C., Caramazza, A. & Yates, J. (1975). Factors influencing assignment of pronoun antecedents. *Cognition*, 3, 227-243.
- Gelfand, A.E. & Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 385-409.
- Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton: Chapman & Hall.

- Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern analysis & Machine Intelligence*, 6, 721-741.
- Gernsbacher, M.A. (1989). Mechanisms that improve referential access. *Cognition*, 32, 99-156.
- Gernsbacher, M.A. (1990). *Language comprehension as structure building*. Hillsdale: Erlbaum.
- Gernsbacher, M.A. & Shroyer, S. (1989). The cataphoric use of indefinite *this* in spoken narratives. *Memory & Cognition*, 17, 536-540.
- Gerrig, R.J. & McKoon, G. (1998). The readiness is all: The functionality of memory-based text processing. *Discourse Processes*, 26, 67-86.
- Gerrig, R.J. & McKoon, G. (2001). Memory processes and experiential continuity. *Psychological Science*, 12, 81-85.
- Gerrig, R.J. & O'Brien, E.J. (2005). The scope of memory-based processing. *Discourse Processes*, 39, 225-242.
- Gethmann, C.F. (1995). Forschungsprogramme. In J. Mittelstraß (Hrsg.), *Enzyklopädie der Philosophie und Wissenschaftstheorie* (Bd. 1, S. 664). Stuttgart: Metzler.
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68, 1-76.
- Gilks, W.R., Richardson, S.E. & Spiegelhalter, D.J. (1996). *Markov chain Monte Carlo in practice*. London: Chapman & Hall.
- Gill, J. (2002). *Bayesian methods for the social and behavioural sciences*. Boca Raton: Chapman & Hall.
- Gillund, G. & Shiffrin, R.M. (1981). Free recall of complex pictures and abstract words. *Journal of Verbal Learning and Verbal Behavior*, 20, 575-592.
- Gillund, G. & Shiffrin, R.M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91, 1-67.
- Glenberg, A.M. & Mathew, S. (1992). When minimalism is not enough: Mental models in reading. *Psychology*, 3. (<http://journals.ecs.soton.ac.uk/resource/psychology>, Datum letzter Zugriff: 17.7.2007).
- Gnedenko, B.V. (1943), Sur la distribution limite du terme maximum d'une serie aleatoire. *Annales Mathematique*, 44, 423-453.
- Gordon, P.C., Camblin, C. & Swaab, T.Y. (2004). On-line measures of coreferential processing. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence*

- comprehension. *Eyetracking, ERPs and beyond* (pp. 139-150). New York: Psychology Press.
- Gordon, P.C., Grosz, B.J. & Gilliom, L.A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17, 311-347.
- Gordon, P.C. & Hendrick, R. (1997). Intuitive knowledge of linguistic co-reference. *Cognition*, 62, 325-370.
- Gordon, P.C., Handrick, R. & Ledoux Foster, K. (2000). Language comprehension and probe-list memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 766-775.
- Goschke, T. (1996). Wille und Kognition: Zur funktionalen Architektur der intentionalen Handlungssteuerung. In J. Kuhl & H. Heckhausen (Hrsg.), *Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie IV: Motivation und Emotion, Bd. 4: Motivation, Volition und Handlung* (S. 583-663). Göttingen: Hogrefe.
- Graesser, A.C. & Haberlandt, K.F. (1986). Research on component processes in reading: Reply to Danks. *Journal of Experimental Psychology: General*, 115, 198-200.
- Graesser, A.C., Hoffman, N.L. & Clark, L.F. (1980). Structural components of reading time. *Journal of Verbal Learning and Verbal Behavior*, 19, 135-151.
- Graesser, A.C., Millis, K.K., Zwaan, R.A. (1997). Discourse comprehension. *Annual Review of Psychology*, 48, 163-189.
- Graesser, A.C., Singer, M. & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371-395.
- Green, D.W., Mitchell, D.C. & Hammond, E.J. (1981). The scheduling of text integration processes in reading. *Quarterly Journal of Experimental Psychology*, 33A, 455-464.
- Greene, S.B. & McKoon, G. (1995). Telling something we can't know: Experimental approaches to verbs exhibiting implicit causality. *Psychological Science*, 6, 262-270.
- Greene, S.B., McKoon, G. & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: learning, Memory, and Cognition*, 18, 266-283.
- Grewendorf, G., Hamm, F. & Sternefeld, W. (1989). *Sprachliches Wissen. Eine Einführung in moderne Theorien der grammatischen Beschreibung* (3. Aufl.). Frankfurt a.M.: suhrkamp.
- Grice, H.P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds.), *Syntax and semantics 3: speech acts* (pp. 41-58). New York: Academic Press.
- Grimshaw, J. (1990). *The structure of argument structure*. Cambridge: MIT Press.

- Grober, Beardsley, W. & Caramazza, A. (1978). Parallel function strategy in pronoun assignment. *Cognition*, 6, 117-133.
- Groeben, N. (1982). *Leserpsychologie: Textverständnis – Textverständlichkeit*. Münster: Aschendorff.
- Groeben, N. (1986). *Handeln, Tun, Verhalten als Einheiten einer verstehend-erklärenden Psychologie*. Tübingen: Francke.
- Groeben, N. (1997). Einleitung: Sozialwissenschaftliche Psychologie-Konzeption zwischen natur- und Geisteswissenschaft. In N. Groeben (Hrsg.), *Zur Programmatik einer sozialwissenschaftlichen Psychologie, Bd. I Metatheoretische Perspektiven, 1. Halbbd. Gegenstandsverständnis, Menschenbilder, Methodologie und Ethik* (S. 1-26). Münster: Aschendorff.
- Groeben, N. (1999). Fazit: Die metatheoretischen Merkmale einer sozialwissenschaftlichen Psychologie. In N. Groeben (Hrsg.), *Zur Programmatik einer sozialwissenschaftlichen Psychologie, Bd. I Metatheoretische Perspektiven, 2. Halbbd. Theoriehistorie, Praxisrelevant, Interdisziplinarität, Methodenintegration* (S. 311-404). Münster: Aschendorff.
- Groeben, N. (2003). Fazit: Problemaufriss einer Theoretischen Psychologie. In N. Groeben (Hrsg.), *Zur Programmatik einer sozialwissenschaftlichen Psychologie, Bd. I Objekttheoretische Perspektiven, 2. Halbbd. Situationsbezug, Reflexivität, Rationalität, Theorieintegration* (S. 317-421). Münster: Aschendorff.
- Groeben, N. (2005). Sprachverarbeitung: Theorienkonkurrenz oder Gegenstandsflexibilität? *Zeitschrift für Psychologie*, 213, 152-158.
- Groeben, N. (2006). Historische Entwicklung der Sprachpsychologie. In J. Funke & P.A. Frensch (Hrsg.), *Handbuch der Allgemeinen Psychologie – Kognition* (S. 575-583). Göttingen: Hogrefe.
- Groeben, N. & Christmann, U. (1997). Sprache. In J. Straub, W. Kempf & H. Werbik (Hrsg.), *Psychologie. Eine Einführung* (S. 345-373). München: dtv.
- Groeben, N. & Scheele, B. (1997). *Argumente für eine Psychologie des reflexiven Subjekts. Paradigmawechsel vom behavioralen zum epistemologischen Menschenbild*. Darmstadt: Steinkopff.
- Grosz, B.J., Joshi, A.K. & Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, 21, 203-225.
- Guéraud, S. & O'Brien, E.J. (Eds.). (2005a). Convergence between memory-based processes and explanation-based processes [Themenheft]. *Discourse Processes*, 39 (2&3).

- Guéraud, S. & O'Brien, E.J. (2005b). Components of comprehension: A convergence between memory-based processes and explanation-based processes. *Discourse Processes*, 39, 123-124.
- Guerry, M., Gimenes, M., Caplan, D. & Rigalleau, F. (2006). How long does it take to find a cause? An online investigation of implicit causality in sentence production. *Quarterly Journal of Experimental Psychology*, 59, 1535-1555.
- Gumbel, E.J. (1958). *Statistics of Extremes*. New York: Columbia University Press.
- Haberlandt, K. (1993). Understanding mental models and inferences. *Psychology*, 4 (<http://journals.ecs.soton.ac.uk/resource/psychology>, Datum letzter Zugriff: 17.7.2007).
- Haberlandt, K. (1994). Methods in reading research. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1-31). San Diego: Academic Press.
- Haberlandt, K.F. & Graesser, A.C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General*, 114, 357-374.
- Haberlandt, K.F., Graesser, A.C. & Schneider, N.J. (1989). Reading strategies of fast and slow readers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 815-823.
- Haberlandt, K.F., Graesser, A.C., Schneider, N.J. & Kiely, J. (1986). Effects of task and new arguments on word reading times. *Journal of Memory and Language*, 25, 314-322.
- Hallinan, A.J. (1993). A review of the Weibull distribution. *Journal of Quality Technology*, 25, 85-93.
- Hankamer, J. & Sag, I.A. (1976). Deep and surface anaphora. *Linguistic Inquiry*, 7, 391-428.
- Harris, T. Wexler, K. & Holcomb, P. (2000). An ERP investigation of binding and coreference. *Brain and Language*, 75, 313-346.
- Harter, H.L. & Moore, A.H. (1967). Asymptotic variances and covariances of maximum-likelihood estimators, from censored samples, of the parameters of Weibull and Gamma populations. *Annals of Mathematical Statistics*, 38, 557-570.
- Härtl, H. (1999). "fürchten" vs. „ängstigen“: Thematische Rollen und Ereignisstrukturen psychischer Verben in einem Modell der Sprachproduktion. In I. Wachsmuth & B. Jung (Eds.), *KogWis 99: Proceeding der 4. Fachtagung der Gesellschaft für Kognitionswissenschaft* (S. 189-194). St. Augustin: Infix.
- Härtl, H. (2001). *CAUSE and CHANGE. Thematische Relationen und Ereignisstrukturen in Konzeptualisierungen und Grammatikalisierungen*. Berlin: Akademie-Verlag.
- Hasher, L. & Zacks, R.T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, 108, 356-388.

- Haviland, S.E. & Clark, H.H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, *13*, 512-521.
- Heathcote, A. (1996). RTSYS: A DOS application for the analysis of reaction time data. *Behavior Research Methods, Instruments & Computers*, *28*, 427-445.
- Heathcote, A. (2004). Fitting the Wald and ex-Wald distributions to response time data: An example using functions for the S-PLUS package. *Behavior Research Methods, Instruments & Computers*, *36*, 678-694.
- Heathcote, A. & Brown, S. (2004). Reply to Speckman and Rouder: A theoretical basis for QML. *Psychonomic Bulletin & Review*, *11*, 577-578.
- Heathcote, A., Popiel, S.J. & Mewhort, D.J.K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, *109*, 340-347.
- Heathcote, A., Brown, S. & Mewhort, D.J.K. (2002). Quantile Maximum Likelihood Estimation of response time distributions. *Psychonomic Bulletin & Review*, *9*, 394-401.
- Heathcote, A., Brown, S. & Cousineau, D. (2004). QMPE: Estimating lognormal, Wald, and Weibull RT distributions with a parameter-dependent lower bound. *Behavior Research Methods, Instruments, & Computers*, *36*, 277-290.
- Heckhausen, H. (1989). *Motivation und Handeln* (2., überarb. Aufl.). Berlin: Springer.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Heider, F. & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*, 243-259.
- Heim, I. (1983). File change semantics and the familiarity theory of definiteness. In R. Bäuerle, C. Schwarze & A. v. Stechow (Eds.), *Meaning, use, and interpretation of language* (pp. 164-189). Berlin: de Gruyter.
- Hempel, C.G. & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, *15*, 135-175.
- Herrmann, T. (1976). *Die Psychologie und ihre Forschungsprogramme*. Göttingen: Hogrefe.
- Herrmann, T. (1979). *Psychologie als Problem*. Stuttgart: Klett-Cotta.
- Herrmann, T. (1994). Forschungsprogramme. In T. Herrmann & W. Tack (Hrsg.), *Enzyklopädie der Psychologie, Themenbereich B, Serie I, Bd. 1 Methodologische Grundlagen der Psychologie* (S. 251-294). Göttingen: Hogrefe.
- Herrmann, T. (1995). *Allgemeine Sprachpsychologie. Grundlagen und Problem* (2., überarbeitete Aufl.). Weinheim: Beltz.
- Herrmann, T. (2005). *Sprache verwenden. Funktionen – Evolution – Prozesse*. Stuttgart: Kohlhammer.

- Hintzman, D. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review*, 93, 411-428.
- Hirst, W. & Brill, G.A. (1980). Contextual aspects of pronoun assignment. *Journal of Verbal Learning and Verbal Behavior*, 19, 168-175.
- Hobbs, J.R. (1978). Resolving pronoun references. *Lingua*, 44, 311-338.
- Hobbs, J.R. (1979). Coherence and coreference. *Cognitive Science*, 3, 67-90.
- Hockley, W.E. (1982). Retrieval processes in continuous recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 497-512.
- Hockley, W.E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental psychology: Learning, Memory, and Cognition*, 10, 598-615.
- Hoffman, C. & Tchir, M.A. (1990). Interpersonal verbs and dispositional adjectives: the psychology of causality embodied in language. *Journal of Personality and Social Psychology*, 58, 765-778.
- Hohle, R.H. (1965). Inferred components of reaction times as functions of foreperiods duration. *Journal of Experimental Psychology*, 69, 382-386.
- Holla, M.S. (1966). Bayesian estimates of reliability function. *Australian Journal of Statistics*, 8, 32-35.
- Huang, Y. (2000). *Anaphora. A cross-linguistic study*. Oxford: Oxford University Press.
- Hudson, S.B., Tanenhaus, M.K., & Dell, G.S. (1986). The effect of the discourse center on the local coherence of a discourse. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society* (pp. 96-101). Hillsdale: Erlbaum.
- Ida, M. (1980). The application of the Weibull distribution to the analysis of the reaction time data. *Japanese Psychological Research*, 22, 207-212.
- Jackendoff, R.S. (1972). *Semantic interpretation in generative grammar*. Cambridge: MIT Press.
- Jackendoff, R.S. (1983). *Semantics and cognition*. Cambridge: MIT Press.
- Jackendoff, R.S. (1990). *Semantic structures*. Cambridge: MIT Press.
- Jackendoff, R.S. (2002). *Foundations of language*. Oxford: Oxford University Press.
- Jackendoff, R.S. (2007). *Language, consciousness, culture. Essays on mental structure*. Cambridge, MA: MIT Press.
- Jarvela, R.J., Lundberg, I. & Bromley, H.J. (1989). How immediate is language understanding? Investigating reading in real time. *Reading and Writing*, 2, 103-122.

- Johnson, N.L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions* (2nd ed., vol. 2). New York: Wiley.
- Juhel, J. (1993). Should we take the shape of reaction time distributions into account when studying the relationship between RT and psychometric intelligence? *Personality and individual differences*, 15, 357-360.
- Just, M.A. & Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329-354.
- Just, M.A. & Carpenter, P.A. (1987). *The psychology of reading and language comprehension*. Boston: Allyn & Bacon.
- Just, M.A. & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.
- Just, M.A., Carpenter, P.A. & Woolley, J.D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228-238.
- Kager, R. (1999). *Optimality theory*. Cambridge: Cambridge University Press.
- Kailuweit, R. (2004). Protorollen und Makrorollen. In R. Kailuweit & M. Hummel (Hrsg.), *Semantische Rollen* (S. 83-103). Tübingen: Narr.
- Kamp, H. & Reyle, U. (1993). *From discourse to logic. Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory. Part 1*. Dordrecht: Kluwer.
- Kant, I. (1787/1998). *Kritik der reinen Vernunft. Nach der 1. und 2. Originalausgabe*. Leipzig: Meiner.
- Kasof, J. & Lee, J.Y. (1993). Implicit causality as implicit salience. *Journal of Personality and Social Psychology*, 65, 877-891.
- Keenan, J.M. (1993). Thoughts about the minimalist hypothesis. *Psychology*, 4 (<http://journals.ecs.soton.ac.uk/resource/psychology>, Datum letzter Zugriff: 17.7.2007).
- Kehler, A. (1997). Current theories of centering for pronoun interpretation: A critical review. *Computational Linguistics*, 23, 467-475.
- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford: CSLI.
- Kehler, A. (2004). Discourse topics, sentence topics, and coherence. *Theoretical Linguistics*, 30, 227-240.
- Kelley, H.H. (1967). Attribution theory in social psychology. *Nebraska Symposium on Motivation*, 15, 192-238.
- Kelley, H.H. (1972). *Causal schemas and the attribution process*. New York: General Learning Press.

- Kelly, G.A. (1955). *The psychology of personal constructs*. New York: Norton.
- Kennedy, A. & Murray, W.S. (1984). Inspection times for words in syntactically ambiguous sentences under three presentation conditions. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 833-849.
- Kieras, D.E. (1981). Component processes in the comprehension of simple prose. *Journal of Verbal Learning and Verbal Behavior*, 20, 1-23.
- Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale: Erlbaum.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163-182.
- Kintsch, W. (1994). The psychology of discourse processing. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 721-739). San Diego: Academic Press.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Kintsch, W. (2005). An overview of top-down and bottom-up effects in comprehension: The CI perspective. *Discourse Processes*, 39, 125-128.
- Klein, K. & Kutscher, S. (2005). *Lexical economy and case selection of psych-verbs in German*. Manuskript, Universität Bochum.
- Klin, C.M. & Myers, J.L. (1993). Reinstatement of causal information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 554-560.
- Knoeferle, P. & Crocker, M.W. (in press). The influence of recent scene events on spoken comprehension: evidence from eye movements. *Journal of Memory and Language*.
- Koornneef, A.W. & van Berkum, J.J.A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, 54, 445-465.
- Kuhn, T.S. (1973). *Die Struktur wissenschaftlicher Revolutionen*. Frankfurt a.M.: Suhrkamp.
- Kutas, M., van Petten, C.K. & Kluender, R. (2006). Psycholinguistics electrified II (1994-2005). In M.J. Traxler & M.A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 659-724). San Diego: Academic Press.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgraves (Eds.), *Criticism and the growth of knowledge* (pp. 91-196). Cambridge: Cambridge University Press.
- Langston, M. & Trabasso, T. (1999). Modeling causal integration and availability of information during comprehension of narrative texts. In H. van Oostendorp & S.R.

- Goldman (Eds.), *The construction of mental representations during reading* (pp. 29-70). Mahwah: Erlbaum.
- Lee, P.M. (2004). *Bayesian statistics. An introduction* (3rd ed.). London: Hodder Arnold.
- Leonard, C.L., Waters, G.S. & Caplan, D. (1997). The use of contextual information by right brain-damaged individuals in the resolution of ambiguous pronouns. *Brain and Language*, 57, 309-342.
- Leth-Steensen, C., King Elbaz, Z. & Douglas, V.I. (2000). Mean response times, variability, and skew in the responding of ADHD children: a response time distributional approach. *Acta Psychologica*, 104, 167-190.
- Levin, B. (1993). *English verb classes and alternations. A preliminary investigation*. Chicago: Chicago University Press.
- Levine, W.H., Guzmán, A.E. & Klin, C.M. (2000). When anaphor resolution fails. *Journal of Memory and Language*, 43, 594-617.
- Linderholm, T., Virtue, S., Tzeng, Y. & van den Broek, P. (2004). Fluctuations in the availability of information during reading: Capturing cognitive processes using the Landscape Model. *Discourse Processes*, 37, 165-186.
- Lindley, D.V. (1980). Bayesian approximation methods. *Trabajos de estadística*, 31, 223-237.
- Logan, G.D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492-527.
- Logan, G.D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 883-914.
- Logan, G.D. (1995). The Weibull distribution, the power law, and the instance theory of automaticity. *Psychological Review*, 102, 751-756.
- Lohnstein, H. (1996). *Formale Semantik und natürliche Sprache*. Wiesbaden: Westdeutscher Verlag.
- Long, D.L. & De Ley, L. (2000). Implicit causality and discourse focus: The interaction of text and reader characteristics in pronoun resolution. *Journal of Memory and Language*, 42, 545-570.
- Long, D.L. & Lea, R.B. (2005). Have we been searching for meaning in all the wrong places? Defining the “search after meaning” principle in comprehension. *Discourse Processes*, 39, 279-298.
- Luce, R.D. (1986). *Response times. Their role in inferring elementary mental organization*. Oxford: Oxford University Press.

- Luce, R.D. (1995). Four tensions concerning mathematical modelling in psychology. *Annual Review of Psychology*, *46*, 1-26.
- Lutz, M.F. & Radvansky, G. (1997). The fate of completed goal information in narrative comprehension. *Journal of Memory and Language*, *36*, 293-310.
- MacDonald, M.C. (1994). Probabilistic constraints and syntactic ambiguity. *Language and Cognitive Processes*, *9*, 157-201.
- MacDonald, M.C., Pearlmutter, N.J. & Seidenberg, M.S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676-703.
- MacLeod, C.M. (1991). Half a century of research on the Stroop effect: An integrative view. *Psychological Bulletin*, *109*, 163-203.
- Magliano, J.P. & Radvansky, G.A. (2001). Goal coordination in narrative comprehension. *Psychonomic Bulletin & Review*, *8*, 372-376.
- Magliano, J.P. & Graesser, A.C. (1991). A three-pronged method for studying inference generation in literary text. *Poetics*, *20*, 193-232.
- Magliano, J.P., Graesser, A.C., Eymard, L.A., Haberlandt, K. & Gholson, B. (1993). Locus of interpretive and inference processes during text comprehension: A comparison of gaze durations and word reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 704-709.
- Majid, A., Sanford, A.J. & Pickering, M.J. (2006). The linguistic description of minimal social scenarios affects the extent of causal inference making. *Journal of Experimental Social Psychology*.
- Malinverno, A. & Briggs, V.A. (2004). Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and Empirical Bayes. *Geophysics*, *69*, 1005-1016.
- Malt, B.C. (1985). The role of discourse structure in understanding anaphora. *Journal of Memory and Language*, *24*, 271-289.
- Mani, I., Pustejovsky, J. & Gaizauskas, R. (Eds.). (2005). *The language of time. A reader*. Oxford: Oxford University Press.
- Marley, A.A.J. & Colonius, H. (1992). The "horse race" random utility model for choice probabilities and reaction times, and its competing risks interpretation. *Journal of Mathematical Psychology*, *36*, 1-20.
- Marslen-Wilson, W., Levy, E. & Tyler, L. (1982): Producing interpretable discourse: The establishment and maintenance of reference. In R.J. Jarvella & W. Klein (Eds.), *Speech, place, and action* (pp. 339-378). New York: Wiley.

- Marslen-Wilson, W.D. & Tyler, L.K., 1980. The temporal structure of spoken language understanding. *Cognition*, 8; 1-71.
- Marslen-Wilson, W.D., Tyler, L.K. & Koster, C. (1993). Integrative processes in utterance resolution. *Journal of Memory and Language*, 32, 647-666.
- Maslow, A. (1966). *The psychology of science: A reconnaissance*. New York: Harper & Row.
- Matthews, A. & Chodorow, M.S. (1988). Pronoun resolution in two-clause sentences: Effects of ambiguity, antecedent location, and depth of embedding. *Journal of Memory and Language*, 27, 245-260.
- McArthur, L.Z. (1972). The how and what of why: Some determinants and consequences of causal attributions. *Journal of Personality and Social Psychology*, 22, 171-188.
- McCawley, J.D. (1968). Lexical insertion in a transformational grammar without deep structure. In B. Darden, C.-J. N. Bailey & A. Davidson (Eds.), *Papers from the fourth meeting of the Chicago Linguistic Society* (pp. 56-75). Chicago: Chicago University Press.
- McClelland, J.L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86, 287-330.
- McDonald, J.L. & BacWhinney, B. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language*, 34, 543-566.
- McElree, B. & Bever, T.G. (1989). The psychological reality of linguistically defined gaps. *Journal of Psycholinguistics*, 18, 21-35.
- McGill, W. J. (1963). Stochastic latency mechanisms. In R.D. Luce & R.R. Bush (Eds.), *Handbook of mathematical psychology* (vol. 1, pp. 309-360). New York: Wiley.
- McKoon, G., Greene, S.B. & Ratcliff, R. (1993). Discourse models, pronoun resolution, and the implicit causality of verbs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1040-1052.
- McKoon, G. & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 82-91.
- McKoon, G. & Ratcliff, R. (1988). Contextually relevant aspects of meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 331-343.
- McKoon, G. & Ratcliff, R. (1989a). Inferences about contextually defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 1134-1146.
- McKoon, G. & Ratcliff, R. (1989b). Semantic associations and elaborative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 326-338.

- McKoon, G. & Ratcliff, R. (1992). Inferences during reading. *Psychological Review*, 99, 440-466.
- McKoon, G. & Ratcliff, R. (1995). The minimalist hypothesis: directions for research. In Weaver, C.A., Mannes, S. & Fletcher, C.R. (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 97-116). Hillsdale, NJ: Erlbaum.
- McKoon, G. & Ratcliff, R. (1998). Memory-based language processing: Psycholinguistic research in the 1990s. *Annual Review of Psychology*, 49, 25-42.
- McKoon, G., Gerrig, R. & Greene, S.B. (1996). Pronoun resolution without pronouns: The consequences of memory-based processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 919-932.
- McLeod, P. (1978). Does probe RT measure central processing demand? *Quarterly Journal of Experimental Psychology*, 30, 83-89.
- Mewhort, D.J.K., Braun, J.G. & Heathcote, A. (1992). Response time distributions and the Stroop task: A test of the Cohen, Dunbar, and McClelland (1990) model. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 872-882.
- Miller, J. (1988). A warning about median reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 539-543.
- Millis, K.K. & Just, M.A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*, 33, 128-147.
- Mitchell, D.C. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. In D. Kieras & M.A. Just (Eds.), *New methods in reading comprehension research* (pp. 69-89). Hillsdale, NJ: Erlbaum.
- Mitchell, D.C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 601-618). Hillsdale, NJ: Erlbaum.
- Mitchell, D.C. (1994). Sentence parsing. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 375-409). New York: Academic Press.
- Mitchell, D.C. (2004). On-line methods in language processing: Introduction and historical review. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension. Eyetracking, ERPs and beyond* (pp. 15-32). New York: Psychology Press.
- Mitchell, D.C., Cuetos, F., Corley, M.M.B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, 24, 469-488.

- Mitchell, D.C. & Green, D.W. (1978). The effects of context and content on immediate processing in reading. *Quarterly Journal of Experimental Psychology*, 30, 609-636.
- Moens, M. & Steedman, M. (1988). Temporal ontology and temporal reference. *Computational Linguistics*, 14, 15-28.
- Moyer, R.S. & Landauer, T.K. (1967). Time required for judgements of numerical inequality. *Nature*, 215, 1519-1520.
- Murphy, G.L. (1985). Processes of understanding anaphora. *Journal of Memory and Language*, 24, 290-303.
- Myers, J.L. & O'Brien, E.J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, 26, 131-157.
- Myers, J.L., O'Brien, E.J., Albrecht, J.E. & Mason, R.A. (1994). Maintaining global coherence during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 876-868.
- Myers, J.L., Shinjo, M. & Duffy, S.A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, 26, 453-465.
- Myung, I.J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47, 90-100.
- Myung, I.J. & Pitt, M.A. (2002). Mathematical modelling. In H. Pashler & J. Wixted (Ed.), *Stevens handbook of experimental psychology, Vol. 4: Methodology in experimental psychology* (pp. 429-460). New York: Wiley.
- Newell, A. & Rosenbloom, P.S. (1981). Mechanisms of skill acquisition and the law of practice. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale: Erlbaum.
- Newmeyer, F.J. (1996). *Generative linguistics. A historical perspective*. London: Routledge.
- Nicol, J.L. & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18, 5-19.
- Nicol, J.L. & Swinney, D.A. (2002). The psycholinguistics of anaphora. In A. Barss (Ed.), *Anaphora* (pp. 72-104). Boston: Blackwell.
- Nieuwland, M.S. & van Berkum, J.J.A. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research*, 1118, 155-167.
- Noordman, L.G.M. & Vonk, W. (1993). A more parsimonious version of minimalism in inferences. *Psychology*, 4 (<http://journals.ecs.soton.ac.uk/resource/psychology>, Datum letzter Zugriff: 17.7.2007).

- Noordman, L.G.M., Vonk, W. & Kempff, H.J. (1992). Causal inferences during reading of expository texts. *Journal of Memory and Language*, 31, 573-590.
- Oakhill, J.V., Garnham, A. & Vonk, W. (1989). The online construction of discourse models. *Language and Cognitive Processes*, 4, 263-286.
- O'Brien, E.J. (1987). Antecedent search processes and the structure of text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 278-290.
- O'Brien, E.J. & Albrecht, J.E. (1991). The role of context in accessing antecedents in text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 94-102.
- O'Brien, E.J. & Albrecht, J.E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 777-784.
- O'Brien, E.J., Albrecht, J.E., Hakala, C.M. & Rizzella, M.L. (1995). Activation and suppression of antecedents during reinstatement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 626-634.
- O'Brien, E.J., Duffy, S.A. & Myers, J.L. (1986). Anaphoric inference during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 346-352.
- O'Brien, E.J. & Myers, J.L. (1999). Text comprehension: A view from the bottom up. In S.R. Goldman, A.C. Graesser & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence* (pp. 35-53). Mahwah: Erlbaum.
- O'Brien, E.J., Plewes, P. & Albrecht, J.E. (1990). Antecedent retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 241-249.
- O'Brien, E.J., Rizzella, M.L., Albrecht, J.E. & Halleran, J.G. (1998). Updating a situation model: a memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1200-1210.
- Osgood, C. E. (1970). Interpersonal verbs and interpersonal behavior In J L Cowan (Ed.) *Studies in thought and language* (pp 133-228). Tucson: University of Arizona Press.
- Osterhout, L., Bersick, M. & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, 25, 273-285.
- Oxford's Advanced Learners Dictionary (1989). *Oxford advanced learner's dictionary of current English* (4th ed.). Oxford: Oxford University Press.
- Papoulis, A. & Pillai, S.U. (2002). *Probability, random variables and stochastic processes* (4th ed.). New York: McGraw-Hill.
- Perner, J. (1991). *Understanding the representational mind*. Cambridge: MIT Press.

- Pesetsky, D. (1987). Binding problems with experiencer verbs. *Linguistic Inquiry*, 18, 126-140.
- Pesetsky, D. (1995). *Zero syntax. Experiencers and cascades*. Cambridge: MIT Press.
- Pickering, M.J. (1999). Sentence comprehension. In S. Garrod & M.J. Pickering (Eds.), *Language processing* (pp. 123-153). Hove: Psychology Press.
- Pickering, M.J. & Maijd, A. (in press). What are implicit causality and consequentiality? *Language and Cognitive Processes*.
- Pickering, M.J. & van Gompel, P.G. (2006). Syntactic parsing. In M.J. Traxler & M.A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed.) (pp. 455-503). Amsterdam: Academic Press.
- Pitt, M.A. & Myung, I.J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, 6, 421-425.
- Pittner, K. (1994). Psychologische Verben und ihre Argumentstrukturen: drei Erklärungsansätze zur syntaktischen Variabilität der Experiencer-Rolle. *Sprache und Sprachen*, 14/15, 96-101.
- Plourde, C.E. & Besner, D. (1997). On the locus of the word frequency effect in visual word recognition. *Canadian Journal of Experimental Psychology*, 51, 181-194.
- Poesio, M., Sturt, P., Artstein, R. & Filik, R. (2006). Underspecification and anaphora: theoretical issues and preliminary evidence. *Discourse Processes*, 42, 157-175.
- Popper, K. (2002). *Logik der Forschung* (10. verb. und vermehrte Aufl.). Tübingen: Mohr.
- Posner, M.I. (1978). *Chronometric explorations of mind*. Hillsdale: Erlbaum.
- Posner, M.I. (2005). Timing the brain: Mental chronometry as a tool in neuroscience. *PLoS Biology*, 3, 204-206.
- Posner, M.I. & Snyder, Snyder, C.R.R. (1975). Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition: the Loyola symposium* (pp. 55-85). Hillsdale, NJ: Erlbaum.
- Postal, P. (1970). *Cross over phenomena*. New York: Holt, Rinehart & Winston.
- Primus, B. (1999a). Rektionsprinzipien. In H. Wegener (Hrsg.), *Deutsch kontrastiv* (S. 135-169). Tübingen: Stauffenberg.
- Primus, B. (1999b). *Cases and thematic roles – Ergative, accusative and active*. Tübingen: Niemeyer.
- Primus, B. (2002). Proto-Roles and Case Selection in Optimality Theory. *Theorie des Lexikons* 122. Arbeiten des Sonderforschungsbereichs 282. Düsseldorf: Universität.

