Sentence Processing in a Second Language: Ambiguity Resolution in German Learners of English

Inauguraldissertation
zur Erlangung des Doktorgrades
an der Philosophischen Fakultät
der Universität zu Köln

vorgelegt von
Anne Rah
aus Aurich

am 06.04.2009

Gutachter:
Prof. Dr. Dany Adone
Prof. Dr. Christiane M. Bongartz

Disputation: 01.07.2009

Table of Contents

List of Tables

List of Figures

Acknowledgements

First of all, I would like to thank my supervisor, Prof. Dr. Dany Adone, for all her support and instruction, for always finding some time for me, even in the busiest of times, and for introducing me to the life as a linguist.

Many thanks also go to my team at the University of Cologne, above all Maxi Limbach, Antje Pillunat, and Susanne Mohr, for providing me with advice, support and much appreciated distraction. Apart from these fellow Ph.D. students, Kristina Trams, Rebecca Werner, Sarah Krauß, Alexandra Röllke, Sarah Wilewsiki, Katharina Haesemann and Emily Kaufmann have contributed very much to our team to create a fun and productive working environment. Furthermore, I have greatly benefited from the discussions with Iman Laversuch and the Ellipses-members, whom I would like to thank a lot.

Back in Göttingen, my particular thanks goes to Dr. Hildegard Farke, who really initiated my interest and experience in psycholinguistics and without whom I would probably not have come to Cologne to write this dissertation.

Parts of the research presented in this dissertation have been presented at the Second Language Research Forum 2007 at the University of Illinois at Champaign, at EuroSLA 2008 at the University of Aix-en-Provence, at the Second Language Research Forum 2008 at the University of Hawai'i at Manoa. I am particularly grateful to the audiences at these conferences for their helpful comments and discussions, and to the Deutsche Forschungsgemeinschaft (DFG) for the travel grants for my journeys to Champaign and to Hawai'i.

None of the research presented here would have been possible without the study participants who generously offered their time for the experiments. I would like to thank all these participants, and in particular the staff of the Junior Year Programme at the University of Bonn and the English Language Teacher's Association Rhine e.V., who helped me to find the native speaker control groups.

I owe many thanks to my friends and family for all their support and encouragement. But most of all I am grateful to my partner, not only for his love, care, and patience, but also his efforts to sometimes drag me away from my work. Finally, I would like to thank my parents for their constant support and encouragement, and for all they have done and are still doing for me.

1 Introduction

In research on second language acquisition (SLA), scientific interest has recently turned to language comprehension in real time. Within such a psycholinguistic approach to SLA, it is particularly interesting to examine the mechanisms that underlie sentence processing or parsing, i.e. the assignment of linguistic structure to a sentence.[1] The aim of the present thesis is to investigate whether there are fundamental differences between first language (L1) and second language (L2) sentence processing. To this purpose, various theoretical approaches are assessed on the basis of empirical data from German ESL (English as a Second Language) learners as compared to native speakers of English. The psycholinguistic data are from the processing of various syntactic ambiguities in off-line tasks, i.e. tasks without time pressure, as well as on-line tasks, which measure processing in real time. The two learner groups tested were at different proficiency levels but all learners had started learning their L2 'late', i.e. after childhood.[2] Furthermore, in contrast to other studies in the field, the present studies tested learners who did not acquire the L2 in an immersion context but at school by direct instruction. This choice of participant groups can provide insights into the following issues: (a) whether it is possible for non-immersed learners to acquire native-like processing mechanisms and (b) how L2 processing mechanisms develop with increasing proficiency.

The investigated structures are syntactic ambiguities, in which one element could structurally belong to two elements in the sentence. Although syntactic ambiguities are comparatively rare in everyday language use, they are widely used to explore processing mechanisms in general and the relationship between processing and mental grammar. One of the constructions under investigation is the so-called PP attachment ambiguity. An illustrative example is provided in (1). Here the PP with the binoculars could either attach to the verb saw or to the noun the boy:

(1) The man saw the boy with the binoculars.

[1] Note that in other works the term 'processing' refers to the segmentation of the speech stream into words and morphemes (e.g. Weissenborn & Höhle 2001). The present thesis focuses on the assignment of syntactic structure during (silent) reading.

[2] 'Late' L2 learners can be distinguished from 'early' learners, who have started to learn an L2 during childhood, and from 'bilinguals', who grew up with two languages simultaneously. The exact age lines are a matter of debate in the literature (e.g. Genessee & Nicoladis 2006 for an overview). It must be noted that some researchers (e.g. Fernández 2003:68) define 'bilinguals' more generally as all persons with two languages. This broader definition will not be adopted in the present thesis.

Another well-know syntactic ambiguity is the relative clause (RC) attachment ambiguity, which is illustrated in (2).[3]

(2) Someone shot the servant of the actress who was on the balcony.
   a. The servant was on the balcony.
   b. The actress was on the balcony.

In this example the RC could be attached to the first noun *the servant*, as in (2a) or to the second noun *the actress*, as in (2b), although in both readings the linear succession of words is identical. The ambiguity is especially intriguing because attachment preferences were found to vary cross-linguistically. For instance, native speakers of English were found to prefer attachment to the second noun (Cuetos, Mitchell & Corley 1996), whereas native speakers of German tend to prefer attachment to the first noun (Hemforth et al. 1998). These cross-linguistic differences are relevant for L2 research because they indicate whether learners transfer L1 processing strategies to their L2. Since this line of research could also provide information on the reasons for the observed cross-linguistic differences, research on L2 processing could help to obtain a better understanding of general processing mechanisms.

Late L2 learners' and native speakers' processing could differ in several ways. First, the learners' access to grammatical representations in real-time processing might be restricted (e.g. Clahsen & Felser 2006b; DeKeyser 2003; Paradis 2004; Ullman 2005). Second, L2 processing might be constrained by the unavailability or deficiency of the required processing mechanisms (e.g. Epstein, Flynn & Martohardjono 1996; Sorace 2006). Third, learners might have limited processing capacity, either as a consequence of particular L1-L2 combinations or as a general effect of L2 processing (e.g. Hopp 2006; McDonald 2006; Prévost & White 2000). These three accounts make distinct predictions for the patterns that might be observed in L2 processing studies. Under the assumption of learners' reduced access to grammatical representations, learners should rely comparatively more heavily on lexical and semantic than on syntactic cues (Clahsen & Felser 2006a,b). Alternatively, they might prefer short-distance over long distance relations. In case L2 processing is constrained by the unavailability or deficiency of processing mechanisms, learners would be predicted to employ similar default strategies. If learners' processing capacities are reduced in comparison to native speakers, learners would be predicted to be slower in real-time processing and to show difficulties to recover from initial misanalyses. Furthermore, reduced L2 processing capacity

[3] Since commas could disambiguate many syntactic ambiguities, they are omitted in most studies on syntactic ambiguity resolution. Accordingly, they are omitted here and in the following examples.

might result in difficulties to process syntactic dependencies or to consider several interpretations of an ambiguity in parallel. Under all three accounts it is also conceivable that learners transfer processing strategies from their L1 to their L2, thus relying on already familiar ways of processing syntactic structures. The present study tests the predictions of these accounts and thus seeks to gain a better understanding of the mechanisms that underlie L2 processing.

This line of research has implications for general SLA theories, considering that the ability to process linguistic input in real time is important for successful acquisition. Moreover processing failure draws attention to a specific structure and thus triggers the acquisition of that structure. Therefore, the way learners process the input determines how they access and acquire grammatical structures (e.g. Berwick & Weinberg 1984; Carroll 2001; Gibson & Wexler 1994; Schwartz & Sprouse 1996; Truscott & Sharwood-Smith 2004; White 1987, 2003). Accordingly, the question arises whether incomplete acquisition might be related to non native-like processing (Fernández 2003; Klein 1999; Juffs & Harrington 1995, 1996). A better understanding of how learners at various proficiency levels integrate words into a sentence and how they assign a structural and semantic organization to the input can thus contribute to a transition theory for SLA, i.e. a theory of how the L2 system develops (Gregg 2003; Juffs 2001).

The present thesis is organised as follows. Chapter 2 gives an overview of the current processing theories that account for native speakers' syntactic ambiguity resolution. It will be discussed whether the theories can convincingly account for the observed psycholinguistic data. Chapter 3 focuses on sentence processing in L2 learners. After introducing some underlying theoretical concepts that account for L1-L2 processing differences, the chapter presents previous behavioural studies on several syntactic ambiguities and discusses the theoretical implications of the respective results. Moreover, the studies are compared and evaluated with regard to their methodology and generalisability. The analysis of previous studies presented in Chapters 2 and 3 raises a few theoretical and methodological problems for behavioural psycholinguistic studies. Therefore, preceding the experimental part of this thesis, methodological issues concerning the choice of participants, materials, procedures and data analysis for the present studies are outlined and discussed in Chapter 4. Chapter 5, the first experimental chapter, presents one off-line and two on-line studies on RC attachment with two groups of German ESL learners. The results are interpreted with regard to L1 influence, L2-specific processing strategies and proficiency influence. Chapter 6 presents three studies on PP attachment.

Since PP attachment is not influenced by cross-linguistic variation in the same way as RC attachment preferences, these studies do not focus on L1 transfer effects but on structural influences on L2 learners' processing strategies. Chapter 7 focuses on attachment preferences in adjunct wh-questions. Since these structures involve syntactic movement, they are well adapted to test whether L2 learners generate deep syntactic structures in the same way as native speakers. The final chapter summarises and discusses the findings of the studies and their possible theoretical implications for the research areas of L2 processing, L2 acquisition and development, and monolingual processing. On the basis of this discussion, open questions for future research are outlined and a final conclusion is drawn.

2 Sentence processing models

In this chapter I introduce the basic concepts that are important for the present study. Since the main research question for the present study is how native-like L2 processing might be, theories of processing in the first language have to be considered as a theoretical background for the present study. Therefore, I describe various general processing models that have been developed to account for empirical data on syntactic ambiguity resolution. The models differ in their assumptions about the working mechanisms and the language universality of the parser. In the following, three types of models are distinguished: language universal, parameterised, prosody-based and exposure-based models. Before the three types of models are discussed, the first section looks at more general questions concerning the nature of the processing mechanism.

2.1 The interface between grammar and processing

The nature of the human language processing mechanism is difficult to investigate, considering that current technology does not permit us to look into the human brain and to directly identify the components which determine how language is represented in the brain and how it is put to use. However, behavioural data can provide insights into the nature of the parser. Testing the processing of syntactic ambiguities can be particularly revealing in this regard. When faced with ambiguous material, the parser opts for one initial analysis without considering all the available information, which can lead to misanalyses and conscious surprise effects, the so-called "garden-path" effects. These effects can be empirically measured by psycho- and neurolinguistic techniques. In self-paced reading tasks, for example, participants read sentences word-wise or chunk-wise and trigger the appearance of the following element by pressing a button (e.g. Just, Carpenter and Woolley, 1982, Marinis, 2003). In eye-tracking tasks, a high-resolution camera measures movements of the pupil as individuals read texts (e.g. Frenck-Mestre, 2005). Electromagnetic resonance potentials (ERPs) are recorded synchronously for the study of brain activity (e.g. Coles and Rugg, 1995; Hahne, 2001). In these measures increased processing difficulty is reflected in elevated reading times, longer fixations, or an ERP effect at the disambiguating region.

One central question about the nature of the human processing mechanism is the relationship between mental grammar and processing. On the one hand, grammatical knowledge influences processing, but on the other hand the way the grammar is constructed depends on how linguistic structures are processed. The scientific interest in processing arose relatively late in comparison to the strong interest in grammatical

knowledge in generative linguistic theory. The strict separation of the two lines of research goes back to Chomsky (1965: 3-15), who distinguished between competence and performance. The former refers to the language knowledge of an ideal hearer or speaker, whereas the latter refers to the concrete language use.

The prominence of competence in linguistics had the effect that many researchers equated the grammaticality with the processability of a sentence (e.g. Fodor, Bever and Garrett, 1974: 313). However, such an approach is problematic because ungrammatical sentences can be successfully processed, which implies that they can be assigned syntactic structure. Other evidence for dissociation between grammar and processing system comes from ambiguous sentences such as (1) (Bever, 1970).

(1) The horse raced past the barn fell.

Although grammatically correct, such a sentence is difficult to process because the verb *raced* tends to be initially interpreted as a main verb. This assumption subsequently proves incompatible with the final word *fell*, which turns out to be the main verb of the sentence. The processing difficulty on the main verb indicates that sentence processing is not only influenced by the grammar. At the same time, such examples show that the parser is influenced to some extent by the syntactic structure of a sentence because the simplest structure is preferred initially (e.g. Fodor, Bever and Garrett, 1974: 221-274 for further evidence). Chomsky (2000: 9) thus concludes that the grammar and the parser must be independent but related systems: "The faculty of language is embedded within the broader architecture of the mind / brain. It interacts with other systems, which impose conditions that language must satisfy if it is to be usable at all".

Most psycholinguistic models thus distinguish between mechanisms for sentence comprehension and production (i.e. performance systems) and an underlying system of linguistic knowledge (i.e. the competence system) (e.g. Crocker, Pickering and Clifton, 2000; Phillips 1996). The null assumption is the PIG (Parser Is Grammar) model (Phillips, 1996: 13), which assumes the grammar plus a finite set of resources such as working memory, past experience and world knowledge that the grammar uses. It is also conceivable, however, that the parser employs specific ambiguity-resolution strategies, phrase-boundary heuristics and other parser-specific principles. The question underlying this debate is whether human languages are optimal in the sense that they perfectly match the processing mechanisms (Chomsky, 1995: 9).

So far, I have treated 'the grammar' as one system. But given the existence of distinct modules of linguistic knowledge (Fodor, 1983), the question arises whether syntax

has a special role in sentence processing, in that syntactic principles apply before semantic and pragmatic constraints. Such a specialisation of modules would allow efficient processing at each level. Other models assume that the parser computes all types of information interactively (e.g. Swinney and Hakes, 1976; Traxler and Pickering, 1996; Tyler and Marslen-Wilson, 1977). In any case, adult native speakers were found to access distinct linguistic modules and to integrate both bottom-up and top-down information (e.g. Gibson and Pearlmutter, 2000; Gibson, 2006). With regard to the time course of processing decisions, it is relatively uncontested that the parser works incrementally, i.e. new material is integrated as soon as it is encountered. However, it remains a matter of debate whether the parser produces possible analyses serially or holds in mind various interpretations of an ambiguity in parallel (e.g. Gibson and Pearlmutter, 2000; Meng and Bader, 2000).

Another important question about the nature of the human parser that has implications for L2 processing is whether processing mechanisms are language specific or universal. Since language universal processing would be the most parsimonious option, the first processing models were language universal (see Abney, 1989; Frazier, 1978, 1987; Frazier and Fodor, 1978; Gorrell, 1995; and the discussion in chapter 2.1.1). However, this view was seriously challenged by a study reported by Cuetos and Mitchell (1988), which examined relative clause attachment preferences for English and Spanish with off-line and on-line tasks for sentences as the following:

(2) Someone shot the maid of the actress who was on the balcony.

Whereas the English participants preferred to attach the RC low, i.e. to the second of the two DPs[4], the Spanish participants preferred high attachment. The cross-linguistic differences observed by Cuetos and Mitchell (1988) were confirmed by studies on a variety of languages, tasks and materials. A general trend towards low attachment was observed in the following languages: Arabic (Abdelghany and Fodor, 1999; Quinn, Abdelghany and Fodor, 2000), Brazilian Portuguese (Miyamoto, 1998; but see Finger and Zimmer, 2000; Maia, Costa, Fernández and Lourenço-Gomes, 2002, who found a high attachment preference), English (Cuetos and Mitchell, 1988; Fernández, 1999, 2003; Frazier and Clifton, 1996; Gilboy, Sopena, Clifton and Frazier, 1995; Henstra, 1996;

[4] For the structural analysis of the sentences, I adopt the Government and Binding (GB) version of generative grammar. In line with the sentence-level phrase-structure assumption (Chomsky (1986a)), I analyse sentences as an I(inflectional) phrase (IP) dominated by a C(omplementiser) phrase. Moreover, I follow Abney (1987) in analysing the determiner of a noun phrase as the head of its own phrasal projection, the DP, taking the NP as a complement. It must be noted, however, that none of the conclusions I draw in this study crucially depend on these assumptions.

Mitchell and Cuetos, 1991; but see Carreiras and Clifton, 1993; 1999), Italian[5] (Baccino, De Vincenci and Job, 2000; De Vincenzi and Job, 1993; 1995), Japanese[6] (Kamide and Mitchell, 1997; Miyamoto, Nakamura and Takahashi, 2004), Norwegian / Swedish / Romanian (Ehrlich, Fernández, Fodor, Stenshoel andVinereanu, 1999). In contrast, a high attachment tendency was reported in the following languages: Afrikaans (Mitchell, Brysbaert, Grondelaers and Swanepoel, 2000), Bulgarian (Sekerina, Fernández and Petrova, 2004); Croatian (Lovrić 2002; Lovrić and Fodor, 2000), Dutch (Brysbaert and Mitchell, 1996; Frazier and Vonk, 1997; Mitchell and Brysbaert, 1998; Wijnen, 1998), French (Frenck-Mestre, 1997; Frenck-Mestre and Pynte, 2000; Pynte and Colonna, 2001; Zagar, Pynte and Rativeau, 1997; but see Baccino, De Vincenzi and Job, 2000), Galician (García-Orza, Fraga, Tejido and Acuña, 2000), German (Hemforth, Konieczny, Scheepers and Strube, 1998; Hemforth, Konieczny, Seelig and Walter, 2000; Hemforth, Konieczny and Scheepers, 2000a, 2000b; Konieczny and Hemforth, 2000; but see Murray, Rahman and Heydel, 2000), Greek (Papadopoulou, 2006; Papadopoulou and Clahsen, 2003, 2006), Hebrew (Shaked, Bradley and Fernández, 2004), Hindi (Vasishth et al., 2004, 2005), Polish (Nowak, 2000), Russian (Sekerina, 1997, 2004), Spanish (Carreiras and Clifton, 1993, 1999; Cuetos and Mitchell, 1988; Cuetos, Mitchell and Corley, 1996; Gibson, Pearlmutter and Torrens, 1998; Gilboy et al., 1995; Igoa, Carreiras and Meseguer, 1998; Thornton, MacDonald and Gil, 1999).

In addition to the variation between languages, conflicting tendencies were also found within one and the same language. German, for example, was first classified as high attaching (e.g. Hemforth et al., 2000a, 2000b), but a low attachment preference was reported recently (see Augurtzky, 2006; Murray, Rahmann and Heydel, 2000). Similarly, some studies did not find any preference in English (e.g. Carreiras and Clifton, 1993; Clifton, 1988, cited in Frazier, 1990), although English is generally classified as a low attaching language. Conflicting findings were also reported for Brazilian Portuguese, i.e. a low attachment preference by Miyamoto (1998) and a high attachment preference by Maia et al. (2004) and Finger and Zimmer (2001). These intra-language differences might be related to dialectal variation (Augurtzky, pers. communication; Corley and Cuthbert, 1996). Moreover, data from studies on Japanese (Kamide and Mitchell, 1997), Italian (Baccino, De Vincenzi and Job, 2000; Frenck-Mestre and Pynte, 2000) and Spanish (Fernández, 2003) suggest that preferences depend on the experimental method, as well as on the disambiguating region. For example, different results

[5] These studies report an initial low and a final high attachment preference

[6] These studies found an early low attachment preference and a later high attachment preference.

might be obtained for early vs. late disambiguation within the RC because post-syntactic factors influence the ultimate but not the initial preferences (see also Frazier, 1990, for a similar argument). Furthermore, the way of disambiguation might influence attachment preferences to some extent. Previous studies used disambiguation via gender, number or case agreement (Hemforth et al., 1998; Sauerland and Gibson, 1999), or via world knowledge (e.g. Carreiras and Clifton, 1993, 1999) (e.g. Fernández, 2003). In addition, the segmentation (e.g. Bradley, Fernández and Lovrić, 2003; Fernández, 2003; Gilboy and Sopena, 1996; Lovrić, 2002; Pynte and Colonna, 2001; Quinn, Abdelghani and Fodor, 2000; Vasishth et al., 2005; Wijnen, 2004) might influence participants' preferences. Finally, it has to be noted that preferences are often defined based on comparatively weak statistical effects.

Besides between- and within-language variation, inter-individual differences have been reported, for example with regard to working memory as measured by the reading span task (Daneman and Carpenter, 1980). This task measures the number of words presented in sentence contexts that a person can remember. In studies with English monolinguals, participants with a low reading span showed a high attachment preference, whereas high span participants tended to attach the RC low (Mendelsohn and Pearlmutter, 1999; Omaki, 2005; Swets et al., 2007; but see Hocking 2003, who did not find a relation of preferences and reading span, and Traxler, 2007, who found that low span readers preferred low attachment).

To sum up, the differences in RC attachment preferences that were observed between languages, within one and the same language, and between individual participants are difficult to explain within a comprehensive processing model. Especially the cross-linguistic variations are controversial because they contradict standard universal parsing models. There are different possibilities to account for these data:

1. The cross-linguistic differences are not due to syntactic but to pragmatic factors (Baccino, De Vincenci and Job, 2000; Frazier and Clifton, 1996)
2. There is no cross-linguistic variation; the observed effects are caused by differences in methodology, materials, or to a confusion of initial and final preferences (Gilboy, et al., 1995; De Vincenzi and Job, 1995; Fernández, 2003).
3. In addition to universal parsing principles, there are parameterised principles which vary cross-linguistically, as in Modifier Straddling (Cuetos and Mitchell, 1988), Anaphor Binding (Hemforth et al. 1998), or Recency and Predicate Proximity (Gibson et al., 1996).

4. Attachment preferences depend on prosodic cross-linguistic differences that apply even in silent reading (Fodor, 1998, 2002; Quinn, Abdelghany and Fodor, 2000).
5. Attachment preferences depend on statistical preferences within a specific language rather than universal principles as in Tuning (Cuetos, Mitchell, and Corley, 1996; Mitchell et al., 1995).

Hence, we can distinguish between accounts that maintain a language universal processing mechanism, parameterised accounts, prosody-based accounts and experience-based accounts of sentence processing. These four approaches will be discussed in turn in the following chapters.

2.2 Language universal models

The initial assumption in sentence processing research was that the human parser is the same across languages and that parsing routines are determined by a set of principles which are constrained by the grammar (see Abney, 1989; Frazier, 1978, 1987; Frazier and Fodor, 1978; Gorrell, 1995; Kimball, 1973) and by processing capacity restrictions. The principles assumed by the Garden Path Model (Frazier, 1978, 1987; Frazier and Fodor, 1978; Frazier and Rayner, 1982) and its successor Construal (Frazier and Clifton, 1996, 1997) are called Minimal Attachment and Late Closure and they require the most parsimonious phrase-structure that is possible without violating grammatical rules. Since the principles are assumed to be universal, no cross-linguistic transfer is predicted according to this model.

The Garden Path Model is seriously challenged by the cross-linguistic differences found for the RC attachment ambiguity (see chapter 2.1). Therefore, proponents of this model developed the Late Closure Theory (De Vincenzi and Job, 1993, 1995), which assumes that initial parsing is cross-linguistically guided by the syntactic Late Closure principle, whereas the final off-line interpretation is determined by additional pragmatic factors which may vary cross-linguistically. However, even a study which used exact translation equivalents for the tested languages (Fernández 2003) found a tendency towards high attachment for the Spanish participants and a tendency towards low attachment for the English materials. Moreover, the studies by Fernández (2003) and Bacciono, De Vincenzi and Job (2000) might have been biased towards low attachment by methodological factors: First, the studies used different nouns in the experimental conditions so that the conditions are difficult to compare. Second, the accuracy rates for the comprehension questions that followed the critical sentences in these studies were comparatively low, so that the statistical analyses are based on a comparatively small data

set. Third, the incorrect responses to these questions were not excluded from the analyses, so that the observed effects might have been biased by sentences that caused comprehension difficulties and possibly processing breakdowns. To conclude, although methodological differences between studies affect participants' attachment preferences, there are still too many on-line findings of cross-linguistic differences to simply attribute them to methodological noise. These results are difficult to explain within the Late Closure framework.

Another attempt to save the principles of the Garden Path Model resulted in Construal (Frazier and Clifton, 1996, 1997), according to which Minimal Attachment and Late Closure only operate on primary phases, i.e. phrases that depend on the main predicate of the sentence. All other phrases are optional and hence called non-primary phrases. These are associated to the other constituents through semantic and pragmatic principles. The integration of non-primary phrases does not occur in the initial parse as for the primary phrases. Instead, the attachment of these elements depends on thematic relations and discourse principles. The application of these may vary cross-linguistically, as in the case of RCs. In English, there is an alternative to the complex Norman genitive construction, i.e. the Saxon genitive (*the actor's servant* instead of *the servant of the actor*), which unambiguously requires attachment to *the servant*. The use of the Norman genitive would imply, following Grice's (1975) maxim of manner (i.e. avoid obscurity and ambiguity), that the RC is attached to *the actor* if possible.[7] For RCs, which are non-primary phrases, Construal predicts association to the current thematic processing domain, i.e. the maximal projection of the most recently processed theta-assigner. The model also offers an explanation for the cross-linguistic low attachment preference if the two DPs are linked by a thematic preposition (e.g. Gilboy et al., 1995; Traxler, Pickering and Clifton, 1998, Traxler et al., 2000, for English; Gilboy et al., 1995, for Spanish; Hemforth et al., 1998, for German; Frenck-Mestre and Pynte, 2001, for French; Baccino et al., 2000; De Vincenci and Job, 1993, 1995, for Italian).

Although Construal can convincingly explain many findings on RC attachment preferences, some issues remain unresolved. Languages like Afrikaans (Mitchell et al., 2000), Croatian (Lovrić, 2003), Dutch (Brysbaert and Mitchell, 1996; Frazier and Vonk, 1997; Mitchell and Brysbaert, 1998; Wijnen, 1998) and German (Hemforth et al., 1998, 1999), for example, have an alternative genitive construction, but speakers of these lan-

[7] Supporting evidence is provided by production studies (Costantino et al., 1999; Oria-Merino et al., 2000) in which the English participants preferred the Norman genitive when the relative clause was forced to attach low, while the Saxon genitive was preferred in the forced high attachment condition.

guages were found to prefer high attachment. Furthermore, languages like Brazilian Portuguese (Miyamoto, 1998) or Romanian (Ehrlich et al., 1999) show evidence of a low attachment preference without having a pre-nominal genitive like English.

Another problem for Construal is that it does not specify the time course for the attachment of non-primary phrases, which makes it difficult to formulate empirically testable predictions. Moreover, the distinction between primary and non-primary phrases is problematic. The classification of prepositional phrases, for example, seems to depend on the sentence context. It remains unclear whether the same PP can be primary in certain contexts and non-primary in others. This vagueness affects studies which test the predictions of Construal by comparing PP and RC attachment (e.g. Thornton, Gil and MacDonald, 1998; Traxler, Pickering and Clifton, 1998). Further specification of the distinction between primary and non-primary relations is needed in order to make empirically testable predictions. Furthermore, Construal cannot convincingly explain the observed cross-linguistic differences in RC attachment preferences.

2.3 Parameterised models

The cross-linguistic differences in RC attachment preferences suggest that processing principles might not be universal but language-specific. This idea is accounted for by so-called parameterised models like Anaphor Binding (Hemforth et al., 1998) or Recency and Predicate Proximity (Gibson et al., 1996). These models assume language universal processing principles the strength of which may vary cross-linguistically.

Anaphor Binding was developed within the framework of Parameterized Head Attachment (PHA) by Konieczny et al. (1994, see also Hemforth et al. 1998; Konieczny 1996). The model assumes that initial attachment decisions are guided not only by structural principles but also by the availability of lexical heads in the syntactic structure and by the lexical properties of these heads. Cross-linguistic differences in RC attachment are explained with language-specific properties of relative pronouns, which are anaphors and thus have to be bound by an antecedent. Since it is assumed that anaphors preferentially associate with salient discourse entities (e.g. Brennan, Friedman and Pollard, 1987; Garrod and Sanford, 1985; Grosz, Joshi and Weinstein, 1995; Sanford and Garrod, 1981, but see Wolf et al,. 2004), RCs are expected to attach high because the first DP is the internal argument of the verb and thus the more salient discourse entity. The strength of Anaphor Binding is assumed to vary cross-linguistically. In English, for example, anaphor binding is not very strong because relative pronouns are not obligatory and can be replaced by the complementiser *that*. Therefore, a low attachment prefe-

rence is expected because of the language universal principle to attach to the most recently processed head. In German, in contrast, relative pronouns are obligatory. Hence, Anaphor Binding is predicted to be strong, which accounts for the high attachment preference found in German.

This account correctly predicts cross-linguistic differences in RC attachment preferences for many languages, for example a low attachment preference in English, Swedish and Norwegian, in which RCs do not have to be introduced by a relative pronoun and a high attachment preference for languages with obligatory relative pronouns like German, Dutch and Russian. The predictions of this account also hold for other ambiguities. If a PP is attached instead of a RC, the model predicts low attachment because the PP does not include an anaphor. This prediction was tested and confirmed by Hemforth et al. (1998), who found a low attachment preference for German PPs in off-line and on-line tasks.

Nevertheless, there is some empirical evidence against this model, for example from Greek RC attachment preferences (Papadopoulou and Clahsen, 2003). Although Greek relative pronouns are complementisers as in English, Greek readers preferred high attachment. Conversely, in languages with obligatory relative pronouns such as Arabic (Abdelghany and Fodor, 1999) or Romanian (Ehrlich et al., 1999), a low attachment preference was found. Data from Spanish, Italian and Brazilian Portuguese are also problematic for Anaphor Binding. Although the three languages use similar complementisers to introduce RCs, Spanish readers showed a high attachment preference whereas low attachment was preferred in the two other languages. Apart from these problems with the experimental data, Anaphor Binding can be criticised for theoretical reasons. It is not entirely clear, for example, how the exact strength of Anaphor Binding in a language is determined. Moreover, generative syntax theory assumes a requirement for pronoun binding, even if a pronoun is not overtly realised (see Chomsky, 1981: 190; Chomsky, 1986b: 181ff.). In this case, an empty operator is assumed to occupy the structural position of the relative pronoun (i.e. the Spec-CP position). Since such an empty element is subject to binding requirements (e.g. Chomsky, 1981; Rizzi, 1990), there should be no difference between overt and empty anaphors.

Gibson et al. (1996) proposed an alternative account based on the parameterisation of syntactic principles. This account is located within the parsing model termed Syntac-

tic Prediction Locality (SPL) by Gibson (1998).[8] The model assumes two competing syntactic principles:

(4) Recency: Preferentially attach structures for incoming lexical items to structures built more recently.

(5) Predicate Proximity: Attach as close as possible to the head of a predicate phrase.

As in other models, Recency is assumed to result from the architecture of the human language parser and thus does not underlie cross-linguistic variation. In contrast, Predicate Proximity is parameterised, i.e. its strength varies across languages, depending on the average distance between a verb and its arguments in a language. This prediction is based on the assumption that the higher the distance between a verb and its arguments is, the more strongly the verb has to be activated throughout the sentence in order to control the attachment of its arguments. If the activation of the verb is strong, a violation of Predicate Proximity will be costly. Therefore, languages with comparatively high average distance between predicates and their arguments are expected to show a high attachment preference because of the relative strength of Predicate Proximity. Word order might be the decisive factor for the relative strength of the principle. Languages with a strict word order in which the arguments are close to the verb, such as SVO or OVS languages, are expected to show a low attachment preference. In contrast, for languages in which arguments can be far away from the verb, such as VOS, VSO, SOV or OSV languages, a high attachment preference is expected. Following these criteria, the model correctly predicts a low attachment preference for English as a strict SVO language and a high attachment preference for German, which has both SVO and SOV patterns and comparatively free word order. However, for Brazilian Portuguese as a language with rigid word order where the verb has to be adjacent to its objects, the model predicts a low attachment preference at all processing stages but recent studies (Finger and Zimmer, 2000; Maia et al., 2004) reported an ultimate preference for high attachment. Moreover, a study by Ehrlich et al. (1999) found an off-line low attachment preference in languages with flexible word order like Swedish, Norwegian, and Romanian, contrary to the predictions of the Recency and Predicate Proximity model.

[8] SPL focuses on the relation between the parser and the available computational resources. It is interactive in that it assumes that lexical, plausibility and discourse information contribute to the interpretation of the sentence, similar as in constraint-based models (see chapter 2.1.4). Furthermore, it is assumed that several possible interpretations can be held in memory and compared to each other.

The model can explain the low attachment preference for sentences in which the two DP attachment sites are linked by a lexical preposition like *with* with discourse structure (Gibson, 1998). Attachment therefore depends on the number of discourse referents between the RC and its potential attachment site. Contrary to genitives without prepositions and non-thematic prepositions, thematic prepositions introduce a new referent to the discourse. Therefore, the presence of a thematic preposition further increases the distance between the higher DP and the RC. A more serious challenge for the model arises from the observed differences between the attachment of PPs and RCs (see Traxler, Pickering and Clifton, 1998) because the model predicts similar preferences for both structures. Moreover, the cross-linguistic differences in RC attachment are problematic because Recency and Predicate Proximity should apply universally in a specific language. Another challenge for the model is a study by Hemforth et al. (1998) on German three-place attachment preferences. The participants' preference for attachment to the first DP contradicts Predicate Proximity that is predicted to apply in German because the first DP is the element with the highest distance to the predicate in the German verb-final construction. To sum up, although this account can convincingly explain many of the observed attachment preferences, the assumption that cross-linguistic differences in word order rigidity are the decisive factor for RC attachment preferences needs further support and refinement.

2.4 Prosody-based models

The Implicit Prosody Hypothesis (Fodor, 1998, 2002; Quinn, Abdelghany and Fodor, 2000) assumes parameterisation on a prosodic level. According to this model, processing is guided by prosodic factors, even in silent reading. The anti-gravity law (Fodor, 1998) predicts that prosodically heavier constituents tend to attach higher in the hierarchical sentence structure than lighter ones. Cross-linguistic differences in RC attachment preferences are attributed to prosodic differences across languages (Fodor, 1998; Jun, 2003). For prosodically light elements, an overall low attachment preference is predicted. Attachment preferences for prosodically heavy elements are assumed to depend on the weight of the constituent relative to its potential attachment sites. Prosodic breaks might also influence attachment preferences, especially in RC attachment ambiguities. A break preceding the RC is expected to trigger high attachment. If there is no break before the RC, the RC should attach to the second DP.

Two distinct bodies of evidence support the claim that attachment preferences are guided by Implicit Prosody. One body of studies investigated the effects of prosodic

manipulations on attachment preferences. The underlying assumption is that a short RC is preferably included in the preceding phrase. By contrast, a longer RC is more likely separated from the preceding material by a prosodic phrase boundary, as illustrated in (6) (Augurtzky 2006: 67).

(6) a. (DP1 of DP2) (long relative clause)

b. (DP1 of DP2 short relative clause) or: (DP1) (of DP2 short relative clause)

Evidence for such an effect is provided by Carreiras (1992) and Gilboy and Sopena (1996), who induced a prosodic break by inserting a comma between the second DP and the RC. Effects of length manipulations on RC attachment preferences were reported for a number of languages (see the studies on Arabic, Brazilian Portuguese, Croatian, Dutch, English, French, German, Hindi, Russian and Spanish reported in chapter 2.1.). Carreiras and Clifton (1999), however, report a high attachment preference for relatively short Spanish RCs like *que estaba enferma* (‘who was sick’).

The results of such studies suggest that length manipulations can explain some of the empirical findings on syntactic ambiguity resolution. But an account of RC attachment preferences must be able to explain the cross-linguistic differences that were found in previous studies. This issue is addressed by the second body of studies on prosodic influences, which focuses on the specific prosodic characteristics of a language. For example, Quinn, Abdelghani and Fodor (2000) investigated whether there are different prosodic default patterns in a high attaching language like French and low attaching languages like Arabic and English. They recorded native speakers of these languages reading RCs in forced high or low disambiguations. Whereas no cross-linguistic differences were found for short RCs, long RCs elicited distinct prosodic patterns in the languages tested. The first stress-bearing unit in the RC is prominent in French but not in Arabic for forced high and low attachment. In English, in contrast, the RC seems to be prominent only in forced high attachments. Although the two low attaching languages displayed slightly divergent patterns, the stress patterns were in line with the attachment preferences observed for the respective languages. The findings are thus in line with Implicit Prosody in that they suggest that prosodically prominent constituents are more likely to attach high than less prominent constituents. Recent studies on other languages report similar results (Edmonds, Killam and Liljestrand, 2006; Jun, 2003). However, it is unclear whether these patterns from reading aloud also apply to silent reading.

Although Implicit Prosody can account for some attachment phenomena observed in empirical studies, it has to be noted that this account might only be descriptively ade-

quate. The observed attachment preferences are in line with prosodic factors. However, it is conceivable that the preferences might be guided by underlying structural or discourse factors rather than by prosody. Moreover, the account cannot explain cross-linguistic differences found with materials that were matched for length and other prosodic manipulations (Cuetos and Michell, 1988; Fernández, 2003). More research in this comparatively new framework is necessary in order to obtain a clearer picture.

2.5 Experience-based models

In contrast to the previously discussed models, experience-based parsing models do not assume a set of fixed parsing principles. The parser is held to be statistics-driven, i.e. it keeps records of the way structures are resolved in a particular language. Ambiguities should thus be resolved in the way that is most frequent in that particular language. Whereas Tuning (Cuetos, Mitchell, and Corley, 1996; Mitchell et al., 1995) focuses on syntactic structures, more fine-grained models assume detailed statistical records which include lexical and contextual information. These latter models are also termed Lexicalist or Constraint-satisfaction Models (e.g. MacDonald, Pearlmutter, and Seidenberg, 1994; Trueswell, Tanenhaus and Garnsey, 1994).

Tuning predicts that parsing preferences correspond to the statistical patterns for the respective constructions that are found in corpora. Therefore, one way of testing this account is to compare corpus data with the results of psycholinguistic studies. A corpus analysis for RC attachment in English and Spanish reported by Mitchell, Cuetos and Corley (1992, cited in Cuetos, Mitchell, and Corley, 1996) revealed a low attachment preference for English and a high attachment for Spanish in the respective corpora, which is in line with the findings of psycholinguistic studies. Corresponding patterns for corpus analyses and behavioural data were also found in French (Balthazar and Kister, 1995; Mitchell et al., 1995; Pynte, 1998; Zagar, Rativeau and Pynte, 1997). However, some studies found a divergence between corpora and psycholinguistic data that contradicts Tuning. For example, Mitchell and Brysbaert (1998) discuss evidence from Dutch which indicates that low attachment of RCs is more frequent in corpora, whereas high attachment prevails in on-line data.[9] Although Desmet, Brysbaert and De Baecke (2002) relativise these findings by showing that the divergence between corpus and experimental data is partly related to lexical properties of the attachment sites,[10] some divergence

[9] It has to be noted that the comparison of experimental and corpus data might not be suitable to test a processing model. Corpus data might be influenced by discourse factors which are absent in single experimental sentences. Moreover, ambiguities are rare in corpora that include edited written texts.

[10] The authors predict a high attachment preference for two human DPs and if the first DP is human and the second non-human. If the DPs are in the inverse order or are both non-human, a low attachment preference is predicted. It must be noted

between the comprehension and production data reported by Mitchell et al. (2000) remains. According to the sentence completion and corpus studies reported by Desmet, Brysbaert and De Baecke (2002), sentences in which the first attachment site is non-human and the second is human should elicit a low attachment preference, but Mitchell et al. (2000) found a high attachment preference.

Further evidence against Tuning comes from constructions with three potential DP attachment sites (Gibson et al., 1996, for Spanish; Gibson, Schütze and Salomon, 1996; Gibson and Schütze, 1999, for English). For these constructions, on-line and off-line studies (Gibson et al., 1996; Gibson and Schütze, 1999) found a preference to attach to the third DP whereas the second DP was the least preferred attachment site. This result contrasts with the findings from English corpora, in which the third DP was also preferred and the first DP was the least preferred attachment site. Mixed evidence also comes from a study reported by Cuetos et al. (1996), which tested how Spanish seven-year-old children's attachment preferences were influenced by exposure to certain constructions. After a pre-test to determine the children's attachment preferences, one group was confronted with RCs disambiguated for high attachment; the other groups' materials were disambiguated for low attachment. In the latter group, the initial high attachment preference did not significantly change, whereas in the first group the initial high attachment preference became even stronger. Hence, despite a frequency effect in favour of the initial preference, the participants' preference did not change when they were exposed to materials that contradicted their initial preference.[11]

In addition to the empirical problems, the explanatory adequacy of Tuning is debatable. Although Tuning captures some empirical observations correctly on a descriptive level, there might be a deeper reason for the observed preferences. Thus, it is possible that speakers of a specific language show the attachment preferences that are most frequently employed, but that the reason for these preferences is not frequency but rather some structural, prosodic or discourse factor that varies cross-linguistically and is the underlying reason for the observed attachment preferences. In response to such criticism, Cuetos et al. (1996) acknowledge that Tuning accounts for rapid on-line processing decisions without considering any possible underlying grammatical properties that might influence these decisions.

---

that a fine-grained corpus analysis considering information below the sentence level deviates from the original assumption that coarse-grained analyses reflect comprehension results (Mitchell et al., 1995).

[11] The authors point out that the exposure to low attachment might not have changed the participants' preference because they were still exposed to everyday language, which favours high attachment. However, it is doubtful that seven-year-old children are confronted with many of these constructions in the one week that lay between the two points of testing. Thus, no clear conclusion can be drawn from these data, which shows how difficult it is to experimentally test effects of exposure.

Like Tuning, constraint-based parsing models (MacDonald, Pearlmutter and Seidenberg 1994; Trueswell, Tanenhaus and Garnsey 1994) ascribe a high importance to experience. In contrast to Tuning, however, they emphasise the importance of non-syntactic information. Three broad types of constraints can be distinguished: (a) lexical information (Garnsey et al., 1997; MacDonald, Pearlmutter and Seidenberg, 1994; Trueswell, Tanenhaus and Garnsey, 1994), (b) structural information (Juliano and Tanenhaus, 1994), (c) pragmatic / discourse information (Altmann and Steedman, 1988; Spivey-Knowlton and Sedivy, 1995). Although in principle various constraints are assumed to apply in parallel, some constraints might affect sentence interpretation earlier than others. These effects are attributed to locality and frequency of a form and to the nature of the activation and competition mechanisms (McRae, Spivey-Knowlton and Tanenhaus, 1998).

With regard to RC attachment preferences, MacDonald, Pearlmutter and Seidenberg (1994: 697-698) attribute cross-linguistic variation to the modification frequency of the nouns in the complex DP. If the first noun is modified more often than the second, this should result in a high attachment preference, and vice versa. This claim, however, was not supported by an English completion study (Corley and Corley, 1995), which experimentally manipulated the position of the nouns. This study found an overall low attachment preference that was independent of the respective positions of the two DPs. Furthermore, Gibson, Pearlmutter and Torrens (1998) compared Spanish sentences with two and three possible attachment DPs. Although the same DP occupied the lowest position in both conditions, there was a low attachment preference when there were three attachment sites and a high attachment preference when there were two possible attachment sites. An alternative explanation for cross-linguistic variation in RC attachment within the framework of constraint-based models is suggested by Thornton, Gil and MacDonald (1998) and Thornton, MacDonald and Gil (1999). These authors claim that attachment preferences depend on the degree of the nouns' availability for modification.12 If a noun has already been modified, it should not be as readily available for further modification as a noun that has not been modified (Thornton MacDonald and Gil, 1999: 1349). Evidence supporting this account comes from completion and self-paced reading tasks with Spanish and English participants, which found that the modifiability of the two DP attachment sites influenced PP attachment preferences. Since in

[12] This idea goes back to Referential Theory (RT) (Altmann and Steedman, 1988; Crain and Steedman, 1985), a discourse-driven processing theory which assumes three processing principles: (1.) the principle of a priori plausibility, which favours the most plausible reading, (2.) the principle of referential success, which favours attachment to an already established referent, and (3.) the principle of parsimony, which favours the reading that requires the fewest unsatisfied presuppositions.

the RC attachment construction the first DP is always modified by the second, a cross-linguistic low attachment preference is expected. The high attachment preference found in some languages is explained with the availability of an alternative construction like the Saxon genitive (see Frazier and Clifton, 1996, and the discussion in chapter 2.1.1).

2.6 Summary

To sum up the discussion of this chapter, none of the presented models could explain all the data on the processing of attachment ambiguities. Whereas most proposals successfully predict specific and isolated observations, none of them is able to explain the whole range of empirical findings on RC attachment ambiguities in numerous languages. This task becomes even more challenging because experimental data from different tasks and materials cannot be directly compared.

Despite these difficulties, the reported findings allow some conclusions about the mechanisms that guide attachment preferences. First, the findings suggest that non-syntactic factors influence initial attachment preferences. For example, the influence of the human vs. non-human DPs or the theta-assigning properties of the linking preposition on RC attachment preferences underlines the importance of discourse factors. However, these factors alone cannot explain the observed structural effects. Therefore, further research is necessary to determine to what extent structural, lexical and pragmatic factors interact. Second, methodological factors such as the length of the attached elements, the way of disambiguation and the requirements of the task were found to influence readers' attachment preferences. These findings underline the need for further empirical research which carefully controls for these factors in order to obtain new insights into the mechanisms that guide attachment preferences.

Research on L2 learners' processing is no less interesting but it is even more subject to experimental noise. Apart from methodological factors, learners' attachment decisions might be influenced by their first language or other foreign languages and by individual differences such as proficiency in and exposure to the respective languages. Furthermore, it is possible that learners do not employ the same mechanisms as native speakers. In spite of these complicating factors, L2 processing research is helpful because it can provide a better understanding of how learners access grammatical structures. Furthermore, this research can also shed new light on first language processing, for example by evaluating the predictions of universal, parameterised or exposure-based models. The following chapter addresses these issues by looking at previous research on L2 sentence processing of various types of syntactic ambiguities.