- Primus, B. (2004). Proto-Rollen und Verbttyp: Kasusvariation bei psychischen Verben. In M. Hummel & R. Kaluweit (Hrsg.), *Semantische Rollen* (S. 377-401). Tübingen: Narr.
- Primus, B. (2006). Mismatches in semantic-role hierarchies and the dimensions of role semantics. In I. Bornkessel, M. Schlesewsky, B. Comrie & A.D. Friederici (Eds.), *Semantic role universals and argument linking. Theoretical typological and psycholinguistic perspectives* (pp. 53-87). Berlin: de Gruyter.
- Pustejovsky, J. (1988). The geometry of events. In C. Tenny (Ed.), *Studies in generative approaches to aspect*. Lexicon Project Working Papers 24, Cambridge: MIT.
- Pustejovsky, J. (1991). The syntax of event structure. *Cognition*, 41, 47-81.
- R Development Core Team. (2004). *R 2.01: A language and environment for statistical computing*. Wien: R Foundation for Statistical Computing.
- Raaijmakers, J.G.W., Schrijnemakers, J.M.C. & Gremmen, F. (1999). How to deal with “The language-as-fixed-effect-fallacy”: Common misconceptions and alternative solutions. *Journal of Memory and Language*, 41, 416-426.
- Radford, A. (1997). *Syntactic theory and the structure of English. A minimalist approach*. Cambridge: Cambridge University Press.
- Ramsay, J.O. & Silverman, B.W. (2002). *Applied functional data analysis. Methods and case studies*. New York: Springer.
- Ramsay, J.O. & Silverman, B.W. (2005). *Functional data analysis* (2nd ed.). New York: Springer.
- Rapp, D.N. & van den Broek, P. (2005). Dynamic text comprehension. *Current Directions in Psychological Science*, 14, 276-279.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59-108.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, 86, 446-461.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114, 510-532.
- Ratcliff, R., Gomez, P. & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, 111, 159-182.
- Ratcliff, R. & Murdock, B.B. (1976). Retrieval processes in recognition memory. *Psychological Review*, 83, 190-214.
- Ratcliff, R. & Rouder, J.N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9, 347-356.

- Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical linear models. Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage.
- Rayner, K., Kambe, G. & Duffy, S.A. (2000). The effect of clause wrap-up on movements during reading. *The Quarterly Journal of Experimental Psychology*, 53A, 1061-1080.
- Rayner, K., Sereno, S.C., Morris, R.K., Schmauder, A.R., & Clifton, C. (1989). Eye movements and online language comprehension processes. *Language and Cognitive Processes*, 4, 21-49.
- Read, T.R.C. & Cressie, N.A.C. (1988). *Goodness-of-fit statistics for discrete multivariate data*. New York: Springer.
- Richards, E. & Singer, M. (2001). Representation of complex goal structures in narrative comprehension. *Discourse Processes*, 31, 111-135.
- Richter, T. (2003). *Epistemologische Einschätzungen beim Textverstehen*. Lengerich: Pabst.
- Richter, T. & van Holt (2005). ELVES. Ein computergestütztes Diagnostikum zur Erfassung der Effizienz von Teilprozessen des Leseverstehens. *Diagnostica*, 51, 169-182.
- Rigalleau, F. & Caplan, D. (2000). Effects of gender marking in pronominal coindexation. *Quarterly Journal of Experimental Psychology*, 53, 23-52.
- Rigalleau, F., Caplan, D. & Baudiffier, V. (2004). New arguments in favour of an automatic gender pronominal process. *Quarterly Journal of Experimental Psychology*, 57, 893-933.
- Rizzalla, M.L. & O'Brien, E.J. (1996). Accessing global causes during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1208-1218.
- Roberts, S. & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, 107, 358-367.
- Rohrer, D. & Wixted, J.T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition*, 22, 511-524.
- Rotter, J.B. (1954). *Social learning and clinical psychology*. New York: Prentice-Hall.
- Rouder, J.N. (2005a). Are unshifted distributional models appropriate for response time? *Psychometrika*, 71, 377-381.
- Rouder, J.N. (2005b). *R syntax for the single population model of hierarchical parametric Bayesian framework*. Syntax routine courtesy provided by Jeff Rouder.
- Rouder, J.N. & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, 12, 573-604.

- Rouder, J.N. & Speckman, P.L. (2004). An evaluation of the Vincentizing method of forming group-level response time distributions. *Psychonomic Bulletin & Review*, *11*, 419-427.
- Rouder, J.N., Lu, J., Speckman, P., Sun, D. & Jiang, Y. (2005). A hierarchical model for estimation response time distributions. *Psychonomic Bulletin & Review*, *12*, 195-223.
- Rouder, J.N., Sun, D., Speckman, P.L., Lu, J. & Zhou, D. (2003). A hierarchical Bayesian statistical framework for response time distributions. *Psychometrika*, *68*, 589-606.
- Rudolph, U. (1997). Implicit verb causality: Verbal schemas and covariation information. *Journal of Language and Social Psychology*, *16*, 132-158.
- Rudolph, U. & Försterling, F. (1997a). The psychological causality implicit in verbs: A review. *Psychological Bulletin*, *121*, 192-218.
- Rudolph, U. & Försterling, F. (1997b). Zur impliziten Kausalität in Sprache: Kriterien zur Selektion von Stimulusmaterial in Studien zur Verbkausalität. *Zeitschrift für Experimentelle Psychologie*, *44*, 293-304.
- Rudolph, U. & Spörrle, M. (1999). Alter, Attraktivität und Intelligenz von Vornamen: Wortnormen für Vornamen im Deutschen. *Zeitschrift für Experimentelle Psychologie*, *46*, 115-128.
- Rudolph, U. & von Hecker, U. (1997). Die Erklärung interpersonalen Ereignisse: Zur Bedeutung von Balanciertheit und Kausalität. *Zeitschrift für Experimentelle Psychologie*, *44*, 246-265.
- Rudolph, U. & von Hecker, U. (2006). Three principles of explanation. Verb schemas, balance, and imbalance repair. *Journal of Language and Social Psychology*, *25*, 1-29.
- Rudolph, U., Böhm, R. & Lummer, M. (2006). Ein Name sagt mehr als 1000 Worte: Zur sozialen Wahrnehmung von Vornamen. *Zeitschrift für Sozialpsychologie*, *38*, 17-32.
- Salthouse, T.A. & Hedden, T. (2002). Interpreting reaction time measures in between-group comparisons. *Journal of Clinical and Experimental Neuropsychology*, *24*, 858-872.
- Sanders, T.J.M. & Gernsbacher, M.A. (2004). Accessibility in text and discourse processing. *Discourse Processes*, *37*, 79-89.
- Sanford, A.J. & Garrod, S.C. (1981). *Understanding written language*. Chichester: Wiley.
- Sanford, A.J. & Garrod, S.C. (1989). What, when, and how? Questions of immediacy in anaphoric reference resolution. *Language and Cognitive Processes*, *4*, 235-262.
- Sanford, A.J. & Garrod, S.C. (1994). Selective processing in text understanding. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 699-719). San Diego: Academic Press.

- Sanford, A.J. & Garrod, S.C. (1998). The role of scenario mapping in text comprehension. *Discourse Processes*, 26, 159-190.
- Sanford, A.J. & Garrod, S.C. (2005). Memory-based approaches and beyond. *Discourse Processes*, 39, 205-224.
- Sanford, A.J., Garrod, S., Lucas, A. & Henderson, R. (1983). Pronouns without explicit antecedents? *Journal of Semantics*, 2, 303-318.
- Sanford, A.J. & Graesser, A.C. (2006). Shallow processing and underspecification. *Discourse Processes*, 42, 99-108.
- Sanford, A.J., Moar, K. & Garrod, S.C. (1988). Proper names as controllers of discourse focus. *Language and Speech*, 31, 43-56.
- Sanford, A.J. & Graesser, A.C. (2005). Shallow processing and underspecification. *Discourse Processes*, 42, 99-108.
- Schneider, W., Eschman, A. & Zuccolotto, A. (2002). E-Prime Reference Guide. Pittsburgh: Psychology Software Tools.
- Schlesinger, I.M. (1992). The experiencer as an agent. *Journal of Memory and Language*, 31, 315-332.
- Schneider, W. & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1-66.
- Schwarz, W. (2001). The ex-Wald distribution as a descriptive model of response times. *Behavior Research Methods, Instruments & Computers*, 33, 457-469.
- Semin, G.R. & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology*, 54, 558-568.
- Semin, G.R. & Marsman, J.G. (1994). „Multiple inference-inviting properties“ of interpersonal verbs: event instigation, dispositional inference, and implicit causality. *Journal of Personality and Social Psychology*, 67, 836-849.
- Shiffrin, R.M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127-190.
- Shillcock, R. (1982). The on-line resolution of pronominal anaphora. *Language and Speech*, 25, 385-401.
- Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.

- Springston, J. (1976). *Some cognitive aspects of presupposed coreferential anaphora*. Unpublished PhD thesis, Stanford University.
- Simner, J. & Pickering, M.J. (2005). Planning causes and consequences in discourse. *Journal of Memory and Language*, 52, 226-239.
- Singer, M. (1993). Minimalism: a hedged analysis of restricted inference processing. *Psychology*, 4. (<http://journals.ecs.soton.ac.uk/resource/psychology>, Datum letzter Zugriff: 17.7.2007).
- Singer, M. (1994). Discourse inference processes. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 479-515). San Diego: Academic Press.
- Singer, M., Graesser, A.C. & Trabasso, T. (1994). Minimal or global inferences during reading. *Journal of Memory and Language*, 33, 421-441.
- Singer, M., Halldorson, M., Lear, J.C. & Andrusiak, P. (1992). Validation of causal bridging inferences. *Journal of Memory and Language*, 31, 507-524.
- Singer, M. & Richards, E. (2005). Representing complex narrative goal structures: competing memory-based and situational influences. *Discourse Processes*, 39, 189-204.
- Sinha, S.K. & Sloan, J.A. (1988). Bayes estimation of the parameters and reliability function of the 3-parameter weibull distribution. *IEEE Transactions on Reliability*, 37, 364-369.
- Sivia, D.S. & Skilling, J. (2006). *Data analysis. A Bayesian tutorial* (2nd ed.). Oxford: Oxford University Press.
- Skinner, B.F. (1938). *The behavior of organisms: An experimental analysis* New York: Appelton-Century-Croft.
- Smith, R.L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72, 67-90.
- Smith, E. & Hancox, P. (2001). Representation, coherence and inference. *Artificial Intelligence Review*, 15, 295-323.
- Smith, R.L. & Naylor, J.C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Applied Statistics*, 36, 358-369.
- Speckman, P.L. & Rouder, J.N. (2004). A comment on Heathcote, Brown and Mewhort's QMLE method for response time distributions. *Psychonomic Bulletin & Review*, 11, 574-576.
- Spieler, D.H., Balota, D.A. & Faust, M.E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 461-479.

- Spieler, D.H., Balota, D.A. & Faust, M.E. (2000). Levels of selective attention revealed through analyses of response time distributions. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 506-526.
- Soland, R.M. (1968). Bayesian analysis of the Weibull process with unknown scale parameters and its applications to acceptance sampling. *IEEE Transactions on Reliability*, 17, 84-90.
- Soland, R.M. (1969). Bayesian analysis of the Weibull process with unknown shape and scale parameters. *IEEE Transactions on Reliability*, 18, 181-184.
- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, 153, 652-654.
- Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 57, 421-457.
- Stevenson, R., Crawley, R.A. & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, 9, 519-548.
- Stevenson, R., Knott, A., Oberlander, J. & McDonald, S. (2000). Interpreting pronouns and connectives: Interactions among focusing, thematic roles and coherence relations. *Language and Cognitive Processes*, 15, 225-262.
- Stevenson, R.J., & Vitkovitch, M. (1986). The comprehension of onophoric relations. *Language and Speech*, 29, 335-360.
- Stewart, A.J., Pickering, M.J. & Sanford, A.J. (1998). Implicit consequentiality. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 1031-1036). Mahwah: Erlbaum.
- Stewart, A.J., Pickering, M.J. & Sanford, A.J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory and Language*, 42, 423-443.
- Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Sturt, P. (2002a). A new look at the syntax-discourse interface: The use of binding principles in sentence processing. *Journal of Psycholinguistic Research*, 32, 125-139.
- Sturt, P. (2003b). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542-562.
- Suh, S. & Trabasso, T. (1993). Inferences during reading: Converging evidence from discourse analysis, talk-aloud protocols, and recognition priming. *Journal of Memory and Language*, 32, 279-300.

- Talmy, L. (1976). Semantic causative types. In M. Shibatani (Ed.), *The grammar of causative constructions* (pp. 43-116). New York: Academic Press.
- Tanenhaus, M.K. (2004). On-line sentence processing: past, present, and future. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension. Eyetracking, ERPs and beyond* (pp. 371-393). New York: Psychology Press.
- Taraban, R. & McClelland, J.R. (1988). Constituent attachment and thematic role assignment in sentence processing: Influence of content-based expectations. *Journal of Memory and Language*, 27, 597-632.
- Thibadeau, R., Just, M.A. & Carpenter, P.A. (1982). A model of the time course and content of reading. *Cognitive Science*, 6, 157-203.
- Thomas, E.A.C., & Ross, B. (1980). On appropriate procedures for combining probability distributions within the same family. *Journal of Mathematical Psychology*, 21, 136-152.
- Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, 22, 1701-1728.
- Townsend, J.T. (1972). Some results concerning the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology*, 25, 168-199.
- Townsend, J.T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science*, 1, 46-52.
- Traxler, M.J., Bybee, M.D. & Pickering, M.J. (1997a). Influence of connectives on language comprehension: Eye-tracking evidence for incremental interpretation. *Quarterly Journal of Experimental Psychology*, 50A, 481-497.
- Traxler, M.J., Sanford, A.J., Asked, J.P. & Moxey, L.M. (1997b). Processing causal and diagnostic statements in discourse. *Journal of Experimental Psychology: Learning, memory, and Cognition*, 23, 88-101.
- Traxler, M.J., Pickering, M.J., & Clifton, C. Jr. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, 39, 558-592.
- Trueswell, J., Tanenhaus, M.K. & Garnsey, S.M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285-313.
- Trueswell, J., Tanenhaus, M.K. & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 528-553.

- Tyler, L.K. & Marslen-Wilson, W.D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, 16, 683-692.
- Tyler, L.K. & Marslen-Wilson, W.D. (1982). Processing of utterances in discourse contexts: On-line resolution of anaphors. *Journal of Semantics*, 1, 297-314.
- Ulrich, R. & Miller, J. (1993). Information processing models generating lognormally distributed reaction times. *Journal of Mathematical Psychology*, 37, 513-525.
- Ulrich, R. & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General*, 123, 38-80.
- Ulrich, R. & Schröter, H. (2006). Mentale Chronometrie. In J. Funke & P.A. Frensch (Hrsg.), *Handbuch der Allgemeinen Psychologie – Kognition* (S. 685-690). Göttingen: Hogrefe.
- van Berkum, J.J.A. (2004). Sentence comprehension in a wider discourse: An we use ERPs to keep track of things? In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension. Eyetracking, ERPs and beyond* (pp. 229-270). New York: Psychology Press.
- van Berkum, J.J.A., Koornneef, A.W., Otten, M. & Nieuwland, M.S. (2006). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*, 1146, 158-171.
- vanBerkum, J.J.A., Brown, C.M. & Hagoort, P. (1999a). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language*, 41, 147-182.
- van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999b). When does gender constrain parsing? Evidence from ERPs. *Journal of Psycholinguistic Research*, 28(5), 555–571.
- van Breukelen, G.J.P. (1995). Parallel information processing models compatible with lognormally distributed response times. *Journal of Mathematical Psychology*, 39, 396-399.
- van den Broek, P. (1994). Comprehension and memory of narrative texts. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 539-588). San Diego: Academic Press.
- van den Broek, P., Rapp, D.N. & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes*, 39, 299-316.
- van den Broek, P., Ridsen, K., Fletcher, C.R. & Thurlow, R. (1996). A “landscape“ view of reading: Fluctuating patterns activation and the construction of a stable memory

- representation. In B.K. Britton & A.C. Graesser (Eds.), *Models of understanding text* (pp.165-187). Hillsdale: Erlbaum.
- van den Broek, P., Tzeng, Y. & Kendeou, P. (2004). *The computational implementation of the landscape model. Version 1.0. User's Manual*. Minneapolis: Minnesota.
- van den Broek, Young, M., Tzeng, Y. & Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of a memory representation. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71-98). Mahwah: Erlbaum.
- van Dijk, T.A. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- van Valin, R.D. & LaPolla, R. (1997). *Syntax: structure, meaning and function*. Cambridge: Cambridge University Press.
- van Gompel, R.P.G., Pickering, M.J. & Traxler, M.J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, 45, 225-258.
- van Gompel, R.P.G. & Liversedge, S.P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29,128-139.
- van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, 7, 424-465.
- van Zandt, T. (2002). Analysis of response time distributions. In H. Pashler & J. Wixted (Ed.), *Stevens handbook of experimental psychology, Vol. 4: Methodology in experimental psychology* (pp. 461-516). New York: Wiley.
- Vanables, W.N. & Ripley, B.D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.
- Vendler, Z. (1967). *Linguistics and Philosophy*. Ithaca: Cornell University Press.
- Vonk, W. (1984). Eye movements during comprehension of pronouns. In A.G. Gale & F. Johnson (Eds.), *Theoretical and applied aspects of eye movement research* (pp. 203-212). Amsterdam: Elsevier.
- Vonk, W. (1985a). On the purpose of reading and the immediacy of processing pronouns. In R. Groner, G.W. McKonkie & C. Menz (Eds.), *Eye movements and human information processing* (pp. 207-215). Amsterdam: North-Holland.

- Vonk, W. (1985b). The immediacy of inferences in the understanding of pronouns. In G. Rickheit & H. Strohner (Eds.), *Inferences in text processing* (pp. 205-217). Amsterdam: Elsevier.
- Vonk, W., Hustinx, L.G. & Simons, W.H.G. (1992). The use of referential expressions in structuring discourse. *Language and Cognitive Processes*, 7, 301-333.
- Walker, M.A., Joshi, A.K. & Prince, E.F. (1997). *Centering in discourse* Oxford: Oxford University Press.
- Wegener, H. (1999). Zum Bedeutungs- und Konstruktionswandel bei psychischen Verben. In H. Wegener (Hrsg.), *Deutsch kontrastiv* (S. 171-210). Tübingen: Stauffenberg.
- Weibull, W. (1939). *A statistical theory of strength of materials*. Stockholm: Proceedings of the Royal Academy of Engineering Science.
- Weibull, W. (1951). A statistical function of wide applicability. *Journal of Applied Mechanics*, 18, 293-297.
- Westermann, R. (2000). *Wissenschaftstheorie und Experimentalmethodik: Ein Lehrbuch zur psychologischen Methodenlehre*. Göttingen: Hogrefe.
- Wilcox, R.R. (1992). Why can methods for comparing means have relatively low power, and what can you do to correct the problem? *Current Directions in Psychological Science*, 1, 101-105.
- Wiley, J. & Myers, J.L. (2003). Availability and accessibility of information and causal inferences from scientific text. *Discourse Processes*, 36, 109-129.
- Wilks, Y. (1975). A preferential, pattern-seeking, semantics for natural language.inference. *Artificial Intelligence*, 6, 53-74.
- Winograd, T. (1972). *Understanding natural language*. New York: Academic Press.
- Wixted, J.T. & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1024-1039.
- Wunderlich, D. (1997). Cause and the structure of verbs. *Linguistic Inquiry*, 28, 27-68.
- Yap, M.J. & Balota, D.A. (2007). *Additive and interactive effects on response time distributions in visual word recognition*. Manuscript submitted for publication.
- Yap, M.J., Balota, D.A., Cortese, M.J. & Watson, J.M. (2006). Single- versus dual-process models of lexical decision performance: Insights from response time distributional analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 1324-1344.

- Zucchini, W. (2000). An introduction on model selection. *Journal of Mathematical Psychology*, 44, 41-61.
- Zwaan, R.A. & Graesser, A.C. (1993a). Reading goals and situation models. *Psychology*, 4. (<http://journals.ecs.soton.ac.uk/resource/psychology>, Datum letzter Zugriff: 17.7.2007).
- Zwaan, R.A. & Graesser, A.C. (1993b). There is no empirical evidence that some inferences are automatically or partially encoded in text comprehension. *Psychology*, 4. (<http://journals.ecs.soton.ac.uk/resource/psychology>, Datum letzter Zugriff: 17.7.2007).
- Zwaan, R.A. & Rapp, D.N. (2006). Discourse comprehension. In M.J. Traxler & M.A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed.) (pp. 725-764). Amsterdam: Academic Press.

Appendix A: Statistische Erläuterungen zu Kap. 3

A.1 Location-Scale-Verteilungen: Beispiel Normalverteilung

Eine der bekanntesten Verteilungen ist die sog. Normalverteilung. Weniger bekannt ist die korrespondierende Formel (A.1.1), nämlich ihre Wahrscheinlichkeitsdichtefunktion (pdf):

$$(A.1.1) \quad f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

π steht dabei für die bekannte (Rudolfsche) Kreiszahl (3.14...) und e für die Eulersche Zahl als Basis des natürlichen Logarithmus (2.71...). Die Parameter der Verteilung sind der Mittelwert μ und die Standardabweichung σ . Mathematisch schreibt man $X_i \sim N(\mu, \sigma)$, was sich folgendermaßen paraphrasieren lässt: X ist normalverteilt mit den Parametern μ und σ . Die bekannteste dieser Parameterkombinationen ist die sog. „Standardnormalverteilung“, bei der $\mu = 0$ und $\sigma = 1$. Durch (lineare) z -Transformation ($z = (x - \mu) / \sigma$) lassen sich alle anderen Normalverteilungen in diese Form bringen. Abb. A.1.1 zeigt verschiedene Normalverteilungen, die sich durch unterschiedliche Parameterkombinationen ergeben:

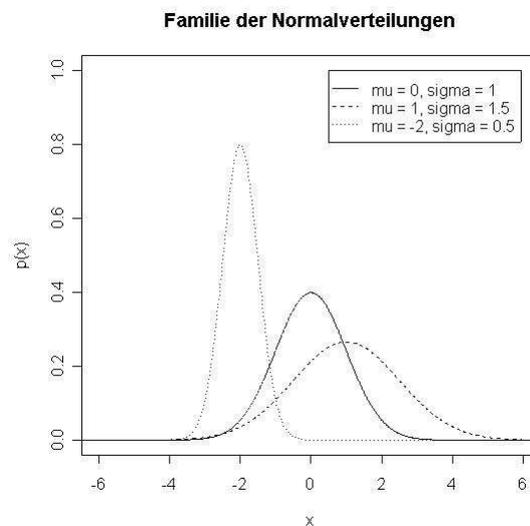


Abb. A.1.1: Verschiedene Normalverteilungen.

Da die Normalverteilung zwei freie Parameter (μ und σ) hat, gehört sie zur Familie der zweiparametrischen Verteilungen. Darüber hinaus ist sie eine sog. *location-scale-Verteilung*, weil einer der Parameter den „Ort“ der Verteilung in Relation zur x -Achse angibt (Lokationsparameter) und der andere ihre Skalierung, d.h. die „Breite“ der Verteilung (Skalierungsparameter). Im Fall der Normalverteilung ist μ der Lokationsparameter, da sich mit einer Veränderung von μ die gesamte Verteilung auf der x -Achse verschiebt (vgl. Abb. A.1.2).

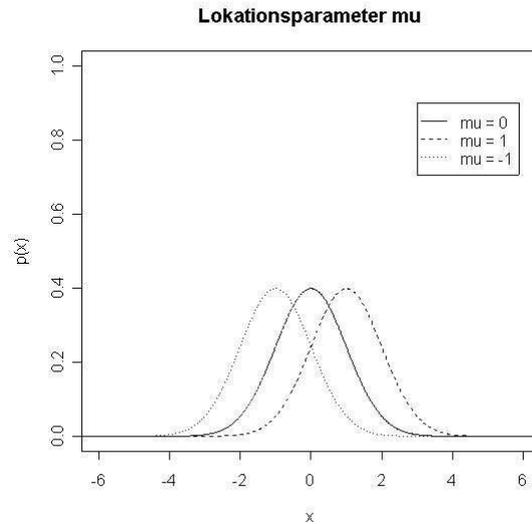


Abb. A.1.2: Einfluss von μ auf die Lokation einer Normalverteilung (bei fixiertem $\sigma = 1$).

Die Standardabweichung σ der Normalverteilung dient hingegen als ihr Skalierungsparameter. Je nachdem wie groß σ ausfällt, wird die Normalverteilung entweder „breiter“ oder „schmäler“ (vgl. Abb. A.1.3).

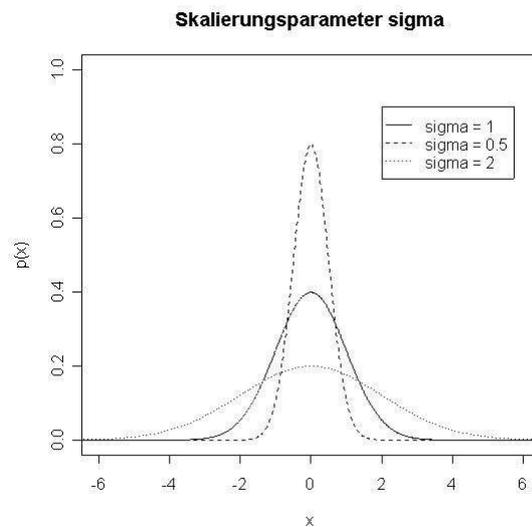


Abb. A.1.3: Einfluss von σ auf die Skalierung der Normalverteilung (bei fixiertem $\mu = 0$).

Den Begriffen „Lage“ und „Skalierung“ lässt sich auch eine präzise mathematische Bedeutung geben (vgl. Jiang et al., 2004, Rouder et al., 2005, Thomas & Ross, 1980):

Definition Location-Scale-Form: Es seien X_i verschiedene Zufallsvariablen, deren Verteilungsfunktionen jeweils durch zwei Parametern ψ_i und θ_i so bestimmt wird, dass gilt: $f(x_i | \psi_i, \theta_i)$. Die Verteilungen von X_i gehören dann und nur dann zu einer gemeinsamen Familie der location-scale-Verteilungen, wenn die Bedingungen a) und b) gleichzeitig erfüllt sind:

a) Es gibt eine Transformation $h(x_i) = z_i = \frac{x_i - \psi_i}{\theta_i}$ (wobei $-\infty < \psi_i < \infty$ und $\theta_i > 0$), die die Ausprägungen

einer Zufallsvariablen in eine standardisierte Normalform bringt. In diesem Fall nennt man ψ den *Lageparameter* und θ den *Skalierungsparameter* der Verteilungsfamilie.

b) Die normalisierten Zufallsvariablen X_i haben alle eine identische Wahrscheinlichkeitsdichtefunktion $f(x_i | \psi_i, \theta_i) = \theta_i^{-1} g(z_i)$ (vgl. Thomas & Ross, 1980, Theorem 1).

Erläuterung: Bedingung a) ist trivial und spezifiziert lediglich das Grundprinzip der location-scale-Form. Denn es ist unschwer zu erkennen, dass es sich bei der geforderten Transformation $h(x)$ lediglich um eine generalisierte Form der bekannten z -Transformation für die Normalverteilung handelt: Man nehme einen Datenpunkt und „zentriere“ ihn an der Lage der Verteilung und relativiere ihn an der Skalierung der Verteilung. Bei der z -Transformation ist der Lageparameter der Mittelwert und der Skalierungsparameter die Standardabweichung. Die oben genannte, generelle Definition für beliebige Verteilungsformen lässt jedoch auch ganz andere Parameter einer Verteilung als Lage- und Skalierungsparameter zu. Bedingung b) spezifiziert nun, dass die einzelnen Verteilungen der Zufallsvariablen wirklich *nur* in ihren Lage- und Skalierungsparametern divergieren dürfen. Oder anders formuliert: Die einzelnen Zufallsverteilungen müssen eine einheitliche Form haben, die durch die Funktion $g(z)$ bestimmt wird. Nur wenn diese Bedingung erfüllt ist, dann gehören die einzelnen Verteilungen zur gleichen Verteilungs-Familie.

Beispiele: Bei der Normalverteilung ist der Mittelwert μ der Lageparameter und die Standardabweichung σ der Skalierungsparameter, die Funktion $g(z)$ ist durch $g(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$ gegeben. Auch die Exponentialverteilung ist eine location-scale-Verteilung, bei der allerdings der Lageparameter auf 0 gesetzt wurde. Der Skalierungsparameter ist jedoch durch den inversen Rate-Parameter λ der Verteilung gegeben und die Verteilungsfunktion ist $g(z) = e^{-z}$.

Die in der oben genannten Definition genannten Bedingungen sind ziemlich allgemein und werden von vielen zweiparametrischen Funktionen erfüllt. Wie das Beispiel der Exponentialverteilung zeigt, lassen sich jedoch auch die meisten einparametrischen Funktionen als location-scale-Verteilung formulieren, wenn man annimmt, dass einer der beiden Parameter gleich 0 ist. Weitere Funktionen lassen sich durch die Fixierung einzelner Parameter u.U. auch in location-scale-Format bringen.

A.2 Location-Scale-Shape-Verteilungen: Beispiel Weibull-Verteilung

Das Konzept der „Form“ einer Verteilung scheint zwar intuitiv plausibel zu sein, ist jedoch mathematisch und statistisch gar nicht so leicht zu fassen. Bezeichnender Weise definieren z.B. Mosteller und Tukey (1977, pp. 8ff.; vgl. auch Ratcliff, 1978) den Begriff der Form lediglich ex negativo: Die Form einer Verteilung ist das, was übrig bleibt, wenn man von ihrer genauen Lokation und Skalierung abstrahiert.

Beispiel: Wenn man in Abb. A.1.1 alle Zahlen unter der x-Achse entfernt, dann wissen wir nicht mehr, wo diese Verteilung genau zu „verorten“ ist (ob es z.B. ein Intelligenztest ist, der auf den Wert 50 oder auf den Wert 100 normiert ist), und wir wissen auch nicht, wie breit diese Verteilung „in Wirklichkeit“ ist (ob z.B. der Intelligenztest eine Standardabweichung von 10 oder von 15 hat). Man kann also sowohl den Lokations- als auch den Skalierungsparameter beliebig ändern, ohne dass eine Funktion aufhört, eine Normalverteilung zu sein. Das, was aber für die Normalverteilung charakteristisch ist (sozusagen das „ \mathcal{N} “ in dem Ausdruck „ $\mathcal{N}(\mu, \sigma)$ “) ist ihre Form, die vom gesamten „Rest“ von Formel (1) bestimmt ist, nämlich wie μ und σ miteinander verknüpft sind.

Während also die Form der Normalverteilung gewissermaßen fixiert ist, gibt es auch Verteilungen, deren Form variabel ist, die sich – unabhängig von Lage und Skalierung – in Abhängigkeit von weiteren Parametern ändert. Solche Verteilungen verletzen die Bedingung b) der oben genannten Definition der location-scale-Verteilungen. Denn diese besagt, dass sich solche Verteilungen nach Abzug von Lage- und Skalierungsinformationen nicht mehr voneinander unterscheiden dürfen. Durch eine Liberalisierung der Bedingung b) ergibt sich dann die Klasse der sog. location-scale-shape-Verteilungen (vgl. Rouder et al., 2005, Fn. 3):

Definition location-scale-shape-Verteilung: Es seien X_i verschiedene Zufallsvariable, deren Verteilungsfunktionen jeweils durch die Parameter $\psi_i, \theta_i, \beta_{i1} \dots \beta_{in}$ so bestimmt wird, dass gilt: $f(x_i | \psi_i, \theta_i, \beta_{i1} \dots \beta_{in})$. Die Verteilungen von X_i gehören dann und nur dann zu einer gemeinsamen Familie von location-scale-shape-Verteilungen, wenn die Bedingungen a) und b) gleichzeitig erfüllt sind:

a) Es gibt eine Transformation $h(x_i) = z_i = \frac{x_i - \psi_i}{\theta_i}$ (wobei $-\infty < \psi_i < \infty$ und $\theta_i > 0$), die die Ausprägungen

einer Zufallsvariablen in eine standardisierte Normalform bringt. In diesem Fall nennt man ψ den *Lageparameter* und θ den *Skalierungsparameter* der Verteilungsfamilie.

b) Die normalisierten Zufallsvariablen X_i haben alle die gleiche bedingte Wahrscheinlichkeitsdichtefunktion $f(x_i | \psi_i, \theta_i, \beta_{i1} \dots \beta_{in}) = \theta_i^{-1} g(z_i | \beta_{i1} \dots \beta_{in})$. In diesem Fall nennt man $\beta_{i1} \dots \beta_{in}$ die *Formparameter* der Verteilung.