3 Sentence processing in the second language

During the past few years, an increasing number of psycholinguists have studied L2 sentence processing. The most important questions for this comparatively young field of research are to what extent L1 and L2 processing differ and which factors are responsible for these differences. Discrepancies between L1 and L2 sentence processing could be either transitional or fundamental. Transitional discrepancies, which are possibly related to a relative lack of processing capacity or automaticity, might simply disappear when the learners gain proficiency and security. Fundamental discrepancies, in contrast, would be related to diverging processing mechanisms and thus would not simply disappear over time. An example of such fundamentally different processing is transfer of ambiguity resolution preferences from the learners' L1 to their L2. It is also conceivable that learners are influenced by processing mechanisms which are independent of their L1 and their L2. Such L2-specific processing might entail that learners rely comparatively more heavily on lexical and semantic rather than syntactic cues (Clahsen and Felser, 2006,a,b). Alternatively, they might prefer simple short-distance relations or be unable to consider several possible interpretations of an ambiguity in parallel. All these mechanisms might compensate for a relative lack of (access to) grammatical competence, processing mechanisms, processing automaticity or processing capacity. Hence, the following questions have to be addressed by research on L2 syntactic processing:

1. Is L2 processing fundamentally different from L1 processing?
2. If L2 processing differs from L1 processing, what are the reasons for this difference?
3. Does consistent or prolonged exposure to L2 input help to develop native-like processing mechanisms?

The aim of the present study is to address these questions. Therefore, the following chapters consider theories and experimental data on L2 processing. Following an overview over theoretical approaches to the mental representation of a L2, I will introduce possible explanations for observed differences between L1 and L2 processing. More precisely, I discuss transfer from the learners' L1 to their L2 (chapter 3.2) and L2-specific processing (chapter 3.3) and their respective theoretical backgrounds. In chapter 3.4 I focus on an alternative explanation for possible differences between L1 and L2 processing, i.e. learners' limited processing resources, which might be related to interference of the L1, problems of lexical access or a relative lack of processing automaticity or working memory capacity. In chapter 3.5 I discuss possible age of acquisition and proficiency effects on L2 processing mechanisms.

3.1 The mental representation of a second language

Before addressing the question whether L2 learners employ processing mechanisms from their L1 when processing L2 input, I will consider the general question how learners' two linguistic systems are represented and controlled. The discussion will be limited to language comprehension (see Green 1998 and Poulisse 1997 on L2 production and Kroll & DeGroot 1997 on the bilingual mental lexicon). Advanced learners apparently master two interrelated linguistic systems. Studies found that learners can switch between their two languages in interactions with other learners (Bhatia & Ritchie 1998; Myers-Scotton 1993). They can talk about events or situations they have encountered in one language environment with speakers of their other language (Francis 1999; Francis, Romo & Gelman 2002; Verhoeven 1992). Provided they are sufficiently proficient in both languages, learners can also keep to one of their languages if this is pragmatically required, and they do not find it hard to suppress one language when using the other (Green 1986, 1998; Grosjean 1982).[13]

In L2 sentence processing the matter is less straightforward, also because it is still unresolved whether processing is language-specific or universal (see Chapter 2). This controversy is difficult to resolve on the basis of behavioural data.[14] Reading speed might be an indicator because the distribution of processing resources to two distinct parsers might slow down processing. However, the effect could also be reversed because one single parser might be overstrained and therefore less efficient. Moreover, a processing slowdown is not necessarily related to the nature of the parser. Instead, it might be caused by syntactic, lexical or pragmatic factors or by L2-specific processing.

Assuming that processing is language universal, it is possible that learners use the same mechanisms for their L1 and L2 or that L1 and L2 processing rely on different mechanisms. The first possibility in this respect is that learners only use their L1 for processing L1 input, whereas they employ specific mechanisms when processing L2 input. The second possibility is that learners have two identical parsers, which, however, would not be very economical and cannot explain language contact phenomena like code-switching or borrowing. For a language-specific parser, there are also two possibilities, i.e. either the same parser or different parsers are operative in L1 and L2 processing. If the same parser is used for L1 and L2 processing, a high amount of trans-

[13] Similar evidence comes from studies with bilingual children, who seem to be able to separate their two languages from an early age on (Paradis & Genesee 1996). This finding suggests that the two linguistic systems can be accessed independently.

[14] Evidence from neuroimaging studies has also been controversial. Whereas Kim et al. (1997) found that early and late L2 learners showed different patterns of activation in L2 processing, Illes et al. (1999) claim that the two languages of a person access a common semantic system. Neurolinguistic research on patients with localised brain damage provides some evidence that late learners have different morphosyntactic processing mechanisms for their two languages (e.g. Fabbro 2001).

fer from the learners' L1 to their L2 is predicted. Under the assumption of different parsers for the L1 and L2, it is possible that each of these parsers adopts the specific mechanisms for one of the two languages, which would lead to a strict separation between the two languages. It is also conceivable that learners still apply the parsing mechanisms for their L1 when processing input their L1, while they employ second-language specific mechanisms when processing input in their L2.

To sum up the discussion up to this point, L2 processing theories are influenced by theories of monolingual processing. To resolve the question of the universality of the human parser, research on L2 learners can provide interesting insights, but studies have to carefully consider the predictions of different theories for the respective set-up. Such research focuses on how cross-linguistic differences between the learners' L1 and L2 affect the learners' performance in the target language. Such "influence resulting from the similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired" (Odlin, 1989: 27) is also called transfer. Most previous L2 studies concentrated on vocabulary and syntactic transfer (e.g. Odlin, 2003, for an overview). Syntactic transfer is accounted for by the Full Transfer Full Access (FTFA) hypothesis (e.g. Schwartz and Sprouse, 1996; White, 2003), which claims that L2 learners have full access to UG and hence parameter resetting is possible in L2 acquisition (e.g. Epstein, Flynn and Martohardjono, 1996, 1998; Flynn, 1996; Flynn and Martohardjono, 1994, for a similar viewpoint). Another claim of the model is that there is full transfer in that the learner's L1 grammar, including the respective parameter settings, constitutes the initial state of L2 acquisition. Empirical evidence in support of this hypothesis comes for example from studies by Slabakova (2000) and Yuan (1998), in which beginning learners of different L1s behaved differently with respect to the investigated linguistic structures, i.e. reflexives and aspect.[15] These findings indicate that the learners' initial stages differed from each other. Although the model focuses on transfer of grammatical knowledge, it is also applicable to processing. Therefore, L2 learners should start out with the processing preferences from their L1, given that processing is assumed to be language specific.

3.2 Influence of the first language

Theoretical positions with regard to L1 transfer in L2 processing depend on the processing model that is adopted. In chapter 2.1 we have seen five different theoretical

[15] However, FTFA has also been challenged. For example, the Processability Hypothesis (Pienemann, 1998) claims that transfer is only possible if learners are able to process the respective constructions. Accordingly, no transfer is predicted in the initial state of L2 acquisition, but only in later states. Since FTFA and the Processability Hypothesis differ primarily in the predictions for beginning learners, the distinction is not in the focus of the present study.

positions on cross-linguistic variation in sentence processing. Processing models which assume language universal principles like the Garden Path Model (see chapter 2.1.1) do not predict transfer in L2 processing. However, revised versions of these models allow for some cross-linguistic variation and accordingly for L1 transfer in L2 processing. Parameterised models assume language universal processing principles of cross-linguistically varying strength. Hence, the strength of the principles in the learners' L1 would be expected to be transferred to the L2 according to such models. The same is true for prosodically and experience based models (see chapters 2.1.3 and 2.1.4).

An experience-based processing model that directly accounts for L1 transfer effects is the Competition Model (CM) (Bates and MacWhinney, 1987, 1989; MacWhinney, 1997). The authors describe syntactic processing as the assignment of form-function-mappings, which is guided by word order, agreement, animacy or case marking cues. The strength of these cues is assumed to vary cross-linguistically. According to the CM, the process of L1 or L2 acquisition comprises determining the cue strengths for the language that is to be acquired. In the beginning, L2 learners are predicted to transfer the cues from their L1 to their L2, before cue strengths change with increasing experience. Evidence for the CM mostly comes from off-line Agent-identification tasks, in which participants read sentences and are asked to identify the entity that performs the action (e.g. Year, 2003). Studies with L2 learners from different language backgrounds and at different proficiency levels found evidence for forward transfer from the L1 to the L2 (Harrington, 1987; Sasaki, 1994), backward transfer from the L2 to the L1 (Liu, Bates and Li, 1992), amalgamation of L1 and L2 cues (Hernandez, Bates and Avila, 1994) and differentiation of language-specific cue strengths (Liu, Bates and Li, 1992; Sasaki, 1994). Other studies found that increased use of the L2 enhances sensitivity to L2 cues (Kilborn, 1989; Kilborn and Cooreman, 1987; McDonald, 1987). However, the type of transfer found in these studies did not always correspond to the learners' proficiency level and to the quantity of input they had been exposed to. A study by Sasaki (1994), for example, found forward transfer in proficient Japanese learners of English, whereas even beginning English learners of Japanese already showed evidence for a differentiation of their two languages. Hernandez, Bates and Avila (1994) observed amalgamation in almost native-like Spanish learners of English, who would be expected to show differentiation at this stage of proficiency. One possible reason for such conflicting findings is that some cues might be more susceptible to transfer than others. Furthermore, the quality rather than the quantity of language exposure might play a role, so that individual differences are expected. Another problem lies in the methodology of these stud-

ies. Some of the investigated sentences are ungrammatical in one language and grammatical in the other. Since participants might be biased to accept sentences in grammaticality judgements, this might confound the results (see Gibson, 1992). In the light of the mixed evidence, it is unclear whether the CM can reliably account for L1 transfer effects in L2 processing. Furthermore, it must be noted that the tasks used in this research are generally off-line tasks and the results might not be valid for on-line tasks (MacWhinney, 2004: 132). Hence, the predictions of the CM do not necessarily hold for the on-line experiments of the present study.

The main implication of these theories for the present study is that many researchers assume some cross-linguistic transfer in L2 learners, which is expected to subside with increasing L2 proficiency and / or exposure. Although models accounting for transfer effects are often embedded within a specific framework, in principle any non-universal processing model can account for the transfer of attachment preferences and processing mechanisms from the learners' L1 to their L2.

3.3 A fundamental difference?

Differences between L1 and L2 processing might be more fundamental than assumed by the previously discussed transfer accounts, in that they do not simply subside with increasing exposure to and / or proficiency in the target language. In the present chapter I focus on two possibilities of second-language specific processing that are widely discussed in the literature.

The first possibility entails a preference to attach incoming material to the most recently encountered element of a sentence. Hence, the learners would over-rely on a structural principle like Late Closure (Frazier and Clifton, 1996, see also chapters 2.1.1) or Recency (Gibson et al., 1996, see also chapter 2.1.2). Although this idea has not been formulated in a comprehensive framework, it is conceivable that learners prefer attachment to the most recently encountered element, especially in complex structures and in demanding on-line tasks. The only study that found direct evidence for such a preference was the study on PP attachment by Frenck-Mestre and Pynte, (1997). The scarcity of evidence is partly related to the fact that native speakers show a low attachment preference in many constructions, which makes it difficult to empirically distinguish native-like from a second-language specific processing.

The most controversially discussed account of second-language specific processing is the Shallow Structure Hypothesis ('SSH') by Clahsen and Felser (2006b), according to which L2 morphological and syntactic processing is fundamentally different from L1

processing. Instead of analysing hierarchical syntactic structures when processing language input, learners are assumed to build up a shallow structure representation, which is based on lexical information such as verb argument structure as well as on pragmatic and contextual information. Although in principle L2 learners have acquired structural knowledge of the target language, this information is not accessible in initial processing. The pattern is reversed for children acquiring a L1, who are assumed to rely on structural information while ignoring lexical-semantic and contextual cues. Clahsen and Felser (2006a,b,c) argue that all individuals technically have access to the full and the shallow structure processing route. Hence, even native speakers of a language resort to shallow structures in certain situations.[16]

Some evidence in support of the SSH comes from ERP studies by Hahne (2001) and Hahne and Friederici (2001) with Russian and Japanese learners of German. In these studies, learners and native speakers showed similar negative peaks at around 400 ms after the stimulus onset (N400), which are usually associated with lexical-semantic effects. But the learners did not show native-like early anterior negativities (ELAN) associated with first-pass parsing (Hahne, 2001) or P600 effects that have been associated with later syntactic processes (Hahne and Friederici, 2001). Similar results were reported by recent neuroimaging studies.[17] For example, Wartenburger et al. (2003) studied the neural responses to syntactic and semantic-pragmatic violations in Italian learners of German at different proficiency levels and found effects of age of acquisition only in the syntactic but not the semantic condition. For syntactic judgements, the two late learner groups showed increased neural activity in language-sensitive areas, i.e. Broca's area and subcortical structures. The bilingual group did not show any language-specific differences in cerebral activation. Rüschemeyer, Zysset and Friederici (2006) scanned proficient Russian learners of German when reading German sentences with syntactic and semantic violations. Whereas no differences were found in activation loci or strengths between L1 and L2 participants in the semantic condition, the learners did not show the same differences in the superior temporal cortex between grammatical and ungrammatical sentences as the native speakers. Instead, these regions were activated

[16] This assumption is in line with L1 sentence processing models which assume two distinct processing routes, i.e. the full route which considers all available information, and the less complex heuristic processing route which may lead to incomplete but 'good enough' (Ferreira, 2003), 'pseudosyntactic' (Townsend and Bever, 2001) or 'underspecified' (Sanford and Sturt, 2002) representations (e.g. Christianson et al., 2006; Christianson and Slattery, 2005 for some empirical evidence).

[17] Neuroimaging techniques such as PET (Positron Emission Tomography) and fMRI (functional magnetic resonance imaging) measure changes in neuronal activity between two processing conditions as indicated by changes in blood flow to particular brain areas. Since changes of hemodynamic states take relatively long compared to the fast pace of speech or reading, the comparatively poor temporal resolution of these methods requires studying comparatively extended tasks. Nevertheless, these techniques allow the precise identification of the brain areas involved in processing.

irrespective of the grammaticality of the stimuli. This finding indicates that less proficient learners commit more resources to syntactic processing of even simple grammatical structures than native speakers. Taken together, the findings from neuroimaging and electrophysiological studies suggest that L2 learners' ability to make use of phrase structure information during L2 processing might be reduced in comparison to native speakers. However, such activation patterns do not provide conclusive evidence for shallow syntactic processing in L2 learners.

Direct behavioural evidence in support of the SSH is equally difficult to obtain. One line of research has focused on learners' sensitivity to extra-linguistic information that is predicted by the SSH. This prediction is supported by studies on subject-object ambiguities as in *After Bill drank the water proved to be poisoned* by Frenck-Mestre and Pynte (1997, Experiment 2), Juffs and Harrington (1996) and Juffs (1998b), which found that learners were sensitive to verb argument structure information. In an experiment on the same structures including an additional plausibility manipulation, Felser and Roberts (2004) found that advanced Greek learners of English were more strongly influenced by plausibility information than native speakers, and that they had more difficulty recovering from an initial misanalysis. Williams, Möbius and Kim (2001) also investigated L2 learners' sensitivity to plausibility manipulations, using a SPR task with sentences that involved adjunct extractions such as *Which friend did the gangster hide the car for late last night?*, where the fronted *wh*- phrase was either a plausible or an implausible object of the verb *hide*. The results indicate that the learners, like the native speaker controls, initially analyzed the *wh*- filler as the direct object of the verb, and that the subsequent reanalysis was more difficult if the *wh*- filler was a plausible direct object. However, the plausibility effects seemed to be more immediate for the native speakers than for the learners, indicating that the groups differed in their ability to recover from misanalysis.

Although these findings suggest that learners are sensitive to extra-linguistic factors, this is not necessarily evidence for a fundamentally difference between L1 and L2 processing. So in order to directly test the predictions of the SSH, we have to test learners' processing of structures that involve deep structural processing. The structures for which the SSH makes the strongest predictions involve syntactic dependencies, i.e. structural relations between two separate elements within a sentence. One such type of structure would be the so-called wh-dependencies illustrated below:

(1) [Which man]$_i$ did Mary kiss t_i yesterday?

Following the minimalist grammar framework (Chomksy, 1995; Radford, 2004), it is assumed that the wh-operator which man is moved out of the complement position of the verb into the SpecCP position, leaving behind a copy or a trace of itself in its original position. Although the psychological reality of these traces is still widely discussed in the L1 processing literature (e.g. Phillips, 2006, for an extensive review), cross-modal priming studies found reactivation of the operator at the gap position (MacDonald, 1989; Nichol and Swinney, 1989, but see McKoon, Albritton and Ratcliff, 1996). However, such an effect in a structure like (1) is not necessarily the result of a complex movement operation but could rather be related to direct lexical association between the operator and its trace. Structures that involve a so-called intermediate gap (see (2)) thus might provide more insights into structural processing.

(2) [Which man]$_i$ did Mary say t_i that she kissed t_i yesterday?

Even for native speakers, the processing of filler-gap dependencies is a challenge because a moved element such as the wh-operator *which man* in (2) must be kept in memory until it can be associated its subcategorizer. Hence, an increased distance between operator and trace should increase processing load (Gibson, 1998). The parser is therefore expected to exhibit a preference of short over long distance dependencies, making use of a principle like the Active Filler Strategy (Clifton and Frazier, 1989) or the Minimal Chain Principle (De Vincenzi, 1991). Although the results of many studies that provide evidence for the psychological reality of such intermediate traces of wh-movement (e.g. Nichol, Fodor and Swinney, 1994; Nichol and Swinney, 1989; Swinney et al., 1988) or of DP-movement (e.g. Bever and Sanz, 1997; MacDonald, 1989; McElree and Bever, 1989) could alternatively be explained by lexically driven filler-gap association (e.g. Pickering and Barry, 1991; Sag and Fodor, 1994), some recent studies from head-final languages (Aoshima, Phillips and Weinberg, 2004; Clahsen and Featherston, 1999; Fiebach, Schlesewsky, and Friederici, 2002; Nakano, Felser, and Clahsen, 2002) and English long-distance wh-movement (Gibson andWarren, 2004) provide evidence that L1 processing of filler-gap structures is driven by syntactic dependencies.

For L2 learners, such an assumption implies that they face various difficulties when processing filler-gap dependencies. First, if learners' processing capacity is reduced in comparison to native speakers (see chapter 3.4), they might have difficulties to associate the operator to its trace, especially when the distance between the two elements is long. In this case, filler integration is predicted to be delayed or even unsuccessful. Second, if learners do not process hierarchical structure information as it is assumed by the SSH,

they might try to lexically associate the operator and the trace without assuming structural mediators like intermediate gaps, or the association might be unsuccessful. Third, it is possible that transfer effects from the learners' L1 occur. Therefore, it would be particularly difficult for learners whose L1 is a wh-in situ language to successfully process filler-gap dependencies on-line.

Nevertheless, there are some conceptual problems with the SSH, as pointed out for example by Dekydtspotter, Schwartz and Sprouse (2006). First, they note that a fundamental difference between L1 and L2 processing would imply that "the entire relationship between language and other domains of cognition is fundamentally different" (Dekydtspotter, Schwartz and Sprouse (2006: 35). Second, the authors point to a potential learnability problem. Since the ability to process syntactic structures is crucial for the successful acquisition of these structures, it would be impossible for learners who only employ the shallow processing route to ever acquire grammatical structures. These criticisms show that further clarifications are needed to fully evaluate the SSHs. For example, Clahsen and Felser (2006a, b) do not specify in which situations or stages of acquisition learners apply shallow processing. Hence, shallow processing might be restricted to specific structures, to on-line processing, to inexperienced learners, or to learners who started to acquire their L2 after a certain age. In order to clarify these points, further research with a variety of linguistic structures and learner groups of various proficiency levels and language constellations is necessary. Furthermore, limited processing capacity might be one of the underlying reasons for shallow structure processing. Clahsen and Felser argue against such an explanation, pointing out that if processing capacity limitations were a problem, "we would expect L2 learners to prioritise on grammatical information in the same way as children do" (Clahsen and Felser 2006a: 28). However, this is not necessarily the case, considering that processing capacity limitations might lead native speakers and L2 learners to different compensation mechanisms, depending on the availability of lexical, semantic and pragmatic knowledge. Further research with structures of varying complexity might be fruitful in this respect. Another point that requires specification is whether the SSH assumes a qualitative or a quantitative difference between L1 and L2 processing. Although the differentiation between children acquiring their L1 and older learners acquiring their L2 suggests a fundamental difference between the two processes, the parallel that is drawn between learners' shallow and native speakers' 'good enough' representations suggests some continuity between L1 and L2 processing.

To sum up the findings on the L2 processing of filler-gap dependencies, there is some evidence that advanced learners try to associate the wh-element to its subcategoriser. It remains unclear, however, whether this is a purely lexical effect in the learners and a syntactic effect in native speakers, as it would be predicted by the SSH, or whether learners make use of the same syntactic processing mechanisms as native speakers. Another possible confound lies in the methodology of the respective studies. Whereas studies that used on-line reading methods combined with metalinguistic tasks like grammaticality judgements did not find fundamental differences between L1 and L2 processing, on-line studies which required reading for comprehension alone found that L2 learners process wh-dependencies differently than native speakers. This discrepancy suggests that learners are able to process wh-dependencies in a native-like way if they are forced to concentrate on the linguistic properties of the sentences but prefer shallow processing if they focus on contents. More experimental evidence is needed in order to further investigate this issue.

To sum up, it is conceivable that learners use specific processing preferences that are independent of their L1 and L2 tocompensate for a relative lack of access to grammatical representations, as suggested by the SSH, or to reduce processing load. The latter possibility is addressed in more detail in the following chapter.

3.4 Cognitive resource limitations

Cognitive resource limitations might be responsible for differences between L1 and L2 processing because learners are faced with several factors that induce additional processing load. Such factors might be parallel activation of several languages and inhibition of the L1. It is also possible that learners' access to the relevant grammatical representations or lexical knowledge is inhibited in comparison to native speakers. Evidence for possible cognitive resource limitations in L2 learners comes from the observation that non-native processing is slower than native processing. Learners manifest comparatively longer reaction times in SPR tasks (Juffs, 2001), in grammaticality judgement tasks (Bialystok and Miller, 1999; Mayberry and Lock, 2003; McDonald, 2000; Murphy, 1997), in lexical decisions (Scherag et al., 2004) and in conceptual processing (McElree, Jia and Litvak, 2000). However, these results do not necessarily indicate that L2 processing is qualitatively different from L1 processing (e.g. Frenck-Mestre, 2002; McDonald, 2006). This issue was tested explicitly in a study by Hoover and Dwivedi (1998). Although slower readers spent more time to process the end of a sentence, their on-line reading-time patterns did not significantly differ from those of

the faster readers. Conversely, fast L2 readers do not necessarily demonstrate native-like processing. This was shown, for example, in a study by Felser et al. (2003, see chapter 2.4.4), in which German learners of English read the experimental sentences faster than native speakers but nevertheless showed non native-like processing patterns.

Hence, although processing speed alone cannot explain all the differences observed between L1 and L2 processing, it might be an indicator of processing automaticity, which might be a more helpful concept for explaining non native-like L2 processing. Automaticity means that tasks do not involve attentional control, and less efficient, parallel or redundant processes are replaced by more efficient and integrative processes (e.g. DeKeyser, 2001; Segalowitz, 2003; Segalowitz and Hustijn, 2005, for an extensive discussion). Differences in automaticity of L2 versus L1 processing are mostly investigated by ERP studies because this method can provide insights into the time-course of processing. But the results are difficult to interpret, especially since L1 and L2 patterns might differ. ERP studies on L2 sentence processing (e.g. Müller, 2005, for an overview) found similar P600 effects as in native speakers. Peak latencies were sometimes delayed and the learners did not show the early anterior negativity (ELAN) that was found in native speakers. In neuroimaging studies, differences between adult learners and native speakers in strength and spread of qualitatively overlapping neural substrates point to limited efficiency and increased effort in L2 processing (e.g. Hasegawa et al., 2002; Stowe and Sabourin, 2005; Wartenburger et al., 2003).