Location-scale-shape-Verteilungen unterscheiden sich von den location-scale-Verteilungen also eigentlich nur darin, dass Bedingung b) nun auch *bedingte* Wahrscheinlichkeitsdichtefunktionen zulässt. Das heißt, die Verteilungsformen der (um Lage- und Skalierungsinformationen bereinigten) normalisierten Zufallsvariablen müssen nicht mehr einheitlich verlaufen,

sondern dürfen sich in Abhängigkeit von bestimmten Formparametern unterscheiden. Dadurch ist diese Klasse von Verteilungen zwar wesentlich flexibler als die restringiertere Klasse der location-scale-Verteilungen. Dafür verliert sie jedoch die teilweise sehr günstigen statistischen Eigenschaften dieser Verteilungs-Familie (z.B. was die Aggregation von Einzelverteilungen betrifft, vgl. Jiang et al., 2004).

Beispiel: Die dreiparametrische Weibull-Funktion besitzt folgende kumulative Wahrscheinlichkeitsdichte (vgl.

Abschnitt 3.3): $F(x; \psi, \theta, \beta) = 1 - e^{-\left(\frac{x-\psi}{\theta}\right)^\beta}$. Ganz offensichtlich findet sich eine Transformation, sodass

$z = \frac{x-\psi}{\theta}$. Damit ist Bedingung a) der obigen Definition erfüllt und ψ ist der Lage- und θ der Skalierungsparameter der Verteilung. Die resultierende Verteilungsfunktion ist dann $g(z) = 1 - e^{-z^\beta}$.

Diese Verteilungsfunktion verstößt gegen die Bedingung b) in der Definition der Familie der location-scale-Verteilungen, weil $g(z)$ nicht nur von z , sondern auch vom Parameter β abhängig ist. Solange man also β nicht kennt, kann man wenig über die Form von $g(z)$ sagen: Die Form dieser Funktion ist flexibel, es ist eigentlich eine Funktionenschar. Eine Möglichkeit wäre, β auf einen bestimmten Wert (sagen wir 2) zu fixieren. Die Verteilungsfunktion der dreiparametrischen Weibull-Funktion verstößt jedoch nicht gegen Bedingung b) der Definition der Familie der location-scale-shape-Verteilungen. Denn diese lässt zusätzliche Parameter, welche die Funktion $g(z)$ mit bestimmen ausdrücklich zu. Da β damit sozusagen die Form der Verteilung steuert, wird dieser Parameter vollkommen zu Recht Formparameter genannt.

Obwohl die Klasse der location-scale-shape-Verteilungen vergleichsweise hochinklusiv aussieht, gibt es Funktionen, die sich nicht oder nicht ohne weiteres in dieses Format bringen lassen. Ein Beispiel hierfür ist die Ex-Gauss-Verteilung, die in ihrer Original-Parametrisierung nicht location-scale- und damit auch nicht location-scale-shape-kompatibel ist (vgl. Abschnitt 3.3.2). Allerdings ist es möglich, für diese Funktion eine alternative Parametrisierung zu finden, die dem location-scale-Format genügt.

A.3 Momente: Schiefe und Kurtosis

Wenn man die Standard-Funktion $g(z)$ und damit die Form einer Verteilung kennt, dann weiß man noch nicht, wie diese genau aussieht, also ob sie z.B. symmetrisch oder asymmetrisch ist etc. Die Frage ist nun, wie sich die Form einer Verteilung adäquat beschreiben lässt. Die traditionelle Statistik verwendet eine Reihe von Begriffen (wie z.B. uni- vs. bimodal, u-förmig, abfallend etc.), um bestimmte Charakteristika einer Verteilung zunächst einmal rein qualitativ zu beschreiben (vgl. z.B. Bortz, 1999; Hartung, 2005). Darüber hinaus existiert eine große Menge statistischer Kennwerte, die bestimmte Teilaspekte der Form einer Verteilung beschreiben soll. Genauso wie die bekannten Maße der zentralen Tendenz (und hier v.a. Modalwert, Median und Arithmetisches Mittel) und der Dispersion (und hier v.a. der Interquartilsbereich) Aspekte der Lage respektive Skalierung abbilden sollen, gibt es für die höherwertigen Formeigenschaften einer Verteilung die beiden statistischen Kennwerte *Schiefe* und *Kurtosis* (manchmal auch Exzess oder Wölbung genannt). Dabei soll die Schiefe (engl. *skewness*) einer Verteilung angeben, ob eine Verteilung symmetrisch ist oder aber rechts- oder links-schief ausfällt. Die Kurtosis einer Verteilung gibt an, ob eine Verteilung breit- oder schmal-gipflig ist.

Zur Berechnung von Schiefe und Kurtosis wird häufig auf die sog. *Momente* einer Zufallsverteilung zurückgegriffen, die auch für die Bestimmung des Erwartungswerts und der Varianz einer Verteilung relevant sind (de Groot, 1970; Papoulous & Pillai, 2002):

Definition Moment: Es sei X eine Zufallsvariable, k eine natürliche und r eine reelle Zahl. Dann ist das k -te Moment bezüglich r durch $m_k(r) = E((X - r)^k)$ definiert (nach Georgii, 2004; Papoulous & Pillai, 2002).

Das heißt, das k -te Moment ist nichts anderes als der Erwartungswert der k -ten Potenz einer auf r zentrierten Zufallsvariable. Der Erwartungswert berechnet sich dabei genauso wie oben gezeigt. Allerdings kann der Momentbegriff nicht so allgemein verwendet werden, da man ja zunächst gezwungen ist, für r eine bestimmte Größe einzusetzen. Zwei Varianten sind hier besonders prominent. Die eine ist das sog. *Moment um Null* oder *Gewöhnliches Moment*. Hier wird für r der Wert 0 eingesetzt, d.h. das erste Moment ($m_1 = E(X)$) entspricht einfach dem Erwartungswert einer Zufallsvariablen, das zweite Moment dem Erwartungswert der quadrierten Zufallsvariablen ($m_2 = E(X^2)$) etc.

Die andere wichtige Klasse von Momenten sind die so genannten *zentralen Momente*. Bei ihnen wird für r der Erwartungswert einer Variablen selbst eingesetzt, d.h. es gilt $\mu_k = E((X - E(X))^k)$. Das 1. zentrale Moment einer Verteilungsfunktion, $\mu_1 = E((X - E(X)))$, ist ein wenig informationsarm, da es immer den Wert 0 annimmt (die Abweichungen aller Werte

einer Variablen zu ihrem Mittelwert summieren sich zwangsläufig zu 0 auf). Stattdessen wird meist das 1. gewöhnliche Moment angegeben, nämlich der Erwartungswert einer Variablen selbst. Er bestimmt offensichtlich die Lage oder Lokation der Verteilung in Relation zur x-Achse. Das 2. zentrale Moment einer Zufallsvariablen, $\mu_2 = E((X - E(X))^2)$ entspricht nun der Varianz der Verteilung, worauf auch die Ähnlichkeit zur Formel zur Berechnung der Stichprobenvarianz hinweist. Wie bereits gesehen, dient die Varianz (bzw. die Standardabweichung) als Skalierungsparameter. Interessanter sind im vorliegenden Kontext allerdings das 3. und 4. zentrale Moment einer Zufallsverteilung. Denn das 3. standardisierte, zentrale Moment einer Zufallsvariable dient als Maß der Schiefe einer Verteilung:

$$(5) \quad \beta_1 = \frac{\mu_3(X)}{\sigma^3(X)} = \frac{E((X - E(X))^3)}{\text{Var}(X)^{\frac{3}{2}}}$$

Dabei bezieht sich der Begriff „standardisiert“ auf die Tatsache, dass das zentrale Moment durch die Standardabweichung gleicher Ordnung geteilt wird. Dadurch entsteht eine skalenfreie, „dimensionslose“ Größe, die als Maß der Symmetrie dienen kann. Negative Schiefe-Werte stehen für eine links-, positive Schiefe-Werte für eine rechtsschiefe Verteilung.

In ähnlicher Weise ist das 4. standardisierte, zentrale Moment

$$(6) \quad \beta_2 = \frac{\mu_4(X)}{\sigma^4(X)} [-3]$$

ein Maß für den „Wölbungsgrad“ einer Verteilung, der als Kurtosis bezeichnet wird (vgl. De-Carlo, 1997). Da die Normalverteilung eine Kurtosis von 3 hat, wird in vielen Lehrbüchern eine modifizierte Formel angegeben, die vom 4. standardisierten Moment noch eine 3 abzieht (was dann manchmal in Abgrenzung zur reinen Kurtosis „Exzess“ genannt wird und mit γ_2 abgekürzt wird). Werte größer 0 weisen dann auf eine (in Relation zur Normalverteilung) steilgipflige, Werte kleiner 0 auf eine flachgipflige Verteilung hin.

Im Prinzip ist eine Verteilungsfunktion durch die Angabe all ihrer Momente (zumindest soweit sie überhaupt existieren) genau bestimmbar (in der Praxis geht man jedoch selten über den 4. Moment hinaus). So ist z.B. eine Normalverteilung genau dadurch definiert, dass ihr 3. zentrales Moment gleich 0 und das 4. standardisierte, zentrale Momente gleich 3 ist. Allgemein lässt sich nachweisen, dass bei einer normalverteilten Zufallsvariable alle ungeraden Momente den Wert 0 annehmen und alle geraden ein Vielfaches der Varianz der Verteilung sind. Eine Normalverteilung ist also allein durch die Angabe des 1. und 2. Momentes präzise charakterisiert.

A.4 Methoden zur Schätzung der Weibull-Verteilung

A.4.1 Moment-Methode

Die sog. Moment-Methode ist das historisch gesehen älteste Verfahren, das zur Schätzung der Weibull-Parameter vorgeschlagen wurde. Es erhält seinen Namen dadurch, dass die Schätzer auf den verschiedenen Momenten der empirisch beobachtbaren Verteilung aufbauen (vgl. hierzu Appendix A.3). Obwohl sehr viele unterschiedliche Moment-Schätzer existieren (vgl. im Überblick Zanakis, 1979), ist das prinzipielle Vorgehen ungefähr so: Man berechnet die ersten drei Stichproben-Momente, die alle drei in einer bestimmten Beziehung zu den Parametern der Weibull-Verteilung stehen. Für das erste Moment (Erwartungswert: A1) und das zweite Moment (Varianz: A2) wurde dies bereits dargestellt (vgl. Kap. 3.3.1). Für das dritte Moment ist die Beziehung in (A3) gegeben.

$$(A.4.1) \quad M_1(X) = E(X) = \Gamma\left(1 + \frac{1}{c}\right) \times \sigma + \theta$$

$$(A.4.2) \quad M_2(X) = \text{Var}(X) = \left(\Gamma\left(1 + \frac{2}{c}\right) - \Gamma\left(1 + \frac{1}{c}\right)^2 \right) \times \sigma^2$$

$$(A.4.3) \quad M_3(X) = \frac{\Gamma\left(1 + \frac{3}{c}\right) - 3 \times \Gamma\left(1 + \frac{2}{c}\right)^2 \times \left(1 + \frac{1}{c}\right) + 2 \times \left(1 + \frac{2}{c}\right)^3}{\sqrt{\left(\Gamma\left(1 + \frac{2}{c}\right) - 3 \times \Gamma\left(1 + \frac{1}{c}\right)^2 \right)^3}}$$

Der Punkt ist nun, dass das erste Moment von allen drei Parametern der Verteilung abhängig ist, das zweite von shape- und scale-Parameter und das dritte jedoch nur noch vom shape-Parameter. Von daher ist es eine recht einfache Lösungsstrategie, die Gleichung des dritten Moments nach dem shape-Parameter aufzulösen, und so einen Schätzer für ihn zu erhalten. Dieser wird dann in die Gleichung des zweiten Moments eingesetzt und auf die gleiche Art ein Schätzer des scale-Parameters gewonnen. Die Gleichung für das erste Moment liefert dann den shift-Parameter. Alles sehr schön und sehr einfach.

Zu einfach. Denn das Dumme ist, dass dieses Schätzverfahren voraussetzt, dass die empirischen Stichproben-Momente selbst fehlerfrei geschätzt wurden. Wenn dies nicht der Fall ist, dann werden alle Parameter entsprechend verzerrt, und das mit jedem Schritt immer stärker. Und die Fehleranfälligkeit der empirischen Momente durch reinen Stichprobenfehler und Ausreißer (die gerade für das dritte Moment als Startpunkt der Kette am stärksten ausfallen wird!) war ja gerade eines der Hauptargumente dafür, überhaupt ein parametrisches Modell zu

benutzen (vgl. Kap. 3.2.2). Denn eine deskriptive Beschreibung der Verteilung liefern die Momente ja auch selbst. Wenn wir diese verlässlich hätten, warum sollten wir dann überhaupt noch Parameter schätzen? Kurz: Die Methode löst das Problem der Parameterschätzung nicht, sondern setzt seine Lösung bereits voraus. Ansonsten sind alle Parameter verzerrt und die Schätzung ausgesprochen ineffizient.

A.4.2 Maximum-Likelihood-Schätzung

Die Maximum-Likelihood-Methode (im Folgenden ML genannt), die von Fisher (1922) formalisiert wurde, lässt sich vielleicht am besten verstehen, wenn man sie als *reverse engineering* der Wahrscheinlichkeitsdichtefunktion einer Verteilung begreift (vgl. allgemein Myung, 2003; eine sehr ausführliche Darstellung der ML-Schätzung der Weibull-Parameter vgl. z.B. Offinger, 1996). Genauso wie beim Ablesen eines t -Wertes aus einer Lehrbuch-Tabelle sagt diese für einen bestimmten Datenpunkt (z.B. 1.96), wie wahrscheinlich er wäre, wenn die Verteilung mit einer bestimmten Parameterkonstellation (z.B. $df = 2$) versehen ist. Das Konzept des *Likelihood* (für das es m.W. keine adäquate deutsche Übersetzung gibt), funktioniert genau umgekehrt. Es gibt nicht die Wahrscheinlichkeit eines Datenpunktes bei gegebener Parameterkonstellation an ($p(\text{Daten} \mid \text{Parameter})$), sondern die Wahrscheinlichkeit für einen bestimmten *Parameter* gegeben die Daten ($p(\text{Parameter} \mid \text{Daten})$). In einer sog. Likelihood-Funktion werden deshalb Daten- und Parametervektor in der Wahrscheinlichkeitsdichtefunktion gegeneinander ausgetauscht. Ein bestimmter Parameter wird dann als bester Schätzer ausgewählt, wenn er diese Likelihood-Funktion maximiert (was der Methode ihren Namen gibt). Aus numerischen Gründen wird allerdings meist nicht der Likelihood maximiert, sondern vielmehr sein negativer Logarithmus minimiert, was aber mathematisch gesehen auf das Gleiche hinausläuft. Wenn der ML-Schätzer mit dem globalen Minimum des Parameterraums zusammenfällt, dann hat er eine Reihe positiver Eigenschaften, z.B. ist er konsistent, asymptotisch erwartungstreu, effizient (d.h. es gibt kein effizienteres Schätzverfahren), suffizient (d.h. er nutzt alle Informationen maximal aus) und invariant gegenüber Reparametrisierung.

Konkret wird die Likelihood-Funktion separat nach allen Parametern gleichzeitig abgeleitet und simultan ein Minimum in dem sich ergebenden Parameterraum gesucht (mittels der üblichen Methoden der Differentialrechnung: 1. Ableitung = 0, 2. Ableitung < 0). Für die Weibull-Verteilung lässt sich dieses Gleichungssystem nicht analytisch lösen, sondern muss numerisch mit einem iterativen, nicht-linearen Optimierungsalgorithmus approximiert werden. Die mathematischen Einzelheiten dieses Optimierungsprozesses (die nicht trivial sind, vgl. z.B. Nocedal & Wright, 1999) müssen uns hier nicht interessieren. Wichtig ist lediglich,

dass es durchaus sein kann, dass es keine oder keine singuläre Lösung für diese Likelihood-Minimierung gibt.

Und da fangen die Probleme auch schon an. Denn es ist gut bekannt, dass die Weibull-Verteilung in vielen Fällen überhaupt keine oder nur eine „irregulär“ ML-Lösung hat, da sie nicht die Regularitätsbedingungen der ML-Schätzung erfüllt (vgl. Johnson et al., 1994). Zwei Probleme sind es v.a., die die Weibull-Verteilung zu einem „schwierigen Fall“ machen: das sog. „unbounded likelihood problem“ und das sog. „embedded models problem“ (vgl. Cheng & Amin, 1983; Cheng & Iles, 1987, 1990; Cheng & Traylor, 1995; Smith, 1985).

Das *unbounded likelihood problem* besteht darin, dass Verteilungen, deren Range von ihren Parametern abhängig ist, überhaupt nicht die Voraussetzungen der ML-Schätzung erfüllen, da diese – abgesehen von der funktionalen Beziehung zwischen Parameter und Daten – von deren grundsätzlichen Unabhängigkeit ausgeht. Diese Voraussetzung ist jedoch offensichtlich bei allen Verteilungen verletzt, die einen shift-Parameter haben, der als Minimum der Verteilung ja gerade festlegt, dass es keine niedrigeren Datenpunkte als ihn selbst geben darf. Entsprechend ist die ML-Schätzung für alle shift-Verteilungen (wie auch die „geshiftete“ Exponential-, Gamma- oder Wald-Verteilung) insgesamt erschwert (vgl. Heathcote et al., 2004).

Das eigentliche Problem entsteht dadurch, dass die Likelihood-Funktion der Verteilung genau dann maximal wird, wenn der shift-Parameter mit dem empirischen Minimum der Stichprobe zusammenfällt. Denn klarerweise ist der kleinste Stichprobenwert in der Tat der beste Schätzer für den Ansatzpunkt der Verteilung. Allerdings ist er sogar zu gut! Denn da die Funktion nur ein einziges Minimum hat, geht ihre Dichte an dieser Stelle gegen Null und komplementär der Likelihood gegen unendlich. Es liegt also eine Singularität vor und der Likelihood ist „unbounded“. Wenn aber schon allein durch den shift-Parameter ein unendlicher Likelihood garantiert ist, dann ist auch völlig egal, welche Werte die anderen beiden Parameter annehmen: Ihre Schätzung wird also sehr variabel und inkonsistent ausfallen.

Man kann zeigen (vgl. Cheng & Amin, 1983; Cheng & Traylor, 1995), dass dieses Problem umgangen werden kann, wenn es entweder a) ein lokales Maximum gibt, das eine stabile Parameterschätzung gewährleistet (dies gilt z.B. für die geshiftete Lognormal-Verteilung) oder aber b) die Singularität „disconnected“ ist (was bedeutet, dass der Likelihood kontinuierlich zunimmt, wenn sich der Parameter von unten dem kleinsten Stichprobenwert nähert; dies ist z.B. bei der geshifteten Wald-Verteilung der Fall).

Leider kann das die Weibull-Verteilung nur partiell retten. Denn es ist ja eine Eigenart der Weibull-Verteilung, bei einem shape-Wert von 1 von einem umgekehrt J-förmigen Verlauf (mit einem sehr „steilen“ Ansatz) in eine allmählich ansteigende Glockenform zu wechseln (vgl. Kap. 3.3.1). In diesem Fall kann sich die Verteilung nicht (wie in Bedingung b) ge-

nannt) langsam von unten her an das Minimum herantasten, sondern muss „schlagartig“ anfangen. Dies ist in Abb. A.4.1 noch einmal dargestellt.

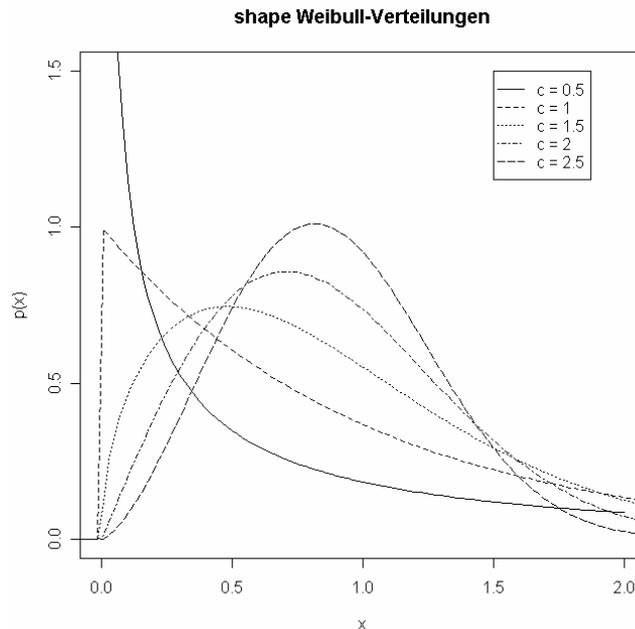


Abb. A.4.1: Abhängigkeit der Form der Weibull-Verteilung vom shape-Parameter c ($\theta = 0$, $\sigma = 1$).

Entsprechend existiert für den Bereich $c < 1$ gar keine ML-Lösung, da es weder ein lokales Maximum gibt, noch die Singularität *disconnected* ist. Im Intervall $1 < c < 2$ hat die Verteilung immer noch einen sehr „abrupten“ Ansatz, weshalb die Singularität immer noch nicht *disconnected* ist. Allerdings existiert hier manchmal ein lokales Maximum. Für diesen Bereich ist eine ML-Lösung zumindest definiert, allerdings fällt sie sehr instabil aus. Ab einem shape-Parameter von 2 (bzw. aus numerischen Gründen eher 2.2, vgl. Cohen & Whitten, 1982), ist die Singularität dann endlich *disconnected*, was sich graphisch darin niederschlägt, dass die Verteilung (wie die Normalverteilung auch) nun einen linken „Ausläufer“ hat.

Als Zwischenfazit für das *unbounded likelihood problem* lässt sich festhalten, dass eine vernünftige ML-Schätzung der Weibull-Parameter erst ab einem shape-Wert von ca. 2 möglich ist, also wenn die Verteilung keine allzu große (positive) Schiefe aufweist. Unglücklicherweise ist das zweite große Problem für die Maximum Likelihood-Schätzung, das *embedded model problem*, dem ersten genau entgegengesetzt und tritt deshalb exakt dann auf, wenn die Verteilung *nicht genügend* schief ist oder sogar eine negative Schiefe hat. Auch hier ist die Weibull-Verteilung nicht allein: Schlechte Parameterschätzungen bei geringer Schiefe wird auch für die Lognormal- und die Wald-Verteilung berichtet, v.a. bei kleinen Stichproben.

Das *embedded model problem* (vgl. Cheng & Iles, 1990) besteht darin, dass alle geshifteten Verteilungen eigentlich zweiparametrische Modelle sind, die lediglich zusätzlich durch

einen dritten Parameter auf der x-Achse hin- und herbewegt werden. Das führt dazu, dass alle diese Verteilungen einen impliziten zweiparametrischen Spezialfall enthalten, ein sog. „eingebettetes Modell“. Dieses muss nicht unbedingt zur gleichen Familie gehören wie die Ausgangsverteilung, ja sie muss noch nicht einmal zur gleichen Klasse von Funktionen gehören. Für die Weibull-Verteilung ist diese eingebettete Verteilung z.B. die Gumbel-Verteilung, die zum location-scale-Typ gehört.

Das eingebettete Modell ist also immer einfacher strukturiert als das volle Modell und ist immer sein am wenigsten rechtsschiefer Grenzfall. Der Punkt ist nun der, dass bei diesem Zwei-Parameter-Modell gewissermaßen ein Parameter „überflüssig“ ist, da nicht genügend Informationen zur Schätzung aller drei Parameter zur Verfügung stehen. Konsequenterweise wird der überflüssige Parameter dann versuchen, „sich unsichtbar zu machen“ und gegen minus Unendlich streben (bei der Weibull-Verteilung ist dies der shift-Parameter). Wenn nun das eingebettete Modell die Verteilung genauso gut oder vielleicht sogar noch besser fittet als die Originalverteilung, dann übernimmt das eingebettete Modell das Kommando und „drängt“ einen der Parameter ins Unendliche. Und das passiert v.a. dann, wenn die empirische Verteilung dem eingebetteten Modell möglichst ähnlich ist, also eine niedrige Schiefe aufweist. Da eingebettete Modelle also gerade darüber definiert sind, dass sie einen unendlichen Parameterwert haben, haben iterative Optimierungsalgorithmen mit ihnen ein Problem, weil sie dann nicht oder nur sehr schlecht konvergieren. Die Schätzungen für alle anderen Parameter sind dann wieder sehr instabil und variabel (vgl. auch das Fallbeispiel bei Rouder et al., 2005).

Man kann das Problem auch so betrachten: Bei einem shape-Wert von $c = 3.6$ „kippt“ die Weibull-Verteilung in ihrer Form. Wenn sie vorher einen scharfen linken Ansatzpunkt hatte, dann bricht sie bei negativer Schiefe plötzlich sehr abrupt ab (vgl. Abb. A.4.2). Das heißt, die Verteilung hat keinen „Schwanz“ mehr auf der rechten Seite, der die notwendigen Informationen zur Schätzung eines eigenen shape-Parameters bereitstellen könnte. Deswegen muss stattdessen das eingebettete Modell gefittet werden (vgl. Coles & Powell, 1996).

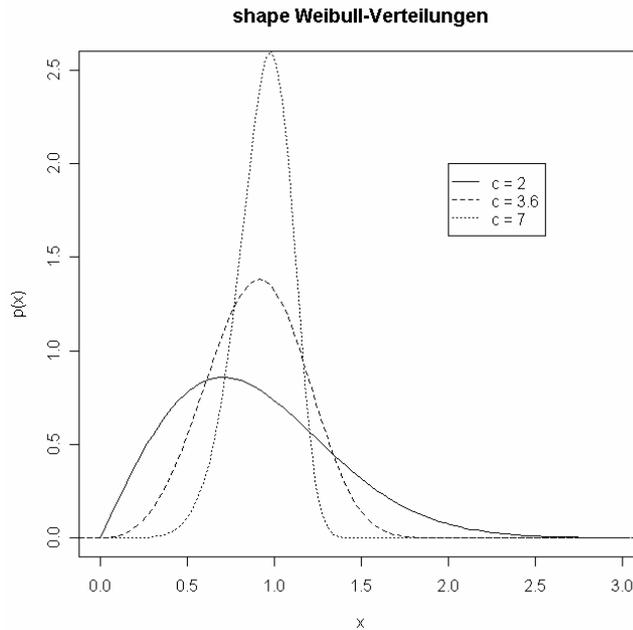


Abb. A.4.2: Abhängigkeit der Form der Weibull-Verteilung vom shape-Parameter c ($\theta = 0$, $\sigma = 1$).

Das Problem entsteht zumindest teilweise dadurch, dass die shift-Verteilung wie die Weibull-Funktion eine ungünstige Parametrisierung aufweist: Sie weisen genau dann die geringste Schiefe auf, wenn der shift-Parameter gegen minus Unendlich strebt, also einem infiniten Wert. Bei der Ex-Gauss-Verteilung ist dies z.B. anders: Hier ist die Verteilung dann am symmetrischsten, wenn ein Parameter (nämlich τ) gegen einen finiten Wert (nämlich 0) konvergiert. Cheng und Iles (1990) stellen enttäuscht fest, dass eine Reparametrisierung der Verteilung keine Verbesserung bringt. Stattdessen plädieren sie für ein sog. „soft bounding“ durch die Einführung einer hierarchischen Komponente bei der Parameterschätzung, bei der eine explizite Elternverteilung den Unendlichkeitstrend des shift-Parameters „ausbremst“.

Zusammengenommen bedeutet das, dass die Weibull-Verteilung allein unter dem Gesichtspunkt der Parameterschätzung eine denkbar schlechte Wahl zur RT-Modellierung ist. Dabei verursacht das, was theoretisch ihre große Stärke ist, nämlich die Tatsache, dass sie als geshiftete Verteilung eine untere Grenze besitzt, schätztechnisch die größten Probleme. Die diversen Manifestationen dieses Grundübels (*unbounded likelihood problem*, *embedded models problem*) führen nachfolgend dazu, dass keine sinnvollen ML-Schätzer existieren, wenn der shape-Parameter der Verteilung unter einen kritischen Wert von 2 fällt, aber auch wenn die Verteilung zu wenig Schiefe aufweist (Rouder et al., 2005, geben hier einen kritischen shape-Wert von ca. 4 an). Das Fenster, in dem die ML-Methode problemlos eingesetzt werden kann, ist also ziemlich schmal. Hier ließe sich einwenden, dass es aber exakt der Bereich ist, der für menschliche RT-Verteilungen sinnvoller Weise angenommen werden sollte. Zumindest theoretisch sollte es also überhaupt gar kein Problem geben (vgl. z.B. Heathcote,

2004). Dieses Argument ist jedoch nur bedingt stichhaltig. Zwar ist richtig, dass es Quatsch wäre, eine Theorie aufzustellen, die RT-Verteilung mit einem sehr steilen Ansatzpunkt vorhersagt. Und selbstverständlich sind die meisten empirischen RT-Verteilungen eher rechts- als linksschief. Nichtsdestotrotz ergeben sich gerade dann, wenn die Verteilung nur wenige Datenpunkte umfasst, allein durch die normale Stichprobenvariabilität durchaus sehr häufig auch Verteilungen, die durch shape-Werte weit unter 2 oder über 4 charakterisierbar sind. Wenn diese Verteilungen dann sehr extreme Parameterschätzungen liefern, dominieren diese fast ausschließlich die Parameterschätzung und das Verfahren wird durch diese zusätzliche massive Fehlervarianz ausgesprochen ineffizient.

A.4.3 Bayes-Schätzung

Aufgrund der erheblichen und massiven Probleme der ML-Methode (und ihren diversen Modifikationen, vgl. Cohen & Whitten, 1982) mit der Weibull-Verteilung ist es kein Wunder, dass sich die Literatur zur Weibull-Verteilung bereits ausgesprochen früh alternativen Verfahren zugewendet hat, und hier insbesondere non-parametrischen und Bayes-Verfahren (vgl. Basu, 1964; Holla, 1966; Bhattacharya, 1967; s. auch Shimi & Tsokos, 1977). Allerdings konnten sich diese Verfahren lange Zeit nicht durchsetzen, und zwar aufgrund berechnungstechnischer Probleme, die – wie gleich zu sehen – nicht unerheblich sind. Erst in den 1980er Jahren wurden die konzeptuellen Ansätze für den vollen 3-Parameter-Fall formuliert (vgl. Smith & Naylor, 1987; Sinha & Guttman, 1988; Sinha & Sloan, 1988), die seit der Einführung des *Monte Carlo Markov Chain*- (MCMC-) Verfahren in den 1990er Jahren auch mathematisch handhabbar wurden. Insbesondere durch Softwarepakete wie (Win)BUGS ist es heute möglich, Bayes-Schätzer mittels MCMC-Sampling (vergleichsweise) schnell und effizient zu berechnen. Dies hat wesentlich zur Verbreitung des Ansatzes beigetragen (vgl. Coles & Powell, 1996).

Wie der Name schon sagt, basiert der Bayes-Ansatz für die Schätzung der Weibull-Parameter auf dem Bayes-Theorem (vgl. einführend zur Bayes-Statistik z.B. Lee, 2004; Sivia & Skilling, 2006). Im Gegensatz zu allen Schätzverfahren der klassischen bzw. frequentistischen Statistik wird der Schätzer der Parameter einer bestimmten Funktion ($p(\text{Daten} \mid \theta)$) nicht als ein einziger, fixer Wert angesehen. Vielmehr ist der Schätzer selbst eine Zufallsvariable, die variieren kann und damit selbst eine bestimmte Verteilung generiert. Dies ist die sog. Posterior-Wahrscheinlichkeit ($p(\theta \mid \text{Daten})$), die gesucht ist und (wie die Likelihood-Funktion bei ML-Verfahren) angibt, wie wahrscheinlich eine bestimmte Parameterkonstellation bei vorliegenden Daten ist.

Im Gegensatz zur klassischen Statistik ist für die Bayes-Schätzung die Spezifizierung einer eigenen Prior-Verteilung notwendig, die Annahmen darüber enthält, wie die sich Parameter *a priori*, also vor Betrachtung der Daten, verteilen ($p(\theta)$).

Die Annahme einer eigenen Prior-Verteilung wird häufig kritisiert, weil sie ja – scheinbar – die „Objektivität“ des Verfahrens gefährdet und Vorannahmen des/der Forschers/in Einfluss auf die Ergebnisse haben. Das ist zwar grundsätzlich richtig, aber kein Anlass zur Besorgnis. Denn wie noch zu zeigen ist, ist die Berücksichtigung einer expliziten Prior-Verteilung gerade keine Schwäche, sondern vielmehr die große Stärke des Verfahrens ist und ist nicht „subjektiver“ als viele andere Entscheidungen der klassischen Statistik auch.

Der Clou ist nun der, dass sich (nach dem Bayes-Theorem) die Dichte der Posterior-Verteilung einfach aus der Multiplikation der Priorverteilung mit der Dichteverteilung der zu schätzenden Funktion ergibt, die zusätzlich noch an der Grundwahrscheinlichkeit der Daten ($p(\text{Daten})$) relativiert wird (vgl. (A4.4)) (für eine umfassende Darstellung von Bayes-Schätzverfahren im Allgemeinen vgl. z.B. Bolstad, 2004; Congdon, 2001; Gill, 2002; für die Weibull-Verteilung im Speziellen Sinha & Sloan, 1988; Berger & Sun, 1993).

$$(A.4.4) \quad p(\theta \mid \text{Daten}) = \frac{p(\text{Daten} \mid \theta) \times p(\theta)}{p(\text{Daten})}$$

Wenn die Daten-Funktion ($p(\text{Daten} \mid \theta)$) als eine Funktion der Daten bei gegebenem Parameter-Vektor angesetzt wird, dann nennt man sie (*prädiktive*) *Stichproben-Verteilung* (und ist mit der Wahrscheinlichkeitsdichtefunktion der Funktion identisch). Die Posterior-Verteilung dient dann der *Hypothesenprüfung*. Wenn die Daten-Funktion hingegen als eine Funktion des Parameter-Vektors bei gegebenen Daten betrachtet wird, dann nennt man diese Teilverteilung *Likelihood-Funktion* (die mit der Likelihood-Funktion der ML-Schätzung identisch ist). Die Posterior-Verteilung dient dann zur *Schätzung* der Funktionsparameter.

Bei Modellen, die lediglich auf einem einzigen Parameter basieren, gibt die Posterior-Verteilung den Schätzer dieses Parameters direkt an, bei Wahrscheinlichkeitsmodellen, die (wie die Weibull-Verteilung) auf mehreren Parametern gleichzeitig beruhen, jedoch lediglich eine sog. *gemeinsame Posterior-Verteilung* (*joint posterior distribution*). Diese sagt uns noch nichts über die Schätzung der einzelnen Parameter, sondern nur etwas über die aller Parameter zusammen. Um Schätzer für jeden der drei Parameter der Weibull-Verteilung einzeln zu erhalten, braucht man deshalb die *marginale* oder *Rand-Posterior-Verteilungen*. Diese erhält man, indem die jeweils nicht relevanten („nuisance“-) Parameter aus dem gemeinsamen Posterior „herausintegriert“. Ebenso wie die Randverteilungen einer Kreuztabelle ist die verbleibende Marginalverteilung eine univariate Verteilung, für die sich in gewohnter Manier entweder der Erwartungswert als Punktschätzer oder auch ein Konfidenzintervall als Intervallschätzer bestimmen lässt.

Konkret ist das Vorgehen bei der Weibull-Verteilung so, dass man ihre – bekannte – Wahrscheinlichkeitsdichtefunktion nimmt und wie bei der ML-Schätzung in eine Likelihood-Funktion umwandelt. Als Prior-Verteilungen werden dann meist sog. „uninformative“ Prior gewählt, die keine wirkliche Information zu den Parametern liefern, sondern alle Parameterwerte als mehr oder minder gleich wahrscheinlich ansetzen. (Das ist nämlich auch eine Art, sein „Vorwissen“ explizit zu machen: Klar zu sagen, dass man nichts weiß!). Nach diesen beiden Setzungen lässt sich dann die gemeinsame Posterior-Verteilung vergleichsweise einfach bestimmen.

Aber danach gehen die Probleme los: Lange Zeit „hing“ die Entwicklung geeigneter Bayes-Schätzer für die Weibull-Verteilung aufgrund der schlichten Tatsache, dass selbst mit sehr einfachen und uninformativen Prior-Verteilung die Posterior-Verteilung so komplex ist, dass sich analytisch nicht die notwendigen marginalen Posterior-Verteilungen angeben ließen. Stattdessen musste man sich etwas einfallen lassen: Sinha und Sloan (1988) verwendeten das sog. *Bayes Linear Estimates*-Verfahren von Lindley (1980), welches jedoch nicht die gesamten marginalen Posterior-Verteilungen liefert, sondern nur ihre ersten beiden Momente. Smith und Naylor (1987) entwickelten ein numerisches Quadratur-Verfahren und Singpurwalla und Song (1988) verwendeten die Laplace-Approximation. Alle diese Ansätze liefern jedoch nicht, was man eigentlich haben will, nämlich die marginale Posterior-Verteilung selbst.

Das änderte sich erst, als Anfang der 1990er Jahre sog. *Monte Carlo Markov Chain*-Verfahren entwickelt wurden (vgl. Gelfand & Smith, 1990). Bei diesen handelt es sich um eine Klasse von numerischen Verfahren, die das Problem der Integration der marginalen Posterior-Verteilung dadurch „umgehen“, dass sie die Dichtefunktion nicht berechnen, sondern vielmehr *simulieren*. Grundprinzip ist, dass eine einzige schwierige Berechnung durch eine Sequenz einfacherer ersetzt wird, die jedoch zusammen zu dem gleichen Ergebnis kommen. Die Aufgabe, zu erklären wie MCMC-Verfahren genau funktionieren, ist nicht trivial und würde hier zu viel Platz in Anspruch nehmen (vgl. hierzu z.B. Chib & Greenberg, 1995; Gilks, Richardson, & Spiegelhalter, 1996). Wichtig ist nur zu verstehen, dass es sich bei ihnen (und speziell bei dem im Folgenden zur Anwendung kommenden Spezialfall des sog. *Gibbs Sampler*, vgl. Geman & Geman, 1984; Casella & George, 1992) um iterative Verfahren handelt, bei dem wechselseitig nacheinander Stichproben aller Parameter aus der gemeinsamen Posterior-Verteilung gezogen werden und alle anderen dann berechnet werden. Die resultierenden Werte werden dann wiederum nachfolgend als Startwerte einer erneuten Stichprobenziehung benutzt etc. Unter sehr allgemeinen Annahmen wird die Verteilung der einzelnen Parameter zu seiner marginalen Randverteilung konvergieren (vgl. Tierney, 1994).

Mathematisch gesehen ist entscheidend, dass Bayes-Methoden kein Maximum oder Minimum in einem gegebenen Parameterraum suchen und deshalb auch nicht auf den asymptotischen Annahmen der ML-Schätzung beruhen (Smith & Naylor, 1987). Stattdessen modellie-

ren sie schlicht die Verteilung der Parameter direkt. Da sie deswegen bei der Parameterschätzung auch nicht in eine unendliche Singularität laufen können, sind sie nicht vom *unbounded likelihood problem* betroffen, welches die ML-Schätzung so quält. Das liegt daran, dass die Parameterschätzung bei den beiden Ansätzen unterschiedlich ist (nicht jedoch ihre theoretische Fundierung: beide basieren auf dem Likelihood-Prinzip!): Beim Bayes-Verfahren findet man die Schätzer nicht über Maximierung, sondern Integrierung einer Funktion. Obwohl diese beiden mathematischen Verfahren meist zu ähnlichen Ergebnissen führen, können sie u.U. divergieren (vgl. z.B. Smith & Naylor, 1987): Maximierungsalgorithmen suchen nach dem „Gipfelpunkt“ in einer Parameterebene (Zweidimensionalität vorausgesetzt), das Integral einer Funktion wird jedoch genau an dem Punkt maximal, an dem eine Ebene die meiste Masse besitzt (und dies ist bei flachen, aber breiten Profilen halt nicht unbedingt der Gipfel!). Oder anders formuliert: Wenn der Parameterraum nicht asymptotisch quadratisch ist (Gipfel und Masseschwerpunkt also auseinander fallen), dann ist eine Voraussetzung der ML-Schätzung verletzt (vgl. Cheng & Traylor, 1995). Bayes-Schätzer kommen mit dieser Irregularität besser zurecht, weil sie nach dem Masseschwerpunkt und nicht nach dem Gipfel der Verteilung suchen. Wenn die Verteilungsebene hingegen in der Tat quadratisch ist, dann kommen ML- und Bayes-Schätzung tendenziell auch zu den gleichen Ergebnissen.

A.5 Vergleich verschiedener Verfahren zur Schätzung der Weibull-Verteilung

In Kap. 3.4 wurde behauptet, dass Bayes-Verfahren allgemein und das Hierarchische Modell von Rouder et al. speziell günstig zur Schätzung der Parameter der Weibull-Verteilung sind und anderen Möglichkeiten (hier insbesondere der ML-Schätzung) überlegen. Der generelle Vorteil der Bayes- gegenüber der ML-Schätzung lässt sich jedoch auch quantifizieren. Dies sei am Beispiel einer kleinen Simulationsstudie von Sinha und Sloan (1988) dargestellt (vgl. Tab. A.5.1). In der Studie wurde das Schätzverhalten der Moment-, ML- und (linear-approximierten) Bayes-Schätzer für eine bestimmte Parameterkonstellation der Weibull-Verteilung (nämlich shift = 30, scale = 100 und shape = 3) miteinander verglichen. Dargestellt ist jeweils der relative Bias und die relative Effizienz eines Parameters als klassische Indikatoren der Schätzgüte (der Bias ist ein Maß der Erwartungstreue und gibt an, wie sehr und in welche Richtung ein Schätzer den wahren Wert verschätzt; die Effizienz bezieht sich darauf, wie präzise die Schätzung ist; niedrige Werte stehen dabei für eine hohe Effizienz).

N	Verfahren ^a	shift		scale		shape	
		Bias ^b	Effizienz ^c	Bias ^b	Effizienz ^c	Bias ^b	Effizienz ^c
10	M	+ 129	-----	- 46	-----	- 1	-----
	ML	+ 131	69	- 55	24	- 19	55
	B	+ 5	Neg.	- 14	Neg.	+ 58	Neg.
20	M	+ 144	-----	- 50	-----	- 38	-----
	ML	+ 156	64	- 54	20	- 45	37
	B	+ 123	Neg.	- 41	Neg.	- 29	5
40	M	+ 55	-----	- 17	-----	- 8	-----
	ML	+ 33	54	- 10	18	+ 1	25
	B	- 33	Neg.	+ 12	Neg.	+ 24	8
100	M	- 216	-----	+ 71	-----	+ 19	-----
	ML	+ 1	47	- 1	17	+ 9	24
	B	- 42	38	+ 13	13	+ 24	189
200	M	+ 2011	-----	- 99	-----	- 48	-----
	ML	+ 22	30	- 9	10	- 1	13
	B	+ 5	25	- 3	9	+ 4	12
300	M	- 14	-----	+ 2	-----	+ 9	-----
	ML	+ 16	27	- 7	9	- 0.3	13
	B	+ 5	24	- 3	9	+ 4	11
400	M	- 5	-----	+ 3	-----	+ 0.3	-----
	ML	+ 25	20	- 7	7	- 7	9
	B	+ 16	19	- 4	6	- 4	8

^aM = Moment, ML = Maximum Likelihood, B = Bayes (vgl. Sinha & Guttman, 1988), ^bDifferenz (geschätzt – wahr) relativiert an der Originalausprägung des Parameters in %, ^cML = SD Parameterschätzer in %, B = Posterior-Varianz in %, M = k.A.; Neg. = Negative Varianzschätzung

Tab. A.5.1: Relativer Bias und Effizienz verschiedener Schätzmethoden in der Studie von Sinha & Sloan (1988) (Originalverteilung: shift = 30, scale = 100, shape = 3)

Es ist klar zu erkennen, dass die Parameter der Moment-Methode (für die keine Effizienz-Maße berichtet werden) einen teilweise dramatischen Bias haben. Auch zeigt sich, dass die Bayes-Schätzung in den meisten Fällen der ML-Schätzung sowohl hinsichtlich des Bias

als auch bezüglich der Effizienz überlegen ist. Allerdings ergeben sich bei Stichproben mit weniger als 100 Datenpunkten deutliche Probleme, da hier die Posterior-Varianz negativ ist, was darauf schließen lässt, dass das Verfahren in diesen Fällen versagt. Man beachte, dass diese Ergebnisse für eine Parameterkonstellation simuliert wurden, die noch vergleichsweise günstig für die ML-Schätzung ist ($2 < \text{shape} < 4$).

Eine weitere Simulationsstudie von Rouder et al. (2005) vergleicht die Effizienz (Bias-Werte werden nicht berichtet) verschiedener Schätzverfahren für typische RT-Verteilungen direkt und zwar sowohl für die Individualschätzung als auch für die Schätzung von Populationsparametern (vgl. Tab. A.5.2).

<i>N</i>	Ebene	Verfahren ^a	shift ^b	scale ^b	shape ^b
80	Individual	HB	8	14	19
		ML	11	17	24
		QML	12	19	27
		QLS	19	28	38
	Population	HB + PA	1	2	4
		ML + PA	2	4	3
		QML + PA	1	2	3
		QLS + PA	3	4	4
		V + QML	4	5	7
		V + QLS	5	6	12
20	Individual	HB	14	22	27
		ML	21,153,818	29,086,500	56,039,300
		QML	32,516,363	44,710,000	74,593,400
		QLS	34,140,727	46,938,500	51,722,150
	Population	HB + PA	3	4	7
		ML + PA	2,728,000	3,751,000	7,136,950
		QML + PA	4,157,818	5,717,000	9,463,050
		QLS + PA	3,867,636	5,318,000	5,897,550
		V + QML	6	9	10
		V + QLS	5	7	16

^aHB = Hierarchical Bayes (Rouder et al., 2003), ML = Maximum Likelihood, QML = Quantil Maximum Likelihood (Heathcote et al., 2002), QLS = Quantil Least Squares (Rouder et al., 2004), PA = Parameter Averaging, V = Vincentizing; ^bRMSE relativiert an der Originalausprägung des Parameters in %

Tab. A.5.2: Relative Effizienz der ML- und Hierarchical Bayes-Schätzer in der Studie von Rouder et al. (2005) (Originalverteilung: shift = 275, scale = 200, shape = 2; ML-Schätzer als Originalwerte)

Wichtig ist hier v.a. die Tatsache, dass Rouder et al. zwei unterschiedliche Stichprobengrößen miteinander verglichen haben, nämlich RT-Verteilungen, die auf 80 Items pro Person beruhen, und solche, die lediglich 20 Items umfassen. Die Tabelle zeigt deutlich, dass das Hierarchical Bayes-Verfahren das einzige auf der Individualebene ist, dass bei Untersuchungsbedingungen mit lediglich 20 Items überhaupt sinnvoll eingesetzt werden kann. Die Unreliabilität der anderen Verfahren ist so dramatisch, dass ihr Einsatz keinen Sinn macht.

Bezüglich der Populationsschätzung können Vincentizing-Methoden in diesem Fall noch mit der Hierachical-Bayes-Schätzung konkurrieren, sind aber insgesamt ineffizienter.

Zusammenfassend lässt sich feststellen, dass die Verwendung der Hierarchical Bayes-Schätzung die Schätzung von individuellen RT-Verteilungsparametern mit geringen Itemstichproben überhaupt erst möglich macht. Das Verfahren ist mit 20 Items ungefähr so effektiv wie die ML-Schätzung mit 80 Items, d.h. es erlaubt eine Itemreduktion um den Faktor 4 pro Untersuchungsbedingung. Ein Vergleich mit den Ergebnissen der Simulationsstudie von Dolan et al. (2002) macht darüber hinaus deutlich, dass die Schätzung der Weibull-Parameter auch mit kleinen Itemstichproben ungefähr so effektiv vorgenommen werden kann wie für die Parameter der Ex-Gauss-Verteilung, die sich wesentlich einfacher schätzen lässt (vgl. Kap. 3.3.2).

Appendix B: Stimulusmaterial

A-P Verben *Subjekt: Agens – Objekt: Patiens; Kausalität auf Agens*

AP 1 befragte

GenMinKonPlus: Andreas befragte Gunnar, weil er über das neue Hochschulgesetz alles genau wissen wollte.

GenMinKonMin: Andreas befragte Gunnar, weil er über das neue Hochschulgesetz immer genau informiert war.

GenPlusKonPlus: Andreas befragte Gudrun, weil er über das neue Hochschulgesetz alles genau wissen wollte.

GenPlusKonMin: Katja befragte Gunnar, weil er über das neue Hochschulgesetz immer genau informiert war.

AP 2 beherrschte

GenMinKonPlus: Helmut beherrschte Clemens, weil er in dieser schwierigen Sportart nun einmal stärker war.

GenMinKonMin: Helmut beherrschte Clemens, weil er in dieser schwierigen Sportart nun einmal schwächer war.

GenPlusKonPlus: Helmut beherrschte Hannah, weil er in dieser schwierigen Sportart nun einmal stärker war.

GenPlusKonMin: Manuela beherrschte Clemens, weil er in dieser schwierigen Sportart nun einmal schwächer war.

AP 3 behinderte

GenMinKonPlus: Dirk behinderte Gregor, weil er schon seit der Schulzeit ziemlich unfaire Strategien einsetzte.

GenMinKonMin: Dirk behinderte Gregor, weil er schon seit der Schulzeit nur unfair überwindbar war.

GenPlusKonPlus: Dirk behinderte Herta, weil er schon seit der Schulzeit ziemlich unfaire Strategien einsetzte.

GenPlusKonMin: Elke behinderte Gregor, weil er schon seit der Schulzeit nur unfair überwindbar war.

AP 4 bekämpfte

GenMinKonPlus: Stefan bekämpfte Willi, weil er wegen der letzten Auseinandersetzung wirklich äußerst wütend war.

GenMinKonMin: Stefan bekämpfte Willi, weil er wegen der letzten Auseinandersetzung sehr gefährlich werden konnte.

GenPlusKonPlus: Stefan bekämpfte Sabine, weil er wegen der letzten Auseinandersetzung wirklich äußerst wütend war.

GenPlusKonMin: Lotte bekämpfte Willi, weil er wegen der letzten Auseinandersetzung sehr gefährlich werden konnte.

AP 5 beleidigte

GenMinKonPlus: Katja beleidigte Gudrun, weil sie bei dieser persönlichen Angelegenheit wirklich ausgesprochen unsensibel war.

GenMinKonMin: Katja beleidigte Gudrun, weil sie bei dieser persönlichen Angelegenheit die Kritik unangebracht fand.

GenPlusKonPlus: Katja beleidigte Gunnar, weil sie bei dieser persönlichen Angelegenheit wirklich ausgesprochen unsensibel war.

GenPlusKonMin: Andreas beleidigte Gudrun, weil sie bei dieser persönlichen Angelegenheit die Kritik unangebracht fand.

AP 6 belog

GenMinKonPlus: Manuela belog Hannah, weil sie trotz des guten Verhältnisses Angst vor Vorwürfen hatte.

GenMinKonMin: Manuela belog Hannah, weil sie trotz des guten Verhältnisses sonst verletzt gewesen wäre.

GenPlusKonPlus: Manuela belog Clemens, weil sie trotz des guten Verhältnisses Angst vor Vorwürfen hatte.

GenPlusKonMin: Helmut belog Hannah, weil sie trotz des guten Verhältnisses sonst verletzt gewesen wäre.

AP 7 beschwindelte

GenMinKonPlus: Elke beschwindelte Herta, weil sie für diesen teuren Einkauf übermäßig Schulden gemacht hatte.

GenMinKonMin: Elke beschwindelte Herta, weil sie für diesen teuren Einkauf kein Verständnis gehabt hätte.

GenPlusKonPlus: Elke beschwindelte Gregor, weil sie für diesen teuren Einkauf übermäßig Schulden gemacht hatte.

GenPlusKonMin: Dirk beschwindelte Herta, weil sie für diesen teuren Einkauf kein Verständnis gehabt hätte.

AP 8 bezwang

GenMinKonPlus: Lotte bezwang Sabine, weil sie bei dem letzten Schachturnier eine geniale Strategie verfolgte.

GenMinKonMin: Lotte bezwang Sabine, weil sie bei dem letzten Schachturnier einen dummen Fehler macht.

GenPlusKonPlus: Lotte bezwang Willi, weil sie bei dem letzten Schachturnier eine geniale Strategie verfolgte.

GenPlusKonMin: Stefan bezwang Sabine, weil sie bei dem letzten Schachturnier einen dummen Fehler macht.

AP 9 bremste

GenMinKonPlus: Adam bremste Niklas, weil er bei dem gemeinsamen Sprachunterricht zu wenig Anstrengung investierte.

GenMinKonMin: Adam bremste Niklas, weil er bei dem gemeinsamen Sprachunterricht immer Rücksicht nehmen musste.

GenPlusKonPlus: Adam bremste Ruth, weil er bei dem gemeinsamen Sprachunterricht zu wenig Anstrengung investierte.

GenPlusKonMin: Birgit bremste Niklas, weil er bei dem gemeinsamen Sprachunterricht immer Rücksicht nehmen musste.

AP 10 denunzierte

GenMinKonPlus: Wolf denunzierte Albert, weil er schon aus politischem Prinzip alle systemkritischen Äußerungen meldete.

GenMinKonMin: Wolf denunzierte Albert, weil er schon aus politischem Prinzip permanent das System kritisierte.

GenPlusKonPlus: Wolf denunzierte Olivia, weil er schon aus politischem Prinzip alle systemkritischen Äußerungen meldete.

GenPlusKonMin: Dolores denunzierte Albert, weil er schon aus politischem Prinzip permanent das System kritisierte.

AP 11 drangsalierte

GenMinKonPlus: Friedrich drangsalierte Leonard, weil er seit der gemeinsamen Schulzeit eine starke Antipathie verspürte.

GenMinKonMin: Friedrich drangsalierte Leonard, weil er seit der gemeinsamen Schulzeit immer erfolgreicher gewesen war.

GenPlusKonPlus: Friedrich drangsalierte Svenja, weil er seit der gemeinsamen Schulzeit eine starke Antipathie verspürte.

GenPlusKonMin: Ella drangsalierte Leonard, weil er seit der gemeinsamen Schulzeit immer erfolgreicher gewesen war.

AP 12 drohte

GenMinKonPlus: Damian drohte Marcel, weil er bei dieser existenziellen Finanzinvestition sonst alles verloren hätte.

GenMinKonMin: Damian drohte Marcel, weil er bei dieser existenziellen Finanzinvestition sonst nicht zugestimmt hätte.

GenPlusKonPlus: Damian drohte Regina, weil er bei dieser existenziellen Finanzinvestition sonst alles verloren hätte.

GenPlusKonMin: Renate drohte Marcel, weil er bei dieser existenziellen Finanzinvestition sonst nicht zugestimmt hätte.

AP 13 folgte

GenMinKonPlus: Birgit folgte Ruth, weil sie in dieser fremden Stadt nicht den Weg kannte.

GenMinKonMin: Birgit folgte Ruth, weil sie in dieser fremden Stadt offenbar den Weg kannte.

GenPlusKonPlus: Birgit folgte Niklas, weil sie in dieser fremden Stadt nicht den Weg kannte.

GenPlusKonMin: Adam folgte Ruth, weil sie in dieser fremden Stadt offenbar den Weg kannte.

AP 14 hänselte

GenMinKonPlus: Dolores hänselte Olivia, weil sie nach der misslungenen Diät nichts als Schadenfreude fühlte.

GenMinKonMin: Dolores hänselte Olivia, weil sie nach der misslungenen Diät noch stärker zugenommen hatte.

GenPlusKonPlus: Dolores hänselte Albert, weil sie nach der misslungenen Diät nichts als Schadenfreude fühlte.

GenPlusKonMin: Wolf hänselte Olivia, weil sie nach der misslungenen Diät noch stärker zugenommen hatte.

AP 15 hinderte

GenMinKonPlus: Ella hinderte Svenja, weil sie wegen der angespannten Lage neue Probleme kommen sah.

GenMinKonMin: Ella hinderte Svenja, weil sie wegen der angespannten Lage sich sonst geschadet hätte.

GenPlusKonPlus: Ella hinderte Leonard, weil sie wegen der angespannten Lage neue Probleme kommen sah.

GenPlusKonMin: Friedrich hinderte Svenja, weil sie wegen der angespannten Lage sich sonst geschadet hätte.

AP 16 hinderte

GenMinKonPlus: Renate hinterging Regina, weil sie bei der nächsten Sitzung den Plan durchsetzen wollte.

GenMinKonMin: Renate hinterging Regina, weil sie bei der nächsten Sitzung den Plan ablehnt hätte.

GenPlusKonPlus: Renate hinterging Marcel, weil sie bei der nächsten Sitzung den Plan durchsetzen wollte.

GenPlusKonMin: Damian hinterging Regina, weil sie bei der nächsten Sitzung den Plan ablehnt hätte.

AP 17 hofierte

GenMinKonPlus: Felix hofierte Werner, weil er in der nächsten Legislaturperiode davon nur profitieren konnte.

GenMinKonMin: Felix hofierte Werner, weil er in der nächsten Legislaturperiode bestimmt mächtig werden würde.

GenPlusKonPlus: Felix hofierte Jasmin, weil er in der nächsten Legislaturperiode davon nur profitieren konnte.

GenPlusKonMin: Trudi hofierte Werner, weil er in der nächsten Legislaturperiode bestimmt mächtig werden würde.

AP 18 jagte

GenMinKonPlus: Mario jagte Georg, weil er trotz des gerichtlichen Freispruchs persönlich Rache üben wollte.

GenMinKonMin: Mario jagte Georg, weil er trotz des gerichtlichen Freispruchs des Verbrechens schuldig war.

GenPlusKonPlus: Mario jagte Marie, weil er trotz des gerichtlichen Freispruchs persönlich Rache üben wollte.

GenPlusKonMin: Petra jagte Georg, weil er trotz des gerichtlichen Freispruchs des Verbrechens schuldig war.

AP 19 kontaktierte

GenMinKonPlus: Matthias kontaktierte Heiko, weil er auf die offene Frage endlich Antwort haben wollte.

GenMinKonMin: Matthias kontaktierte Heiko, weil er auf die offene Frage eine Antwort geben konnte.

GenPlusKonPlus: Matthias kontaktierte Cornelia, weil er auf die offene Frage endlich Antwort haben wollte.

GenPlusKonMin: Luise kontaktierte Heiko, weil er auf die offene Frage eine Antwort geben konnte.

AP 20 kränkte

GenMinKonPlus: Günther kränkte Tim, weil er bei jeder denkbaren Gelegenheit in ein Fett-
nöpfchen trat.

GenMinKonMin: Günther kränkte Tim, weil er bei jeder denkbaren Gelegenheit leicht sehr
beleidigt reagierte

GenPlusKonPlus: Günther kränkte Melanie, weil er bei jeder denkbaren Gelegenheit in ein
Fettnöpfchen trat.

GenPlusKonMin: Klara kränkte Tim, weil er bei jeder denkbaren Gelegenheit leicht sehr be-
leidigt reagierte.

AP 21 neckte

GenMinKonPlus: Trudi neckte Jasmin, weil sie nach einem Glas Wein in alberner Stimmung war.

GenMinKonMin: Trudi neckte Jasmin, weil sie nach einem Glas Wein immer alle Begriffe verwechselte.

GenPlusKonPlus: Trudi neckte Werner, weil sie nach einem Glas Wein in alberner Stimmung war.

GenPlusKonMin: Felix neckte Jasmin, weil sie nach einem Glas Wein immer alle Begriffe verwechselte.

AP 22 quälte

GenMinKonPlus: Petra quälte Marie, weil sie wegen der abgelehnten Bewerbung in bössartiger Stimmung war.

GenMinKonMin: Petra quälte Marie, weil sie wegen der abgelehnten Bewerbung ein leichtes Opfer war.

GenPlusKonPlus: Petra quälte Georg, weil sie wegen der abgelehnten Bewerbung in bössartiger Stimmung war.

GenPlusKonMin: Mario quälte Marie, weil sie wegen der abgelehnten Bewerbung ein leichtes Opfer war.

AP 23 schadete

GenMinKonPlus: Luise schadete Cornelia, weil sie bei allen wichtigen Entscheidungen wieder keine Mitsprache duldete.

GenMinKonMin: Luise schadete Cornelia, weil sie bei allen wichtigen Entscheidungen immer den Kürzeren zog.

GenPlusKonPlus: Luise schadete Heiko, weil sie bei allen wichtigen Entscheidungen wieder keine Mitsprache duldete.

GenPlusKonMin: Matthias schadete Cornelia, weil sie bei allen wichtigen Entscheidungen immer den Kürzeren zog.

AP 24 schikanierte

GenMinKonPlus: Klara schikanierte Melanie, weil sie gerade auf diesem Gebiet ungewünschte Konkurrenz vergraulen wollte.

GenMinKonMin: Klara schikanierte Melanie, weil sie gerade auf diesem Gebiet so viele Angriffspunkte bot.

GenPlusKonPlus: Klara schikanierte Tim, weil sie gerade auf diesem Gebiet ungewünschte Konkurrenz vergraulen wollte.

GenPlusKonMin: Günther schikanierte Melanie, weil sie gerade auf diesem Gebiet so viele Angriffspunkte bot.

AP 25 schlug

GenMinKonPlus: Jochen schlug Pascal, weil er nach dieser letzten Aktion die Kontrolle verloren hatte.

GenMinKonMin: Jochen schlug Pascal, weil er nach dieser letzten Aktion nichts anderes verdient hatte.

GenPlusKonPlus: Jochen schlug Susanne, weil er nach dieser letzten Aktion die Kontrolle verloren hatte.

GenPlusKonMin: Christiane schlug Pascal, weil er nach dieser letzten Aktion nichts anderes verdient hatte.

AP 26 schmeichelte

GenMinKonPlus: Ludwig schmeichelte Hans, weil er für das geplante Tennisspiel dringend einen Partner brauchte.

GenMinKonMin: Ludwig schmeichelte Hans, weil er für das geplante Tennisspiel ein optimaler Partner war.

GenPlusKonPlus: Ludwig schmeichelte Uta, weil er für das geplante Tennisspiel dringend einen Partner brauchte.

GenPlusKonMin: Heike schmeichelte Hans, weil er für das geplante Tennisspiel ein optimaler Partner war.

AP 27 schüttelte

GenMinKonPlus: Artur schüttelte Achim, weil er in dieser gefährlichen Situation keinen anderen Ausweg wusste.

GenMinKonMin: Artur schüttelte Achim, weil er in dieser gefährlichen Situation zur Salzsäule erstarrt war.

GenPlusKonPlus: Artur schüttelte Dagmar, weil er in dieser gefährlichen Situation keinen anderen Ausweg wusste.

GenPlusKonMin: Ulrike schüttelte Achim, weil er in dieser gefährlichen Situation zur Salzsäule erstarrt war.

AP 28 täuschte

GenMinKonPlus: Frank täuschte Tobias, weil er für eine offene Konfrontation einfach zu feige war.

GenMinKonMin: Frank täuschte Tobias, weil er für eine offene Konfrontation einfach zu mächtig war.

GenPlusKonPlus: Frank täuschte Ursula, weil er für eine offene Konfrontation einfach zu feige war.