There is also behavioural evidence that resource limitations constrain the automaticity of L2 sentence processing, even in highly proficient learners. For instance, differences between off-line and on-line processing indicate that learners have difficulties to apply their grammatical knowledge in real time, or in situations of increased processing load. Evidence for such off-line / on-line asymmetries comes from reading studies testing grammatical agreement by Emmorey et al. (1995) with learners of American Sign Language and Jiang (2004, 2007) with Chinese learners of English. In both studies, some of the learners showed a lack of sensitivity to morphosyntactic information in on-line processing whereas they demonstrated native-like mastery of the same phenomena in corresponding off-line tasks.

Other studies found evidence for non-native like L2 processing by manipulating task demands. In study reported by Kilborn (1992), for example, German learners of English had to identify target words embedded in sentences in three conditions: (a) ‘normal’ sentences that included syntactic and semantic information, (b) ‘syntactic’ sentences in

which all lexical words were replaced with random words of the same word class, so that syntactic relations were maintained but semantic information was lost, and (c) random strings of words without syntactic or semantic relations. In contrast to the control group, who used all types of information, the learners showed facilitation for syntactic information compared to random sentences, but no significant further processing benefit for additional semantic information. Interestingly, the native speakers' patterns under increased demands (i.e. white noise) resembled the learners' patterns. Similar findings are reported by McDonald (2006) from an off-line grammaticality judgement task with intermediate to advanced learners from various L1 backgrounds. The learners' judgement accuracy was significantly correlated with lexical decoding ability and working memory ('WM') span. The native speaker control group, who performed at ceiling level in the unstressed conditions, paralleled the learners' pattern of decreased judgement accuracies under distracters like concurrent memorisation tasks, white noise, stimuli in compressed speech, and response time pressure.[18] Another study that compared learners with native speakers under increased task demands is reported by Hopp (in press). The study tested case marking, subject-verb agreement and gender concord in English, Dutch and Russian advanced to near-native L2 learners of German with different off-line and on-line tasks. The results show that native-like ultimate attainment of L2 inflection in off-line and on-line processing is possible for late learners. Non-target-like L2 inflection was found to be systematically related to transfer from the L1 and reduced processing efficiency in the L2. Moreover, in an auto-paced reading task with increased presentation rate, native speakers showed similar accuracy patterns as the learners under normal presentation speed.

Taken together, the findings indicate firstly that L2 learners' off-line and on-line performance is influenced by resource limitations and secondly that additional processing load in native speakers can lead to L2-like performance patterns. Hence, L2 processing mechanisms were similar to those employed by native speakers under increased task demands. The comparatively lower automaticity in L2 processing points to restrictions in the computational resources available to L2 learners. However, studies investigating the grammatical performance of native speakers under memory load obtained heterogeneous results. Whereas Blackwell and Bates (1995) found such an effect in a grammaticality judgement task, Dick et al. (2001), Waters, Caplan and Rochon (1995) and Wa-

[18] It has to be noted, however, that an account based on processing difficulty alone cannot explain the differences between distinct morphosyntactic phenomena in this study. Furthermore, the learner group was quite heterogeneous, and the structures tested were not matched for number of items, length or the point of the sentence at which the violation occurred. These methodological shortcomings question the generalisability of the results.

ters, Caplan and Yampolski (2003) did not find any WM effects in different types of comprehension tasks. Thus, WM capacity may be more crucial in grammaticality judgements than in comprehension tasks. Furthermore, the results by Blackwell and Bates (1995) and McDonald (2006) indicate that the importance of WM capacity might depend on the grammatical construction under investigation. These findings suggest that the processing difficulties observed in L2 learners are not necessarily related to limited resources but to difficulties in allocating the resources efficiently (e.g. Hopp, 2006; Wilson, Sorrace and Keller, 2009).

The assumption of either reduced resource limitations or reduced efficiency in controlling these resources in L2 learners is in line with capacity models of L1 sentence processing such as the Late Assignment of Syntax Theory (Townsend and Bever, 2001), the model of 'good enough processing' (e.g. Ferreira, 2003; Ferreira, Ferraro and Bailey, 2002; Ferreira and Patson, 2007) or the underspecification account (Sanford and Sturt, 2002). According to these models, native speakers have access to a full and a reduced processing route, depending on task demands and WM capacity (Christianson et al., 2006). If the application of 'good enough' instead of full representations is to some extent dependent on computational load, they are expected to be more frequent in learners than in native speakers.[19]

Another factor that has been investigated with regard to cognitive resource limitations is whether learners' WM capacity is reduced in comparison to native speakers. This line of research is based on the assumption that learners with higher WM capacity are able to store and analyze more chunks of language and, hence, are able to acquire phonological and morphosyntactic competence more easily than learners whose capacity is lower (Ellis, 2001). Generally, memory is referred to as the ability to store, arrange and recall received information and stimuli via the neural system, either in a conscious or in a subconscious, automatic way (Baddeley, 2007; Just and Carpenter, 1992). Standard memory models (e.g. Fabbro, 2001: 97) distinguish short-term or working memory and long-term memory. The function of short-term memory is to temporarily contain and process information. Since it is mainly responsible for reasoning, comprehension, learning and consciousness, the question arises to what extent it is involved in L1 and L2 acquisition. This line of research focuses on the phonological WM by measuring a person's capacity to remember a series of unrelated words or non-words of different length. Corresponding studies provided extensive evidence that phonological

[19] Interestingly, this corresponds to the predictions of the Shallow Structure Hypothesis, albeit for different reasons (see Clahsen and Felser, 2006b: 34, and chapter 3.2).

WM plays a role in learning new words in children and adults (e.g. Baddeley, Gathercole and Papagno, 1998; Ellis, 2001; Gathercole, 2006; Papagno and Valar, 1995; Service et al., 1992). The second approach investigates reading span, i.e. the simultaneous capacity of processing and storing information. This type of memory capacity is generally measured with the reading span task (see Daneman and Carpenter, 1980), in which participants read lists of sentences aloud and then to recall the final word of each sentence without explicit rehearsal.

The role of WM in native speakers' sentence processing as measured by reading span tasks is highly contested in the literature. According to the parsing model developed by Just and Carpenter (1992) individual differences in WM capacity influence processing in that high-span readers compute multiple analyses of ambiguous structures and keep them in memory, whereas low-span readers quickly commit to the simpler analyses (but see Caplan and Waters, 1999; Waters and Caplan, 1996a). ERP studies on syntactic ambiguity resolution by Friederici, Steinhauer, Mecklinger and Meyer (1998) and Bornkessel, Fiebach and Friederici (2004) also found qualitative and quantitative differences between high-span and low-span native speakers. Whereas the former showed P600 effects for syntactic reanalysis in ambiguous sentences, the latter showed N400 effects. This discrepancy might be interpreted in such a way that the low-span participants do not compute syntactically based reanalyses (see also Vos, Gunter, Schriefers and Friederici, 2001). Similarly, an ERP study on German wh-questions by Fiebach, Schlesewski and Friederici (2002) found a sustained left anterior negativity in object-extractions. Since the effect only appeared in questions with long filler-gap distance and was modulated by individual WM capacity differences, the authors interpret it as an indicator of memory processes required for maintaining the dislocated object in memory. For L1 processing, inter-individual differences thus might be partially related to limited processing resources. This finding is confirmed by a recent off-line questionnaire study by Swets et al. (2007), who found that WM influenced native speakers' RC attachment. Interestingly, their high span participants showed a stronger preference for low attachment than the low span participants. The authors interpret the result in such a way that individuals with high WM capacity divide the sentence into larger prosodic chunks, which results in a low attachment preference (see chapter 2.1.3). Traxler (2007), however, found exactly the contrary in an eye-tracking study. Low span readers had the strongest low attachment preference, possibly because they prefer attachment to the structurally closest element in the sentence. Taken together, the results of these studies suggest a relationship between WM capacity and processing performance in L1

reading, in such a way that low span individuals have more difficulty with long-distance relations within a sentence. The consequences of WM differences depend on the specific structures and tasks involved.

Assuming that L2 sentence processing implies a high burden on the processing mechanism since retrieval of lexical and grammatical knowledge is more difficult, it is possible that individual differences in WM capacity influence learners' processing. Thus, learners might behave like low span native speakers. It is also conceivable that learners behave native-like only if their WM capacity is sufficiently large, or if they are highly proficient in their L2. With regard to off-line processing, positive correlations have been reported between reading span scores and performance in comprehension (e.g., Harrington and Sawyer, 1992) and grammaticality judgment tasks (Robinson, 2002). In studies investigating reading span differences between native speakers and learners, learners generally achieve lower reading spans in their L2 as compared to their L1. Hence, reading spans for high span learners overlap with low span natives, and not with high span natives (Felser and Roberts, 200720; Juffs, 2005; McDonald, 2006; Omaki, 2005; Williams, 2006). Other studies investigating reading span influence on L2 processing obtained more heterogeneous results. Whereas global proficiency (Harrington and Sawyer, 1992; Miyake and Friedman, 1998), lexical retrieval and vocabulary use (Kroll, 2006) as well as text comprehension (Walter, 2004) were found to be affected by inter-individual reading-span differences, no robust effects of WM on L2 syntactic processing have thus far been attested (e.g. Juffs, 2004, for an overview). This might be at least partly attributable to methodological problems because the studies employed different WM tests or different scoring procedures (e.g. Conway et al., 2005; Friedman and Miyake, 2005; Waters and Caplan, 2004, for a discussion). Besides, it is debatable whether these tests measure the WM capacity that is needed for syntactic processing.[21] Moreover, task demands seem to affect the influence of WM, as suggested by a SPR study investigating the online processing of subject-object ambiguities in Dutch RC constructions (Havik et al., 2009). The results suggest that the high span learners patterned like the low span native speakers only when task demands were high. However, when reading for comprehension alone and when only structural information

[20] Note, however, that the learners in Felser and Roberts (2007) are compared with native speakers in Roberts, Marinis, Felser and Clahsen (2006), who took a different reading span test. Hence, comparisons between these studies have to be considered with caution.

[21] For example, MacDonald and Christiansen (2002) suggest that working memory effects might be related to language experience and exposure to text (e.g. Acheson, Wells and MacDonald, 2008; but see Just and Varma, 2002). Furthermore, working memory capacity seems to covariate with factors such as lexical decoding skill, print exposure, and ability to shift attention (Engle, 2002; La Pointe and Engle, 1990).

was available, the learners' processing differed from the native speakers' processing independent of WM span. To investigate to what extent cognitive resource limitations influence RC attachment preferences, Omaki (2005) conducted a SPR task and an additional reading span test with Japanese learners of English. In contrast to the previously reported SPR tasks that presented the sentences phrase by phrase, Omaki chose a word-by-word presentation, which might increase WM load. Although no relationship was found between reading span and attachment preferences in the learners, the study provides tentative evidence that the learners' attachment preferences resemble those of low-span native speakers of English. L2 learners' non-native processing patterns might therefore be related to processing capacity limitations. However, further research is necessary in order to resolve this issue.

Thus it is not clear whether WM capacity as it is measured in current reading-span tests has an influence on learners' attachment preferences that is measurable in behavioural studies. There is some evidence that resource limitations are at least partly responsible for differences between L1 and L2 processing. However, it remains open whether these differences are related to parallel activation and inhibition by the L1, to higher processing load, to reduced automaticity or to an interaction of all or several of these factors. Thus, future studies must address this issue by manipulating one or several of these variables in order to find out more about how and to what extent cognitive resource limitations influence L2 processing. Furthermore, it is probable that the degree of WM effects is at least partly dependent on task demands as well as on learner-internal variables such as maturation and proficiency.

3.5 Age of acquisition and proficiency

It is conceivable that differences between L1 and L2 processing are related to maturational changes so that age of acquisition plays a crucial role for successful L2 acquisition. This idea was first developed in the Critical Period Hypothesis (Lenneberg, 1967; Penfield and Roberts, 1959).[22] The concept was resumed and empirically supported in the 1990ies by Newport and colleagues (e.g. Johnson and Newport, 1989). According to a strong version of this hypothesis, the neural system supporting L1 processing becomes unavailable for late L2 learners as a consequence of neurocognitive changes. This assumption would imply that late L2 learners recruit different neural regions and distinct neurophysiological mechanisms in language processing as compared to native speakers.

[22] Evidence for critical periods in L1 acquisition mainly came from feral children (e.g. Curtiss, 1988), congenitally deaf children acquiring sign language later in life (e.g. Mayberry and Lock, 2003), and differences in auditory discrimination abilities of non-native phonemic contrasts in children at different ages (e.g. Werker and Tees, 2005).

Hence, native-like language acquisition would be impossible for learners who started learning a L2 after the critical period. Given a modular organisation of language, there might be different critical periods for linguistic subcomponents, with some subcomponents being unaffected by critical periods and others being subject to critical periods of distinct onsets and offsets (e.g. Long, 2005; Singleton, 2005). The acquisition of grammar has often been claimed to be affected by critical period effects (e.g. Newport, Bavelier and Neville, 2001). It is conceivable that such effects also appear in grammatical processing (Ullman, 2004; Wartenburger, et al. 2003).

Evidence for such a maturational account is provided by studies on RC attachment by Dussias (2001, 2003) and Fernández (2003) (see chapter 2.4.4.1), which found differences between early and late L2 learners. However, it has to be noted that the choice of age lines between the learner groups was rather arbitrary, which makes the results difficult to interpret. Moreover, the finding of Dussias (2001) that some of the late learners were able to adopt native-like processing for their L2 challenges her overall conclusion of a fundamental difference between early and late learners. Furthermore, recent neuroimaging and electrophysiological studies (e.g. Indefrey, 2006; Stowe and Sabourin, 2005, for neuroimaging studies; and Hahne, 2001; Ojima, Nakata and Kakigi, 2005; Rossi et al., 2006, for electrophysiological studies) found that highly proficient late learners showed native-like automatic brain responses in the L2 processing of syntax and semantics and activated similar language areas as the native speakers. Thus, it seems to be possible for learners to reach native-like processing performance in their L2. Age of acquisition effects in the reported studies were gradual, starting in early childhood and continuing beyond the end of puberty (Abutalebi, Cappa and Perani, 2001; Pallier, 2007). These results are in line with a processing account that is not based on a critical period, but rather on influence of the first langage. Learners with higher ages of acquisition had increased exposure to their L1, which might constrain L2 processing with regard to lexical retrieval (Scherag et al., 2004), grammatical access (McDonald, 2006), and phonological decoding (Flege, Yeni-Komshian and Liu, 1999; Mayberry and Fischer, 1989).

Whereas L2 processing develops gradually without being impeded by a critical period of language acquisition, it is not identical to L1 processing, especially in less proficient learners. Therefore, other accounts assume two distinct processing routes for accessing grammatical knowledge (DeKeyser, 2000; Ellis, 2002; Jiang, 2004; Paradis, 1997, 2004; Ullman, 2005). The first type of knowledge is conscious and subject to at-

tentional control, whereas the second type is unconscious and automatic. Although these models use different names for these types of knowledge, such as explicit versus implicit knowledge (DeKeyser, 2000, 2003; Hulstijn, 2002; Paradis, 2004), non-integrated versus integrated knowledge (Jiang, 2004, 2007), or declarative versus procedural knowledge (Ullman, 2005), their basic assumptions are similar. Concerning L1 versus L2 syntactic processing, these approaches assume that L1 processing is mostly automatic, and is based on implicit, integrated knowledge. In contrast, L2 processing is restricted to the use of explicit, non-integrated knowledge. Nevertheless, learners can achieve relatively high levels of L2 proficiency and fluency.

As an example of such a model, the Declarative-Procedural Model (Ullman, 2001, 2004) is introduced in some detail. The model distinguishes the declarative memory system, which stores memorised words and phrases, from the procedural system, which processes combinatorial rules of language. It is assumed that late learners over-rely on the declarative system, even in sentence processing, where native speakers use the procedural system. This qualitative difference in syntactic processing might be related to a weakening of the latter system in the course of maturation. The maturational changes are assumed to occur during childhood and adolescence, possibly triggered by increasing estrogen levels in both genders (Ullman, 2005). The predictions of the model for L2 sentence processing resemble those of the Shallow Structure Hypothesis. Both accounts predict that learners have no difficulty with lexical-semantic and plausibility information during processing. In contrast, they are expected to under-use rule-based syntactic information. In the case of simple syntactic structures, it cannot be excluded that associative memorisation controlled by the declarative memory system leads to successful syntactic processing. But this processing route would still be predicted to be less automatic than the native-like application of the rule-based grammar system controlled by procedural memory. With regard to the development of native-like L2 processing, the Declarative-Procedural Model assumes that the procedural system becomes less available in the course of maturation. However, the model needs further specification concerning the availability of the procedural system to late L2 learners at different proficiency levels. Moreover, an fMRI study with learners of German by Rüschemyer et al. (2005) found high levels of activation in frontal brain structures that are generally associated with procedural processing. Hence, the learners' significantly lower off-line judgement scores in comparison to the native speaker control group in this study cannot be attributed to the learners' under-reliance on procedural processing. Furthermore, the localisation assumed by the Declarative-Procedural Model could not be confirmed by

brain imaging studies which found that overlapping cortical regions were involved both in L1 and L2 processing (Indefrey, 2006; Perani and Abutalebi, 2005).

Thus it is conceivable that there is no fundamental maturational difference between L1 and L2 processing. Instead, L2 processing might develop gradually towards more native-like patterns with increasing proficiency. For example, Osterhout et al. (2006) report corresponding effects in English learners of French after only several months of exposure. Both semantic and morphosyntactic violations in subject-verb and determiner-noun agreement were presented to the learners at one month, four and eight months of university classroom instruction in French. Semantic violations were found to elicit N400s after one month of instruction. At the same time point, the learners evinced N400s in response to morphosyntactic violations. This indicates that they differentiated between well-formed and ill-formed syntax, albeit not in a native-like manner. After four months, the N400 was replaced by a native-like P600 as it is normally found for syntactic violations. A neuroimaging study on native Mandarin Chinese speakers acquiring Dutch reported by Indefrey (2006) provides evidence that learners activate syntax-specific brain regions that are involved in L1 processing after only a few months of exposure to the target language. The reported studies all focus on comprehension in natural languages, but similar results were obtained for artificial language learning (Opitz and Friederici, 2003) and sentence production (Golestani et al., 2006). Moreover, proficiency effects on processing patterns were also observed in more proficient L2 learners (Hahne, 2001; Ojima et al., 2005; Rossi et al., 2006).

Similar evidence comes from behavioural studies. For example, in a study on the RC attachment ambiguity with English learners of French, Frenck-Mestre (2002) found that only the less proficient learners showed the same low attachment preference as in their L1, whereas the more proficient learners showed a native like high attachment preference. Similarly, a study on ultimate attainment with regard to morphosyntactic phenomena like case marking, subject-verb agreement and gender agreement, Hopp (to appear) found that it was possible for late L2 learners of different L1 backgrounds to attain native-like proficiency, both in terms of grammatical knowledge and processing.

The reported studies thus suggest that proficiency might be more decisive for the development of native-like processing than the age of first exposure to the L1. Such a view implies that the difference between L1 and L2 processing is gradual rather than fundamental. A theory of L2 processing thus has to account for a gradual development between stages of proficiency, be it towards less L1 influence, towards deep rather than

shallow syntactic processing, more involvement of the procedural memory system or more automatic processing.

3.6 Summary

In the previous sections, we have seen that L2 processing theories make different assumptions about learners' and native speakers' processing. These differences are either attributed to L2-specific processing, to L1 interference or to reduced processing capacity or automaticity. With regard to L2-specific processing, it is possible that learners do not generate deep syntactic structures when processing L2 input on-line or that they prefer structurally close syntactic dependencies. It is also conceivable that learners are to some extent influenced by preferences from their L1, which however might be a default adopted by the learners because of shallow syntactic processing or a relative lack of processing capacity.[23] Cognitive resource limitations in L2 learners might be either directly responsible for differences between L1 and L2 processing, or they might indirectly lead to L1 transfer or to second-languagespecific processing. Concerning the nature of these processing limitations, it remains open whether they are related to the concept of short-term WM or to the long-term procedural WM system. Furthermore, L2 learners' processing difficulties might result from restricted access to lexical or grammatical representations, or from parallel activation of and inhibition by the learners' L1. As these summarising observations show, it is well possible that various factors are interrelated in non-native like L2 processing. This possibility makes it difficult to experimentally test the theoretical accounts. Another problem is that "the mere fact that there is an observed non-isomorphy between natives and learners does **not** entail that the natives and the learners deploy fundamentally different mechanisms" (Dekydtspotter, Schwartz and Sprouse, 2006: 33). Moreover, the inconclusive evidence from previous studies is at least partly attributable to differences in participants' age, proficiency, amount of L2 exposure and age of first contact with the L2 across studies. This heterogeneity makes it difficult to compare the results of the studies and almost impossible to draw conclusions about learners' processing development over time. The present studies try to overcome these difficulties by systematically testing and comparing the off-line and on-line processing behaviour of learners at two proficiency levels with different but related syntactic constructions.

[23] Although L2-specific processing or transfer could be related to learners' difficulties with grammar modules or interfaces, this aspect will be neglected in the following because the present study focuses on comparatively proficient learners who have already acquired the grammatical structures in question.

4 Methodological issues

The previous research outlined in the preceding chapters shows that it is not easy to develop L2 processing experiments that test what they are supposed to be testing without being biased by other factors. Moreover, it is difficult to control for every possible influencing factor and at the same time to have natural experimental sentences and representative participant groups. For the studies of the present thesis, care was taken to avoid as many methodological flaws as possible. The present chapter outlines how this was accomplished with regard to the choice of participant groups, experimental procedures, development and evaluation of materials and methods of data analysis.

4.1 Participants

The study of sentence processing focuses on performance rather than on competence (see Chapter 2.1). But especially in the case of L2 learners it is not easy to completely separate these issues because competence can only be measured by performance testing, and performance is always influenced by competence. In order to exclude competence as an influencing factor as far as possible, the learners who participated in the present experiments were sufficiently proficient to have acquired the structures in question. Therefore, it is assumed that they are able to understand and produce such sentences in off-line tasks. In order to double-check this, the learners completed grammar tasks in the respective experiments to ensure that they had the grammatical knowledge that was needed to accomplish these tasks. The choice of learners who only differ from native speakers in their on-line processing behaviour permits to further investigate the reasons for the learners' non native-like processing strategies.