GenPlusKonMin: Tanja täuschte Tobias, weil er für eine offene Konfrontation einfach zu mächtig war.

AP 29 trotzte

GenMinKonPlus: Christiane trotzte Susanne, weil sie bei dieser komplizierten Frage es einfach besser wusste

GenMinKonMin: Christiane trotzte Susanne, weil sie bei dieser komplizierten Frage es einfach keine Ahnung hatte.

GenPlusKonPlus: Christiane trotzte Pascal, weil sie bei dieser komplizierten Frage es einfach besser wusste

GenPlusKonMin: Jochen trotzte Susanne, weil sie bei dieser komplizierten Frage es einfach keine Ahnung hatte.

AP 30 überflügelte

GenMinKonPlus: Heike überflügelte Uta, weil sie beim letzten Marathon-Lauf einfach besser vorbereitet war.

GenMinKonMin: Heike überflügelte Uta, weil sie beim letzten Marathon-Lauf einfach schlechter vorbereitet war.

GenPlusKonPlus: Heike überflügelte Uta, weil sie beim letzten Marathon-Lauf einfach besser vorbereitet war.

GenPlusKonMin: Heike überflügelte Uta, weil sie beim letzten Marathon-Lauf einfach schlechter vorbereitet war.

AP 31 überlistete

GenMinKonPlus: Ulrike überlistete Dagmar, weil sie in diesem besonderen Fall anders nicht gewonnen hätte.

GenMinKonMin: Ulrike überlistete Dagmar, weil sie in diesem besonderen Fall sich besonders dumm anstellte.

GenPlusKonPlus: Ulrike überlistete Achim, weil sie in diesem besonderen Fall anders nicht gewonnen hätte.

GenPlusKonMin: Artur überlistete Dagmar, weil sie in diesem besonderen Fall sich besonders dumm anstellte.

AP 32 überredete

GenMinKonPlus: Tanja überredete Ursula, weil sie am Ende der Diskussion die besseren Argumente hatte.

GenMinKonMin: Tanja überredete Ursula, weil sie am Ende der Diskussion keine Argumente mehr hatte.

GenPlusKonPlus: Tanja überredete Tobias, weil sie am Ende der Diskussion die besseren Argumente hatte.

GenPlusKonMin: Frank überredete Ursula, weil sie am Ende der Diskussion keine Argumente mehr hatte.

AP 33 überrumpelte

GenMinKonPlus: Emil überrumpelte Oswald, weil er in dieser undurchsichtigen Situation einen schnellen Erfolg witterte.

GenMinKonMin: Emil überrumpelte Oswald, weil er in dieser undurchsichtigen Situation das Denken eingestellt hatte.

GenPlusKonPlus: Emil überrumpelte Ingeborg, weil er in dieser undurchsichtigen Situation einen schnellen Erfolg witterte.

GenPlusKonMin: Bettina überrumpelte Oswald, weil er in dieser undurchsichtigen Situation das Denken eingestellt hatte.

AP 34 übertraf

GenMinKonPlus: Eugen übertraf Gottfried, weil er gerade bei komplizierten Aufgaben sehr viel besser war.

GenMinKonMin: Eugen übertraf Gottfried, weil er gerade bei komplizierten Aufgaben sehr schnell verkrampft wurde.

GenPlusKonPlus: Eugen übertraf Gottfried, weil er gerade bei komplizierten Aufgaben sehr viel besser war.

GenPlusKonMin: Eugen übertraf Gottfried, weil er gerade bei komplizierten Aufgaben sehr schnell verkrampft wurde.

AP 35 verhörte

GenMinKonPlus: Adrian verhörte Jonathan, weil er wegen des fehlenden Alibis immer noch misstrauisch war.

GenMinKonMin: Adrian verhörte Jonathan, weil er wegen des fehlenden Alibis immer noch verdächtig war.

GenPlusKonPlus: Adrian verhörte Doris, weil er wegen des fehlenden Alibis immer noch misstrauisch war.

GenPlusKonMin: Edina verhörte Jonathan, weil er wegen des fehlenden Alibis immer noch verdächtig war.

AP 36 verleitete

GenMinKonPlus: Boris verleitete Nils, weil er in fast allen Dingen eine mitreißende Persönlichkeit hatte.

GenMinKonMin: Boris verleitete Nils, weil er in fast allen Dingen sehr leicht beeinflussbar war.

GenPlusKonPlus: Boris verleitete Patrizia, weil er in fast allen Dingen eine mitreißende Persönlichkeit hatte.

GenPlusKonMin: Nina verleitete Nils, weil er in fast allen Dingen sehr leicht beeinflussbar war.

AP 37 verletzte

GenMinKonPlus: Bettina verletzte Ingeborg, weil sie bei diesem persönlichen Thema ohne Rücksicht aggressiv war.

GenMinKonMin: Bettina verletzte Ingeborg, weil sie bei diesem persönlichen Thema ganz besonders überempfindlich reagierte.

GenPlusKonPlus: Bettina verletzte Oswald, weil sie bei diesem persönlichen Thema ohne Rücksicht aggressiv war.

GenPlusKonMin: Emil verletzte Ingeborg, weil sie bei diesem persönlichen Thema ganz besonders überempfindlich reagierte.

AP 38 verleumdete

GenMinKonPlus: Mathilde verleumdete Rita, weil sie während der gemeinsamen Klassenfahrt viele gemeine Lügen erzählte.

GenMinKonMin: Mathilde verleumdete Rita, weil sie während der gemeinsamen Klassenfahrt selbst Gerüchte verbreitet hatte.

GenPlusKonPlus: Mathilde verleumdete Rita, weil sie während der gemeinsamen Klassenfahrt viele gemeine Lügen erzählte.

GenPlusKonMin: Mathilde verleumdete Rita, weil sie während der gemeinsamen Klassenfahrt selbst Gerüchte verbreitet hatte.

AP 39 verriet

GenMinKonPlus: Edina verriet Doris, weil sie nach Abschluss des Unternehmens überflüssige Mitwisser loswerden wollte.

GenMinKonMin: Edina verriet Doris, weil sie nach Abschluss des Unternehmens nur eine Belastung war.

GenPlusKonPlus: Edina verriet Jonathan, weil sie nach Abschluss des Unternehmens überflüssige Mitwisser loswerden wollte.

GenPlusKonMin: Adrian verleumdete Rita, weil sie während der gemeinsamen Klassenfahrt nur eine Belastung war.

AP 40 widerstand

GenMinKonPlus: Nina widerstand Patrizia, weil sie bei dieser wichtigen Angelegenheit eigentlich keine Versuchung verspürte.

GenMinKonMin: Nina widerstand Patrizia, weil sie bei dieser wichtigen Angelegenheit viel zu risikoreich handelte.

GenPlusKonPlus: Nina widerstand Nils, weil sie bei dieser wichtigen Angelegenheit eigentlich keine Versuchung verspürte.

GenPlusKonMin: Boris widerstand Patrizia, weil sie bei dieser wichtigen Angelegenheit viel zu risikoreich handelte.

A-E Verben *Subjekt: Agens – Objekt: Evocator, Kausalität auf Evocator*

AE 1 bedrängte

GenMinKonMin: Bastian bedrängte Ferdinand, weil er in der unvorhergesehenen Notlage das Geld dringend benötigte.

GenMinKonPlus: Bastian bedrängte Ferdinand, weil er in der unvorhergesehenen Notlage am besten Helfen konnte.

GenPlusKonMin: Bastian bedrängte Ina, weil er in der unvorhergesehenen Notlage das Geld dringend benötigte.

GenPlusKonPlus: Selma bedrängte Ferdinand, weil er in der unvorhergesehenen Notlage am besten Helfen konnte.

AE 2 beglückwünschte

GenMinKonMin: Konstantin beglückwünschte Edgar, weil er bei diesem anspruchsvollen Lehrer die herausragende Leistung respektierte.

GenMinKonPlus: Konstantin beglückwünschte Edgar, weil er bei diesem anspruchsvollen Lehrer herausragende Leistungen gezeigt hatte.

GenPlusKonMin: Konstantin beglückwünschte Annika, weil er bei diesem anspruchsvollen Lehrer die herausragende Leistung respektierte.

GenPlusKonPlus: Elisabeth beglückwünschte Edgar, weil er bei diesem anspruchsvollen Lehrer herausragende Leistungen gezeigt hatte.

AE 3 begünstigte

GenMinKonMin: Horst begünstigte Vinzenz, weil er für die jahrelange Treue eine Gegenleistung erbringen wollte.

GenMinKonPlus: Horst begünstigte Vinzenz, weil er für die jahrelange Treue endlich eine Entschädigung verdiente.

GenPlusKonMin: Horst begünstigte Vinzenz, weil er für die jahrelange Treue eine Gegenleistung erbringen wollte.

GenPlusKonPlus: Horst begünstigte Vinzenz, weil er für die jahrelange Treue endlich eine Entschädigung verdiente.

AE 4 belächelte

GenMinKonMin: Karl belächelte Guido, weil er in ziemlich unkonventioneller Art schrullige Originale sympathisch fand.

GenMinKonPlus: Karl belächelte Guido, weil er in ziemlich unkonventioneller Art häufig schrulliges Verhalten zeigte.

GenPlusKonMin: Karl belächelte Agnes, weil er in ziemlich unkonventioneller Art schrullige Originale sympathisch fand.

GenPlusKonPlus: Martha belächelte Guido, weil er in ziemlich unkonventioneller Art häufig schrulliges Verhalten zeigte.

AE 5 belohnte

GenMinKonMin: Selma belohnte Ina, weil sie wegen des anstrengenden Umzuges etwas gut machen wollte.

GenMinKonPlus: Selma belohnte Ina, weil sie wegen des anstrengenden Umzuges viel Zeit geopfert hatte.

GenPlusKonMin: Selma belohnte Ferdinand, weil sie wegen des anstrengenden Umzuges etwas gut machen wollte.

GenPlusKonPlus: Bastian belohnte Ina, weil sie wegen des anstrengenden Umzuges viel Zeit geopfert hatte.

AE 6 beschimpft

GenMinKonMin: Elisabeth beschimpfte Annika, weil sie wegen des vergessenen Babysittertermins schrecklich stark aufgebracht war.

GenMinKonPlus: Elisabeth beschimpfte Annika, weil sie wegen des vergessenen Babysittertermins keinerlei Entschuldigung vorgebracht hatte.

GenPlusKonMin: Elisabeth beschimpfte Annika, weil sie wegen des vergessenen Babysittertermins schrecklich stark aufgebracht war.

GenPlusKonPlus: Elisabeth beschimpfte Annika, weil sie wegen des vergessenen Babysittertermins keinerlei Entschuldigung vorgebracht hatte.

AE 7 beschuldigte

GenMinKonMin: Lara beschuldigte Karin, weil sie seit dem letzten Krankheitsschub unter starkem Verfolgungswahn litt.

GenMinKonPlus: Lara beschuldigte Karin, weil sie seit dem letzten Krankheitsschub nicht einmal angerufen hatte.

GenPlusKonMin: Lara beschuldigte Vinzenz, weil sie seit dem letzten Krankheitsschub unter starkem Verfolgungswahn litt.

GenPlusKonPlus: Horst beschuldigte Karin, weil sie seit dem letzten Krankheitsschub nicht einmal angerufen hatte.

AE 8 beschwichtigte

GenMinKonMin: Martha beschwichtigte Agnes, weil sie während des gemeinsamen Urlaubs keinen Streit haben wollte.

GenMinKonPlus: Martha beschwichtigte Agnes, weil sie während des gemeinsamen Urlaubs unnötig wütend geworden war.

GenPlusKonMin: Martha beschwichtigte Agnes, weil sie während des gemeinsamen Urlaubs keinen Streit haben wollte.

GenPlusKonPlus: Martha beschwichtigte Agnes, weil sie während des gemeinsamen Urlaubs unnötig wütend geworden war.

AE 9 bespitzelte

GenMinKonMin: Johannes bespitzelte Anton, weil er wegen des fehlenden Geldes eine böse Vermutung hegte.

GenMinKonPlus: Johannes bespitzelte Anton, weil er wegen des fehlenden Geldes schon mehrfach gelogen hatte.

GenPlusKonMin: Johannes bespitzelte Anton, weil er wegen des fehlenden Geldes eine böse Vermutung hegte.

GenPlusKonPlus: Johannes bespitzelte Anton, weil er wegen des fehlenden Geldes schon mehrfach gelogen hatte.

AE 10 bestärkte

GenMinKonMin: Egon bestärkte Till, weil er aufgrund besonders großer Fachkenntnis die Vorschläge ausgezeichnet fand.

GenMinKonPlus: Egon bestärkte Till, weil er aufgrund besonders großer Fachkenntnis immer ausgezeichnete Vorschläge machte.

GenPlusKonMin: Egon bestärkte Laura, weil er aufgrund besonders großer Fachkenntnis die Vorschläge ausgezeichnet fand.

GenPlusKonPlus: Simone bestärkte Till, weil er aufgrund besonders großer Fachkenntnis immer ausgezeichnete Vorschläge machte.

AE 11 bestrafte

GenMinKonMin: Udo bestrafte Oliver, weil er zur Aufrechterhaltung der Disziplin Regelverletzungen streng ahnden musste.

GenMinKonPlus: Udo bestrafte Oliver, weil er zur Aufrechterhaltung der Disziplin als Sündenbock dienen sollte.

GenPlusKonMin: Udo bestrafte Diana, weil er zur Aufrechterhaltung der Disziplin Regelverletzungen streng ahnden musste.

GenPlusKonPlus: Agathe bestrafte Oliver, weil er zur Aufrechterhaltung der Disziplin als Sündenbock dienen sollte.

AE 12 dankte

GenMinKonMin: Fabian dankte Hartmut, weil er bei der nachgeholtten Geburtstagsfeier das Geschenk wunderbar fand.

GenMinKonPlus: Fabian dankte Hartmut, weil er bei der nachgeholtten Geburtstagsfeier ein Geschenk mitgebracht hatte.

GenPlusKonMin: Fabian dankte Yvonne, weil er bei der nachgeholtten Geburtstagsfeier das Geschenk wunderbar fand.

GenPlusKonPlus: Hannelore dankte Hartmut, weil er bei der nachgeholtten Geburtstagsfeier ein Geschenk mitgebracht hatte.

AE 13 empfahl

GenMinKonMin: Andrea empfahl Nora, weil sie von allen befragten Personen nur Gutes gehört hatte.

GenMinKonPlus: Andrea empfahl Nora, weil sie von allen befragten Personen am besten geeignet war.

GenPlusKonMin: Andrea empfahl Anton, weil sie von allen befragten Personen nur Gutes gehört hatte.

GenPlusKonPlus: Johannes empfahl Nora, weil sie von allen befragten Personen am besten geeignet war.

AE 14 engagierte

GenMinKonMin: Simone engagierte Laura, weil sie nach dem öffentlichen Bewerbungsvortrag alle Anforderungen erfüllt sah.

GenMinKonPlus: Simone engagierte Laura, weil sie nach dem öffentlichen Bewerbungsvortrag eindeutig am qualifiziertesten war.

GenPlusKonMin: Simone engagierte Till, weil sie nach dem öffentlichen Bewerbungsvortrag alle Anforderungen erfüllt sah.

GenPlusKonPlus: Egon engagierte Laura, weil sie nach dem öffentlichen Bewerbungsvortrag eindeutig am qualifiziertesten war.

AE 15 entließ

GenMinKonMin: Agathe entließ Diana, weil sie im Hochsicherheitstrakt des Atomreaktors Schlampigkeiten nicht dulden konnte.

GenMinKonPlus: Agathe entließ Diana, weil sie im Hochsicherheitstrakt des Atomreaktors wiederholt schlampig gearbeitet hatte.

GenPlusKonMin: Agathe entließ Oliver, weil sie im Hochsicherheitstrakt des Atomreaktors Schlampigkeiten nicht dulden konnte.

GenPlusKonPlus: Udo entließ Diana, weil sie im Hochsicherheitstrakt des Atomreaktors wiederholt schlampig gearbeitet hatte.

AE 16 entschuldigte

GenMinKonMin: Hannelore entschuldigte Yvonne, weil sie als einzige eingeladene Person von der Erkrankung wusste.

GenMinKonPlus: Hannelore entschuldigte Yvonne, weil sie als einzige eingeladene Person überraschend schwer erkrankt war.

GenPlusKonMin: Hannelore entschuldigte Hartmut, weil sie als einzige eingeladene Person von der Erkrankung wusste.

GenPlusKonPlus: Fabian entschuldigte Yvonne, weil sie als einzige eingeladene Person überraschend schwer erkrankt war.

AE 17 erkannte

GenMinKonMin: Mirko erkannte Simon, weil er trotz der langen Jahre ein ausgezeichnetes Gedächtnis hatte.

GenMinKonPlus: Mirko erkannte Simon, weil er trotz der langen Jahre praktisch völlig unverändert aussah.

GenPlusKonMin: Mirko erkannte Gisela, weil er trotz der langen Jahre ein ausgezeichnetes Gedächtnis hatte.

GenPlusKonPlus: Debora erkannte Simon, weil er trotz der langen Jahre praktisch völlig unverändert aussah.

AE 18 förderte

GenMinKonMin: Holger förderte Eike, weil er für die zu besetzende Stelle schnell eine Nachfolge brauchte.

GenMinKonPlus: Holger förderte Eike, weil er für die zu besetzende Stelle wirklich ideale Voraussetzungen hatte.

GenPlusKonMin: Holger förderte Elsa, weil er für die zu besetzende Stelle schnell eine Nachfolge brauchte.

GenPlusKonPlus: Nena förderte Eike, weil er für die zu besetzende Stelle wirklich ideale Voraussetzungen hatte.

AE 19 glaubte

GenMinKonMin: Harald glaubte Michael, weil er in all den Jahren viel Vertrauen gefasst hatte.

GenMinKonPlus: Harald glaubte Michael, weil er in all den Jahren noch nie gelogen hatte.

GenPlusKonMin: Harald glaubte Ida, weil er in all den Jahren viel Vertrauen gefasst hatte.

GenPlusKonPlus: Frida glaubte Michael, weil er in all den Jahren noch nie gelogen hatte.

AE 20 gratulierte

GenMinKonMin: Philip gratulierte Lothar, weil sie in der abschließenden Klausur jede bessere Leistung bewunderte.

GenMinKonPlus: Philip gratulierte Lothar, weil sie in der abschließenden Klausur die meisten Punkte hatte.

GenPlusKonMin: Philip gratulierte Jana, weil sie in der abschließenden Klausur jede bessere Leistung bewunderte.

GenPlusKonPlus: Angela gratulierte Lothar, weil sie in der abschließenden Klausur die meisten Punkte hatte.

AE 21 half

GenMinKonMin: Debora half Gisela, weil sie kurz vor der Klausur den Stoff gut konnte.

GenMinKonPlus: Debora half Gisela, weil sie kurz vor der Klausur den Stoff nicht konnte.

GenPlusKonMin: Debora half Simon, weil sie kurz vor der Klausur den Stoff gut konnte.

GenPlusKonPlus: Mirko half Gisela, weil sie kurz vor der Klausur den Stoff nicht konnte.

AE 22 kritisierte

GenMinKonMin: Nena kritisierte Elsa, weil sie nach der langen Prüfungsvorbereitung diese Ängste irrational fand.

GenMinKonPlus: Nena kritisierte Elsa, weil sie nach der langen Prüfungsvorbereitung dauernd schlechte Laune hatte.

GenPlusKonMin: Nena kritisierte Eike, weil sie nach der langen Prüfungsvorbereitung diese Ängste irrational fand.

GenPlusKonPlus: Holger kritisierte Elsa, weil sie nach der langen Prüfungsvorbereitung dauernd schlechte Laune hatte.

AE 23 lobte

GenMinKonMin: Frida lobte Elsa, weil sie bei dem entscheidenden Endspiel solchen Einsatz wirklich bewunderte.

GenMinKonPlus: Frida lobte Elsa, weil sie bei dem entscheidenden Endspiel solchen Einsatz gezeigt hatte.

GenPlusKonMin: Frida lobte Michael, weil sie bei dem entscheidenden Endspiel solchen Einsatz wirklich bewunderte.

GenPlusKonPlus: Harald lobte Elsa, weil sie bei dem entscheidenden Endspiel solchen Einsatz gezeigt hatte.

AE 24 rügte

GenMinKonMin: Angela rügte Jana, weil sie auch nach einer Woche das Geschirr ungespült vorfand.

GenMinKonPlus: Angela rügte Jana, weil sie auch nach einer Woche noch nicht gespült hatte.

GenPlusKonMin: Angela rügte Jana, weil sie auch nach einer Woche das Geschirr ungespült vorfand.

GenPlusKonPlus: Angela rügte Jana, weil sie auch nach einer Woche noch nicht gespült hatte.

AE 25 rühmte

GenMinKonMin: Moritz rühmte Ulrich, weil er bei der diesjährigen Meisterschaft wahre Größe anerkennen konnte.

GenMinKonPlus: Moritz rühmte Ulrich, weil er bei der diesjährigen Meisterschaft großen Mut bewiesen hatte.

GenPlusKonMin: Moritz rühmte Margot, weil er bei der diesjährigen Meisterschaft wahre Größe anerkennen konnte.

GenPlusKonPlus: Alina rühmte Ulrich, weil er bei der diesjährigen Meisterschaft großen Mut bewiesen hatte.

AE 26 tadelte

GenMinKonMin: Eduard tadelte Bodo, weil er gerade bei wichtigen Verhandlungen Nachlässigkeiten nicht gutheißen konnte.

GenMinKonPlus: Eduard tadelte Bodo, weil er gerade bei wichtigen Verhandlungen häufig peinliche Fehler machte.

GenPlusKonMin: Eduard tadelte Flora, weil er gerade bei wichtigen Verhandlungen Nachlässigkeiten nicht gutheißen konnte.

GenPlusKonPlus: Kerstin tadelte Bodo, weil er gerade bei wichtigen Verhandlungen häufig peinliche Fehler machte.

AE 27 tröstete

GenMinKonMin: Wim tröstete Herbert, weil er bei dieser unerwarteten Pechsträhne sehr großes Mitleid empfand.

GenMinKonPlus: Wim tröstete Herbert, weil er bei dieser unerwarteten Pechsträhne einfach etwas Mitgefühl verdiente.

GenPlusKonMin: Wim tröstete Brigitte, weil er bei dieser unerwarteten Pechsträhne sehr großes Mitleid empfand.

GenPlusKonPlus: Paula tröstete Herbert, weil er bei dieser unerwarteten Pechsträhne einfach etwas Mitgefühl verdiente.

AE 28 umwarb

GenMinKonMin: Leon umwarb Volker, weil er für die nahende Statistik-Klausur eine kompetente Beratung brauchte.

GenMinKonPlus: Leon umwarb Volker, weil er für die nahende Statistik-Klausur die richtigen Lösungen hatte.

GenPlusKonMin: Leon umwarb Monika, weil er für die nahende Statistik-Klausur eine kompetente Beratung brauchte.

GenPlusKonPlus: Silke umwarb Volker, weil er für die nahende Statistik-Klausur die richtigen Lösungen hatte.

AE 29 unterstützte

GenMinKonMin: Alina unterstützte Margot, weil sie nach der nächsten Wahl auf eine Gegenleistung hoffte.

GenMinKonPlus: Alina unterstützte Margot, weil sie nach der nächsten Wahl eine Gegenleistung versprochen hatte.

GenPlusKonMin: Alina unterstützte Ulrich, weil sie nach der nächsten Wahl auf eine Gegenleistung hoffte.

GenPlusKonPlus: Moritz unterstützte Margot, weil sie nach der nächsten Wahl eine Gegenleistung versprochen hatte.

AE 30 verbesserte

GenMinKonMin: Kerstin verbesserte Flora, weil sie in der überarbeiteten Hausarbeit immer noch Fehler entdeckte.

GenMinKonPlus: Kerstin verbesserte Flora, weil sie in der überarbeiteten Hausarbeit den Fehler übersehen hatte.

GenPlusKonMin: Kerstin verbesserte Bodo, weil sie in der überarbeiteten Hausarbeit immer noch Fehler entdeckte.

GenPlusKonPlus: Eduard verbesserte Flora, weil sie in der überarbeiteten Hausarbeit den Fehler übersehen hatte.

AE 31 verdächtigte

GenMinKonMin: Paula verdächtigte Brigitte, weil sie durch die offensichtlichen Lügen zunehmend misstrauisch geworden war.

GenMinKonPlus: Paula verdächtigte Brigitte, weil sie durch die offensichtlichen Lügen etwas zu verbergen versuchte.

GenPlusKonMin: Paula verdächtigte Herbert, weil sie durch die offensichtlichen Lügen zunehmend misstrauisch geworden war.

GenPlusKonPlus: Wim verdächtigte Brigitte, weil sie durch die offensichtlichen Lügen etwas zu verbergen versuchte.

AE 32 vergab

GenMinKonMin: Silke vergab Monika, weil sie bei dem klärenden Gespräch die Entschuldigung akzeptieren konnte.

GenMinKonPlus: Silke vergab Monika, weil sie bei dem klärenden Gespräch um Verzeihung gebeten hatte.

GenPlusKonMin: Silke vergab Volker, weil sie bei dem klärenden Gespräch die Entschuldigung akzeptieren konnte.

GenPlusKonPlus: Leon vergab Monika, weil sie bei dem klärenden Gespräch um Verzeihung gebeten hatte.

AE 33 verhöhnzte

GenMinKonMin: Arne verhöhnzte Alexander, weil er auch nach langen Jahren immer noch empört war.

GenMinKonPlus: Arne verhöhnzte Alexander, weil er auch nach langen Jahren nichts dazu gelernt hatte.

GenPlusKonMin: Arne verhöhnzte Marion, weil er auch nach langen Jahren immer noch empört war.

GenPlusKonPlus: Rosa verhöhnzte Alexander, weil er auch nach langen Jahren nichts dazu gelernt hatte.

AE 34 verklagte

GenMinKonMin: Christian verklagte Fred, weil er aufgrund der klaren Beweislage einen vollständigen Prozesserfolg erwartete.

GenMinKonPlus: Christian verklagte Fred, weil er aufgrund der klaren Beweislage den Prozess verlieren würde.

GenPlusKonMin: Christian verklagte Claudia, weil er aufgrund der klaren Beweislage einen vollständigen Prozesserfolg erwartete.

GenPlusKonPlus: Helga verklagte Fred, weil er aufgrund der klaren Beweislage den Prozess verlieren würde.

AE 35 verließ

GenMinKonMin: Urs verließ Julian, weil er nach den Jahren des Zusammenlebens plötzlich allein sein wollte.

GenMinKonPlus: Urs verließ Julian, weil er nach den Jahren des Zusammenlebens plötzlich immer Streit anfang.

GenPlusKonMin: Urs verließ Gerda, weil er nach den Jahren des Zusammenlebens plötzlich allein sein wollte.

GenPlusKonPlus: Erika verließ Julian, weil er nach den Jahren des Zusammenlebens plötzlich immer Streit anfang.

AE 36 verspottete

GenMinKonMin: Erich verspottete Ivo, weil er durch die aufgeflogene Täuschung die ideale Gelegenheit hatte.

GenMinKonPlus: Erich verspottete Ivo, weil er durch die aufgeflogene Täuschung als Betrüger entlarvt wurde.

GenPlusKonMin: Erich verspottete Uschi, weil er durch die aufgeflogene Täuschung die ideale Gelegenheit hatte.

GenPlusKonPlus: Gina verspottete Ivo, weil er durch die aufgeflogene Täuschung als Betrüger entlarvt wurde.

AE 37 verteidigte

GenMinKonMin: Rosa verteidigte Marion, weil sie bei dem großen Stress den Fehler entschuldbar fand.

GenMinKonPlus: Rosa verteidigte Marion, weil sie bei dem großen Stress nicht allein verantwortlich war.

GenPlusKonMin: Rosa verteidigte Alexander, weil sie bei dem großen Stress den Fehler entschuldbar fand.

GenPlusKonPlus: Arne verteidigte Marion, weil sie bei dem großen Stress nicht allein verantwortlich war.

AE 38 vertraute

GenMinKonMin: Helga vertraute Claudia, weil sie aufgrund der bisherigen Zusammenarbeit einfach nicht anders konnte.

GenMinKonPlus: Helga vertraute Claudia, weil sie aufgrund der bisherigen Zusammenarbeit am besten Bescheid wusste.

GenPlusKonMin: Helga vertraute Fred, weil sie aufgrund der bisherigen Zusammenarbeit einfach nicht anders konnte.

GenPlusKonPlus: Christian vertraute Claudia, weil sie aufgrund der bisherigen Zusammenarbeit am besten Bescheid wusste.

AE 39 verzieh

GenMinKonMin: Erika verzieh Gerda, weil sie durch das gute Klausurergebnis heute gutmütig gestimmt war.

GenMinKonPlus: Erika verzieh Gerda, weil sie durch das gute Klausurergebnis alles wieder gut machte.

GenPlusKonMin: Erika verzieh Julian, weil sie durch das gute Klausurergebnis heute gutmütig gestimmt war.

GenPlusKonPlus: Urs verzieh Gerda, weil sie durch das gute Klausurergebnis alles wieder gut machte.

AE 40 würdigte

GenMinKonMin: Gina würdigte Uschi, weil sie gerade in schwierigen Zeiten die Erfolge neidlos anerkannte.

GenMinKonPlus: Gina würdigte Uschi, weil sie gerade in schwierigen Zeiten wirklich Herausragendes geleistet hatte.

GenPlusKonMin: Gina würdigte Ivo, weil sie gerade in schwierigen Zeiten die Erfolge neidlos anerkannte.

GenPlusKonPlus: Erich würdigte Uschi, weil sie gerade in schwierigen Zeiten wirklich Herausragendes geleistet hatte.

S-E Verben *Subjekt: Stimulus – Objekt: Experiencer, Kausalität auf Stimulus*

SE1 amüsierte

GenMinKonPlus: Bernhard amüsiert Dieter, weil er nach dem peinlichen Misserfolg einen unverwüstlichen Humor hatte.

GenMinKonMin: Bernhard amüsiert Dieter, weil er nach dem peinlichen Misserfolg dringend etwas Aufmunterung brauchte.

GenPlusKonPlus: Bernhard amüsiert Jennifer, weil er nach dem peinlichen Misserfolg einen unverwüstlichen Humor hatte.

GenPlusKonMin: Berta amüsiert Dieter, weil er nach dem peinlichen Misserfolg dringend etwas Aufmunterung brauchte.

SE2 ärgerte

GenMinKonPlus: August ärgerte Richard, weil er während der schweren Klausur überhaupt niemanden abschreiben ließ.

GenMinKonMin: August ärgerte Richard, weil er während der schweren Klausur auf Geräusche empfindlich reagiert.

GenPlusKonPlus: August ärgerte Ines, weil er während der schweren Klausur überhaupt niemanden abschreiben ließ.

GenPlusKonMin: Rabea ärgerte Richard, weil er während der schweren Klausur auf Geräusche empfindlich reagiert.

SE3 bedrohte

GenMinKonPlus: Heinz bedrohte Peter, weil er auf dem leeren Schulhof einen klaren Vorteil hatte.

GenMinKonMin: Heinz bedrohte Peter, weil er auf dem leeren Schulhof ein leichtes Opfer war.

GenPlusKonPlus: Heinz bedrohte Beate, weil er auf dem leeren Schulhof einen klaren Vorteil hatte.

GenPlusKonMin: Beate bedrohte Peter, weil er auf dem leeren Schulhof ein leichtes Opfer war.

SE4 beeindruckte

GenMinKonPlus: Sebastian beeindruckte Hannes, weil er in jeder denkbaren Situation einen coolen Spruch wusste.

GenMinKonMin: Sebastian beeindruckte Hannes, weil er in jeder denkbaren Situation ein cooles Vorbild suchte.

GenPlusKonPlus: Sebastian beeindruckte Tina, weil er in jeder denkbaren Situation einen coolen Spruch wusste.

GenPlusKonMin: Rachel beeindruckte Hannes, weil er in jeder denkbaren Situation ein cooles Vorbild suchte.

SE5 befremdete

GenMinKonPlus: Berta befremdete Jennifer, weil sie in der vorliegenden Situation vom gewohnten Vorgehen abwich.

GenMinKonMin: Berta befremdete Jennifer, weil sie in der vorliegenden Situation verbindliche Absprachen missachtet sah.

GenPlusKonPlus: Berta befremdete Dieter, weil sie in der vorliegenden Situation vom gewohnten Vorgehen abwich.

GenPlusKonMin: Bernhard befremdete Jennifer, weil sie in der vorliegenden Situation verbindliche Absprachen missachtet sah.

SE6 begeisterte

GenMinKonPlus: Rabea begeisterte Ines, weil sie durch diese neue Gesangstechnik gewaltige Fortschritte gemacht hatte.

GenMinKonMin: Rabea begeisterte Ines, weil sie durch diese neue Gesangstechnik schnell zu beeindrucken war.

GenPlusKonPlus: Rabea begeisterte Richard, weil sie durch diese neue Gesangstechnik gewaltige Fortschritte gemacht hatte.

GenPlusKonMin: August begeisterte Ines, weil sie durch diese neue Gesangstechnik schnell zu beeindrucken war.

SE7 belästigte

GenMinKonPlus: Beate belästigte Emma, weil sie während der ganzen Bahnfahrt nur dummes Zeug redete.

GenMinKonMin: Beate belästigte Emma, weil sie während der ganzen Bahnfahrt eigentlich konzentriert arbeiten wollte.

GenPlusKonPlus: Beate belästigte Peter, weil sie während der ganzen Bahnfahrt nur dummes Zeug redete.

GenPlusKonMin: Heinz belästigte Emma, weil sie während der ganzen Bahnfahrt eigentlich konzentriert arbeiten wollte.

SE8 beschämte

GenMinKonPlus: Rachel beschämte Tina, weil sie bei den schwierigen Aufgaben sehr viel besser war.

GenMinKonMin: Rachel beschämte Tina, weil sie bei den schwierigen Aufgaben sehr viel schlechter war.

GenPlusKonPlus: Rachel beschämte Hannes, weil sie bei den schwierigen Aufgaben sehr viel besser war.

GenPlusKonMin: Sebastian beschämte Tina, weil sie bei den schwierigen Aufgaben sehr viel schlechter war.

SE9 bestürzte

GenMinKonPlus: Max bestürzte Olaf, weil er bis zum letzten Ende nichts davon erzählt hatte.

GenMinKonMin: Max bestürzte Olaf, weil er bis zum letzten Ende völlig ahnungslos gewesen war.

GenPlusKonPlus: Max bestürzte Nathalie, weil er bis zum letzten Ende nichts davon erzählt hatte.

GenPlusKonMin: Astrid bestürzte Olaf, weil er bis zum letzten Ende völlig ahnungslos gewesen war.

SE10 betrückte

GenMinKonPlus: Uwe betrückte Alfred, weil er bei diesem wichtigen Problem nicht genügend Anstrengung investierte.

GenMinKonMin: Uwe betrückte Alfred, weil er bei diesem wichtigen Problem einfach mehr Anstrengung erwartete.

GenPlusKonPlus: Uwe betrückte Jessica, weil er bei diesem wichtigen Problem nicht genügend Anstrengung investierte.

GenPlusKonMin: Anita betrückte Alfred, weil er bei diesem wichtigen Problem einfach mehr Anstrengung erwartete.

SE11 beunruhigte

GenMinKonPlus: Elmar beunruhigte Luca, weil er besonders bei langen Autofahrten häufig die Konzentration verlor.

GenMinKonMin: Elmar beunruhigte Luca, weil er besonders bei langen Autofahrten ein großes Sicherheitsbedürfnis hatte.

GenPlusKonPlus: Elmar beunruhigte Julia, weil er besonders bei langen Autofahrten häufig die Konzentration verlor.

GenPlusKonMin: Lea beunruhigte Luca, weil er besonders bei langen Autofahrten ein großes Sicherheitsbedürfnis hatte.

SE12 blamierte

GenMinKonPlus: Jürgen blamierte Lorenz, weil er mitten in der Öffentlichkeit sehr intime Details ausplauderte.

GenMinKonMin: Jürgen blamierte Lorenz, weil er mitten in der Öffentlichkeit den Anschuldigungen ausgeliefert war.

GenPlusKonPlus: Jürgen blamierte Allmut, weil er mitten in der Öffentlichkeit sehr intime Details ausplauderte.

GenPlusKonMin: Valeska blamierte Lorenz, weil er mitten in der Öffentlichkeit den Anschuldigungen ausgeliefert war.

SE13 deprimierte

GenMinKonPlus: Astrid deprimierte Nathalie, weil sie nach all den Nachhilfestunden eine sechs geschrieben hatte.

GenMinKonMin: Astrid deprimierte Nathalie, weil sie nach all den Nachhilfestunden bessere Noten erhofft hatte.

GenPlusKonPlus: Astrid deprimierte Olaf, weil sie nach all den Nachhilfestunden eine sechs geschrieben hatte.

GenPlusKonMin: Max deprimierte Nathalie, weil sie nach all den Nachhilfestunden bessere Noten erhofft hatte.

SE14 empörte

GenMinKonPlus: Anita empörte Jessica, weil sie wegen des geplatzten Geschäftes dauernd ungemein vorwurfsvoll war.

GenMinKonMin: Anita empörte Jessica, weil sie wegen des geplatzten Geschäftes sowieso schon gereizt war.

GenPlusKonPlus: Anita empörte Alfred, weil sie wegen des geplatzten Geschäftes dauernd ungemein vorwurfsvoll war.

GenPlusKonMin: Uwe empörte Jessica, weil sie wegen des geplatzten Geschäftes sowieso schon gereizt war.

SE15 entmutigte

GenMinKonPlus: Lea entmutigte Julia, weil sie bei der letzten Hausarbeit nur noch Kritik äußerte.

GenMinKonMin: Lea entmutigte Julia, weil sie bei der letzten Hausarbeit mehr Hilfe erhofft hatte.

GenPlusKonPlus: Lea entmutigte Luca, weil sie bei der letzten Hausarbeit nur noch Kritik äußerte.

GenPlusKonMin: Elmar entmutigte Julia, weil sie bei der letzten Hausarbeit mehr Hilfe erhofft hatte.

SE16 enttäuschte

GenMinKonPlus: Valeska enttäuschte Allmut, weil sie in fast allen Lebensbereichen hinter den Erwartungen zurückblieb.

GenMinKonMin: Valeska enttäuschte Allmut, weil sie in fast allen Lebensbereichen sehr hohe Erwartungen hatte.

GenPlusKonPlus: Valeska enttäuschte Lorenz, weil sie in fast allen Lebensbereichen hinter den Erwartungen zurückblieb.

GenPlusKonMin: Jürgen enttäuschte Allmut, weil sie in fast allen Lebensbereichen sehr hohe Erwartungen hatte.

SE17 entzückte

GenMinKonPlus: Bernd entzückte Erwin, weil er nach der letzten Prüfung ein schönes Geschenk mitbrachte.

GenMinKonMin: Bernd entzückte Erwin, weil er nach der letzten Prüfung kein Geschenk erwartet hätte.

GenPlusKonPlus: Bernd entzückte Nicole, weil er nach der letzten Prüfung ein schönes Geschenk mitbrachte.

GenPlusKonMin: Franziska entzückte Erwin, weil er nach der letzten Prüfung kein Geschenk erwartet hätte.

SE18 erfreute

GenMinKonPlus: Konrad erfreute Lars, weil er gerade bei schwierigen Arbeiten konzentriert und effizient war.

GenMinKonMin: Konrad erfreute Lars, weil er gerade bei schwierigen Arbeiten Konzentration und Effizienz schätzte.

GenPlusKonPlus: Konrad erfreute Angelika, weil er gerade bei schwierigen Arbeiten konzentriert und effizient war.

GenPlusKonMin: Anke erfreute Lars, weil er gerade bei schwierigen Arbeiten Konzentration und Effizienz schätzte.

SE19 erheiterte

GenMinKonPlus: Manfred erheiterte Lukas, weil er nach der letzten Klausur albern und übermütig war.

GenMinKonMin: Manfred erheiterte Lukas, weil er nach der letzten Klausur traurig und niedergeschlagen war.

GenPlusKonPlus: Manfred erheiterte Anja, weil er nach der letzten Klausur albern und übermütig war.

GenPlusKonMin: Anne erheiterte Lukas, weil er nach der letzten Klausur traurig und niedergeschlagen war.

SE20 erschreckte

GenMinKonPlus: Kurt erschreckte Gerhard, weil er nach dem schweren Unfall oft sehr schlecht aussah.

GenMinKonMin: Kurt erschreckte Gerhard, weil er nach dem schweren Unfall zeitweise sehr ängstlich war.

GenPlusKonPlus: Kurt erschreckte Charlotte, weil er nach dem schweren Unfall oft sehr schlecht aussah.

GenPlusKonMin: Annette erschreckte Gerhard, weil er nach dem schweren Unfall zeitweise sehr ängstlich war.

SE21 erstaunte

GenMinKonPlus: Walter erstaunte Rolf, weil er mitten in der Nacht ohne besonderen Grund anrief.

GenMinKonMin: Walter erstaunte Rolf, weil er mitten in der Nacht eigentlich keinen Anruf erwartete.

GenPlusKonPlus: Walter erstaunte Isabell, weil er mitten in der Nacht ohne besonderen Grund anrief.

GenPlusKonMin: Maria erstaunte Rolf, weil er mitten in der Nacht eigentlich keinen Anruf erwartete.

SE22 erzürnte

GenMinKonPlus: Franziska erzürnte Nicole, weil sie bei allen finanziellen Angelegenheiten oft unerträglich nachlässig war.

GenMinKonMin: Franziska erzürnte Nicole, weil sie bei allen finanziellen Angelegenheiten Nachlässigkeiten nicht Leiden konnte.

GenPlusKonPlus: Franziska erzürnte Erwin, weil sie bei allen finanziellen Angelegenheiten oft unerträglich nachlässig war.

GenPlusKonMin: Bernd erzürnte Nicole, weil sie bei allen finanziellen Angelegenheiten Nachlässigkeiten nicht Leiden konnte.

SE23 faszinierte

GenMinKonPlus: Anne faszinierte Anja, weil sie durch ein selbstbewusstes Auftreten aus der Masse hervorstach.

GenMinKonMin: Anne faszinierte Anja, weil sie durch ein selbstbewusstes Auftreten leicht zu beeindrucken war.

GenPlusKonPlus: Anne faszinierte Lukas, weil sie durch ein selbstbewusstes Auftreten aus der Masse hervorstach.

GenPlusKonMin: Manfred faszinierte Anja, weil sie durch ein selbstbewusstes Auftreten leicht zu beeindrucken war.

SE24 gefiel

GenMinKonPlus: Annette gefiel Charlotte, weil sie für die freie Stelle sehr gut geeignet war.

GenMinKonMin: Annette gefiel Charlotte, weil sie für die freie Stelle genau diesen Typ suchte.

GenPlusKonPlus: Annette gefiel Gerhard, weil sie für die freie Stelle sehr gut geeignet war.

GenPlusKonMin: Kurt gefiel Charlotte, weil sie für die freie Stelle genau diesen Typ suchte.

SE25 imponierte

GenMinKonPlus: Maria imponierte Isabell, weil sie bei den meisten Diskussionen schlagfertig und überzeugend war.

GenMinKonMin: Maria imponierte Isabell, weil sie bei den meisten Diskussionen leicht zu beeindrucken war.

GenPlusKonPlus: Maria imponierte Rolf, weil sie bei den meisten Diskussionen schlagfertig und überzeugend war.

GenPlusKonMin: Walter imponierte Isabell, weil sie bei den meisten Diskussionen leicht zu beeindrucken war.

SE26 interessierte

GenMinKonPlus: Fridolin interessierte Christoph, weil er in der einführenden Blockveranstaltung nur schlaue Fragen stellte.

GenMinKonMin: Fridolin interessierte Christoph, weil er in der einführenden Blockveranstaltung nach neuen Freunden suchte.

GenPlusKonPlus: Fridolin interessierte Miriam, weil er in der einführenden Blockveranstaltung nur schlaue Fragen stellte.

GenPlusKonMin: Iris interessierte Christoph, weil er in der einführenden Blockveranstaltung nach neuen Freunden suchte.

SE27 irritierte

GenMinKonPlus: Bruno irritierte Sven, weil er bei den abschließenden Erläuterungen wieder zu diskutieren anfang.

GenMinKonMin: Bruno irritierte Sven, weil er bei den abschließenden Erläuterungen die Argumentation nicht verstand.

GenPlusKonPlus: Bruno irritierte Maren, weil er bei den abschließenden Erläuterungen wieder zu diskutieren anfang.

GenPlusKonMin: Lisa irritierte Sven, weil er bei den abschließenden Erläuterungen die Argumentation nicht verstand.

SE28 langweilte

GenMinKonPlus: Ingo langweilte Jonas, weil er von dem diesjährigen Spanienurlaub viel zu ausführlich erzählte.

GenMinKonMin: Ingo langweilte Jonas, weil er von dem diesjährigen Spanienurlaub bereits alle Geschichten kannte.

GenPlusKonPlus: Ingo langweilte Chloe, weil er von dem diesjährigen Spanienurlaub viel zu ausführlich erzählte.

GenPlusKonMin: Katrin langweilte Jonas, weil er von dem diesjährigen Spanienurlaub bereits alle Geschichten kannte.

SE29 nervte

GenMinKonPlus: Dominik nervte Ralf, weil er nach der langen Diskussion keine Ruhe geben wollte.

GenMinKonMin: Dominik nervte Ralf, weil er nach der langen Diskussion das Thema leid war.

GenPlusKonPlus: Dominik nervte Daniella, weil er nach der langen Diskussion keine Ruhe geben wollte.

GenPlusKonMin: Ulla nervte Ralf, weil er nach der langen Diskussion das Thema leid war.

SE30 schockierte

GenMinKonPlus: Iris schockierte Miriam, weil sie vor der endgültigen Gerichtsverhandlung keine Vorstrafen erwähnt hatte.

GenMinKonMin: Iris schockierte Miriam, weil sie vor der endgültigen Gerichtsverhandlung keine Vorstrafen vermutet hatte.

GenPlusKonPlus: Iris schockierte Christoph, weil sie vor der endgültigen Gerichtsverhandlung keine Vorstrafen erwähnt hatte.

GenPlusKonMin: Fridolin schockierte Miriam, weil sie vor der endgültigen Gerichtsverhandlung keine Vorstrafen vermutet hatte.

SE31 störte

GenMinKonPlus: Lisa störte Maren, weil sie in der ersten Urlaubsnacht ohne Unterbrechung quatschen wollte.

GenMinKonMin: Lisa störte Maren, weil sie in der ersten Urlaubsnacht gern unbehelligt schlafen wollte.

GenPlusKonPlus: Lisa störte Sven, weil sie in der ersten Urlaubsnacht ohne Unterbrechung quatschen wollte.

GenPlusKonMin: Bruno störte Maren, weil sie in der ersten Urlaubsnacht gern unbehelligt schlafen wollte.

SE32 überraschte

GenMinKonPlus: Katrin überraschte Cloe, weil sie mit einer solchen Taktik entgegen aller Erwartung handelte.

GenMinKonMin: Katrin überraschte Cloe, weil sie mit einer solchen Taktik keinen Erfolg erwartet hätte.

GenPlusKonPlus: Katrin überraschte Jonas, weil sie mit einer solchen Taktik entgegen aller Erwartung handelte.

GenPlusKonMin: Ingo überraschte Cloe, weil sie mit einer solchen Taktik keinen Erfolg erwartet hätte.

SE 33 überzeugte

GenMinKonPlus: Ulla überzeugte Daniela, weil sie bei dieser wichtigen Frage die besseren Argumente hatte.

GenMinKonMin: Ulla überzeugte Daniela, weil sie bei dieser wichtigen Frage die Argumentation bestechend fand.

GenPlusKonPlus: Ulla überzeugte Ralf, weil sie bei dieser wichtigen Frage die besseren Argumente hatte.

GenPlusKonMin: Dominik überzeugte Daniela, weil sie bei dieser wichtigen Frage die Argumentation bestechend fand.

SE 34 umgarnte

GenMinKonPlus: Herrmann umgarnte Thomas, weil er in der fremden Stadt neue geschäftliche Kontakte brauchte.

GenMinKonMin: Herrmann umgarnte Thomas, weil er in der fremden Stadt viele wichtige Kontakte hatte.

GenPlusKonPlus: Herrmann umgarnte Johanna, weil er in der fremden Stadt neue geschäftliche Kontakte brauchte.

GenPlusKonMin: Nadja umgarnte Thomas, weil er in der fremden Stadt viele wichtige Kontakte hatte.

SE 35 verärgerte

GenMinKonPlus: Rainer verärgerte Franz, weil er bei den wöchentlichen Teamsitzungen sehr schnell ausfallend wurde.

GenMinKonMin: Rainer verärgerte Franz, weil er bei den wöchentlichen Teamsitzungen das Thema leid war.

GenPlusKonPlus: Rainer verärgerte Tatjana, weil er bei den wöchentlichen Teamsitzungen sehr schnell ausfallend wurde.

GenPlusKonMin: Barbara verärgerte Franz, weil er bei den wöchentlichen Teamsitzungen das Thema leid war.

.SE 36 verblüfft

GenMinKonPlus: Hubert verblüffte Oskar, weil er seit dem letzten Mal riesige Fortschritte gemacht hatte.

GenMinKonMin: Hubert verblüffte Oskar, weil er seit dem letzten Mal keine Fortschritte erwartet hatte.

GenPlusKonPlus: Hubert verblüffte Inga, weil er seit dem letzten Mal riesige Fortschritte gemacht hatte.

GenPlusKonMin: Edith verblüffte Oskar, weil er seit dem letzten Mal keine Fortschritte erwartet hatte.

SE 37 verschreckte

GenMinKonPlus: Nadja verschreckte Johanna, weil sie wegen des überholenden LKWs plötzlich ohne Vorwarnung schrie.

GenMinKonMin: Nadja verschreckte Johanna, weil sie wegen des überholenden LKWs gerade abgelenkt gewesen war.

GenPlusKonPlus: Nadja verschreckte Torsten, weil sie wegen des überholenden LKWs plötzlich ohne Vorwarnung schrie.

GenPlusKonMin: Herrmann verschreckte Johanna, weil sie wegen des überholenden LKWs gerade abgelenkt gewesen war.

SE 38 verunsicherte

GenMinKonPlus: Anke verunsicherte Angelika, weil sie besonders in neuen Situationen immer völlig unkalkulierbar handelte.

GenMinKonMin: Anke verunsicherte Angelika, weil sie besonders in neuen Situationen auf Kalkulierbarkeit Wert legte.

GenPlusKonPlus: Anke verunsicherte Lars, weil sie besonders in neuen Situationen immer völlig unkalkulierbar handelte.

GenPlusKonMin: Konrad verunsicherte Angelika, weil sie besonders in neuen Situationen auf Kalkulierbarkeit Wert legte.

SE 39 verwirrte

GenMinKonPlus: Barbara verwirrte Tatjana, weil sie wegen der vielen Aussprachefehler schlecht zu verstehen war.

GenMinKonMin: Barbara verwirrte Tatjana, weil sie wegen der vielen Aussprachefehler kaum ein Wort verstand.

GenPlusKonPlus: Barbara verwirrte Franz, weil sie wegen der vielen Aussprachefehler schlecht zu verstehen war.

GenPlusKonMin: Rainer verwirrte Tatjana, weil sie wegen der vielen Aussprachefehler kaum ein Wort verstand.

SE 40 verwunderte

GenMinKonPlus: Edith verwunderte Inga, weil sie auf der gestrigen Sitzung ausgesprochen souverän gewirkt hatte.

GenMinKonMin: Edith verwunderte Inga, weil sie auf der gestrigen Sitzung ganz andere Erwartungen hatte.

GenPlusKonPlus: Edith verwunderte Oskar, weil sie auf der gestrigen Sitzung ausgesprochen souverän gewirkt hatte.

GenPlusKonMin: Hubert verwunderte Inga, weil sie auf der gestrigen Sitzung ganz andere Erwartungen hatte.

E-S Verben *Subjekt: Experiencer – Objekt: Stimulus, Kausalität auf Stimulus*

ES 1 achtete

GenMinKonMin: Arnold achtete Valentin, weil er für die jahrelange Hilfe sehr tiefe Dankbarkeit empfand.

GenMinKonPlus: Arnold achtete Valentin, weil er für die jahrelange Hilfe keine Gegenleistung annehmen wollte.

GenPlusKonMin: Arnold achtete Marlene, weil er für die jahrelange Hilfe sehr tiefe Dankbarkeit empfand.

GenPlusKonPlus: Christa achtete Valentin, weil er für die jahrelange Hilfe keine Gegenleistung annehmen wollte.

ES 2 akzeptierte

GenMinKonMin: Gustav akzeptierte Ulf, weil er während der gemeinsamen Arbeit keine andere Wahl hatte.

GenMinKonPlus: Gustav akzeptierte Ulf, weil er während der gemeinsamen Arbeit wirklich sehr nett war.

GenPlusKonMin: Gustav akzeptierte Linda, weil er während der gemeinsamen Arbeit keine andere Wahl hatte.

GenPlusKonPlus: Elfriede akzeptierte Ulf, weil er während der gemeinsamen Arbeit wirklich sehr nett war.

ES 3 beachtete

GenMinKonMin: Benjamin beachtete Ernst, weil er trotz der vielen Anwesenden jeden Gast kennenlernen wollte.

GenMinKonPlus: Benjamin beachtete Ernst, weil er trotz der vielen Anwesenden nicht zu übersehen war.

GenPlusKonMin: Benjamin beachtete Nadine, weil er trotz der vielen Anwesenden jeden Gast kennenlernen wollte.

GenPlusKonPlus: Ada beachtete Ernst, weil er trotz der vielen Anwesenden nicht zu übersehen war.

ES 4 beargwöhnte

GenMinKonMin: Klaus beargwöhnte Heiner, weil er nach den neuesten Informationen außerordentlich misstrauisch geworden war.

GenMinKonPlus: Klaus beargwöhnte Heiner, weil er nach den neuesten Informationen schon häufiger gestohlen hatte.

GenPlusKonMin: Klaus beargwöhnte Lola, weil er nach den neuesten Informationen außerordentlich misstrauisch geworden war.

GenPlusKonPlus: Pamela beargwöhnte Heiner, weil er nach den neuesten Informationen schon häufiger gestohlen hatte.

ES 5 bedauerte

GenMinKonMin: Christa bedauerte Marlene, weil sie schon seit frühen Kindertagen allen gegenüber mitfühlend war.

GenMinKonPlus: Christa bedauerte Marlene, weil sie schon seit frühen Kindertagen schwer zu kämpfen hatte.

GenPlusKonMin: Christa bedauerte Valentin, weil sie schon seit frühen Kindertagen allen gegenüber mitfühlend war.

GenPlusKonPlus: Arnold bedauerte Marlene, weil sie schon seit frühen Kindertagen schwer zu kämpfen hatte.

ES 6 bemerkte

GenMinKonMin: Elfriede bemerkte Linda, weil sie inmitten der anonymen Masse nach bekannten Gesichtern suchte.

GenMinKonPlus: Elfriede bemerkte Linda, weil sie inmitten der anonymen Masse deutlich zu erkennen war.

GenPlusKonMin: Elfriede bemerkte Ulf, weil sie inmitten der anonymen Masse nach bekannten Gesichtern suchte.

GenPlusKonPlus: Gustav bemerkte Linda, weil sie inmitten der anonymen Masse deutlich zu erkennen war.

ES 7 bemitleidete

GenMinKonMin: Ada bemitleidete Nadine, weil sie seit dem psychiatrischen Praktikum einen starken Helferkomplex hatte.

GenMinKonPlus: Ada bemitleidete Nadine, weil sie seit dem psychiatrischen Praktikum einen starken Waschzwang hatte.

GenPlusKonMin: Ada bemitleidete Ernst, weil sie seit dem psychiatrischen Praktikum einen starken Helferkomplex hatte.

GenPlusKonPlus: Benjamin bemitleidete Nadine, weil sie seit dem psychiatrischen Praktikum einen starken Waschzwang hatte.

ES 8 beneidet

GenMinKonMin: Pamela beneidete Lola, weil sie trotz gleicher beruflicher Tätigkeit viel weniger Geld verdiente.

GenMinKonPlus: Pamela beneidete Lola, weil sie trotz gleicher beruflicher Tätigkeit viel mehr Geld bekam.

GenPlusKonMin: Pamela beneidete Heiner, weil sie trotz gleicher beruflicher Tätigkeit viel weniger Geld verdiente.

GenPlusKonPlus: Klaus beneidete Lola, weil sie trotz gleicher beruflicher Tätigkeit viel mehr Geld bekam.

ES 9 betreute

GenMinKonMin: Heinrich betreute Bert, weil er seit dem tragischen Unfall die einzige Bezugsperson war.

GenMinKonPlus: Heinrich betreute Bert, weil er seit dem tragischen Unfall kaum mehr laufen konnte.

GenPlusKonMin: Heinrich betreute Ingrid, weil er seit dem tragischen Unfall die einzige Bezugsperson war.

GenPlusKonPlus: Vera betreute Bert, weil er seit dem tragischen Unfall kaum mehr laufen konnte.

ES 10 bevorzugte

GenMinKonMin: Jakob bevorzugte Daniel, weil er im entscheidenden letzten Wahlgang alle Alternativen schlimmer fand.

GenMinKonPlus: Jakob bevorzugte Daniel, weil er im entscheidenden letzten Wahlgang die beste Alternative war.

GenPlusKonMin: Jakob bevorzugte Carmen, weil er im entscheidenden letzten Wahlgang alle Alternativen schlimmer fand.

GenPlusKonPlus: Sarah bevorzugte Daniel, weil er im entscheidenden letzten Wahlgang die beste Alternative war.

ES 11 bewunderte

GenMinKonMin: Joseph bewunderte Jan, weil er unter den gegebenen Umständen das Geleistete begeisternd fand.

GenMinKonPlus: Joseph bewunderte Jan, weil er unter den gegebenen Umständen großartigen Weitblick bewiesen hatte.

GenPlusKonMin: Joseph bewunderte Helene, weil er unter den gegebenen Umständen das Geleistete begeisternd fand.

GenPlusKonPlus: Leonie bewunderte Jan, weil er unter den gegebenen Umständen großartigen Weitblick bewiesen hatte.

ES 12 bezichtigte

GenMinKonMin: Hugo bezichtigte Armin, weil er gerade bei entscheidenden Problemen Unzuverlässigkeit nicht ausstehen konnte.

GenMinKonPlus: Hugo bezichtigte Armin, weil er gerade bei entscheidenden Problemen immer wieder wortbrüchig wurde.

GenPlusKonMin: Hugo bezichtigte Lena, weil er gerade bei entscheidenden Problemen Unzuverlässigkeit nicht ausstehen konnte.

GenPlusKonPlus: Irene bezichtigte Armin, weil er gerade bei entscheidenden Problemen immer wieder wortbrüchig wurde.

ES 13 ehrte

GenMinKonMin:

Vera ehrte Ingrid, weil sie für dieses außergewöhnliche Leben Respekt und Anerkennung empfand.

GenMinKonPlus: Vera ehrte Ingrid, weil sie für dieses außergewöhnliche Leben höchsten Respekt verdient hatte.

GenPlusKonMin: Vera ehrte Bert, weil sie für dieses außergewöhnliche Leben Respekt und Anerkennung empfand.

GenPlusKonPlus: Heinrich ehrte Ingrid, weil sie für dieses außergewöhnliche Leben höchsten Respekt verdient hatte.

ES 14 ehrte

GenMinKonMin: Sarah fürchtete Carmen, weil sie schon vor dem Gespräch schlimme Erzählungen gehört hatte.

GenMinKonPlus: Sarah fürchtete Carmen, weil sie schon vor dem Gespräch gerichtliche Schritte angedroht hatte.

GenPlusKonMin: Sarah fürchtete Daniel, weil sie schon vor dem Gespräch schlimme Erzählungen gehört hatte.

GenPlusKonPlus: Jakob fürchtete Carmen, weil sie schon vor dem Gespräch gerichtliche Schritte angedroht hatte.

ES 15 glorifizierte

GenMinKonMin: Leonie glorifizierte Helene, weil sie schon in der Schule immer ein Vorbild brauchte.

GenMinKonPlus: Leonie glorifizierte Helene, weil sie schon in der Schule allen ein Vorbild war.

GenPlusKonMin: Leonie glorifizierte Jan, weil sie schon in der Schule immer ein Vorbild brauchte.

GenPlusKonPlus: Joseph glorifizierte Helene, weil sie schon in der Schule allen ein Vorbild war.

ES 16 grollte

GenMinKonMin: Irene grollte Lena, weil sie nach der gescheiterten Verhandlung allen irgendwelche Vorwürfe machte.

GenMinKonPlus: Irene grollte Lena, weil sie nach der gescheiterten Verhandlung überhaupt keine Anteilnahme zeigte.

GenPlusKonMin: Irene grollte Armin, weil sie nach der gescheiterten Verhandlung allen irgendwelche Vorwürfe machte.

GenPlusKonPlus: Hugo grollte Lena, weil sie nach der gescheiterten Verhandlung überhaupt keine Anteilnahme zeigte.

ES 17 hasste

GenMinKonMin: Mike hasste Roland, weil er besonders gegenüber wehrlosen Untergebenen offenen Sadismus abstoßend fand.

GenMinKonPlus: Mike hasste Roland, weil er besonders gegenüber wehrlosen Untergebenen ohne Grund aggressiv wurde.

GenPlusKonMin: Mike hasste Antonia, weil er besonders gegenüber wehrlosen Untergebenen offenen Sadismus abstoßend fand.

GenPlusKonPlus: Lilli hasste Roland, weil er besonders gegenüber wehrlosen Untergebenen ohne Grund aggressiv wurde.

ES 18 huldigte

GenMinKonMin: Viktor huldigte Robert, weil er durch Barmherzigkeit und Güte menschlich tief berührt war.

GenMinKonPlus: Viktor huldigte Robert, weil er durch Barmherzigkeit und Güte immer opfernd geholfen hatte.

GenPlusKonMin: Viktor huldigte Sonja, weil er durch Barmherzigkeit und Güte menschlich tief berührt war.

GenPlusKonPlus: Lydia huldigte Robert, weil er durch Barmherzigkeit und Güte immer opfernd geholfen hatte.

ES 19 idealisierte

GenMinKonMin: Dietrich idealisierte Justus, weil er schon nach wenigen Augenblicken die große Begabung erkannte.

GenMinKonPlus: Dietrich idealisierte Justus, weil er schon nach wenigen Augenblicken eine überwältigende Anziehungskraft ausstrahlte.

GenPlusKonMin: Dietrich idealisierte Sibylle, weil er schon nach wenigen Augenblicken die große Begabung erkannte.

GenPlusKonPlus: Eva idealisierte Justus, weil er schon nach wenigen Augenblicken eine überwältigende Anziehungskraft ausstrahlte.

ES 20 ignorierte

GenMinKonMin: Florian ignorierte Igor, weil er seit der letzten Auseinandersetzung jeden Kontakt vermeiden wollte.

GenMinKonPlus: Florian ignorierte Igor, weil er seit der letzten Auseinandersetzung immer unverschämter geworden war.

GenPlusKonMin: Florian ignorierte Judith, weil er seit der letzten Auseinandersetzung jeden Kontakt vermeiden wollte.

GenPlusKonPlus: Hildegard ignorierte Igor, weil er seit der letzten Auseinandersetzung immer unverschämter geworden war.

ES 21 mochte

GenMinKonMin: Lilli mochte Antonia, weil sie unabhängig von unterschiedlichen Interessen immer ausgesprochen harmoniesüchtig war.

GenMinKonPlus: Lilli mochte Antonia, weil sie unabhängig von unterschiedlichen Interessen niemals Schwächen ausgenutzt hatte.

GenPlusKonMin: Lilli mochte Roland, weil sie unabhängig von unterschiedlichen Interessen immer ausgesprochen harmoniesüchtig war.

GenPlusKonPlus: Mike mochte Antonia, weil sie unabhängig von unterschiedlichen Interessen niemals Schwächen ausgenutzt hatte.

ES 22 missachtete

GenMinKonMin: Lydia missachtete Sonja, weil sie durch die dauernden Querelen die Nase voll hatte.

GenMinKonPlus: Lydia missachtete Sonja, weil sie durch die dauernden Querelen immer unleidlicher geworden war.