Another issue that is crucial when testing L2 learners is their learning background, which includes factors such as the age of first exposure to the L2, the time of learning, as well as the acquisition context, i.e. natural vs. instructed acquisition, feedback, and motivation. In order to determine the role of each of these factors, they have to be carefully controlled for in an experimental setting. The question how each of these variables influences L2 processing is not in the scope of the present thesis, which focuses on general differences between L1 and L2 on-line processing. Hence, inter-individual variables were excluded to some extent by choosing learners from the same pool. They were all university students of English who had been exposed to English in an instruc-

tional context from approximately the same age on, i.e. around eleven years.[24] In order to test the influence of proficiency, the learners who took part in the present studies were from two groups that were as homogeneous as possible with regard to the time of learning English and proficiency.[25] Participants from the intermediate proficiency group were first or second semester students at German universities (mean age = 21 years). These students had learned English for 11 years on average and none of them had spent more than a few weeks in an English-speaking country. Participants of the advanced group were students preparing for their final examinations of English at German universities. Since it is common for German university students of English to spend at least one term during their studies in an English-speaking country, participants of this group had lived in English-speaking countries for at least three months. They had learned English for 15 years on average and their mean age was 25.

Before the study, all learners took part in a grammar test. This was to ensure that participants within the same group were approximately on the same level. The test was a slightly abbreviated pencil-and-paper version of the on-line Oxford Placement Test. The participants of the intermediate proficiency group gave about 75% correct answers on average, which corresponds to the second highest level in the placement test, i.e. "English for Social and Academic Purposes". Participants of the advanced group gave an average of about 90% correct answers, which corresponds to the highest level in the placement test, i.e. "Advanced".

Nevertheless, there might still be inter-personal variation between these groups. First, some learners might be better at listening and speaking than grammar and thus might have underperformed in the grammar task. Second, the time spent in English-speaking countries might not be a reliable indicator of the learners' proficiency because learners generally differ in how much progress they make during their stay abroad. In order to address these possible difficulties, potential participants who did not clearly fit into one of the groups with regard to the two variables were excluded from the studies.

The control group consisted of native speakers of English who considered British or American English as their L1. Although some of them were living in Germany at the time of testing, they reported that English was still their dominant language which they

[24] At present, German pupils start learning English earlier, but this change was implemented after the learners tested in the present study were at school.

[25] Assuming that the time of learning English represents the L2 exposure of these learners, it is difficult to distinguish proficiency and exposure, unless individual exposure, i.e. the time participants are exposed to English via friends, media, etc. is considered. Since the present studies seek for generalisable differences between L1 and L2 processing, such individual factors are beyond the scope of the present thesis.

used on a regular basis. The native speakers were comparable to the learners in their socio-linguistic background and cognitive abilities. They were all undergraduate and graduate students at German, British or American universities. Native speakers of both British and American English varieties took part in the studies. In order to ensure that the variety of English did not influence participants' attachment preferences, further statistical analyses with Variety as an additional between-subjects variable were run on the results. Although the problem could have been avoided by choosing native speakers of either British or American English, this would have been problematic because the amount of exposure to each variety would be difficult to control for in the learners. Restricting the control group to one variant thus would not have been appropriate to capture possible differences in the learners' input.

All participants were able to accomplish the requirements of the studies. They all had normal or corrected to normal vision and reported that they had never been diagnosed with any learning or behavioural disorders. Each participant took part in only one of the reported studies to ensure that all participants were naïve with respect to the precise purpose of the study.[26] This precaution, however, is not taken by all studies in the field. For example, participants in the Felser et al. (2003) study took part in the off-line and the on-line study. Dussias (2001, 2003) and Fernández (2003) even had their participants take part in studies in their two languages. Such a within-subjects design has the advantage that the patterns of one participant across different tasks or materials can be fully compared. Such a comparison is impossible if the various tasks are accomplished by different participants, even if they are chosen from the same homogeneous pool of participants. But a within-subjects design is unfortunate for the following reasons. If participants take part in two distinct parts of the same experiment, the possibility of interferences cannot be excluded, even if there is some time between the experimental sessions. It must be expected that participants, especially those from university communities, reflect on the experiments they took part in. And even if there are a sufficient number of distracter and filler sentences, it is probable that participants understand the purpose of the study when encountering the same type of sentence in a different language or task. Furthermore, it is natural that they try to remember the answers they gave before to similar sentences or questions, and that they attune to a certain type of

[26] It must be noted that sometimes two experiments were combined in such a way that the experimental sentences of one experiment constituted at the same time the filler items for the other. In these cases participants took part in two experiments at the same time without being aware of it.

materials and lose concentration. Thus, the present experiments aim to exclude such confounding factors by testing each participant in only one of the experiments.

4.2 Choice of experimental procedures

Since the aim of the present thesis is to compare L1 and L2 on-line and off-line processing of various syntactic ambiguities, it is important to have consistent methods for testing participants' preferences. The methods should not be intrusive in order to avoid intimidating the learners or creating an unintended testing situation. For this purpose, it is also important that the instructions are not too difficult to understand and to accomplish. Furthermore, they must allow for testing a large number of participants. The present chapter introduces and discusses the methods that were chosen for the present experiments with respect to these criteria.

A comparison of off-line an on-line data as it was attempted in the present experiments only makes sense if both types of experiments measure the same variable. For example, many off-line studies use sentence completion tasks and compare the results to those obtained via on-line comprehension measures. Such an approach might be problematic insofar that sentence completion tasks elicit production performance, which might draw on different processing mechanisms than comprehension performance. For this reason, comprehension based pencil-and-paper questionnaires were chosen as an off-line measure for the present experiments. In these questionnaires, participants read each sentence and then are presented with two alternative interpretations of the sentence. They are told to make their choice as spontaneously as possible. Nevertheless, they have time to re-read the sentence or parts of it, which makes this task insensitive to immediate parsing decisions. The advantage of the comparatively simple task is that participants do not need extensive meta-linguistic knowledge. A disadvantage is that plausibility or personal preferences can influence participants' decisions. Therefore, these factors should be reduced as far as possible by the choice of the experimental sentences (see Chapter 4.3.). Participants are told that for some sentences both interpretations are possible, which reduces their fear of making mistakes. In order to ensure that participants still pay attention to the task, filler sentences with only one correct interpretation should be included so that participants who do not answer these sentences correctly can be excluded from the analysis.[27]

For the on-line tasks, the SPR paradigm (Just, Carpenter & Woolley 1982) was chosen. In this type of on-line reading task sentences are presented either in a word-by-

[27] It turned out, however, that no participants had to be excluded because of this criterion.

word or phrase-by-phrase fashion. The participants thus read each part of a sentence on a computer screen and trigger the appearance of the next word or segment by pressing a key. The time participants take to press the key to view the next word or phrase is measured. The visual presentation of the stimuli in SPR tasks can be either cumulative or non-cumulative (Marinis 2003). In non-cumulative tasks the appearance of the next segment triggers the disappearance of the previous one, whereas in cumulative presentations the previous words remain displayed on the screen. The advantage of the non-cumulative or moving window technique, which was chosen for the present on-line experiments, is that readers cannot see and refer back to the phrases or words they have previously read. They are therefore encouraged to keep their processing up to date (Mitchell and Green 1978: 610). Hence, this paradigm is supposed to be sensitive to early processing decisions without being influenced by later factors that might occur if participants re-read parts of the sentence.

The assumption underlying such a task is that processing difficulties are reflected in prolonged reaction times. This assumption is however problematic in that reaction times might be influenced by other factors. These factors might be temporary, such as distraction or fatigue, or non-temporary, such as gender, age or testing time. A measure like eye-movement recording might be more exact and less susceptible to external factors. Furthermore, it has to be considered that the phrase-by-phrase or the word-by-word presentation, which makes it impossible to jump back and forth in the sentence, creates a unnatural reading situation in that it decelerates reading and forces participants to remember all parts of the sentence in order to understand the whole. In spite of these possible shortcomings, the results of SPR experiments on syntactic ambiguity resolution were found to be comparable to other measures such as eye-tracking (e.g. Ferreira & Henderson 1990; Garnsey et al. 1997; Just, Carpenter & Woolley 1982). For the on-line experiments presented in this thesis, the SPR paradigm was chosen because it has some advantages over other paradigms. The first advantage is that SPR experiments are comparatively easy to conduct. Thus they permit the testing of reasonably large numbers of participants and thus to obtain comparatively reliable data. Secondly, this method creates a relatively normal and relaxed reading situation in comparison to eye-tracking experiments, in which participants have to wear the equipment and are restricted in their movements. Thirdly, the choice of the SPR paradigm allows for a comparison with previous experiments on L2 learners' attachment preferences (see Chapter 3.2.). Finally, since some of the present studies tested so far unexplored constructions, it makes sense

to first study them with SPR before deciding about the most appropriate task for further investigation.

In order to address the possible shortcomings of the SPR paradigm, care was taken to exclude or to neutralise factors such as interpersonal reaction time differences, firstly by choosing relatively homogeneous participant groups with regard to age, cognitive abilities and computer experience, and secondly by testing comparatively many participants. Furthermore, other factors such as the way of segmentation and disambiguation, the order of presentation, and the comprehension questions or grammaticality judgements following each sentence were carefully considered and controlled, as discussed in the following paragraphs.

In SPR tasks it is possible to present the experimental sentences word by word or phrase by phrase. The type of segmentation can significantly influence the results, especially under the assumption that prosodic aspects of the sentence guide processing decisions (see Chapter 2.3.4.). The word-by-word presentation requires comparatively high concentration and processing capacity on the part of the participants, especially if they are learners. Chunk-wise presentations do not have this disadvantage, but they might bias participants towards artificial prosodic patterns because a short break is forced after each chunk. In order to avoid such biases, a word-by-word presentation was chosen for all but one of the present SPR experiments. Moreover, the possible shortcomings of the way of presentation in the experiments are considered in the respective discussion sections. In all SPR experiments, the presentation of the items was controlled by the Linger software (Rohde 2001).

While in the off-line questionnaires participants read ambiguous sentences and have to choose their preferred interpretations, on-line experiments work with disambiguated sentences. This is necessary because differences between participants and between conditions are expected to show in elevated reading times. For syntactic ambiguity processing, it is assumed that participants are surprised and hence show longer reaction times when encountering a disambiguation towards a dispreferred interpretation of the sentence. Hence, participants must be sensitive to the type of disambiguation in the experimental task. Since this sensitivity cannot be taken for granted, especially in the learner groups, the present experiments include additional off-line grammaticality judgement tasks. In these tasks participants read a sentence and have to indicate whether or not it is grammatically correct. It has to be noted that such a task requires some metalinguistic knowledge from the participants and that it forces them to reflect critically on

the experimental materials. For this reason, it is not advisable that participants accomplish such a task prior to the actual experiment, although such a procedure could help to exclude some participants before testing them with the whole experimental task.

In order to ensure that participants read the sentences carefully and pay attention to the task, most SPR tasks include yes-no comprehension questions after some or all the sentences. Such a measure is also important to detect sentences that have been misinterpreted and consequently might yield different reading time patterns than the sentences that were interpreted in a target-like way. This type of additional on-line comprehension task was chosen for the SPR experiments on RC attachment (Experiments 2 and 3) and the first on-line reading task of PP attachment ambiguities (Experiment 4). Instead of comprehension questions, some researchers (e.g. Juffs 1998a, 1998b, 2006; Juffs & Harrington 1996) inserted a grammaticality judgement instead of a comprehension question after each sentence in order to test whether participants were able to process the experimental sentences. For both comprehension tasks and grammaticality judgement tasks, the time participants take to give the answers can be measured as an additional variable, following the assumption that processing difficulties are reflected in longer response times. Concerning the materials for these tasks, approximately the same number of correct and incorrect response items should be included in order to avoid that participants are confused or biased by an unequal ratio.

A speeded grammaticality judgement might be more sensitive to participants' processing difficulties than comprehension questions because accuracy rates in the latter are generally quite high, even for sentences which are unusual or which cause processing problems. Hence, accuracy rates are not necessarily sensitive to participants' processing difficulties. Speeded grammaticality judgements, in contrast, can vary considerably across groups and experimental conditions and thus provide a further measure of potential processing difficulties. Such a task was used for the second SPR study on PP attachment (Experiment 6) and for the SPR study on wh-questions (Experiment 8). In these experiments, participants were asked after reading each sentence whether or not it was a possible English sentence. This instruction is insofar problematic as definitions of possible and impossible may vary across participants. In order to avoid this confusion or extensive metacognitive reflection on the part of the participants, they were encouraged to rely on their feeling of what might sound good in English (see the Appendix for the precise instructions in the respective tasks). Another potential problem of the judgement task was that it had no option in case participants were undecided. This might have

forced participants towards a decision and thus might have artificially increased the variance (e.g. Schütze 1996 for extensive discussion).[28] A more general problem with grammaticality judgements in L2 learners is that learners' knowledge about some sentences might be indeterminate, and hence their judgements might result from guessing (Schachter, Tyson & Diffley 1976). Moreover, it is difficult to interpret the learners' judgements because they might have a bias to accept sentences they are unsure about or to reject sentences for reasons unrelated to the experimental manipulations (see also Mandell 1999 for a discussion). Nevertheless, several studies (Juffs 1998a, 1998b, 2006; Juffs & Harrington 1996) found that L2 learners are able to perform successfully in these tasks. Moreover, in certain constructions, for example questions, it would be unnatural to construct comprehension question that refer to the target questions. In such cases, speeded grammaticality judgements provide at least indirect comprehension data because incorrect judgements indicate that participants have not processed sentences properly.

It can be concluded that in spite of the potential problems both comprehension questions and grammaticality judgements are important as an addition to the reaction times measured in an SPR task to ensure that participants read the sentences attentively and understood them correctly. Furthermore, if the potential problems are taken into consideration when interpreting the data, these two additional measures can provide additional information on participants' on-line processing mechanisms.

Another important issue in the choice of experimental procedures is whether a participant should see one experimental sentence in several conditions. This procedure is chosen by several researchers because it permits a direct comparison between experimental conditions. However, such a procedure is problematic because one participant sees comparatively many experimental items. Hence, participants might attune to the specific types of sentences and develop strategies of processing them. For this reason, a compromise was chosen for the present experiments. For each experiment, different lists were created in such a way that one group of participants saw a sentence in one condition and the other group saw it in the other condition. The lists have to be matched in such a way that every participant sees the same number of sentences per condition, and it becomes even more important that all sentences are carefully matched for length,

[28] Although methods such as magnitude estimation (e.g. Bard, Robertson & Sorace1996; Cowart 1997) or thermometer judgements (Featherston 2008) would be more appropriate in this respect, they were not employed in the present studies because the grammaticality judgements were used here to examine participants' first intuitions about the sentences without making them reflect too long on their judgements. Therefore, the grammaticality judgement tasks were kept as simple as possible.

naturalness, plausibility and experimental condition. Furthermore, the number of participants should be comparatively high in order to have a sufficient number of participants per list. The sentences on each list were pseudo-randomised for each participant in such a way that no two experimental sentences directly followed one another.

In order to create a comparable experimental situation for all participants, they completed all tasks within an experiment in the same order. While participants were unaware of the precise purpose of the study and the background of the experimental procedures, they were carefully instructed what they were supposed to do in the respective tasks. To this purpose all experiments were preceded by instructions and practice items so that participants could familiarise with the way of presentation and find a reading and clicking speed they were comfortable with. Participants were instructed to read the experimental sentences silently and as quickly and as carefully as possible, but without being afraid of making mistakes. They were also told explicitly if the experiments contained ambiguous sentences (see the instructions in the respective Appendixes). Preceding the experiments, all participants filled in a language questionnaire and signed a letter of agreement to participate (see Appendix A3 and A4). Furthermore, the learners completed the grammar test (see Chapter 4.1.) and they saw a list of all the words that appeared in the SPR task in alphabetical order (see Chapter 4.3.).

4.3 Development of materials

In order to ensure that the experimental materials test what they are supposed to be testing, they have to be carefully developed in such a way that the experimental variables are the only difference between the experimental conditions. This is easiest to achieve if experimental sentences across conditions only differ in one variable. Since a certain number of experimental sentences are needed in order to make the results generalisable and to distract participants from the precise purpose of the experiments, the sentences must be comparable in terms of length and lexical frequency. For SPR tasks as those in the present experiments, it is especially important that the sentences are not only matched for total length but also for the length of the critical words and regions (see the respective appendixes for information on the length of the sentences).

As the discussion in Chapter 2.4. shows, it is not only the length but also the frequency of certain lexical items which might influence how and how fast a word is read and integrated into the sentence. It would be extremely difficult to perfectly match all words in all experimental sentences for frequency. Moreover, it must be considered that frequency is a relative factor because corpora yield different frequency counts. For the

present experiments it would be more important to know the frequency of the words in the input that participants encounter. Especially for the learners this variable is difficult to control for because they are exposed to different language input than native speakers. Furthermore, the requirement of a perfect frequency control could make it extremely difficult to create natural sentences. In the light of these difficulties, it was attempted in the present studies to control at least the potential attachment sites for frequency. This is important because it is conceivable that readers might prefer a certain lexical constellation that they have often heard rather than considering structural attachment preferences. The frequency information used here was provided by the online British National Corpus database. A frequency ratio of at least 0.66 between each pair of critical verbs or nouns was taken as a matching criterion.

Lexical properties of the experimental sentences can also have an unintended effect on the results if participants are not familiar with the lexical items, which is particularly relevant if learners are tested. In order to avoid such effects in the present experiments, the lexical items were all comparatively simple and frequent. Furthermore, preceding the SPR task the learners saw a list of all words and word forms which appeared in the study. The words were listed in their inflected and uninflected forms and they were in alphabetical order. The learners were encouraged to ask the experimenter for the meaning of any words they did not know.

Another issue to be considered in the development of appropriate materials to study syntactic ambiguity resolution is the point of disambiguation. First, it is important that the disambiguating region is not the last region of the sentence because previous on-line reading experiments found so-called wrap-up effects, i.e. longer reading times on the last element of the sentence (Just & Carpenter 1980; Mitchell & Green 1978; Rayner, Kambe & Duffy 2000). To exclude the possibility that an effect in the disambiguating region is in fact a mere wrap-up effect, the disambiguating element should not be the last element of a sentence. Furthermore, it should be considered in the interpretation of the data whether the point of disambiguation is early or late because this manipulation was found to influence processing patterns for instance in RC attachment ambiguities (Brysbaert & Mitchell 1996).

Second, the type of disambiguation is important. In many studies on attachment preferences the sentences were disambiguated via semantic or pragmatic information. The problem is that participants might differ in their sensitivity to this type of disambiguation, depending on their experiences and world knowledge. Although similar ar-

guments might also apply for grammatical disambiguations, especially in studies with L2 learners, it is easier to independently test participants' morphosyntax than their world knowledge. Another problem with semantic or pragmatic disambiguations is that they rarely occur at one specific point of the sentence that is easily recognisable as it is the case with most morphosyntactic disambiguations. For these reasons the present experiments only adopted the semantic / pragmatic disambiguation for attachment ambiguities in PP constructions and wh-questions which cannot be disambiguated otherwise. In the case of RC attachment, the sentences can also be disambiguated via morphological features like gender or number (see Chapter 5.4.2. for a discussion). A general problem with these types of disambiguation when cross-linguistic variance is at issue is that languages may vary in the morphological marking of these features. Thus it is expedient, especially in studies with L2 learners, that participants' sensitivity to the disambiguating factor is independently evaluated.

Apart from the experimental items, filler sentences are an essential part of every psycholinguistic experiment. One function of such fillers is to distract participants from the precise purpose of the study and to prevent them from developing strategies to read the experimental sentences. A second function of fillers is to provide a measure for the participants' overall performance in the task. Thus, they are an additional indicator of the learners' L2 proficiency. This is particularly valuable in on-line tasks, because an additional measure of participants' on-line performance is helpful to interpret their reading of the target sentences. In all experiments reported in the present dissertation, the ratio of fillers to target sentences was 2:1. In two cases experiments were combined so that the experimental sentences of one study served at the same time as fillers of the other. In these cases, there were equal numbers of experimental sentences for the two experiments and the same number of unrelated filler sentences. In contrast to the items within each experiment, which were carefully matched for length, the filler sentences and the words therein were of different length in order to avoid the development of fixed rhythms or patterns of clicking in the SPR tasks.

4.4 Evaluation of materials

Following the development of materials according to the criteria that were discussed in the previous section, the materials have to be checked for effects unforeseen by the experimenter. For the present studies, this revision was accomplished by two linguistically trained native speakers of English, one of them a speaker of American English and the other one a speaker of British English. However, these informants might have been bi-

ased by their linguistic background knowledge, and it is possible that they did not foresee some of the difficulties that participants might have. Therefore, additional naturalness and plausibility rating studies were run with native speakers who were unaware of the purpose of the experiments. In each of the respective rating tasks, a group of ten native speakers of English judged whether the experimental sentences sounded like natural English sentences. They did this in pencil-and-paper questionnaires by judging each sentence on a scale from one (very natural / plausible) to five (very unnatural / implausible). Such a rating scale is problematic in two respects. First, participants might differ in their interpretation of the qualitative values on the scale. Second, such a judgement task only yields data that are ranked on an ordinal scale.[29] Nevertheless, this type of task is widely used in psycholinguistic experiments because it is comparatively easy to conduct and requires little meta-linguistic knowledge on the part of the participants. Moreover, it was found that categorical judgements, gradient judgements on a scale and magnitude estimations yield similar results (Weskott & Fanselow 2008).

The overall ratings in the present judgement tasks were comparatively low for the ambiguous sentences under investigation. This finding is in line with other experiments on syntactic ambiguity processing and is possibly related to the fact that ambiguities are highly marked and normally avoided in everyday language use. It is important, however, that there were no significant differences between participants' ratings of individual sentences because such a finding would indicate that a sentence differed from the other sentences in the respective condition. The objective of the plausibility ratings was to exclude sentences that deviated from the others in their real world plausibility because such differences might bias participants' reading of these sentences in the off-line and on-line tasks. To avoid any potential biases, each participant only took part in one of the naturalness or plausibility judgement tasks.

In the studies with disambiguated materials it is possible that participants' judgements are biased by the respective way of disambiguation. In order to address the possibility of lexical or frequency preferences in favour of one way of disambiguation, two or three lists were created for each experiment. Each participant thus saw every experimental item only in one disambiguation.

[29] According to the theory of scale types (Stevens 1946), all measurement in science uses four different types of numerical scales, i.e. nominal, ordinal, interval and ratio scales. Measurements with ordinal scales are ordered such that higher numbers represent higher values. However, the intervals between the numbers are not necessarily equal. Therefore, "... in the strictest propriety the ordinary statistics involving means and standard deviations ought not to be used with these scales, for these statistics imply a knowledge of something more than the relative rank order of data." (Stevens 1946: 679).

4.5 Data analysis

Preceding the data analysis, it is important to consider possible outliers in the data. In an off-line task, outliers are all instances where no answer or both answers are chosen. In the present experiments these data were excluded from further analysis. In on-line SPR tasks, outliers are wrongly answered comprehension questions or moments where a participant temporarily loses concentration for reasons unrelated to the experiment. The results of the present SPR tasks were screened for outliers following the standard procedure used in the literature (e.g. Felser et al. 2003; Juffs 1998a,b). In a first step means and standard deviations were calculated for each group's reading times of each region of each experimental sentence. Then all the data points more than 2.5 standard deviations below or above the respective means were excluded, a procedure which usually leads to the deletion of two to five percent of the data for a specific group. The resulting reading times for each region of a sentence per group and per experimental condition for the respective experiments are presented in Appendix E.

For reasons of space and clarity, the results sections for each experiment only report the statistical analyses for the critical or disambiguating regions of a sentence. All statistical analyses are run using the SPSS statistics software. As a first statistical analysis on the results of each experiment, a mixed analysis of variance (ANOVA) with group as a between-subjects variable and the respective experimental manipulations as within-subjects variables was conducted. If this analysis yielded a significant interaction between the factors, a next step was to run separated one-factor repeated measures ANOVAs on the results of the individual groups. These analyses were run separately with subjects (F_1) and with items (F_2) as a random variable. In order to investigate whether there was a relationship between participants' proficiency and their attachment choices or their reading patterns, Pearson correlations were calculated for the respective pairs of variables.

As post-hoc tests for between-subjects-variables such as Group, the Tukey-procedure was used. In the case of more than two-level within-subjects variables, planned pair-wise contrasts are reported. For all statistical tests the significance level α was set at 0.05. In cases where the Mauchly-test of sphericity proved significant, the Huynh & Feldt (1976) adjusted probability values were reported with the original, unadjusted degrees of freedom.

5 Experiments on RC attachment

The cross-linguistic differences in RC attachment preferences are a productive area for psycholinguistic research. These structures are particularly interesting for L2 processing because they could help to answer the questions whether, in how far and under which conditions learners transfer processing strategies from their L1 to their L2. The experiments presented in this chapter address these questions by testing off-line and on-line RC attachment preferences in learners at different proficiency levels whose L1 and L2 are generally assumed to elicit different attachment preferences. However, recent studies on German RC attachment preferences obtained controversial results (see Chapter 2.2.1). Therefore, preceding the studies on L2 learners' RC attachment preferences, Chapter 5.1 focuses on RC attachment preferences in German.

5.1 RC attachment preferences in German

While English monolinguals were generally found to prefer low attachment (see Chapter 2.1), German monolinguals' preferences are less clear-cut. Using the eye-movement monitoring technique, Hemforth et al. (1998), for example, found a high attachment preference. The materials included sentences as (45) which were either disambiguated by number on the verb within the RC, by gender on the subject of the RC, or by their pragmatic content.

(45) Klaus traf die Lehrerin der Tochter, die in Deutschland lebte, ...
Klaus met the–ACC teacher the–GEN daughter who in Germany lived-SG
'Klaus met the teacher of the daughter who lived in Germany, …'

The authors report a high attachment preference for all three disambiguation types in this on-line task, a finding that matches results reported by Hemforth, Konieczny & Scheepers (2000) and Groß (2002) from off-line studies.

A study by Augurzky, Alter & Pechmann (2004) tested whether attachment preferences were influenced by the linking preposition. The sentences either contained the linking preposition mit ('with') or the second DP was in genitive case. The authors conducted a visual ERP study, a questionnaire study and a SPR task with materials that were disambiguated via number. In all studies, they observed a low attachment preference for sentences in which DPs were linked by prepositions. This preference showed in a positivity in the activation potentials when high attachment was forced, in a higher percentage of low attachment answers in the questionnaire, and in significantly elevated

reaction times for this condition in the SPR study.[30] This effect of the preposition is in line with findings from English (see Chapter 2.2.2.), which indicate that the preposition has some influence on readers' attachment preferences, either because of its phonological or thematic properties.

It has to be considered, however, that the German construction does not exactly parallel its English counterpart. The German high attachment preference has been observed for genitive constructions as in (46a). But in colloquial German there is also the alternative *von* + dative construction as in (46b):

(46) a. Jemand erschoss die Dienerin der Schauspielerin, die auf dem Balkon war.
'Someone shot the servant of the actress who was on the balcony.'

b. Jemand erschoss die Dienerin von der Schauspielerin, die auf dem Balkon war.
'Someone shot the servant of the actress who was on the balcony.'

Groß (2002) conducted an off-line study to test German native speakers' preferences in these two conditions and in an additional condition with the thematic preposition bei ('next to'). The results of this study are summarised in Table 5-1-1:

gen. high	**gen. low**	**von high**	**von low**	**them. prep. high**	**them. prep. low**
65	35	46	54	35	65

Table 5-1-1: Off-line attachment percentages in German native speakers (adapted from Groß 2002)

As Table 5-1-1 suggests, the study found a significant preference for high attachment in the genitive-condition and for low attachment when the DPs were linked by a thematic preposition. Although no significant preference was found in the *von*-sentences, there was a tendency to attach low in this condition.

In order to reassess this finding, an off-line study with native German university students (n = 30) was designed.[31] For this study, the eight sentences of Experiment 1 were closely translated into German[32] (see Appendix B1 for a list of experimental items). Half of the sentences were linked by the lexical preposition *von* ('of'), whereas in the other half the second DP was in genitive case. In addition to the eight RC attachment choices, the experimental lists included twelve sentences from another control study

[30] It remains open why these studies did not replicate the high attachment preference found by Hemforth and colleagues. One possibility is that length played a role, in that the complex DP is longer when the two DPs are linked by a preposition. Since shorter DPs trigger low attachment, this factor might have weakened the high attachment preference in the comparatively shorter genitive condition.

[31] Their mean age was 24 years (SD = 4.9, range = 18-31), and 77% of them were female.

[32] In some instances, non-literal translations were chosen in order to avoid unnatural expressions. Furthermore, the gender of one of the DPs in each sentence had to be altered in order to avoid disambiguation via the gender of the relative pronoun (*der* or *die*). Hence, in half of the sentences both DPs were masculine; in the other half they were feminine.

(see Chapter 7.1) as well as eight unrelated ambiguous filler sentences. As in the English experiment, participants read the 28 ambiguous sentences and chose their preferred interpretation out of two possibilities. An example sentence is provided below:

(47) Der Arzt untersuchte den Vater des / von dem Jungen der bei dem Unfall verletzt wurde.
'The doctor examined the father of the boy who was injured in the accident.'

This study confirmed the results reported by Groß (2002) in that a significant high attachment preference was found in the genitive-condition ($F(1,29) = 28.066$, $p \leq 0.0001$). But in contrast to the results of Groß (2002), there was also a high attachment preference when the DPs were linked by the preposition *von* ('of') ($F(1,29) = 18.686$, $p \leq 0.0001$). The results are illustrated in the following table:

gen. high	gen. low	von high	von low
74.17 (25.0)	25.8 (25.0)	69.2 (24.3)	30.8 (24.3)

Table 5-1-2: German RC attachment I: mean attachment percentages and (standard deviations)

Since the clear high attachment preference for the *von*-condition deviated substantially from the slight low attachment preference reported by Groß (2002) for the same condition, another study to investigate German off-line attachment preferences was designed. The study used the same materials as the first study, except that the genitive condition was replaced by a condition in which the two possible attachment DPs were linked by a thematic preposition:

(48) Der Arzt untersuchte den Vater bei dem Jungen der bei dem Unfall verletzt wurde.
'The doctor examined the father with the boy who was injured in the accident.'

The rationale behind this study was that the preference for the *von*-condition might be influenced by the other condition of the respective experiment. If this was the case, the low attachment preference for the thematic preposition condition should be transferred to the *von*-condition in the second study. Again, 30 German university students who had not taken part in any of the other experiments completed the questionnaire.[33] The results are illustrated in the following table:

von high	von low	them. prep. high	them. prep. low
32.5 (25.4)	67.5 (25.4)	23.3 (26.4)	76.7 (26.4)

Table 5-1-3: German RC attachment II: mean attachment percentages and (standard deviations)

In this study a significant low attachment preference was found in the thematic preposition condition ($F(1,29) = 13.198$, $p \leq 0.001$) and in the *von*-condition ($F(1,29) = 33.143$,

[33] Their mean age was 23 years (SD = 4.7, range = 19-29), and 70% of them were female.

$p \leq 0.0001$).[34] Whereas the low attachment preference in the former condition is in line with the results of Groß (2002) and Augurtzky (2006), the diverging preferences in the *von*-condition are difficult to explain. One relevant factor suggested by Petra Augurtzky (personal communication) might be a regional difference, considering that Augurtzky's participants came from Eastern Germany, Groß's participants from Western Germany, and the participants tested in the present study were from Northern and Western Germany.[35] The two present studies moreover suggest that the preference for the *von*-condition is relatively unstable and thus might be influenced by the other condition of the respective experiment.[36] This would explain the high attachment preference when the second condition was the genitive-condition, while participants preferred low attachment when the second condition was the thematic preposition condition.

The variable results in the *von*-construction might be related to the unclear grammatical status of the construction. While it is common in colloquial German, it is not accepted in formal High German, at least with definite nouns (Duden Grammatik 2005: 836). Hence it is conceivable that German speakers have diverging intuitions about this construction. In order to test how well accepted the *von*-constructions are and thus how much influence this factor could have on German ESL learners, an additional rating task was conducted with another group of 20 native German university students.[37] In this task, participants rated the acceptability of the sentences on a scale from one (perfectly acceptable) to five (completely unacceptable).[38] The results of this task are summarised in Table 5-1-4:

	genitive-condition	**von-condition**
Mean ratings and (SD)	1.6 (0.4)	2.5 (1.5)

Table 5-1-4: Mean acceptability ratings and (standard deviations) for the German materials

As Table 5-1-4 indicates, sentences in the *von*-condition were judged as being slightly less acceptable than sentences in the genitive-condition. Moreover, the standard devia-

[34] An ANOVA with Preposition and Attachment as within-subjects factors only yielded a significant main effect for Attachment ($F(1,29) = 29.083$, $p \leq 0.001$). However, the interaction approached significance ($F(1,29) = 3.777$, $p \leq 0.062$), which indicates that the low attachment preference in the *von*-condition was not as strong as in the thematic preposition condition.

[35] Although such a dialectal factor is compatible with a phonological account, it is also possible that the alternatives to the genitive are more common in one dialect and thus pragmatically force low attachment, or that the strength of a principle like Predicate Proximity varies regionally. Although not in the scope of the present study, these questions are an interesting area for future research.

[36] A similar influence of the experimental context on attachment preferences was reported for French online attachment preferences (Frenck-Mestre & Pynte 2000). The results depended on whether the *avec*- ('with') condition was presented preceding or following the *de*- ('of') condition.

[37] Participants' mean age was 22 years (SD = 4.2, range = 19-26), and 60% of them were female.

[38] The 16 RC sentences for the present study were intermingled with 16 other ambiguous sentences, half of them for Experiment 7 (see Chapter 7.1.).

tion was higher in the former condition, which indicates some degree of insecurity. The difference between the two conditions was statistically significant ($F(1,19) = 8.393$, $p \leq 0.009$), which suggests that the *von*-sentences are less accepted than the genitive-sentences.

To sum up, the heterogeneous findings for these sentences indicate that the lexical properties of the linking preposition have some influence on attachment preferences in all languages and that there is some variability with regard to the *von*-construction. However, the high attachment preference found in previous studies was replicated by most studies with speakers of standard High German.

5.2 Predictions and research questions

Given the reported cross-linguistic differences between English and German, the question arises whether German ESL learners have adopted the English low attachment preference or whether they transfer the German high attachment preference to their L2. Another interesting question is whether German ESL learners are sensitive to the distinction between the prepositions of (non-theta assigner) and with (theta-assigner). Furthermore, it has been proposed that learners employ L2-specific strategies. In the present studies such strategies might either result in (1.) a preference for attachment to the most recently processed element because of a relative lack of processing resources, or (2.) a lack of any attachment preference as a result of shallow structural processing (Clahsen & Felser 2006b, and the discussion in Chapter 3.1.2.). Hence, the present studies on RC attachment address the following research questions:

A) How do ESL learners process RC attachment ambiguities off-line and on-line, and in how far are these preferences influenced by proficiency?
B) Are ESL learners influenced by their L1 when processing RC attachment ambiguities?
C) Do ESL learners employ L2-specific strategies when processing RC attachment ambiguities?

Similar questions were addressed by a previous study on German ESL learners' attachment preferences reported by Felser et al. (2003). However, there are several problems with this study (see Chapter 3.2.4.2.). The authors did not find any attachment preference in the highly proficient learners who were tested and attributed this finding to the learners' shallow syntactic processing. Nevertheless, the question remains why other studies did find clear preferences in L2 learners (see Chapter 3.2.4.). Furthermore, the lack of attachment preferences reported by Felser et al. (2003) could also be related to a partial influence of transfer or a recency preference. Since the authors do not provide

analyses of the individual participants' preferences, it remains open whether there is really no preference or whether some learners show a high attachment preference and some show a low attachment preference. Hence, the present studies, which test learners at different proficiency levels and also include analyses of the individual participants' data, might provide some new evidence on the mechanisms that guide learners' RC attachment preferences.

5.3 Experiment 1: Off-line RC attachment

The first experiment investigates the ESL learners' attachment preferences in an off-line task. Although participants were told to make their decisions as spontaneously as possible, this task leaves them time to think about their decisions and to re-read certain passages or the whole sentence. Hence, processing capacity limitations should not have a strong effect in this type of task.

Following previous studies on English native speakers' off-line RC attachment preferences (see Chapter 2.1.), the English controls are expected to prefer low attachment in this task. For the two learner groups, the predictions are less straightforward. If they are as sensitive as native speakers to the lexical properties of the linking preposition (see Chapter 2.2.2.), they are expected to prefer low attachment when the two DPs are linked by the preposition with. For the linking preposition of, the respective approaches predict different reading patterns for the learners. A low attachment preference would indicate that the learners either employ an L2-specific recency strategy or similar mechanisms as English monolinguals. If the learners prefer high attachment, this finding would be attributable to L1 transfer. The predictions of the SSH for an off-line task are not straightforward. Although Felser et al. (2003) attribute the absence of attachment preferences in their off-line task to such a strategy, it is also conceivable that shallow processing is restricted to speeded processing in on-line tasks. If the learners' off-line processing is shallow, they are expected to show no attachment preference in this task. In all these cases the subsequent question arises whether the effects are similar for learners at different proficiency levels.

Hence, the RC attachment ambiguity permits to empirically test and compare the predictions of various L2 processing theories. Experiment 1 thus addresses the following research questions:

A) How do ESL learners process RC attachment ambiguities off-line?
B) To what extent are these preferences influenced by the learners' proficiency?

C) Are ESL learners influenced by their L1 or by an L2-specific strategy when processing RC attachment ambiguities?

5.3.1 Participants

The three groups of participants were from the pool described in Chapter 4.1.. The native speaker control group consisted of 15 speakers of British English and 15 speakers of American English. Their mean age was 24.6 years (range = 20-30, SD = 4.8) and 40% of them were female. The biographical data of the two learner groups are summarised in the following table: [39]

Group		Age	Years of learning English	Months in English-speaking countries	Grammar
Intermediate learners (n = 30)	**mean**	21	11	___	77.2
	SD	(1.6)	(1.9)		(6.5)
	range	19-24	9-13		65-82
Advanced learners (n = 30)	**mean**	25	15	10.8	93.4
	SD	(4.1)	(3.9)	(6.3)	(3.9)
	range	23-34	12-23	3-30	87.5-100

Table 5-3-1: Biographical data of the learner groups in Experiment 1

5.3.2 Method

The experimental task was an off-line pencil-and-paper questionnaire in which participants read ambiguous sentences and decided which of the two possible interpretations they preferred. In order to reduce differences unrelated to the experimental manipulations between the sentences, all materials went through an additional norming procedure (see Chapter 4.4.). In this procedure, naturalness and plausibility were judged independently by two groups of English native speakers. 12 sentences were created in which the two possible attachment DPs could either be linked by the preposition with or of. The sentences were assigned to two different lists such that each participant saw a sentence either in the with-condition or in the of-condition. Ten native speakers of English judged the naturalness and ten others the real word plausibility of the sentences on a scale from one (very natural / plausible) to five (very unnatural / implausible). All participants saw the sentences of their respective list in the same randomised order. The sentences were intermingled with 24 unrelated fillers in such a way that no two experimental sentences directly followed one another. The results of the task are illustrated in Table 5-3-2.

[39] In order to test whether these measures were a sufficiently reliable indicator of the learners' proficiency, Pearson correlations between these measures were calculated. There was a significant correlation between years of learning and performance in the grammar task ($r(60) = 0.410$, $p \leq 0.001$), and between months spent in English-speaking countries and grammar ($r(60) = 0.699$, $p \leq 0.0001$).

	Plausibility		Naturalness	
	of	**with**	**of**	**with**
Means and (SD) by subjects	2.25 (1.5)	2.45 (1.78)	3.43 (0.74)	3.83 (0.70)
Means and (SD) by items	2.25 (0.41)	2.46 (0.38)	3.43 (0.40)	3.84 (0.34)

Table 5-3-2: Participants' naturalness and plausibility judgements by subjects and by items in Exp. 1

As Table 5-3-2 indicates, the of-condition was judged as being slightly more natural than the with-condition. This difference reached statistical significance in the one-factor ANOVAs that were run on the results, respectively with subjects and items as random variables ($F_1(1,9) = 5.101$, $p \leq 0.05$; $F_2(1,11) = 7.804$, $p \leq 0.017$). There was no statistically significant difference in the plausibility judgements ($F_1(1,9) = 1.855$, $p \leq 0.266$; $F_2(1,11) = 1.316$, $p \leq 0.276$), which indicates that the conditions did not differ considerably in plausibility. The relatively low standard deviations in the analysis by items suggest that sentences within a condition elicited relatively homogeneous judgements. This finding implies that no factors other than the experimental manipulations had a considerable impact on the naturalness and plausibility judgements of the sentences.

The eight sentences for the off-line study were taken from the body of sentences used in the norming studies. 16 independent filler sentences were added in order to distract participants from the purpose of the study and to prevent the use of reading strategies. Again, two experimental lists were created so that each participant saw each experimental sentence either in the of-condition as in (49a) or in the with-condition as in (49b).

(49) a. The student liked the secretary of the professor who was killed in the robbery.
□ The secretary was killed.
□ The professor was killed.
b. The student liked the professor with the secretary who was killed in the robbery.
□ The secretary was killed.
□ The professor was killed.

All participants saw the sentences of their respective list in the same randomised order, in such a way that two experimental sentences were always divided by at least one filler item (see Appendix B2 for a list of experimental items). In half of the sentences the high attachment interpretation was presented as the first and in the other half as the second alternative. Participants were instructed to read the sentences carefully and to choose the alternative that they preferred, even if the other alternative might be possible as well.

5.3.3 Results

Since the answers to the filler items were comparatively easy, this measure cannot serve as a reliable proficiency measure. But the ceiling level performance (over 98% correct answers) of all participants indicates that they paid attention to the task and completed it to the best of their knowledge. The percentages of attachment choices in the RC attachment conditions are illustrated in Figure 5-3-1.

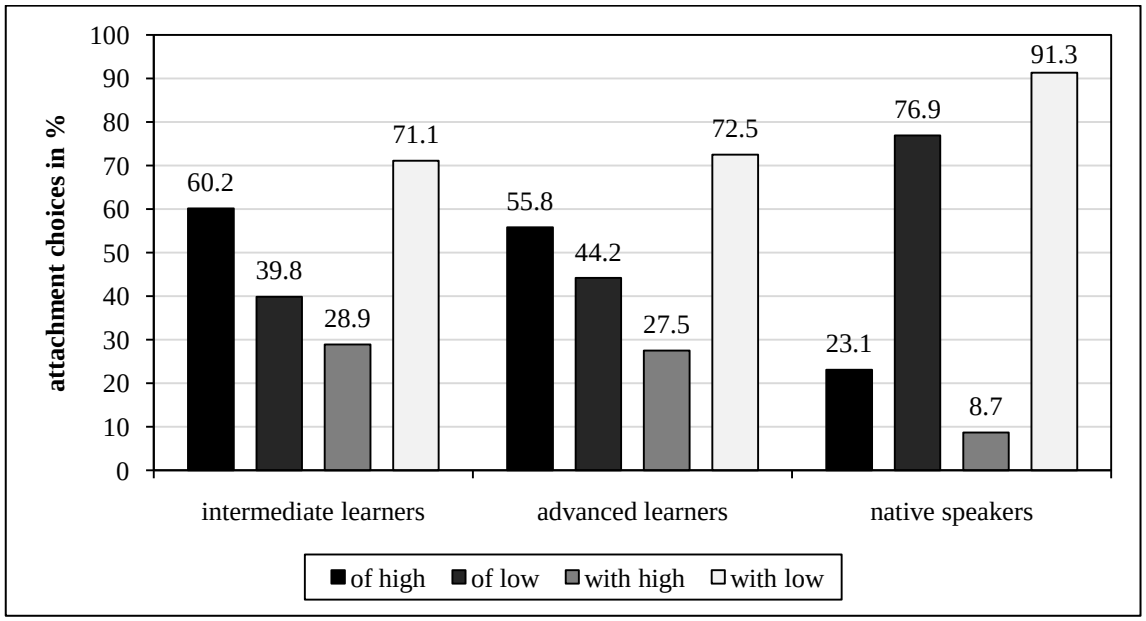

Figure 5-3-1: Participants' attachment preferences per group and per condition in Exp. 1

All results were submitted to a mixed three-factor ANOVA with Preposition (of vs. with) and Attachment (high vs. low) as within-subjects / items variables and Group as a between-subjects / items variable. These analyses did not yield a significant effect for Preposition ($F_1 \leq 1$; $F_2(1,21) = 2.123$, $p \leq 0.160$), but significant effects for Attachment ($F_1(1,87) = 47.119$, $p \leq 0.0001$; $F_2(1,21) = 127.667$, $p \leq 0.0001$) and for Group ($F_1(1,87) = 9.063$, $p \leq 0.0001$; $F_2(2,21) = 3.345$, $p \leq 0.055$). There were also significant interactions of Attachment x Group ($F_1(2,87) = 15.810$, $p \leq 0.0001$; $F_2(2,21) = 38.887$, $p \leq 0.0001$), Attachment x Preposition ($F_1(2,87) = 75.800$, $p \leq 0.0001$; $F_2(1,21) = 28.405$, $p \leq 0.0001$), and Attachment x Preposition x Group ($F_1(2,87) = 4.969$, $p \leq 0.009$; $F_2(2,21) = 13.327$, $p \leq 0.028$).

In order to analyse these complex interactions more closely, the results of the three groups were submitted to separate ANOVAs for the of-conditions and the with-conditions. As Figure 5-3-1 indicates, the native speaker control group preferred low attachment in both conditions. This difference was significant in the analyses by subjects and by items (of-condition: $F_1(1,29) = 57.404$, $p \leq 0.0001$; $F_2(1,7) = 22.588$, $p \leq 0.018$; with-condition: $F_1(1,29) = 278.846$, $p \leq 0.0001$; $F_2(1,7) = 374.999$,

$p \leq 0.0001$). The overall low attachment preference is in line with findings reported in the literature for native speakers of English in this type of off-line task (see Chapter 3.2.4.). As discussed in Chapter 4.1. half of the native speaker control group were speakers of British English, the other half were speakers of American English. Although previous studies found a low attachment preference for both varieties (e.g. Henstra 1996 for British English and Deevy 2000 for American English and the discussion in Chapter 4.1.), a direct comparison of these two groups has never been reported. In order to ensure that the variety of English did not influence participants' attachment preferences and thus might indirectly influence the learners' results, an additional mixed two-factor ANOVA was conducted on the native speakers' results with Condition as a within-subjects factor and Variety as a between-subjects factor. This analysis yielded no significant main effects for Variety ($F \leq 1$). In addition, no significant interaction of Condition x Variety was found ($F \leq 1$). Hence, Variety did not significantly influence participants' attachment preferences.