GenPlusKonMin: Lydia missachtete Robert, weil sie durch die dauernden Querelen die Nase voll hatte.

GenPlusKonPlus: Viktor missachtete Sonja, weil sie durch die dauernden Querelen immer unleidlicher geworden war.

ES 23 missfiel

GenMinKonMin: Eva missfiel Sibylle, weil sie gerade unter guten Freunden immer maßlos angeben musste.

GenMinKonPlus: Eva missfiel Sibylle, weil sie gerade unter guten Freunden ganz besonders kritisch war.

GenPlusKonMin: Eva missfiel Justus, weil sie gerade unter guten Freunden immer maßlos angeben musste.

GenPlusKonPlus: Dietrich missfiel Sibylle, weil sie gerade unter guten Freunden ganz besonders kritisch war.

ES 24 misstraute

GenMinKonMin: Hildegard misstraute Judith, weil sie schon beim letzten Zusammentreffen schlechte Erfahrungen gemacht hatte.

GenMinKonPlus: Hildegard misstraute Judith, weil sie schon beim letzten Zusammentreffen ausgesprochen unfair gewesen war.

GenPlusKonMin: Hildegard misstraute Igor, weil sie schon beim letzten Zusammentreffen schlechte Erfahrungen gemacht hatte.

GenPlusKonPlus: Florian misstraute Judith, weil sie schon beim letzten Zusammentreffen ausgesprochen unfair gewesen war.

ES 25 respektierte

GenMinKonMin: Theo respektierte Patrick, weil er auch in der Konfliktsituation die Beziehung aufrechterhalten wollte.

GenMinKonPlus: Theo respektierte Patrick, weil er auch in der Konfliktsituation Schwächen anderer nicht ausnutzte.

.GenPlusKonMin: Theo respektierte Amelie, weil er auch in der Konfliktsituation die Beziehung aufrechterhalten wollte.

GenPlusKonPlus: Olga respektierte Patrick, weil er auch in der Konfliktsituation Schwächen anderer nicht ausnutzte.

ES 26 schätzte

GenMinKonMin: Otto schätzte David, weil er bei allen wichtigen Problemen einen guten Rat wollte.

GenMinKonPlus: Otto schätzte David, weil er bei allen wichtigen Problemen immer guten Rat wusste.

GenPlusKonMin: Otto schätzte Minna, weil er bei allen wichtigen Problemen einen guten Rat wollte.

GenPlusKonPlus: Heike schätzte David, weil er bei allen wichtigen Problemen immer guten Rat wusste.

ES 27 scheute

GenMinKonMin: Maximilian scheute Fritz, weil er nach dem unschönen Vorfall ein schlechtes Gewissen hatte.

GenMinKonPlus: Maximilian scheute Fritz, weil er nach dem unschönen Vorfall nur noch Vorwürfe machte.

GenPlusKonMin: Maximilian scheute Alice, weil er nach dem unschönen Vorfall ein schlechtes Gewissen hatte.

GenPlusKonPlus: Stefanie scheute Fritz, weil er nach dem unschönen Vorfall nur noch Vorwürfe machte.

ES 28 schmähte

GenMinKonMin: Tom schmähte Wolfgang, weil er von der reißerischen Boulevardzeitung gut dafür bezahlt wurde.

GenMinKonPlus: Tom schmähte Wolfgang, weil er von der reißerischen Boulevardzeitung der Korruption bezichtigt wurde.

GenPlusKonMin: Tom schmähte Käthe, weil er von der reißerischen Boulevardzeitung gut dafür bezahlt wurde.

GenPlusKonPlus: Heidi schmähte Wolfgang, weil er von der reißerischen Boulevardzeitung der Korruption bezichtigt wurde.

ES 29 überschätzte

GenMinKonMin: Olga überschätzte Amelie, weil sie im Bereich persönlicher Beziehungen keine echte Kritikfähigkeit besaß.

GenMinKonPlus: Olga überschätzte Amelie, weil sie im Bereich persönlicher Beziehungen Kommunikationsprobleme geschickt überspielen konnte.

GenPlusKonMin: Olga überschätzte Patrick, weil sie im Bereich persönlicher Beziehungen keine echte Kritikfähigkeit besaß.

GenPlusKonPlus: Theo überschätzte Amelie, weil sie im Bereich persönlicher Beziehungen Kommunikationsprobleme geschickt überspielen konnte.

ES 30 unterschätzte

GenMinKonMin: Heike unterschätzte Minna, weil sie beim hastig durchgeführten Bewerbungsgespräch das Arbeitszeugnis ignoriert hatte.

GenMinKonPlus: Heike unterschätzte Minna, weil sie beim hastig durchgeführten Bewerbungsgespräch sehr bescheiden gewesen war.

GenPlusKonMin: Heike unterschätzte David, weil sie beim hastig durchgeführten Bewerbungsgespräch das Arbeitszeugnis ignoriert hatte.

GenPlusKonPlus: Otto unterschätzte Minna, weil sie beim hastig durchgeführten Bewerbungsgespräch sehr bescheiden gewesen war.

ES 31 verabscheute

GenMinKonMin: Stefanie verabscheute Alice, weil sie gerade bei vertraulichen Absprachen nicht gerne hintergangen wurde.

GenMinKonPlus: Stefanie verabscheute Alice, weil sie gerade bei vertraulichen Absprachen nie den Mund halten konnte.

GenPlusKonMin: Stefanie verabscheute Fritz, weil sie gerade bei vertraulichen Absprachen nicht gerne hintergangen wurde.

GenPlusKonPlus: Maximilian verabscheute Alice, weil sie gerade bei vertraulichen Absprachen nie den Mund halten konnte.

ES 32 verachtete

GenMinKonMin: Heidi verachtete Käthe, weil sie bei dem letzten Gespräch diese Heuchelei erbärmlich fand.

GenMinKonPlus: Heidi verachtete Käthe, weil sie bei dem letzten Gespräch eigentlich nur gelogen hatte.

GenPlusKonMin: Heidi verachtete Wolfgang, weil sie bei dem letzten Gespräch diese Heuchelei erbärmlich fand.

GenPlusKonPlus: Tom verachtete Käthe, weil sie bei dem letzten Gespräch eigentlich nur gelogen hatte.

ES 33 verehrte

GenMinKonMin: Markus verehrte Martin, weil er trotz vieler inhaltlicher Schwächen den Schreibstil überragend fand.

GenMinKonPlus: Markus verehrte Martin, weil er trotz vieler inhaltlicher Schwächen einen überragenden Schreibstil hatte.

GenPlusKonMin: Markus verehrte Jutta, weil er trotz vieler inhaltlicher Schwächen den Schreibstil überragend fand.

GenPlusKonPlus: Katharina verehrte Martin, weil er trotz vieler inhaltlicher Schwächen einen überragenden Schreibstil hatte.

ES 34 vergötterte

GenMinKonMin: Joachim vergötterte Martin, weil er bei der Einschätzung anderer zu starken Übertreibungen neigte.

GenMinKonPlus: Joachim vergötterte Martin, weil er bei der Einschätzung anderer so unglaublich gerecht war.

GenPlusKonMin: Joachim vergötterte Sandra, weil er bei der Einschätzung anderer zu starken Übertreibungen neigte.

GenPlusKonPlus: Vanessa vergötterte Martin, weil er bei der Einschätzung anderer so unglaublich gerecht war.

ES 35 verherrlichte

GenMinKonMin: Jens verherrlichte Thomas, weil er für die lebensgefährliche Rettungsaktion ewig dankbar sein würde.

GenMinKonPlus: Jens verherrlichte Thomas, weil er für die lebensgefährliche Rettungsaktion ausgesprochen viel riskiert hatte.

GenPlusKonMin: Jens verherrlichte Irmgard, weil er für die lebensgefährliche Rettungsaktion ewig dankbar sein würde.

GenPlusKonPlus: Theresa verherrlichte Thomas, weil er für die lebensgefährliche Rettungsaktion ausgesprochen viel riskiert hatte.

ES 36 vermisste

GenMinKonMin: Edmund vermisste Jörg, weil er in der jahrelangen Wohngemeinschaft das Alleinsein verlernt hatte.

GenMinKonPlus: Edmund vermisste Jörg, weil er in der jahrelangen Wohngemeinschaft immer hilfsbereit gewesen war.

GenPlusKonMin: Edmund vermisste Sophie, weil er in der jahrelangen Wohngemeinschaft das Alleinsein verlernt hatte.

GenPlusKonPlus: Veronika vermisste Jörg, weil er in der jahrelangen Wohngemeinschaft immer hilfsbereit gewesen war.

ES 37 verschmähte

GenMinKonMin: Katharina verschmähte Jutta, weil sie während der letzten Ferien bessere Freunde gefunden hatte.

GenMinKonPlus: Katharina verschmähte Jutta, weil sie während der letzten Ferien so gemein gewesen war.

GenPlusKonMin: Katharina verschmähte Martin, weil sie während der letzten Ferien bessere Freunde gefunden hatte.

GenPlusKonPlus: Markus verschmähte Jutta, weil sie während der letzten Ferien so gemein gewesen war.

ES 38 verstand

GenMinKonMin: Vanessa verstand Sandra, weil sie gerade bei komplexen Problemstellungen die Perspektive anderer übernahm.

GenMinKonPlus: Vanessa verstand Sandra, weil sie gerade bei komplexen Problemstellungen sehr gut erklären konnte.

GenPlusKonMin: Vanessa verstand Paul, weil sie gerade bei komplexen Problemstellungen die Perspektive anderer übernahm.

GenPlusKonPlus: Joachim verstand Sandra, weil sie gerade bei komplexen Problemstellungen sehr gut erklären konnte.

ES 39 verteufelte

GenMinKonMin: Theresa verteufelte Irmgard, weil sie bei diesem wichtigen Unternehmen keine Konkurrenz haben wollte.

GenMinKonPlus: Theresa verteufelte Irmgard, weil sie bei diesem wichtigen Unternehmen nichts als gestört hätte.

GenPlusKonMin: Theresa verteufelte Thomas, weil sie bei diesem wichtigen Unternehmen keine Konkurrenz haben wollte.

GenPlusKonPlus: Jens verteufelte Irmgard, weil sie bei diesem wichtigen Unternehmen nichts als gestört hätte.

ES 40 verurteilte

GenMinKonMin: Nadine verurteilte Nina, weil sie nach all den Erfahrungen die Schuldfrage eindeutig fand.

GenMinKonPlus: Nadine verurteilte Nina, weil sie nach all den Erfahrungen sich keineswegs geändert hatte.

GenPlusKonMin: Nadine verurteilte Jörg, weil sie nach all den Erfahrungen die Schuldfrage eindeutig fand.

GenPlusKonPlus: Edmund verurteilte Nina, weil sie nach all den Erfahrungen sich keineswegs geändert hatte.

Appendix C: Zusatzanalysen zu den Experimenten I und II

C.1: Zusatzanalysen Experiment Ia

C.1.1: Pronominale Auflösung. Akkuratheit. Arcussinus-Transformation.

Maßnahme. Alle Anteilswerte wurden arcussinus-transformiert um eine Normalverteilung der Residuen wahrscheinlicher zu machen.

Ergebnisse. Es gab einen Haupteffekt für den Faktor *Gender-Cue* ($F(1,21) = 277.79, p < .001, \eta_p^2 = .930$; $F(1,159) = 136.38, p < .001, \eta_p^2 = .462$) und einen Haupteffekt für den Faktor *IC-Cue* ($F(1,21) = 15.00, p < .001, \eta_p^2 = .417$; $F(1,159) = 108.74, p < .001, \eta_p^2 = .406$). Auch da die Interaktion der beiden Faktoren war signifikant ($F(1,21) = 12.31, p < .002, \eta_p^2 = .369$; $F(1,159) = 72.56, p < .001, \eta_p^2 = .313$), da der IC-Cue nur dann wirksam werden konnte, wenn ein Satz nicht bereits durch den Gender-Cue disambiguiert war (nicht-disambiguiert: $F(1,21) = 34.07, p < .001, \eta_p^2 = .619$; $F(1,159) = 143.38, p < .001, \eta_p^2 = .474$; disambiguiert: beide $F < 1$).

C.1.2: Pronominale Auflösung. Mittlere Reaktionszeit. Nur korrekte Antworten.

Maßnahme. Um zu überprüfen, ob das berichtete Effektmuster durch evtl. Unterschiede zwischen falsch und richtig beantworteten Items moderiert wird, wurde die gleiche Analyse noch einmal nur mit den RT-Mittelwerten für korrekt beantwortete Items berechnet. Bei der *by items*-Analyse ergaben sich durch den Ausschluss falscher Antworten für vier Zellen fehlende Werte (0.6%). Diese wurden jeweils durch den bedingungsspezifischen Mittelwert aller anderen Items ersetzt.

Ergebnisse. Das berichtete Ergebnismuster konnte vollständig repliziert werden: Es gab einen Haupteffekt für den Faktor *Gender-Cue* ($F(1,21) = 56.75, p < .001, \eta_p^2 = .730$; $F(1,159) = 258.67, p < .001, \eta_p^2 = .619$) und einen Haupteffekt für den Faktor *IC-Cue* ($F(1,21) = 30.90, p < .001, \eta_p^2 = .595$; $F(1,159) = 14.49, p < .001, \eta_p^2 = .084$), aber keine Interaktion (beide $F < 1$). Die Größe der beide Effekte ähnlich wie in der nicht-bereinigten Analyse (ca. 680 ms für den Gender-Effekt, ca. 180 ms für den IC-Effekt).

C.1.3: Pronominale Auflösung. Weibull-Parameter. 20 korrekte Items .

Maßnahme. Alle falsch beantworteten Items wurden von der Analyse ausgeschlossen. Von den verbleibenden Items wurden die ersten 20 pro Person und Untersuchungsbedingung zur Analyse ausgewählt. Ansonsten war das methodische Vorgehen identisch zur vollständigen Analyse.

Durchschnittliche Parameter. Die Werte für die einzelnen Parameter fielen ähnlich aus wie bei der vollständigen Analyse (shift: 1830 ms, scale: 1675 ms; shape: 2.01).

shift-Parameter. Es ergab sich ein signifikanter Haupteffekt für den Faktor *Gender-Cue* ($F(1,21) = 37.16, p < .001, \eta_p^2 = .639$): Disambiguiert Sätze konnten ca. 320 ms schneller beantwortet werden als nicht-disambiguierte. Alle anderen Effekte waren nicht signifikant (beide $F < 1$).

scale-Parameter. Es gab sowohl einen Haupteffekt für den Faktor *Gender-Cue* ($F(1,21) = 15.92, p < .001, \eta_p^2 = .431$) als auch einen stärkeren Haupteffekt für den Faktor *IC-Cue* ($F(1,21) = 26.02, p < .001, \eta_p^2 = .553$). Gender-disambiguierte Sätze wiesen einen ca. 220 ms niedrigeren scale-Parameter auf, während Sätze mit kausalitätskongruenten Begründungen einen ca. 150 ms niedrigeren scale-Parameter hatten. Die Interaktion der beiden Faktoren war nicht signifikant ($F < 1$).

shape-Parameter. Es gab weder einen Haupteffekt für den Gender-Cue noch für den IC-Cue (beide $F < 1$), da eine disordinale Interaktion der beiden Faktoren vorlag ($F(1,21) = 35.13, p < .001, \eta_p^2 = .626$). Während der IC-Cue in der Bedingung Gender+ einen leicht positiven Effekt hatte (0.26; $F(1,21) = 14.82, p < .001, \eta_p^2 = .441$), hatte er in der Bedingung Gender- ein leicht negatives (0.26; $F(1,21) = 41.76, p < .001, \eta_p^2 = .665$).

C.1.4: Verifikationsaufgabe. Akkuratheit. Arcussinus-Transformation.

Maßnahme. Alle Anteilswerte wurden arcussinus-transformiert um eine Normalverteilung der Residuen wahrscheinlicher zu machen.

Ergebnisse. Es gab einen Haupteffekt für den Faktor *Gender-Cue* ($F(1,21) = 61.56, p < .001, \eta_p^2 = .746$; $F(1,159) = 71.23, p < .001, \eta_p^2 = .309$). Der Haupteffekt für den Faktor *IC-Cue* was hingegen nicht signifikant ($F(1,21) = 2.75, p = .112, \eta_p^2 = .417$; $F(1,159) = 2.83, p = .095, \eta_p^2 = .017$), weil der IC-Cue nur in der nicht bereits gender-disambiguierten Bedingung wirksam werden konnte (Interaktion: $F(1,21) = 6.79, p < .017, \eta_p^2 = .244$; $F(1,159) = 1.33, p = .250, \eta_p^2 = .008$): Der EHE des IC-Cues war in der nicht-disambiguierten Bedingung signifikant oder marginal signifikant ($F(1,21) = 8.42, p < .009, \eta_p^2 = .286$; $F(1,159) = 2.70, p = .102, \eta_p^2 = .017$), in der disambiguiert jedoch nicht (beide $F < 1$).

C.1.5: Verifikationsaufgabe. Mittlere Reaktionszeit. Nur korrekte Antworten.

Maßnahme. Um zu überprüfen, ob das berichtete Effektmuster durch evtl. Unterschiede zwischen falsch und richtig beantworteten Items moderiert wird, wurde die gleiche Analyse noch einmal nur mit den RT-Mittelwerten für korrekt beantwortete Items berechnet.

Ergebnisse. Das berichtete Ergebnismuster konnte vollständig repliziert werden: Es gab einen Haupteffekt für den Faktor *Gender-Cue* ($F(1,21) = 41.81, p < .001, \eta_p^2 = .666$; $F(1,159) = 59.13, p < .001, \eta_p^2 = .217$) und einen Haupteffekt für den Faktor *IC-Cue* ($F(1,21) = 11.87, p < .002, \eta_p^2 = .361$; $F(1,159) = 4.81, p < .030, \eta_p^2 = .084$). Die Interaktion der beiden Faktoren war nur marginal signifikant ($F(1,21) = 3.38, p = .080, \eta_p^2 = .139$; $F(1,159) = 3.03, p = .084, \eta_p^2 = .019$), war aber jedoch wieder in der Mittelwertskonstellation klar zu erkennen: Der IC-Cue war nur wirksam, wenn ein Satz nicht bereits durch den Gender-Cue disambiguiert wurde.

C.1.6: Verifikationsaufgabe. Weibull-Parameter. 20 korrekte Items .

Maßnahme. Alle falsch beantworteten Items wurden von der Analyse ausgeschlossen. Von den verbleibenden Items wurden die ersten 20 pro Person und Untersuchungsbedingung zur Analyse ausgewählt. Ansonsten war das methodische Vorgehen identisch zur vollständigen Analyse.

Durchschnittliche Parameter. Die Werte für die einzelnen Parameter fielen ähnlich aus wie bei der vollständigen Analyse (shift: 630 ms, scale: 760 ms; shape: 1.60).

shift-Parameter. Es gab keine signifikanten Effekte ($F < 1.02$).

scale-Parameter. Es gab sowohl einen Haupteffekt für den Faktor *Gender-Cue* ($F(1,21) = 35.01, p < .001, \eta_p^2 = .625$) als auch einen stärkeren Haupteffekt für den Faktor *IC-Cue* ($F(1,21) = 4.92, p < .038, \eta_p^2 = .190$). Gender-disambiguierte Sätze wiesen einen um ca. 90 ms niedrigeren scale-Parameter auf. Obwohl die Interaktion der beiden Faktoren nicht signifikant ($F(1,21) = 2.03, p = .169, \eta_p^2 = .088$) war, zeigte eine Analyse der Einfachen Haupteffekte, dass der IC-Effekt lediglich in der Bedingung Gender – signifikant war (ca. 90 ms; $F(1,21) = 4.81, p < .040, \eta_p^2 = .186$), nicht jedoch in der Bedingung Gender + ($F < 1$).

shape-Parameter. Es gab einen Haupteffekt für den Gender-Cue ($F(1,21) = 16.08, p < .001, \eta_p^2 = .434$), während die anderen beiden Effekte nicht signifikant waren (beide $F < 1$). Gender-disambiguierte Sätze wiesen einen um ca. 0.24 Einheiten höheren shape-Parameter auf.

C.2: Zusatzanalysen Experiment Ib

C.2.1: Pronominale Auflösung. Akkuratheit. Arcussinus-Transformation.

Maßnahme. Alle Anteilswerte wurden arcussinus-transformiert.

Ergebnisse. Es gab einen Haupteffekt für den Faktor *Gender-Cue* ($F(1,22) = 128.62, p < .001, \eta_p^2 = .854$; $F(1,159) = 234.03, p < .001, \eta_p^2 = .595$) und einen Haupteffekt für den Faktor *IC-Cue* ($F(1,22) = 51.70, p < .001, \eta_p^2 = .701$; $F(1,159) = 35.70, p < .001, \eta_p^2 = .183$). Die Interaktion der beiden Faktoren war ebenfalls signifikant ($F(1,22) = 14.52, p < .001, \eta_p^2 = .398$; $F(1,159) = 23.33, p < .001, \eta_p^2 = .128$): Der IC-Cue fiel stärker aus, wenn ein Satz nicht bereits durch den Gender-Cue disambiguiert war ($F(1,22) = 65.69, p < .001, \eta_p^2 = .749$; $F(1,159) = 42.80, p < .001, \eta_p^2 = .212$). Wenn hingegen ein Gender-Cue vorhanden war, dann war der Kausalitätseffekt nicht signifikant (beide $F < 1.7$).

Vergleich Aussagetypen. Die einfaktorielle ANOVA wies einen signifikanten Wert auf ($F(2,44) = 8.95, p < .001, \eta_p^2 = .289$; $F(2,318) = 15.06, p < .001, \eta_p^2 = .087$): Der Mittelwert der Bedingung „inkorrekt“ war signifikant ($p < .001$) niedriger als die anderen beiden Mittelwerte, die sich nicht voneinander unterschieden ($p = .208$).

C.2.2: Pronominale Auflösung. Mittlere Reaktionszeit. Nur korrekte Antworten.

Maßnahme. Um zu überprüfen, ob das berichtete Effektmuster durch evtl. Unterschiede zwischen falsch und richtig beantworteten Items moderiert wird, wurde die gleiche Analyse noch einmal nur mit den RT-Mittelwerten für korrekt beantwortete Items berechnet.

Bei der *by items*-Analyse ergab sich durch den Ausschluss falscher Antworten für eine Zelle ein fehlender Wert, der durch den bedingungsspezifischen Mittelwert ersetzt wurde.

Ergebnisse. Das berichtete Ergebnismuster konnte vollständig repliziert werden: Es gab einen Haupteffekt für den Faktor *Gender-Cue* ($F(1,22) = 110.84, p < .001, \eta_p^2 = .834$; $F(1,159) = 295.33, p < .001, \eta_p^2 = .650$) und einen Haupteffekt für den Faktor *IC-Cue* ($F(1,22) = 51.93, p < .001, \eta_p^2 = .702$; $F(1,159) = 36.75, p < .001, \eta_p^2 = .188$), aber keine Interaktion (beide $F < 1.9$). Die Größe der beiden Effekte war ähnlich stark ausgeprägt wie in der nicht-bereinigten Analyse (ca. 700 ms für den Gender-Effekt, ca. 260 ms für den IC-Effekt).

C.2.3: Pronominale Auflösung. Weibull-Parameter. 20 korrekte Items .

Maßnahme. Alle falsch beantworteten Items wurden von der Analyse ausgeschlossen. Von den verbleibenden Items wurden die ersten 20 pro Person und Untersuchungsbedingung zur Analyse ausgewählt. Eine Vp wies in einer Untersuchungsbedingung nur 19 korrekte Antworten auf. Als 20. Wert wurde der bedingungsspezifische Mittelwert der Person eingesetzt. Ansonsten war das methodische Vorgehen mit dem der vollständigen Analyse identisch.

Durchschnittliche Parameter. Die Werte für die einzelnen Parameter fielen ähnlich aus wie bei der vollständigen Analyse (shift: 2660 ms, scale: 1910 ms; shape: 1.88).

shift-Parameter. Es ergab sich ein signifikanter Haupteffekt für den Faktor *Gender-Cue* ($F(1,22) = 53.39, p < .001, \eta_p^2 = .708$): Disambiguiert Sätze konnten ca. 460 ms schneller beantwortet werden als nicht-disambiguierte. Der Haupteffekt des Kausalitäts-Cues war nicht signifikant ($F(1,22) < 1, p = .483, \eta_p^2 = .023$), aber der Interaktionsterm ($F(1,22) = 11.21, p < .003, \eta_p^2 = .337$): Wieder war der IC-Cue lediglich in der bereits disambiguierten Bedingung signifikant ($F(1,22) = 8.60, p < .008, \eta_p^2 = .281$), der ca. 150 ms groß war. In der Bedingung Gender – war der Effekt des Kausalitäts-Cues hingegen nur marginal signifikant ($F(1,22) = 3.83, p = .063, \eta_p^2 = .148$), war jedoch negativ ausgeprägt (ca. 100 ms).

scale-Parameter. Es gab sowohl einen Haupteffekt für den Faktor *Gender-Cue* ($F(1,22) = 30.44, p < .001, \eta_p^2 = .580$) als auch einen stärkeren Haupteffekt für den Faktor *IC-Cue* ($F(1,22) = 25.52, p < .001, \eta_p^2 = .537$). Gender-disambiguierte Sätze wiesen einen um ca. 280 ms niedrigeren scale-Parameter auf, während Sätze mit kausalitätskongruenten Begründungen einen ca. 240 ms niedrigeren scale-Parameter hatten. Die Interaktion der beiden Faktoren war ebenfalls signifikant ($F(1,22) = 14.87, p < .001, \eta_p^2 = .403$): Der Kausalitäts-Cue war nur in der nicht-gender-disambiguierten Bedingung signifikant (ca. 400 ms; $F(1,22) = 38.06, p < .001, \eta_p^2 = .634$), während er in der Bedingung Gender – nicht signifikant war (ca. 80 ms; $F(1,22) = 1.74, p = .200, \eta_p^2 = .073$).

shape-Parameter. Es gab einen schwachen Haupteffekt für den Gender-Cue (ca. 0.17 Skalenpunkte; $F(1,22) = 9.85, p < .005, \eta_p^2 = .309$), der von einem signifikanten Interaktionsterm ($F(1,22) = 6.59, p < .018, \eta_p^2 = .230$) auf die Bedingung IC-Cue + eingeschränkt wurde (ca. 300 Skalenpunkte; $F(1,22) = 11.12, p < .003, \eta_p^2 = .336$), während er in der Bedingung IC-Cue – nicht signifikant war ($F < 1$).

C.2.4: Verifikationsaufgabe. Akkuratheit. Arcussinus-Transformation.

Maßnahme. Alle Anteilswerte wurden arcussinus-transformiert um eine Normalverteilung der Residuen wahrscheinlicher zu machen.

Vergleich Aussagetypen. Die einfaktorielle ANOVA wies einen signifikanten Wert auf ($F(2,44) = 8.95, p < .001, \eta_p^2 = .289$; $F(2,318) = 15.06, p < .001, \eta_p^2 = .087$): Der Mittelwert der Bedingung „inkorrekt“ war signifikant ($p < .001$) niedriger als die andere beiden Mittelwerte, die sich nicht voneinander unterschieden ($p = .208$).

Vergleich Untersuchungsbedingungen. Es gab einen Haupteffekt für den Faktor *Gender-Cue* ($F(1,22) = 61.71, p < .001, \eta_p^2 = .737$; $F(1,159) = 44.66, p < .001, \eta_p^2 = .219$). Der Haupteffekt für den Faktor *IC-Cue* war ebenfalls signifikant ($F(1,22) = 5.84, p < .024, \eta_p^2 = .210$; $F(1,159) = 11.99, p < .001, \eta_p^2 = .070$), jedoch konnte der *IC-Cue* wiederum nur in der nicht bereits *gender-disambiguierten* Bedingung wirksam werden (Interaktion: $F(1,22) = 12.68, p < .002, \eta_p^2 = .366$; $F(1,159) = 9.67, p < .002, \eta_p^2 = .057$): Der EHE des *IC-Cues* war nur in der nicht-disambiguierten Bedingung signifikant ($F(1,22) = 23.60, p < .001, \eta_p^2 = .518$; $F(1,159) = 13.78, p < .001, \eta_p^2 = .080$), in der disambiguiert jedoch nicht oder nur marginal signifikant (beide $F < 1$).

C.2.5: Verifikationsaufgabe. Mittlere Reaktionszeit. Nur korrekte Antworten.

Maßnahme. Um zu überprüfen, ob das berichtete Effektmuster durch evtl. Unterschiede zwischen falsch und richtig beantworteten Items moderiert wird, wurde die gleiche Analyse noch einmal nur mit den RT-Mittelwerten für korrekt beantwortete Items berechnet.

Vergleich Aussagetypen. Die einfaktorielle ANOVA wies einen signifikanten Wert auf ($F(2,44) = 32.11, p < .001, \eta_p^2 = .593$; $F(2,318) = 29.44, p < .001, \eta_p^2 = .156$): Alle drei Mittelwerten trennten klar voneinander, wobei der korrekte Itemtyp den niedrigsten Wert aufwies, der inkorrekte den mittleren und der falsch Itemtyp den höchsten.

Ergebnisse. Das berichtete Ergebnismuster konnte vollständig repliziert werden: Es gab einen Haupteffekt für den Faktor *Gender-Cue* ($F(1,22) = 22.98, p < .001, \eta_p^2 = .511$; $F(1,159) = 22.03, p < .001, \eta_p^2 = .122$) und einen Haupteffekt für den Faktor *IC-Cue* ($F(1,22) = 27.30, p < .001, \eta_p^2 = .554$; $F(1,159) = 29.58, p < .001, \eta_p^2 = .152$). Die Interaktion der beiden Faktoren war ebenfalls signifikant ($F(1,22) = 8.54, p < .008, \eta_p^2 = .280$; $F(1,159) = 19.14, p < .001, \eta_p^2 = .107$): Der *IC-Cue* war stärker ausgeprägt, wenn ein Satz nicht durch den *Gender-Cue* disambiguiert wurde (190 ms vs. 60 ms).

C.3: Zusatzanalysen Experiment IIa

C.3.1: Satzlesezeiten. Mittlere Reaktionszeit. Nur korrekte Antworten.

Maßnahme. Um zu überprüfen, ob das berichtete Effektmuster durch evtl. Unterschiede zwischen falsch und richtig beantworteten Items moderiert wird, wurde die gleiche Analyse noch einmal nur mit den RT-Mittelwerten für korrekt beantwortete Items berechnet. Bei der by-Items-Analyse ergaben sich durch den Ausschluss falscher Antworten für drei Zellen fehlende Werte (0.5%). Diese wurden jeweils durch den bedingungsspezifischen Mittelwert aller anderen Items ersetzt.

Wort-Ebene. Die varianzanalytischen Ergebnisse für die Effekte der experimentellen Faktoren auf der Wortebene finden sich in Tab. C.3.1.

Konstituenten-Ebene. Die Ergebnisse des varianzanalytischen Vergleichs der Konstituenten auf Unterschiede zwischen den vier Untersuchungsbedingungen finden sich in Tab. C.3.2.

By-Participants	HS	Pro	PP	VP
<i>Gender-Cue</i>				
<i>F(1,24)</i>	1.03	2.17	5.57	11.92
<i>p</i>	.319	.153	.027*	.002**
η_p^2	.041	.083	.188	.332
<i>IC-Cue</i>				
<i>F(1,24)</i>	< 1	< 1	< 1	17.29
<i>p</i>	.684	.937	.645	.001**
η_p^2	.007	< .001	.009	.419
<i>Gender- x IC-Cue</i>				
<i>F(1,24)</i>	< 1	3.20	< 1	< 1
<i>p</i>	.986	.086	.609	.483
η_p^2	< .001	.118	.011	.021
By-Items				
<i>Gender-Cue</i>				
<i>F(1,159)</i>	< 1	4.70	12.18	80.57
<i>p</i>	.864	.032*	.001**	.001**
η_p^2	< .001	.029	.071	.336
<i>IC-Cue</i>				
<i>F(1,159)</i>	< 1	< 1	< 1	11.21
<i>p</i>	.633	.546	.434	.001**
η_p^2	.001	.002	.004	.066
<i>Gender- x IC-Cue</i>				
<i>F(1,159)</i>	< 1	< 1	< 1	1.06
<i>p</i>	.836	.780	.532	.305
η_p^2	< .001	< .001	.002	.007

Tab. C.3.2: Ergebnisse des varianzanalytischen Vergleichs der vier Konstituenten auf Unterschiede hinsichtlich der vier Untersuchungsbedingungen.

By-Participants	[NP1]	V	NP2	weil]	[er	PP1	PP2]	[PP3	PP4	VP1]	[VP2	VP3	VP4]
<i>Gender-Cue</i>													
<i>F(1,24)</i>	< 1	< 1	< 1	2.01	< 1	2.75	3.31	4.06	6.77	3.19	13.59	11.61	9.82
<i>p</i>	.624	.576	.340	.170	.562	.110	.081	.055	.016*	.087	.001**	.002**	.005**
η_p^2	.043	< .001	< .001	.093	< .001	.084	.117	.140	.268	.157	.440	.344	.309
<i>IC-Cue</i>													
<i>F(1,24)</i>	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1	23.81
<i>p</i>	.672	.535	.896	.982	.771	.378	.610	.874	.660	.608	.359	.337	.001**
η_p^2	.008	.016	.001	< .001	.004	.033	.011	.001	.008	.011	.035	.039	.498
<i>Gender- x IC-Cue</i>													
<i>F(1,24)</i>	< 1	< 1	< 1	< 1	5.05	< 1	< 1	1.19	< 1	< 1	< 1	< 1	3.23
<i>p</i>	.541	.994	.465	.803	.034*	.723	.617	.286	.959	.865	.551	.592	.085
η_p^2	.002	.073	.071	< .001	.206	.063	.110	.223	.002	< .001	.052	.010	.115
By-Items													
<i>Gender-Cue</i>													
<i>F(1,159)</i>	2.19	< 1	< 1	2.07	1.16	4.10	6.67	5.86	11.47	8.05	14.52	23.61	99.14
<i>p</i>	.141	.806	.663	.153	.283	.045*	.011*	.017*	.001**	.005**	.001**	.001**	.001**
η_p^2	.014	< .001	< .001	.013	.007	.025	.040	.036	.067	.048	.084	.129	.384
<i>IC-Cue</i>													
<i>F(1,159)</i>	< 1	< 1	< 1	< 1	< 1	1.34	< 1	1.08	< 1	< 1	< 1	1.09	14.34
<i>p</i>	.533	.929	.712	.583	.853	.250	.874	.301	.563	.682	.357	.298	.001**
η_p^2	.002	< .001	.001	.002	< .001	.008	< .001	.007	.002	.001	.005	.007	.083
<i>Gender- x IC-Cue</i>													
<i>F(1,159)</i>	1.81	< 1	< 1	< 1	1.04	< 1	< 1	< 1	< 1	< 1	< 1	< 1	3.14
<i>p</i>	.180	.767	.541	.654	.309	.808	.951	.927	.369	.605	.810	.952	.078
η_p^2	.011	.001	.002	.001	.007	< .001	< .001	< .001	.005	.002	< .001	< .001	.019

Tab. C.3.1: Ergebnisse des varianzanalytischen Vergleichs der vier Untersuchungsbedingungen auf der Wortebene.

C.3.2: Satzlesezeiten. Weibull-Parameter. 20 korrekte Items.

Maßnahme. Alle bei der Verifikationsaufgabe falsch beantworteten Items wurden von der Analyse ausgeschlossen. Von den verbleibenden Items wurden die ersten 20 pro Person und Untersuchungsbedingung zur Analyse ausgewählt. Drei Personen wiesen in jeweils einer Bedingung keine 20 korrekt beantworteten Items auf. Die entsprechenden Datenreihen wurden durch die bedingungs- und personenspezifischen Mittelwerte aufgefüllt. Diese Maßnahme führte zur Ersetzung von ca. 0.05% aller Datenpunkte. Ansonsten war das methodische Vorgehen identisch zur vollständigen Analyse.

Wort-Ebene. Die varianzanalytischen Ergebnisse für die Effekte der experimentellen Faktoren auf der Wortebene finden sich in Tab. C.3.3.

Konstituenten-Ebene. Die Ergebnisse des varianzanalytischen Vergleichs der Konstituenten auf Unterschiede zwischen den vier Untersuchungsbedingungen finden sich in Tab. C.3.4.

	[NP1	V	NP2	weil]	[Pro	PP1	PP2]	[PP3	PP4]	[VP1	VP2	VP3	VP4]
shift													
<i>Gender-Cue</i>													
<i>F(1,24)</i>	1.51	6.13	< 1	< 1	< 1	< 1	1.23	2.19	1.69	< 1	5.13	4.49	7.15
<i>p</i>	.231	.021*	.514	.641	.635	.991	.278	.152	.207	.653	.033*	.045*	.013**
η_p^2	.059	.204	.018	.009	.010	< .001	.049	.084	.066	.009	.176	.158	.230
<i>IC-Cue</i>													
<i>F(1,24)</i>	< 1	1.83	< 1	< 1	< 1	1.05	< 1	< 1	< 1	4.00	5.76	< 1	< 1
<i>p</i>	.670	.189	.959	.699	.452	.315	.825	.330	.959	.057	.024*	.475	.389
η_p^2	.008	.071	< .001	.006	.024	.042	.002	.040	< .001	.143	.194	.021	.031
<i>Gender- x IC-Cue</i>													
<i>F(1,24)</i>	3.05	< 1	< 1	2.25	< 1	1.53	< 1	< 1	< 1	< 1	< 1	1.13	1.46
<i>p</i>	.094	.674	.411	.147	.616	.229	.424	.688	.530	.684	.457	.298	.240
η_p^2	.113	.008	.028	.086	.011	.060	.027	.007	.017	.007	.023	.045	.057
scale													
<i>Gender-Cue</i>													
<i>F(1,24)</i>	< 1	8.43	< 1	< 1	< 1	< 1	2.32	2.65	< 1	< 1	< 1	< 1	3.41
<i>p</i>	.788	.008**	.649	.914	.515	.938	.141	.116	.562	.827	.375	.827	.077
η_p^2	.005	.150	.002	.015	.058	.001	.073	.075	.012	.004	.076	.034	.127
<i>IC-Cue</i>													
<i>F(1,24)</i>	< 1	1.11	< 1	< 1	< 1	3.04	< 1	< 1	< 1	1.73	16.95	< 1	12.33
<i>p</i>	.703	.302	.438	.866	.341	.094	.841	.657	.825	.194	.001**	.477	.002*
η_p^2	.006	.044	.025	.001	.038	.112	.002	.008	.002	.069	.414	.021	.339
<i>Gender- x IC-Cue</i>													
<i>F(1,24)</i>	2.48	< 1	< 1	3.15	1.65	< 1	< 1	< 1	< 1	< 1	< 1	< 1	< 1
<i>p</i>	.128	.796	.443	.089	.212	.360	.630	.757	.506	.408	.998	.941	.692
η_p^2	.094	.003	.025	.116	.064	.035	.010	.004	.019	.029	< .001	< .001	.007

	[NP1]	V	NP2	weil]	[er	PP1	PP2]	[PP3	PP4]	[VP1	VP2	VP3	VP4]
shape													
<i>Gender-Cue</i>													
<i>F(1,24)</i>	12.98	8.58	< 1	< 1	1.31	< 1	24.02	< 1	< 1	< 1	5.21	10.07	2.32
<i>p</i>	.001**	.007**	.629	.386	.265	.944	.001**	.848	.616	.615	.032*	.004**	.141
η_p^2	.351	.263	.010	.031	.052	< .001	.500	.002	.011	.011	.178	.296	.088
<i>IC-Cue</i>													
<i>F(1,24)</i>	3.16	< 1	7.91	< 1	2.28	< 1	< 1	9.28	1.98	18.71	28.57	8.72	< 1
<i>p</i>	.088	.352	.010*	.611	.145	.971	.758	.006**	.173	.001**	.001**	.007**	.579
η_p^2	.116	.036	.248	.011	.087	< .001	.004	.279	.076	.438	.543	.266	.013
<i>Gender- x IC-Cue</i>													
<i>F(1,24)</i>	4.03	< 1	2.03	9.98	< 1	1.85	3.18	3.63	< 1	< 1	17.25	2.36	3.32
<i>p</i>	.056	.693	.167	.004**	.977	.187	.087	.069	.394	.330	.001**	.137	.081
η_p^2	.144	.007	.078	.294	< .001	.071	.117	.131	.030	.039	.418	.090	.122

Tab. C.3.3: Varianzanalytischer Vergleich der drei Weibull-Parameter auf Unterschiede hinsichtlich der vier Untersuchungsbedingungen pro Wort.

	<i>shift</i>				<i>scale</i>				<i>shape</i>			
	HS	Pro	PP	VP	HS	Pro	PP	VP	HS	Pro	PP	VP
<i>Gender-Cue</i>												
<i>F(1,24)</i>	< 1	< 1	1.43	9.36	< 1	< 1	< 1	2.25	3.24	4.70	< 1	16.20
<i>p</i>	.939	.833	.244	.005**	.366	.734	.482	.147	.084	.040*	.586	.001**
η_p^2	< .001	.002	.056	.281	.034	.005	.021	.086	.119	.164	.013	.403
<i>IC-Cue</i>												
<i>F(1,24)</i>	< 1	< 1	< 1	1.70	< 1	1.38	< 1	18.60	< 1	< 1	1.71	1.97
<i>p</i>	.510	.532	.566	.204	.524	.251	.502	.001**	.490	.680	.203	.174
η_p^2	.018	.017	.014	.066	.017	.054	.019	.437	.020	.007	.066	.076
<i>Gender- x IC-Cue</i>												
<i>F(1,24)</i>	< 1	1.01	< 1	< 1	< 1	1.36	< 1	< 1	< 1	2.02	< 1	2.34
<i>p</i>	.453	.307	.738	.909	.710	.255	.458	.771	.963	.168	.342	.139
η_p^2	.024	.043	.005	.001	.006	.054	.023	.004	< .001	.078	.038	.089

Tab. C.3.4: Ergebnisse des varianzanalytischen Vergleichs der vier Konstituenten auf Unterschiede hinsichtlich der vier Untersuchungsbedingungen.

C.3.3: Verifikationsaufgabe. Akkuratheit. Arcussinus-Transformation.

Maßnahme. Alle Anteilswerte wurden arcussinus-transformiert um eine Normalverteilung der Residuen zu gewährleisten.

Ergebnisse. Es gab einen extrem starken Haupteffekt für den Faktor *Gender-Cue* ($F(1,24) = 227.92, p < .001, \eta_p^2 = .905; F(1,159) = 483.06, p < .001, \eta_p^2 = .752$). Der Haupteffekt für den Faktor *IC-Cue* was hingegen nur marginal signifikant ($F(1,24) = 3.02, p = .095, \eta_p^2 = .112; F(1,159) = 2.94, p = .088, \eta_p^2 = .017$), weil der IC-Cue nur in der nicht bereits gender-disambiguierten Bedingung wirksam werden konnte (Interaktion: $F(1,24) = 5.76, p < .025, \eta_p^2 = .193; F(1,159) = 3.08, p = .081, \eta_p^2 = .019$): Der EHE des IC-Cues war in der nicht-disambiguierten Bedingung signifikant ($F(1,24) = 12.18, p < .002, \eta_p^2 = .337; F(1,159) = 4.47, p < .036, \eta_p^2 = .027$), in der disambiguiert jedoch nicht (beide $F < 1$).

C.3.4: Verifikationsaufgabe. Mittlere Reaktionszeit. Nur korrekte Antworten.

Maßnahme. Um zu überprüfen, ob das berichtete Effektmuster durch evtl. Unterschiede zwischen falsch und richtig beantworteten Items moderiert wird, wurde die gleiche Analyse noch einmal nur mit den RT-Mittelwerten für korrekt beantwortete Items berechnet. In der by-Items-Analyse ergaben sich durch den Ausschluss für drei Datenpunkte (=0.5%) fehlende Werte. Sie wurden jeweils durch den bedingungsspezifischen Mittelwert ersetzt.

Ergebnisse. Das berichtete Ergebnismuster konnte vollständig repliziert werden: Es gab einen Haupteffekt für den Faktor *Gender-Cue* ($F(1,24) = 45.14, p < .001, \eta_p^2 = .653; F(1,159) = 74.38, p < .001, \eta_p^2 = .319$) und einen Haupteffekt für den Faktor *IC-Cue* ($F(1,24) = 14.06, p < .001, \eta_p^2 = .369; F(1,159) = 7.39, p < .007, \eta_p^2 = .044$). Die Interaktion der beiden Faktoren war hingegen nicht signifikant (beide $F < 1$). Der Gender-Effekt war 220 ms und der IC-Effekt 65 ms groß.

C.3.5: Verifikationsaufgabe. Weibull-Parameter. 20 korrekte Antworten.

Maßnahme. Alle falsch beantworteten Items wurden von der Analyse ausgeschlossen. Von den verbleibenden Items wurden die ersten 20 Datenpunkte pro Person und Untersuchungsbedingung zur Analyse ausgewählt. Drei Vpn wiesen in einer Bedingung nicht die benötigten 20 korrekten Antworten auf. Insgesamt fehlten acht Datenpunkte, die durch den Bedingungsspezifischen Mittelwert pro Person ersetzt wurden.

Durchschnittliche Parameter. Die Werte für die einzelnen Parameter fielen ähnlich aus wie bei der vollständigen Analyse (shift: 731 ms, scale: 732 ms; shape: 1.71).

shift-Parameter. Für diesen Parameter gab keine signifikanten Haupteffekte ($F < 1$). Allerdings war die für die Interaktion der beiden Effekt marginal signifikant ($F(1,24) = 3.80, p = .063, \eta_p^2 = .137$): Während es keinen IC-Effekt für die Bedingung Gender – gab ($F < 1$), war er in der Bedingung Gender + signifikant ($F(1,24) = 4.76, p < .039, \eta_p^2 = .166$): Kongruenz zwischen expliziter und impliziter Kausalität führte in diesem Fall zum einem Verarbeitungsvorteil von ca. 60 ms.

scale-Parameter. Es gab einen Haupteffekt für den Faktor *Gender-Cue* ($F(1,24) = 36.84, p < .001, \eta_p^2 = .606$), während der Haupteffekt des Faktors *IC-Cue* nur marginal signifikant war ($F(1,24) = 3.14, p = .089, \eta_p^2 = .116$). Gender-disambiguierte Sätze wiesen einen um ca. 170 ms niedrigeren scale-Parameter auf. Die Interaktion der beiden Faktoren war marginal signifikant ($F(1,24) = 3.48, p = .074, \eta_p^2 = .127$). Eine Analyse der einfachen Haupteffekte ergab, dass IC-Effekt nur in der noch nicht gender-disambiguierten Untersuchungsbedingung ausgeprägt war ($F(1,24) = 5.59, p < .027, \eta_p^2 = .189$). In der Bedingung Gender + gab es hingegen keinen Effekt für den IC-Cue ($F < 1$).

shape-Parameter. Es gab einen Haupteffekt für den Gender-Cue ($F(1,24) = 12.98, p < .001, \eta_p^2 = .351$), allerdings keinen Effekt für den IC-Cue ($F(1,24) = 2.30, p = .143, \eta_p^2 = .087$). Allerdings war der Interaktionsterm signifikant ($F(1,24) = 13.98, p < .001, \eta_p^2 = .368$), weil der IC-Cue nur dann einen Effekt hatte, wenn der Satz bereits gender-disambiguiert war ($F(1,24) = 11.42, p < .002, \eta_p^2 = .322$). In der Bedingung Gender – war der Kausalitätseffekt hingegen nicht signifikant ($F(1,24) = 2.86, p = .104, \eta_p^2 = .106$).

C.4: Zusatzanalysen Experiment IIb

C.4.1: Satzlesezeiten. Mittlere Reaktionszeit. Nur korrekte Antworten.

Maßnahme. Um zu überprüfen, ob das berichtete Effektmuster durch evtl. Unterschiede zwischen falsch und richtig beantworteten Items moderiert wird, wurde die gleiche Analyse noch einmal nur mit den RT-Mittelwerten für korrekt beantwortete Items berechnet. Bei der by-Items-Analyse ergaben sich durch den Ausschluss falscher Antworten für sches Zellen fehlende Werte (0.6%). Diese wurden jeweils durch den bedingungsspezifischen Mittelwert aller anderen Items ersetzt.

Wort-Ebene. Die varianzanalytischen Ergebnisse für die Effekte der experimentellen Faktoren auf der Wortebene finden sich in Tab. C.4.1.

Konstituenten-Ebene. Die Ergebnisse des varianzanalytischen Vergleichs der Konstituenten auf Unterschiede zwischen den vier Untersuchungsbedingungen finden sich in Tab. C.4.2.

	HS	Pro	PP	VP
By-Participants				
<i>Gender-Cue</i>				
<i>F(1,19)</i>	< 1	1.48	4.41	13.78
<i>p</i>	.922	.239	.049*	.001**
η_p^2	.001	.072	.188	.420
<i>IC-Cue</i>				
<i>F(1,19)</i>	1.10	< 1	< 1	22.16
<i>p</i>	.307	.701	.468	.001**
η_p^2	.055	.008	.028	.539
<i>Gender- x IC-Cue</i>				
<i>F(1,19)</i>	< 1	< 1	5.90	2.67
<i>p</i>	.857	.432	.025*	.119
η_p^2	.002	.033	.237	.123
By-Items				
<i>Gender-Cue</i>				
<i>F(1,159)</i>	< 1	1.48	3.80	32.98
<i>p</i>	.923	.225	.053	.001**
η_p^2	< .001	.009	.023	.172
<i>IC-Cue</i>				
<i>F(1,159)</i>	1.69	< 1	< 1	14.12
<i>p</i>	.196	.644	.816	.001**
η_p^2	.010	.001	< .001	.082
<i>Gender- x IC-Cue</i>				
<i>F(1,159)</i>	< 1	1.18	4.04	< 1
<i>p</i>	.993	.280	.046*	.742
η_p^2	< .001	.007	.025	.001

Tab. C.4.2: Ergebnisse des varianzanalytischen Vergleichs der vier Konstituenten auf Unterschiede hinsichtlich der vier Untersuchungsbedingungen.

	[NP1	V	NP2	weil]	[er	PP1	PP2]	[PP3	PP4	VP1]	[VP2	VP3	VP4]
By-Participants													
<i>Gender-Cue</i>													
<i>F(1,19)</i>	< 1	< 1	< 1	4.15	< 1	< 1	3.70	3.89	6.36	< 1	< 1	4.66	14.33
<i>p</i>	.895	.652	.357	.056	.539	.606	.069	.063	.021*	.377	.397	.044*	.001**
η_p^2	.001	.011	.045	.179	.020	.014	.163	.170	2.51	.041	.038	.197	.430
<i>IC-Cue</i>													
<i>F(1,19)</i>	3.25	< 1	< 1	< 1	< 1	< 1	1.21	< 1	1.03	< 1	2.95	< 1	24.76
<i>p</i>	.088	.898	.424	.797	.724	.840	.286	.659	.323	.631	.102	.634	.001**
η_p^2	.146	.001	.034	.004	.007	.002	.060	.010	.051	.012	.134	.012	.566
<i>Gender- x IC-Cue</i>													
<i>F(1,19)</i>	< 1	1.14	< 1	< 1	< 1	1.17	< 1	< 1	1.40	7.13	4.34	1.13	1.32
<i>p</i>	.764	.299	.406	.813	.676	.293	.760	.523	.251	.015*	.051	.300	.266
η_p^2	.004	.057	.037	.003	.009	.058	.005	.022	.069	.273	.186	.056	.065
By-Items													
<i>Gender-Cue</i>													
<i>F(1,159)</i>	< 1	< 1	< 1	2.36	< 1	< 1	1.60	2.15	4.01	< 1	1.93	12.17	30.32
<i>p</i>	.927	.665	.414	.126	.356	.389	.207	.145	.047*	.384	.166	.001**	.001**
η_p^2	< .001	.001	.004	.015	.005	.005	.010	.013	.025	.005	.012	.071	.160
<i>IC-Cue</i>													
<i>F(1,159)</i>	2.73	< 1	1.07	< 1	< 1	< 1	< 1	1.22	< 1	< 1	2.39	< 1	15.23
<i>p</i>	.100	.706	.302	.377	.753	.965	.483	.271	.569	.511	.124	.758	.001**
η_p^2	.017	.001	.007	.005	.001	< .001	.003	.008	.002	.003	.015	.001	.087
<i>Gender- x IC-Cue</i>													
<i>F(1,159)</i>	< 1	< 1	< 1	< 1	1.54	1.41	< 1	1.16	< 1	6.25	2.16	1.33	< 1
<i>p</i>	.764	.703	.660	.679	.217	.236	.751	.284	.330	.013*	.144	.251	.993
η_p^2	.001	.001	.001	.001	.010	.009	.001	.007	.006	.038	.013	.008	< .001

Tab. C.4.1: Ergebnisse des varianzanalytischen Vergleichs der vier Untersuchungsbedingungen auf der Wortebene.

C.4.2: Satzlesezeiten. Weibull-Parameter. 20 korrekte Items .

Maßnahme. Alle bei der Verifikationsaufgabe falsch beantworteten Items wurden von der Analyse ausgeschlossen. Von den verbleibenden Items wurden die ersten 20 pro Person und Untersuchungsbedingung zur Analyse ausgewählt.

Wort-Ebene. Die varianzanalytischen Ergebnisse für die Effekte der experimentellen Faktoren auf der Wortebene finden sich in Tab. C.4.3.

Konstituenten-Ebene. Die Ergebnisse des varianzanalytischen Vergleichs der Konstituenten auf Unterschiede zwischen den vier Untersuchungsbedingungen finden sich in Tab. C. 4.4.

	[NP1	V	NP2	weil]	[Pro	PP1	PP2]	[PP3	PP4]	[VP1	VP2	VP3	VP4]
shift													
<i>Gender-Cue</i>													
<i>F(1,19)</i>	< 1	13.47	6.60	3.68	1.52	1.08	< 1	1.92	< 1	2.40	< 1	< 1	7.13
<i>p</i>	.649	.002**	.019*	.070	.233	.312	.741	.182	.546	.138	.586	.607	.015*
η_p^2	.011	.415	.258	.162	.074	.054	.006	.092	.020	.112	.016	.014	.273
<i>IC-Cue</i>													
<i>F(1,19)</i>	< 1	< 1	1.06	< 1	< 1	9.26	< 1	< 1	< 1	7.89	16.16	1.95	< 1
<i>p</i>	.907	.556	.316	.967	.443	.007**	.475	.469	.562	.011*	.001**	.178	.756
η_p^2	.001	.019	.053	< .001	.031	.328	.027	.028	.018	.293	.460	.093	.005
<i>Gender- x IC-Cue</i>													
<i>F(1,19)</i>	< 1	< 1	< 1	< 1	1.85	2.34	< 1	< 1	1.16	1.22	< 1	1.52	1.58
<i>p</i>	.553	.650	.973	.496	.190	.143	.833	.487	.294	.284	.505	.232	.224
η_p^2	.019	.011	< .001	.025	.089	.110	.002	.026	.058	.060	.024	.074	.077
scale													
<i>Gender-Cue</i>													
<i>F(1,19)</i>	< 1	18.71	2.67	7.07	< 1	< 1	< 1	< 1	5.95	< 1	< 1	< 1	10.15
<i>p</i>	.762	.001**	.119	.016*	.863	.344	.716	.644	.025*	.354	.653	.345	.005**
η_p^2	.005	.496	.123	.271	.002	.047	.007	.011	.239	.045	.011	.047	.348
<i>IC-Cue</i>													
<i>F(1,19)</i>	< 1	< 1	2.41	< 1	< 1	13.36	< 1	< 1	< 1	2.68	5.41	2.09	11.62
<i>p</i>	.765	.574	.137	.789	.664	.002**	.395	.507	.434	.118	.031*	.165	.003**
η_p^2	.005	.017	.113	.004	.010	.413	.038	.023	.033	.124	.222	.099	.380
<i>Gender- x IC-Cue</i>													
<i>F(1,19)</i>	< 1	< 1	< 1	< 1	1.26	< 1	< 1	< 1	< 1	< 1	< 1	4.37	1.07
<i>p</i>	.643	.505	.408	.349	.275	.576	.973	.748	.959	.682	.689	.050*	.314
η_p^2	.012	.024	.036	.046	.062	.017	< .001	.006	< .001	.009	.009	.187	.053

	[NP1	V	NP2	weil]	[er	PP1	PP2]	[PP3	PP4]	[VP1	VP2	VP3	VP4]
shape													
<i>Gender-Cue</i>													
<i>F(1,19)</i>	< 1	8.72	25.53	< 1	< 1	8.74	24.35	3.73	21.91	1.22	6.24	< 1	5.30
<i>p</i>	.597	.008**	.001**	.463	.739	.008**	.001**	.068	.001**	.283	.022*	.720	.033*
η_p^2	.015	.315	.573	.029	.006	.315	.562	.164	.536	.060	.247	.007	.218
<i>IC-Cue</i>													
<i>F(1,19)</i>	4.67	2.68	4.56	3.28	16.47	12.23	1.88	3.24	< 1	8.45	27.86	1.51	< 1
<i>p</i>	.044*	.118	.046*	.086	.001**	.002**	.187	.088	.960	.009**	.001**	.234	.336
η_p^2	.197	.123	.194	.147	.464	.392	.090	.146	< .001	.308	.595	.074	.049
<i>Gender- x IC-Cue</i>													
<i>F(1,19)</i>	< 1	27.77	< 1	25.94	3.70	6.55	< 1	< 1	5.02	< 1	1.34	17.43	< 1
<i>p</i>	.758	.001**	.546	.001**	.069	.019*	.496	.469	.037*	.447	.262	.001**	.765
η_p^2	.005	.594	.019	.577	.163	.256	.025	.028	.209	.031	.066	.478	.005

Tab. C.4.3: Varianzanalytischer Vergleich der drei Weibull-Parameter auf Unterschiede hinsichtlich der vier Untersuchungsbedingungen pro Wort.

	<i>shift</i>				<i>scale</i>				<i>shape</i>			
	HS	Pro	PP	VP	HS	Pro	PP	VP	HS	Pro	PP	VP
<i>Gender-Cue</i>												
<i>F(1,19)</i>	< 1	1.61	1.74	5.30	< 1	< 1	< 1	7.58	1.54	18.29	6.68	8.23
<i>p</i>	.705	.221	.203	.033*	.483	.709	.643	.013*	.230	.001**	.018*	.010**
η_p^2	.008	.078	.084	.218	.026	.007	.012	.285	.075	.490	.260	.302
<i>IC-Cue</i>												
<i>F(1,19)</i>	< 1	4.27	4.91	< 1	< 1	4.97	2.89	7.17	< 1	7.94	< 1	10.62
<i>p</i>	.687	.053	.039*	.475	.460	.038*	.105	.015*	.693	.011*	.373	.004**
η_p^2	.009	.183	.205	.027	.029	.207	.132	.274	.008	.295	.042	.359
<i>Gender- x IC-Cue</i>												
<i>F(1,19)</i>	< 1	< 1	1.83	1.72	1.85	< 1	< 1	3.32	35.60	< 1	< 1	3.19
<i>p</i>	.758	.915	.192	.206	.189	.788	.938	.084	.001**	.621	.662	.090
η_p^2	.005	.001	.088	.083	.089	.004	< .001	.149	.652	.013	.010	.144

Tab. C.4.4: Ergebnisse des varianzanalytischen Vergleichs der vier Konstituenten auf Unterschiede hinsichtlich der vier Untersuchungsbedingungen.

C.4.3: Verifikationsaufgabe. Akkuratheit. Arcussinus-Transformation.

Maßnahme. Alle Anteilswerte wurden arcussinus-transformiert um eine Normalverteilung der Residuen zu gewährleisten.

Ergebnisse. Es gab einen extrem starken Haupteffekt für den Faktor *Gender-Cue* ($F(1,19) = 91.71, p < .001, \eta_p^2 = .828$; $F(1,159) = 109.01, p < .001, \eta_p^2 = .407$). Der Haupteffekt für den Faktor *IC-Cue* war auch signifikant ($F(1,19) = 20.71, p < .001, \eta_p^2 = .521$; $F(1,159) = 14.05, p < .001, \eta_p^2 = .081$). Die Interaktion der beiden Faktoren war hingegen nur marginal signifikant ($F(1,19) = 3.22, p = .089, \eta_p^2 = .145$; $F(1,159) = 2.81, p = .096, \eta_p^2 = .017$): Wieder war der EHE des IC-Cues in der nicht-disambiguierten Bedingung stärker ausgeprägt als in der disambiguierten (Gender+: $F(1,19) = 8.09, p < .010, \eta_p^2 = .298$; $F(1,159) = 2.19, p = .141, \eta_p^2 = .014$; Gender-: $F(1,19) = 12.98, p < .002, \eta_p^2 = .406$; $F(1,159) = 10.06, p < .002, \eta_p^2 = .059$).

C.4.4: Verifikationsaufgabe. Mittlere Reaktionszeit. Nur korrekte Antworten.

Maßnahme. Um zu überprüfen, ob das berichtete Effektmuster durch evtl. Unterschiede zwischen falsch und richtig beantworteten Items moderiert wird, wurde die gleiche Analyse noch einmal nur mit den RT-Mittelwerten für korrekt beantwortete Items berechnet. In der by-Items-Analyse ergaben sich durch den Ausschluss für sechs Datenpunkte (=0.9%) fehlende Werte. Sie wurden jeweils durch den bedingungsspezifischen Mittelwert ersetzt.

Ergebnisse. Das berichtete Ergebnismuster konnte vollständig repliziert werden: Es gab einen Haupteffekt für den Faktor *Gender-Cue* ($F(1,19) = 22.35, p < .001, \eta_p^2 = .541$; $F(1,159) = 36.32, p < .001, \eta_p^2 = .186$) und einen Haupteffekt für den Faktor *IC-Cue* ($F(1,19) = 14.84, p < .001, \eta_p^2 = .439$; $F(1,159) = 15.88, p < .001, \eta_p^2 = .091$). Die Interaktion der beiden Faktoren war in Abweichung zur Hauptanalyse signifikant, allerdings nur in der by-participants-Analyse ($F(1,19) = 14.34, p < .032, \eta_p^2 = .219$; $F(1,159) < 1, p = .650$). Der Gender-Effekt war 130 ms und der IC-Effekt 80 ms groß.

C.4.5: Verifikationsaufgabe. Weibull-Parameter. 20 korrekte Antworten.

Maßnahme. Alle falsch beantworteten Items wurden von der Analyse ausgeschlossen. Von den verbleibenden Items wurden die ersten 20 Datenpunkte pro Person und Untersuchungsbedingung zur Analyse ausgewählt.

Durchschnittliche Parameter. Die Werte für die einzelnen Parameter fielen ähnlich aus wie bei der vollständigen Analyse (shift: 719 ms, scale: 692 ms; shape: 1.63).

shift-Parameter. Für diesen Parameter gab keine signifikanten Haupteffekte ($F < 1$). Lediglich für den Interaktionsterm deutete sich ein leichter Effekt an, der jedoch nicht signifikant war ($F(1,19) = 2.23, p = .152$). Tendenziell hatte der IC-Cue einen Effekt (von ca. 50 ms) auf den shift-Parameter, allerdings nur in der Bedingung Gender+. Auch dieser Einfache Haupteffekt war jedoch nicht signifikant ($F(1,19) = 1.79, p = .197$).

scale-Parameter. Es gab sowohl einen Haupteffekt für den Faktor *Gender-Cue* ($F(1,19) = 26.20, p < .001, \eta_p^2 = .580$) als auch einen Haupteffekt des Faktors *IC-Cue* ($F(1,19) = 6.53, p < .019, \eta_p^2 = .256$). Der Gender-Effekt war positiv und ca. 130 ms stark, der IC-Cue ebenfalls positiv und betrug ca. 90 ms. Die Interaktion der beiden Faktoren war nicht signifikant ($F < 1$).

shape-Parameter. Es gab einen leichten Haupteffekt für den Faktor *Gender-Cue* ($F(1,19) = 5.12, p < .036, \eta_p^2 = .212$) und einen marginal signifikanten für den Faktor *IC-Cue* ($F(1,19) = 3.03, p = .098, \eta_p^2 = .137$). Der Effekt des Gender-Cues war negativ und ca. 0.16 Einheiten groß, der Effekt des IC-Cues war hingegen positiv und betrug ca. 0.14 Einheiten.