The highly proficient learners showed a clear low attachment preference in the *with*-condition ($F_1(1,29) = 18.209$, $p \leq 0.0001$; $F_2(1,7) = 78{,}818$, $p \leq 0.003$), but no significant preference in the *of*-condition ($F_1(1,29) = 1.092$, $p \leq 0.305$; $F_2 \leq 1$). This result is in line with findings reported by Felser et al. (2003) for learners at a similar proficiency level. However, Figure 5-3-1 shows that the advanced learners in the present study have at least a tendency to attach high. A look at the individual participants' preferences reveals that they did not all show a 50-50 pattern. Instead, only nine of the 30 advanced learners (i.e. 30%) showed no attachment preference. Twelve of these learners (i.e. 43.3%) preferred high attachment, while eight of them (i.e. 26.7%) showed a low attachment preference. This pattern suggests that some of the learners were more strongly influenced by their L1, whereas others had already adopted the English preference, and only a few of these learners were "in between". Nevertheless, the number of low attachment preferences did not significantly correlate with any of the proficiency measures, neither with years of learning ($r(30) = -0.094$, $p \leq 0.622$), nor with time spent abroad ($r(30) = 0.164$, $p \leq 0.422$), nor with participants' performance in the grammar test ($r(30) = 0.017$, $p \leq 0.928$). This finding does not necessarily imply that the proficiency measures chosen to distinguish between groups are invalid, because there were clear differences between the two groups' attachment preferences. Instead, the findings indicate that the relationship between proficiency / exposure and the adoption of native-like processing strategies is not straightforward but might be influenced by the quality of the input.

In contrast to the advanced learners, the intermediate learners did show a significant high attachment preference in the of-condition, at least in the analysis by subjects ($F_1(1,29) = 6.905$, $p \leq 0.014$; $F_2(1,7) = 1.920$, $p \leq 0.260$). The individual participants' analysis shows that 12 of the intermediate learners (i.e. 40%) preferred high and low attachment in 50% of the cases. Only three intermediate learners (i.e. 10%) preferred low attachment, and 15 of these learners (i.e. 50%) preferred high attachment. As for the advanced learners, there were no significant correlations between the number of low attachment preferences and years of learning ($r(30) = 0.009$, $p \leq 0.961$) or performance in the grammar test ($r(30) = -0.018$, $p \leq 0.925$) for this group.

5.3.4 Discussion

The research questions addressed by Experiment 1 were whether ESL learners are influenced by their L1 or by an L2-specific strategy when processing RC attachment ambiguities, and to what extent preferences are influenced by the learners' proficiency. The present chapter discusses the results of the present study with regard to these questions.

In the *with*-condition, all groups preferred low attachment. This finding is in line with previous studies on L1 and L2 RC attachment preferences. Since a cross-linguistic low attachment preference is generally reported for this condition, the learners' low attachment preference in the present study could either be related to L1 transfer or to the employment of native-like processing strategies. None of the learner groups showed a native-like low attachment preference in the *of*-condition, which indicates that the learners are sensitive to the distinction between the theta-assigning preposition with and the non-theta assigning preposition of. While a low attachment preference in the *with*-condition is predicted by all L2 processing approaches discussed here, the approaches differ in their predictions for the *of*-conditions. Whereas the SSH predicts no attachment preference for the learners in the *of*-conditions, a low attachment preference is predicted if learners employ a recency strategy, and a high attachment preference is expected in the case of L1 transfer. These three approaches and their respective predictions for ESL learners' RC attachment preferences are considered in turn in the following paragraphs.

Since none of the two learner groups in the present study showed a clear low attachment preference in the *of*-conditions, a recency strategy cannot be the factor that determines learners' RC attachment preferences. The SSH does not provide a plausible explanation for the findings, either. Although the most proficient learners did not show any significant preference, this finding does not necessarily constitute evidence for the SSH. Instead, the high interpersonal variation rather suggests that these learners have

clear attachment preferences, which, however, are comparatively heterogeneous. Moreover, the less proficient learners' significant high attachment preference is not in line with the SSH. Accordingly, it is unlikely that the advanced learners but not the intermediate learners employ a shallow processing strategy.

The learners' different attachment preferences for the of-condition and the with-condition indicate that they are sensitive to the thematic differences of the linking prepositions. Although this sensitivity is predicted by the SSH, it could also be transferred from the learners' L1. The latter possibility is more plausible, considering that the intermediate learners prefer high attachment in the of-conditions. Since high attachment is expected to impose higher processing load, this finding is not in line with an L2-specific recency strategy. Hence, L1 transfer is the most likely explanation for the intermediate learners' attachment preferences in the present study. Accordingly, the advanced learners' pattern in the of-condition, i.e. no significant overall effect but individual preferences either for high or for low attachment, indicates that some of these learners transfer the attachment preference from their L1 to their L2, while others have adopted a native-like low attachment preference for their L2. In conclusion, the comparison between the two learner groups shows that more proficient learners are less susceptible to L1 transfer. However, the absence of significant correlations of low attachment choices with any of the reported proficiency measures indicates that this relationship is not straightforward but underlies individual variation.

The effects of the preposition linking the two attachment DPs on participants' offline attachment preferences has also implications for general processing theories. Even if the preferences for the German *von* + dative construction are not fully resolved, there is a clear difference between German native speakers' preferences for RCs following DPs linked by the genitive and RCs linked by a thematic preposition such as neben ('next to') (see Chapter 5.1.). This pattern is replicated in the English native speakers' preferences, which are stronger in the with-conditions than in the of-conditions. The finding is in line with previous studies (see Chapter 2.2.2.). Although the influence of lexical factors is most easily explained by constraint-based processing models (see Chapter 2.4.2.), syntactically based models such as Construal (see Chapter 2.2.2.), Recency & Predicate Proximity (see Chapter 2.3.3.) or Anaphor Binding (see Chapter 2.3.2.) have been adapted to account for these findings.

Taken together, the results of the present study are not in line with L2-specific processing strategies like the overuse of a recency strategy or shallow syntactic processing.

Instead, L2 off-line processing is influenced to some extent by the learners' L1. Another important conclusion from the results is that transfer of processing strategies from the learners' L1 subsides with increasing proficiency. While Experiment 1 focuses on off-line processing, the following chapters examine whether similar tendencies can be found for ESL learners on-line RC attachment preferences.

5.4 Experiment 2: On-line RC attachment

Experiment 2 is an on-line SPR task with similar materials as Experiment 1. In contrast to the ambiguous sentences in Experiment 1, the sentences in the present study are disambiguated, i.e. either high or low attachment of the RC is forced. If participants have any attachment preferences, these are expected to show in shorter reading times for the preferred than for the dispreferred disambiguation. Following previous studies on English native speakers' RC attachment preferences, the English native speakers in the present study are expected to show a low attachment preference. As in Experiment 1, the predictions are less clear for the German ESL learners. In their native language, a high attachment preference has been reported (see Hemforth et al. 1998). Hence it is conceivable that the learners are influenced by their L1 and consequently show a high attachment preference in the present study. But it is also possible that they have either adopted the English low attachment preference or that they employ an L2-specific strategy. In Experiment 1, participants were not timed and had the possibility to re-read the sentences. Therefore, their attachment choices do not necessarily reflect initial preferences. As a consequence, the question arises whether learners' on-line processing is influenced by the same factors as their off-line processing. It could be the case, for example, that under time-pressure the learners have a stronger tendency to attach the RC to the most recently encountered DP. It is also possible that they do not construct deep syntactic structures, which might result in the absence of any preference.[40] With regard to the question of L1 influence, the learners might be more strongly influenced by their L1 in on-line processing than in off-line processing. It is also conceivable, however, that L2-specific strategies are more important in this situation and hence have a stronger influence than the learners' L1. Since all these predictions can be tested by an on-line study on German ESL learners' RC attachment preferences, Experiment 2 addresses the following research questions:

[40] Another prediction of the SSH is that learners prefer ambiguous over disambiguated sentences because they do not make any commitments in the absence of structural cues. However, monolingual speakers were also found to prefer completely ambiguous sentences (e.g. Traxler, Pickering & Clifton 1998; Van Gompel et al. 2005), so that a difference between learners and native speakers cannot be expected.

A) How do ESL learners process RC attachment ambiguities on-line?

B) In how far are these preferences influenced by the learners' proficiency?

C) Are ESL learners influenced by their L1 or by an L2-specific strategy when processing RC attachment ambiguities?

5.4.1 Participants

Participants were from the same pool as those who took part in Experiment 1, i.e. a native speaker control group[41] and two groups of German university students of English (see Chapter 4.1.). Participants who had already taken part in Experiment 1 did not take part in Experiment 2. This was to ensure that they could not develop any strategies for processing the experimental sentences and were unaware the precise purpose of the study. Table 5-4-1 summarises the biographical data of the two learner groups: [42]

Group		Age	Years of learning English	Months in English-speaking countries	Grammar
Intermediate learners (n = 30)	**mean**	21.8	11	___	76.9
	SD	(1.3)	(1.8)		(7.2)
	range	19-26	5-14		65-82
Advanced learners (n = 30)	**mean**	25.3	15.4	8.5	91.9
	SD	(4.1)	(4.5)	(6.4)	(3.8)
	range	23-43	13-30	3-36	85-100

Table 5-4-1: Biographical data of the learner groups in Experiment 2

5.4.2. Choice of materials

The experimental sentences were from the same type as in Experiment 1, but in the present on-line study the sentences were disambiguated towards high or low attachment. Since disambiguation via morphosyntactic features is preferable over pragmatic disambiguation in the case of RC attachment (see Chapter 4.3.), for the present study the choice remained between disambiguation via number and gender. The following paragraphs discuss these possibilities in more detail.

Recent studies suggest that disambiguation via gender might cause a high attachment preference that would not be observed otherwise. For instance, Carreiras et al. (2001) conducted an eye-tracking study with Spanish monolinguals with Disambiguation (number vs. gender) as an additional variable. In this study the sentences disambiguated via gender elicited a high attachment preference, whereas number disambigua-

[41] Half of them spoke British English, the other half spoke American English. Their mean age was 26.4 years (range = 20-38, SD = 5.8) and 63% of them were female.

[42] To test whether these measures were a reliable indicator of the learners' proficiency, correlations were calculated. Grammar was found to be correlated with years of learning ($r(60) = 0.441$, $p \leq 0.0001$) and with months in English-speaking countries ($r(60) = 0.608$, $p \leq 0.0001$).

tion triggered a low attachment preference.[43] The authors interpret this finding in such a way that number information is purely syntactic and accordingly is subject to Late Closure, whereas gender information is influenced by semantic factors. But such an account is incompatible with current syntactic theories which assume that gender is a purely syntactic feature of nouns that determines the relationships between elements by the operation of feature checking (e.g. Carstens 2000; Timpli 2003). Furthermore, studies measuring ERPs found that gender violations elicit similar effects as syntactic violations (Carreiras, Gillon-Dowens, Barber & Betancort 2004, Frenck-Mestre 2004, Gunter, Friederici & Schriefers 2000; Van Berkum, Brown & Hagoort 1999). In addition, psycholinguistic studies on RC attachment using gender disambiguation report a low attachment preference for English (e.g. Carreiras & Clifton 1999).

Other studies found problems with the disambiguation via number in that a plural intervening between a singular subject and the verb it must agree with (e.g. the key to the cabinets was...) disrupts processing of the entire clause (e.g. Bock & Miller 1991, Nichol, Forster & Verez 1997; Pearlmutter, Garnsay & Bock 1999 for some evidence from English and Kaan 2002; Pearlmutter 2000; Vigliocco & Nichol 1998 for further discussion).[44] Although these studies focus on relations within sentences while RC attachment ambiguities involve relations between clauses, it is possible that a similar disruption occurs in RC attachment. Hence, a DP2 that does not agree in number with the DP1 and the verb of the RC might disrupt processing in the case of forced DP1 attachment. As a consequence, low attachment would be preferred over high attachment. This prediction was tested by Fernández & Sainz (2004) in two experiments in which Spanish monolinguals read materials that manipulated RC attachment (high vs. low) and the number marking of the DPs. The experimental conditions are illustrated in (50a-d.).[45]

(50) Andrés cenó con...

a. el sobrino de los maestros / que se emborrachó. [DP1-sg DP2-pl RCV-sg]
b. los sobrinos del maestro / que se emborracharon. [DP1-pl DP2-sg RCV-pl]
c. el sobrino de los maestros / que se emborracharon. [DP1-sg DP2-pl RCV-pl]
d. los sobrinos del maestro / que se emborrachó. [DP1-pl DP2-sg RCV-sg]

'Andrew ate with ... the nephew(s) of the teacher(s) / who [refl] got-drunk[sg/pl].)'

[43] The general low attachment preference in the SPR tasks with number-disambiguated materials in Fernández (2003) might be explained in a similar way. However, the Spanish subjects' unusual low attachment preference might also be related to other methodological factors (see Chapter 3.2.4.1.).

[44] Similar results are reported by Vigliocco et al. (1996) for Dutch and French and by Oria-Merino & Sainz (1998, 1999) and Vigliocco, Butterworth & Garrett (1996) for Spanish.

[45] 'RCV' in the examples stands for "relative clause verb", which is morphologically marked for number and hence either agrees with DP1 or DP2.

The first study was a SPR task which presented the sentences in segments as indicated by the slashes in the example. The second study was a response contingent same-different sentence-matching task (Stevenson 1993).[46] In both tasks the authors observed a preference for forced low attachment (50c-d). This pattern, however, does not necessarily reflect participants' attachment preferences. Instead, the preferences could have been triggered by the number configurations because in both tasks sentences with a plural DP2 were found to elicit higher reaction times than those with a singular DP2.[47]

Similar results with English monolinguals are reported by Deevy (2000) with materials as the following:

(51) John was excited to meet . . .

a. the niece of the actors who was recently starring in a very successful play.

b. the nieces of the actor who was recently starring in a very successful play.

c. the nieces of the actor who were recently starring in a very successful play.

d. the niece of the actors who were recently starring in a very successful play.

The results indicate that the processing of singular agreement was disrupted by a plural DP if it stood between the two agreeing singular elements as in (51a). This was not the case if a singular element stood between the two agreeing plural elements as in (51c) or if the two agreeing elements were adjacent as in (51b) or (51d). This finding again suggests that disambiguation via number agreement can be problematic because it might trigger low attachment by decelerating processing speed when high attachment is disrupted by an intervening plural DP. To sum up, these studies found potential difficulties with both ways of disambiguation. Since the processing disruption caused by disambiguation via number was found to be more severe, gender disambiguation was chosen for the present study on RC attachment.

In order to reduce influencing factors other than the experimental manipulations as far as possible, a norming study preceded the actual experiment. Ten English native speakers judged the naturalness and ten others the real-world plausibility of the experimental sentences on a scale from one (very natural / plausible) to five (very unnatural / implausible) (see Chapter 4.4.). Eight experimental sentences were created in such a

[46] In this type of task participants judge whether the two sentences presented to them in turn are the same or different. The task provides two measures: a whole-sentence reading time that is not necessarily sensitive to early processing effects, and a matching time that has been argued to reflect processing load associated with ungrammaticality detection in early processing (Stevenson 1993).

[47] Further support for this assumption comes from Miyamoto (2005), who also found a low attachment preference in his Brazilian Portuguese participants. A reanalysis of the reading times of his SPR study revealed that shorter reading times for low attachment were only found if the second DP was plural and the verb was singular.

way that the two factors Attachment (high vs. low) and Preposition (of vs. with) were manipulated. All sentences were of the structure DP1 – VP – DP2 of / with DP3 – who – VP – himself / herself – wrap-up. The four resulting types of sentences are illustrated in (52a-d) (see Appendix B3 for a list of experimental items):

(52) a.	The student	had liked	the secretary of the professor	who	had killed	herself	in the office.
b.	The student	had liked	the secretary of the professor	who	had killed	himself	in the office.
c.	The student	had liked	the professor with the secretary	who	had killed	himself	in the office.
d.	The student	had liked	the professor with the secretary	who	had killed	herself	in the office.
Region	1	2	3	4	5	6	7

Two lists were created so that each participant saw the of-condition of a sentence in one disambiguation and the corresponding with-condition in the opposite disambiguation. Hence, each participant judged either the naturalness or the plausibility of 16 sentences. Participants' naturalness and plausibility ratings are summarised in Table 5-4-2:

	of high	**of low**	**with high**	**with low**
Naturalness	3.17 (0.58)	1.82 (1.08)	3.72 (0.74)	2.12 (0.78)
Plausibility	2.49 (0.83)	2.12 (1.09)	2.57 (0.83)	2.22 (1.13)

Table 5-4-2: Naturalness and plausibility ratings for Experiment 2

A repeated-measures ANOVA on the naturalness judgements with the factors Preposition (of vs. with) and Attachment (high vs. low) was calculated. This analysis found a significant effect for Attachment ($F_1(1,9) = 19.024$, $p \leq 0.002$; $F_2(1,11) = 69.271$, $p \leq 0.0001$), while the effect for Preposition only approached significance in the analysis by subjects ($F_1(1,9) = 4.655$, $p \leq 0.059$; $F_2 \leq 1$).[48] To conclude, the English native speakers judged low disambiguated sentences as being more natural than high disambiguated sentences, which is in line with previous off-line findings for English (see Chapter 5.1.). Sentences in the with-conditions were judged as being slightly but not significantly less natural than sentences in the of-conditions in the same disambiguation.

The same analysis on the plausibility judgements yielded a significant effect for Attachment only in the analysis by items ($F_1(1,9) = 1.885$, $p \leq 0.203$; $F_2(1,11) = 11.000$,

[48]This was confirmed by a separate one-factor ANOVA with four levels, which yielded a significant main effect ($F_1(3,27) = 13.938$, $p \leq 0.005$; $F_2(3,33) = 17.368$, $p \leq 0.001$). Planned comparisons found a significant difference between the *of*-high and the *of*-low conditions ($F_1(1,9) = 26.731$, $p \leq 0.001$; $F_2(1,11) = 15.843$, $p \leq 0.002$) and between the *with*-high and *with*-low conditions ($F_1(1,9) = 13.157$, $p \leq 0.006$; $F_2(1,11) = 31.354$, $p \leq 0.0001$). Like in the norming study for Experiment 1, judgements seemed to be generally better for the *of*-condition than for the *with*-condition, but this effect did not reach statistical significance (low disambiguation: $F_1(1,9) = 3.329$, $p \leq 0.101$; $F_2(1,11) = 0.040$, $p \leq 0.846$; high disambiguation: $F_1(1,9) = 3.576$, $p \leq 0.091$; $F_2(1,11) = 0.518$, $p \leq 0.487$).

$p \leq 0.007$), and no significant effect for Preposition ($F_1 \leq 1$; $F_2 \leq 1$).[49] Since the weak low attachment preference is in line with the naturalness judgements, plausibility judgements seem to be influenced by the structural properties of the sentence. But the relatively small standard deviations and the lack of a significant effect for Attachment in the plausibility judgements suggest that the sentences did not considerably differ in their real-world plausibility. Hence, this factor should not be responsible for any effect observed in the actual study.

5.4.3 Method

The method used in this study was a word-by-word SPR task (see Chapter 4.2.). The sentences were the same as for the naturalness and plausibility ratings (see examples (57 a-d)). In order to make the sentences within a condition comparable, the sentences and regions of analysis were matched for word and syllable number.[50] The attachment DPs were matched for frequency (see Chapter 4.3.). Furthermore, relatively simple and frequent nouns were chosen in order to avoid confusing or biasing the learners by difficult or unfamiliar words.[51]

Two lists were created so that each participant saw a sentence either in high or low disambiguation. In addition to the 16 experimental sentences, each participant read 16 sentences from Experiment 4 (Chapter 6.1.) and 24 unrelated filler sentences. The experiment was preceded by instructions in English and 8 practice items (see Appendix B3). There was a break after half of the sentences before participants could start the second part of the experiment. The sentences were pseudo-randomised for each participant in such a way that no two experimental sentences directly followed one another. After reading a sentence word by word, participants were instructed to push the keys marked "yes" or "no" to indicate their answers to the comprehension questions.

Following the SPR task, participants completed a pencil-and-paper grammaticality judgement task to check whether they mastered this type of sentence and were sensitive to the disambiguation via gender agreement. The sentences for this task were taken from the pool of sentences for the on-line reading task. There were four types of gram-

[49] A separate one-factor ANOVA with four levels supported this finding, in that it did not yield a significant main effect ($F_1(3,9) = 1.329$, $p \leq 0.285$; $F_2(3,11) = 2.179$, $p \leq 0.109$).

[50] All experimental sentences as well as a table with the word and character counts per region can be found in Appendix B3.

[51] When presented with the vocabulary lists before the actual study, none of the learners asked for the meaning of any of the attachment DPs. Therefore, the observed preferences cannot be attributed to participants not knowing the meaning of any of these DPs.

matical sentences (see (53a-d) and four types of ungrammatical (see (54a-d) sentences, each with two items (see Appendix B4 for a complete list).

(53) a. of-fem.-fem.:
The student had liked the secretary of the woman who had killed herself in the office.
b. of-masc.-masc.:
The passenger smiled at the brother of the pilot who was writing himself a postcard.
c. with-fem.-fem.:
The photographer ignored the duchess with the daughter who introduced herself to the actor.
d. with-masc.-masc.:
The gardener disliked the rich man with the son who bankrupted himself in the casino.

(54) a. of-fem.-masc.:
The woman knew the niece of the actress who was admiring himself a bit too much.
b. of-masc.-fem.:
The policeman arrested the nephew of the salesman who had allowed herself a long holiday.
c. with-fem.-masc.:
The doctor examined the small girl with the mother who had injured himself with a knife.
d. with-masc.-fem.:
The journalist interviewed the politician with the advisor who criticised herself far too often.

The 16 experimental items were mixed with 24 unrelated filler sentences in such a way that no two experimental sentences directly followed one another. 20 of the filler sentences contained various grammatical mistakes which were unrelated to gender agreement and four of the filler sentences were grammatical, so that half of the total 40 sentences were grammatical and half of them were ungrammatical.

5.4.4 Results

Since it is decisive for the interpretation of the results whether participants mastered the particular type of experimental sentences and were sensitive to the disambiguation via gender information on the reflexive pronoun, the results of the off-line grammaticality judgements are reported first. Subsequently, the results from the on-line reading task, namely the answers to the comprehension questions and the word-by-word reading times, are considered.

To begin with, it must be noted that the learners' results in the grammaticality judgements significantly correlated with the other proficiency measures, i.e. the time they had been learning English ($r(60) = 0.292$, $p \leq 0.021$), the time they had spent in English-speaking countries ($r(60) = 0.291$, $p \leq 0.022$), and their performance in the grammar test ($r(60) = 0.325$, $p \leq 0.010$). These correlations show that the grammaticality judgement task was a reliable indicator of the learners' proficiency.

In the filler sentences of the grammaticality judgement task, all three groups performed at ceiling level, i.e. they gave more than 97.5% correct answers. This result implies that all participants understood what they were supposed to do and were paying attention to the task. The judgements of the experimental items were also successful, but more heterogeneous. The results are summarised in Table 5-4-3:

	all exp. items	***Of*-condition**	***With*-condition**	**Fillers**
Intermed. learners	88.33 (11.45)	87.92 (14.11)	88.75 (12.44)	97.50 (4.81)
Advanced learners	93.49 (5.28)	93.33 (6.34)	92.91 (7.10)	98.96 (2.37)
Native speakers	96.88 (3.58)	96.67 (5.62)	97.03 (5.38)	99.58 (1.59)

Table 5-4-3: Accuracy rates and (standard deviations) in the grammaticality judgements in Exp. 2

An ANOVA with Group as a between-subjects variable for all sentences yielded a significant main effect ($F(2,87) = 5.662$, $p \leq 0.005$). Tukey post-hoc tests showed that the difference between the intermediate and the advanced learners was marginally significant ($p \leq 0.045$). The difference between the intermediate learners and the native speakers was highly significant ($p \leq 0.0001$), whereas no significant difference was found between the advanced learners and the native speakers ($p \leq 0.702$). This finding and the comparatively high standard deviations for the intermediate learners suggest that the intermediate learners were not as successful as the other groups in this task. Nevertheless, their judgements were still sufficiently successful to suggest that they mastered the particular sentence type and were sensitive to the disambiguation via gender information on the reflexive pronoun. In order to test whether participants' grammaticality judgements were influenced by the preposition linking the two attachment DPs, the results for these two conditions were submitted to a mixed two-factor ANOVA with Preposition as a within-subjects variable and Group as a between-subjects variable. The analysis yielded a significant main effect for Group ($F(2,87) = 9.593$, $p \leq 0.0001$), but not for Preposition ($F \leq 1$) and no significant interaction ($F \leq 1$). So the preposition did not have a significant influence on participants' grammaticality judgements. The result is in line with the findings from the norming study, in which participants did not show a clear preference for any of the two constructions (see Chapter 5.4.2.).

The SPR task provided three measures: (a) the accuracy rates for the comprehension questions, (b) the time participants took to answer the questions, and (c) the word-by-word reading times. The results of these three measures are presented in turn in the following paragraphs.

The mean percentages of correct answers and standard deviations for each of the three groups are summarised in the following table:[52]

	of high	of low	with high	with low	total
Intermed. learners	89.2 (17.0)	92.5 (11.7)	79.2 (22.8)	95 (10.2)	88.9 (28.2)
Advanced learners	81.7 (19.6)	91.7 (15.2)	88.3 (17.0)	90.8 (19.1)	90.6 (58.8)
Native speakers	89.2 (17.0)	87.5 (18.3)	84.2 (23.2)	86.7 (21.5)	88.1 (10.5)

Table 5-4-4: Accuracy rates and (standard deviations) in the on-line comprehension questions in Exp. 2

The total percentages of correct answers for the experimental items were submitted to a three-factor ANOVA with Group as a between-subjects variable. The analysis did not yield a significant effect ($F \leq 1$), which suggests that all groups were equally accurate in the comprehension questions. Although it may prima facie seem striking that the advanced learners were even slightly more successful than the native speakers in this part of the task, the finding is in line with comparative experiments (e.g. Juffs & Harrington 1996; Rah & Adone to appear).[53]

To further explore the results of the comprehension questions, the four conditions must be considered. A mixed three-factor ANOVA with Preposition (of vs. with) and Attachment (high vs. low) as within-subjects / items variables and Group as between-subjects / items variable yielded a significant effect for Attachment ($F_1(1,87) = 13.147$, $p \leq 0.0001$; $F_2(1,21) = 4.820$, $p \leq 0.018$) but not for Preposition ($F_1 \leq 1$; $F_2(1,21) = 1.678$, $p \leq 0.211$) or Group ($F_1 \leq 1$; $F_2 \leq 1$).[54] However, the interaction of Attachment x Group was significant in the analysis by subjects ($F_1(2,87) = 3.215$, $p \leq 0.045$; $F_2 \leq 1$).[55] The interaction indicates that both learner groups were more successful in the low than in the high attachment conditions, whereas the native speakers did not show such an effect. This finding suggests that it is comparatively easy for the learners to answer a question about the most recently processed element.

To analyse the findings more closely, separate repeated-measures ANOVAs with the factors Preposition and Attachment were conducted for each group. For the intermediate

[52]The total accuracy rates for experimental items and fillers patterned closely together: The intermediate learners answered 93.7% of the questions correctly (SD = 2.5), the advanced learners 93.4% (SD = 5.1) and the native speakers 94.7% (SD = 4.0).

[53]These studies suggest that native speakers read comparatively quickly, without paying as much attention as the learners. Therefore, they might make more mistakes in relation to their actual proficiency.

[54]These findings were supported by a Group x Condition ANOVA, which yielded a significant effect for Condition ($F(3,261) = 3.553$, $p \leq 0.040$), but not for Group ($F(3,87) = 0.280$, $p \leq 0.757$), while the interaction approached significance ($F(6,261) = 2.136$, $p \leq 0.060$). Planned pair-wise comparisons revealed significant differences only between the *of*-high and the *of*-low conditions ($F(1,87) = 5.873$, $p \leq 0.017$) and the *with*-high and the *with*-low conditions ($F(1,87) = 6.269$, $p \leq 0.014$).

[55]None of the other interactions reached statistical significance: Preposition x Group ($F(1,87) = 1.585$, $p \leq 0.211$), Preposition x Attachment ($F(1,87) = 0.537$, $p \leq 0.466$), Preposition x Attachment x Group ($F(1,87) = 1.934$, $p \leq 0.151$).

learners, the analysis yielded a significant effect for Attachment only in the analysis by subjects ($F_1(1,29) = 10.796$, $p \leq 0.003$; $F_2 \leq 1$), and no significant effect for Preposition ($F_1(1,29) = 1.442$, $p \leq 0.240$; $F_2(1,7) = 1.961$, $p \leq 0.204$). But the interaction between these factors was significant, at least in the analysis by subjects ($F_1(1,29) = 5.241$, $p \leq 0.030$; $F_2 \leq 1$).[56] The same analysis of the advanced learners' data only yielded a significant effect for Attachment in the analysis by subjects ($F_1(1,29) = 14.032$, $p \leq 0.001$; $F_2(1,7) = 1.658$, $p \leq 0.239$).[57] For the native speakers, no significant effects were found (Attachment: $F_1 \leq 1$; $F_2(1,7) = 2.775$, $p \leq 0.239$; Preposition: $F_1 \leq 1$; $F_2 \leq 1$; Preposition x Attachment: $F_1 \leq 1$; $F_2 \leq 1$). The results taken together lead to the conclusion that the learners did not have more difficulties with the comprehension questions than the native speakers. Moreover, the respective condition only had a small influence on participants' answers. These findings show that all learners paid attention to the task and were sufficiently advanced to comprehend the experimental sentences. Such a result is particularly important because the word-by-word presentation requires keeping a relatively high number of words in mind until the end of the sentence in order to successfully answer the comprehension question. However, the separate analyses for each group indicate that the learners were more influenced by recency than the native speakers. A similar tendency shows in the time participants took to answer the comprehension questions, as illustrated in Table 5-4-5:[58]

	of high	**of low**	**with high**	**with low**
Intermed. learners	2718.5 (1499.41)	2323.31(495.79)	2581.56 (873.82)	2301.73 (665.03)
Adv. learners	2789.64 (1021.77)	2397.14 (835.94)	2479.36 (790.73)	2339.13 (609.59)
Native speakers	2057.90 (644.56)	1950.27 (600.58)	2466.06 (967.57)	1924.85 (660.27)

Table 5-4-5: Response times in ms and (standard deviations) for correctly answered questions in Exp. 2

A mixed three-factor ANOVA with Preposition (of vs. with) and Attachment (high vs. low) as within-subjects / items variables and Group as a between-subjects / items variable yielded a significant effect for Attachment ($F_1(1,87) = 18.771$, $p \leq 0.0001$;

[56] A one-factor ANOVA with Condition as a four-level factor found similar results, i.e. a significant main effect in the analysis by subjects ($F_1(3,87) = 5.638$, $p \leq 0.001$; $F_2(3,21) = 0.528$, $p \leq 0.668$). In planned comparisons only the difference between the *with*-high and *with*-low conditions turned out to be significant ($F_1(1,29) = 16.644$, $p \leq 0.0001$; $F_2(1,7) = 0.296$, $p \leq 0.603$).

[57] There was neither a significant main effect for Preposition ($F_1(1,29) = 1.613$, $p \leq 0.214$; $F_2(1,7) = 1.023$, $p \leq 0.345$), nor a significant interaction of Preposition x Attachment ($F_1(1,29) = 1.083$, $p \leq 0.307$; $F_2(1,7) = 0.104$, $p \leq 0.756$). An additional one-factor ANOVA with Condition yielded a significant main effect in the analysis by subjects ($F_1(3,87) = 2.292$, $p \leq 0.038$; $F_2(3,21) = 1.113$, $p \leq 0.336$). Planned comparisons found a significant difference between the *of*-high and *of*-low conditions ($F_1(1,29) = 7.250$, $p \leq 0.012$; $F_2(1,7) = 1.724$, $p \leq 0.231$).

[58] The response times in Table 5-4-5 only include correctly answered questions (see Table 5-4-4 above). Outliers, i.e. values deviating more than 2.5 standard deviations from the mean for each group and each item were also excluded from the analysis. This procedure affected 4.9% of the intermediate learners' data, 4.5% of the advanced learners' data, and 3.8% of the native speakers' data.

$F_2(1,21) = 14.206$, $p \leq 0.001$) and a marginally significant effect for Group ($F_1(2,87) = 2.940$, $p \leq 0.058$; $F_2(2,21) = 13.823$, $p \leq 0.0001$). No significant main effect for Preposition was found ($F_1 \leq 1$; $F_2 \leq 1$). But the interactions of Preposition x Group ($F_1(2,87) = 3.324$, $p \leq 0.041$; $F_2(2,21) = 3.049$, $p \leq 0.069$) and Preposition x Attachment x Group were significant ($F_1(2,87) = 3.905$, $p \leq 0.024$; $F_2(2,21) = 4.198$, $p \leq 0.039$).[59]

In order to analyse the findings more closely, separate repeated-measures ANOVAs with Preposition and Attachment were conducted for each group, respectively with subjects and items as random variables. For the intermediate learners, these analyses yielded a significant effect for Attachment ($F_1(1,29) = 6.167$, $p \leq 0.019$; $F_2(1,7) = 5.807$, $p \leq 0.047$), but not for Preposition ($F_1 \leq 1$; $F_2 \leq 1$), and no significant interaction ($F_1(1,29) = 1.026$, $p \leq 0.319$; $F_2(1,7) = 2.224$, $p \leq 0.180$). The same analysis of the advanced learners' data yielded similar results. There was a significant effect for Attachment ($F_1(1,29) = 7.032$, $p \leq 0.013$; $F_2(1,7) = 7.477$, $p \leq 0.029$), but not for Preposition ($F_1(1,29) = 2.745$, $p \leq 0.109$; $F_2(1,7) = 2.186$, $p \leq 0.183$), and no significant interaction between these factors ($F_1 \leq 1$; $F_2 \leq 1$). For the native speakers, there was also a significant effect for Attachment ($F_1(1,29) = 6.427$, $p \leq 0.017$; $F_2(1,7) = 3.300$, $p \leq 0.112$). In contrast to the other groups, the native speakers evinced a marginally significant effect for Preposition ($F_1(1,29) = 3.944$, $p \leq 0.057$; $F_2(1,7) = 2.576$, $p \leq 0.153$) and a significant interaction ($F_1(1,29) = 10.237$; $p \leq 0.004$; $F_2(1,7) = 3.621$, $p \leq 0.099$), at least in the analysis by subjects. These findings imply that only the native speakers' low attachment preference was significantly stronger in the with-condition than in the of-condition. To sum up, both the response times and the accuracy rates show a similar preference for the sentences disambiguated towards low attachment in all three groups. The following paragraphs focus on the third measure provided in this task, i.e. the word-by-word reading times.

Preceding the analyses of the reading times, outliers were excluded following the standard procedure for this type of task (see Chapter 4.5.). To begin with, all results from the comprehension questions were checked for correctness. Since no participant answered less than 80% of all questions incorrectly, no participants were excluded because of this factor. Then, the mean reading times by word were calculated for each participant. The results of two of the 32 intermediate learners had to be excluded from further analyses because their mean reading times by word exceeded the mean of this group by more than 3.0 standard deviations. No participants from any of the other two

[59] There were no significant interactions of Preposition x Attachment ($F_1 \leq 1$; $F_2 \leq 1$) and Group x Attachment ($F_1 \leq 1$; $F_2 \leq 1$).

groups were excluded on account of this criterion. In a next step, all sentences for which the comprehension question had been answered incorrectly were excluded from further analysis. This was important because wrong answers might reveal that participants had not completely understood the sentences or had not paid full attention to the task. The procedure led to the deletion of 16.9% of the intermediate learners' data, 11.5% of the advanced learners' data, and 12.3% of the native speakers' data. The last step was to exclude outliers, i.e. data points where a participant temporarily lost concentration for reasons unrelated to the experiment during the on-line reading task. Each group's responses were screened for data points which were 2.5 standard deviations above or below the mean for each region of every sentence. This procedure led to the deletion of 3.1% of the intermediate learners' data, 2.7% of the advanced learners' data, and 2.9% of the native speakers' data. Thus, a total of 20.0% of the intermediate learners' data, 14.2% of the advanced learners' data, and 15.2% of the native speakers' data were excluded from further analyses. The resulting reading times and standard deviations per group and per region in the four conditions are summarised in Appendix E2.

The first region of interest is the disambiguating region, i.e. the reflexive pronoun himself / herself. Since this region is comparatively short, it is possible that surprise effects are not found on the disambiguating element itself, but also in the following wrap-up region, especially if participants' word-by-word reading times are comparatively fast. Thus, the reading times on the reflexive pronoun (region 7) and on the wrap-up segment (region 8) were submitted to statistical analyses. A mixed three-factor ANOVA with Preposition (of vs. with) and Attachment (high vs. low) as within-subject variables and Group as a between-subjects variable was calculated for each of the two regions. The results are summarised in the following table:

	disambiguating region	**wrap-up region**
Attachment	$F_1(1,87) = 27.826$, $p \leq 0.0001$	$F_1(1,87) = 72.569$, $p \leq 0.0001$
	$F_2(1,21) = 24.395$, $p \leq 0.0001$	$F_2(1,21) = 49.682$, $p \leq 0.0001$
Preposition	$F_1 \leq 1$	$F_1(1,87) = 4.269$, $p \leq 0.042$
	$F_2 \leq 1$	$F_2(1,21) = 3.374$, $p \leq 0.047$
Group	$F_1 \leq 1$; $F_2(2,21) = 2.159$, $p \leq 0.140$	$F_1 \leq 1$; $F_2(2,21) = 2.882$, $p \leq 0.078$
Attachment x Group	$F_1(2,87) = 2.758$, $p \leq 0.069$	$F_1(2,87) = 6.891$, $p \leq 0.002$
	$F_2(2,21) = 5.140$, $p \leq 0.015$	$F_2(2,21) = 3.681$, $p \leq 0.043$
Attachment x Prep.	$F_1 \leq 1$	$F_1(2,87) = 4.660$, $p \leq 0.012$
	$F_2 \leq 1$	$F_2(2,21) = 2.087$, $p \leq 0.056$

Table 5-4-6: Prep x Att x Group ANOVA on the RTs of disambiguating and wrap-up region in Exp. 2

In order to further explore the interactions with the factor Group, the reading times of the three groups in the disambiguating region (see Figure 5-4-1) and the wrap-up region (see Figure 5-4-2) were considered separately.

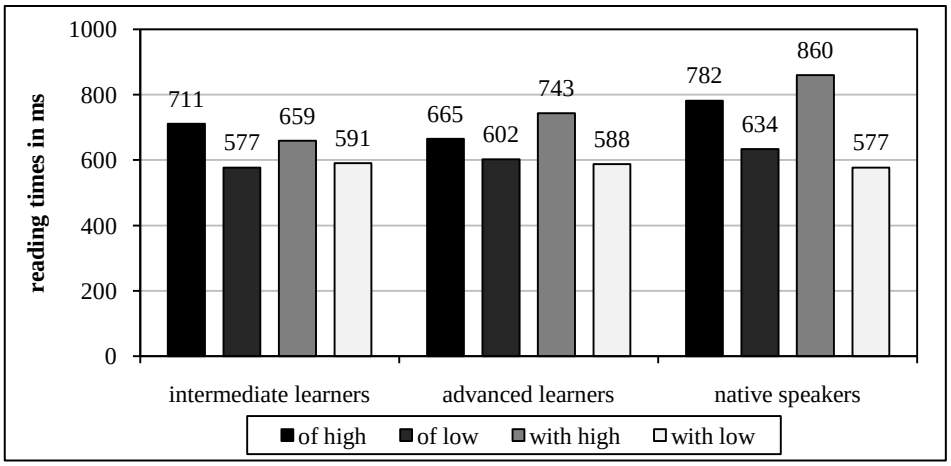

Figure 5-4-1: Reading times in the disambiguating region in Exp. 2

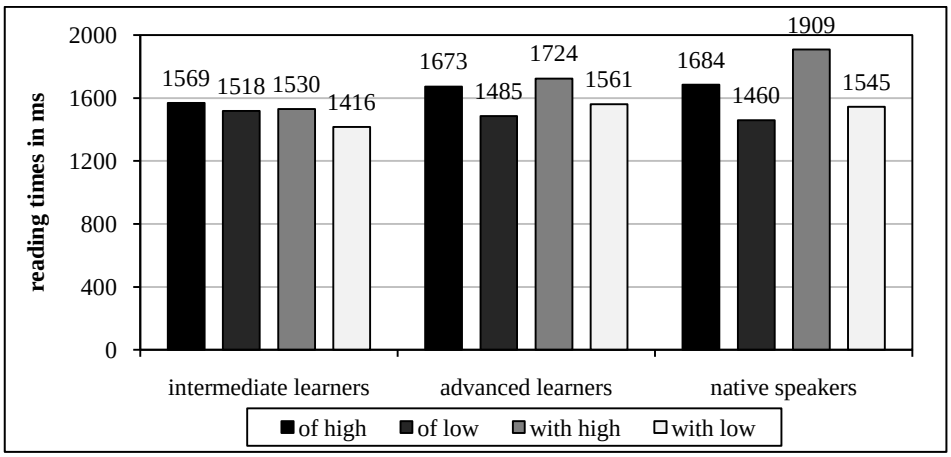

Figure 5-4-2: Reading times in the wrap-up region in Exp. 2

Separate two-factor ANOVAs with Preposition and Attachment as within-subjects factors were conducted for all three groups, respectively with subjects and with items as random variables. For the native speakers, this analysis for the disambiguating region yielded a significant effect for Attachment ($F_1(1,29) = 9.691$, $p \leq 0.004$; $F_2(1,7) = 10.026$, $p \leq 0.016$). There was no significant effect for Preposition ($F_1 \leq 1$; $F_2 \leq 1$) and no significant interaction ($F_1 \leq 1$; $F_2 \leq 1$). The analysis of the wrap-up region also yielded a significant effect for Attachment ($F_1(1,29) = 32.998$, $p \leq 0.0001$; $F_2(1,7) = 18.032$, $p \leq 0.004$). In this region, the effect for Preposition approached significance, at least in the analysis by subjects ($F_1(1,29) = 4.126$, $p \leq 0.053$; $F_2(1,7) = 2.496$, $p \leq 0.158$). As in the disambiguating region, no significant interaction was found ($F_1 \leq 1$; $F_2 \leq 1$). These results imply that the low disambiguations were read faster than high disambiguations in both conditions.

Half of the native controls were speakers of British English, the other half were speakers of American English. Although a low attachment preference was found for both varieties in the off-line study (see Chapter 5.3.), it is possible that differences appear in the on-line study. Therefore, an additional mixed three-factor ANOVA was conducted on the native speakers' results with Preposition and Attachment as within-subjects factors and Variety as a between-subjects factor. The analyses yielded no significant main effects for Variety, neither in the disambiguating region ($F(1,25) = 0.767$, $p \leq 0.390$) nor in the wrap-up region ($F(1,25) = 0.366$, $p \leq 0.550$), and no significant interactions with this factor (disambiguating region: Preposition x Variety: $F \leq 1$, Attachment x Variety: $F \leq 1$, Preposition x Attachment x Variety: $F \leq 1$; wrap-up region: Preposition x Variety: $F \leq 1$, Attachment x Variety: $F \leq 1$, Preposition x Attachment x Variety: $F \leq 1$).

Figures 5-4-1 and 5-4-2 imply that the two learner groups' patterns resemble those of the English native speaker control group. For the advanced learners, two-factor ANOVAs with Preposition and Attachment as within-subject factors and as within-items factors yielded a significant effect for Attachment in the disambiguating region ($F_1(1,29) = 13.391$, $p \leq 0.001$; $F_2(1,7) = 6.573$, $p \leq 0.037$). However, there was no significant effect for Preposition ($F_1(1,29) = 1.614$, $p \leq 0.214$; $F_2(1,7) = 1.515$, $p \leq 0.258$) and no significant interaction ($F_1(1,29) = 1.275$, $p \leq 0.268$; $F_2(1,7) = 1.160$, $p \leq 0.317$). The analyses for the wrap-up region also yielded a significant effect for Attachment ($F_1(1,29) = 32.715$, $p \leq 0.0001$; $F_2(1,7) = 26.253$, $p \leq 0.001$), but here the effect for Preposition was also significant, at least in the analysis by subjects ($F_1(1,29) = 6.253$, $p \leq 0.019$; $F_2 \leq 1$). No significant interaction was found here ($F_1 \leq 1$; $F_2 \leq 1$).

The analysis of the intermediate learners' reading times in the disambiguating region also yielded a significant effect for Attachment ($F_1(1,29) = 25.900$, $p \leq 0.0001$; $F_2(1,7) = 12.591$, $p \leq 0.009$), but no significant effect for Preposition ($F_1 \leq 1$; $F_2 \leq 1$) and no significant interaction ($F_1 \leq 1$; $F_2(1,7) = 2.781$, $p \leq 0.143$). Similarly, the analyses for the wrap-up region found a significant effect for Attachment ($F_1(1,29) = 6.675$, $p \leq 0.015$; $F_2(1,7) = 17.665$, $p \leq 0.004$), but no significant effect for Preposition ($F_1(1,29) = 2.588$, $p \leq 0.119$; $F_2 \leq 1$). No significant interaction was found between the two factors ($F_1 \leq 1$; $F_2(1,7) = 1.593$, $p \leq 0.247$).

5.4.5 Discussion

Since the results of Experiment 2 differ from the off-line results of Experiment 1 (see Chapter 5.3.) and some previous studies (see Chapter 3.2.4.), the present chapter dis-

cusses the reliability and generalisability of the results as well as their implications for the research questions formulated in Chapter 5.2..

The low attachment preference in the reading times was in line with the tendency found in the comprehension questions. Especially for the learners it seemed to be easier to answer the questions for sentences in the low attachment conditions. This finding was unexpected, considering that in high disambiguated sentences the first DP is more prominent in discourse and thus might be easier to remember. It is also noteworthy that the error rates were relatively high in the experimental sentences in comparison to the fillers, even in the native speaker control group. In order to better understand these results of the comprehension questions, the implications of the respective response options must be considered. If a comprehension question is answered with "yes", this indicates that the participant processed the sentence without difficulty on the first parse. A "no" answer, however, could be interpreted in two ways. First, it might point to a (temporary) processing difficulty because the sentence was not in line with the participant's preferred attachment. The second possibility is that the participant understood the sentence incorrectly. In either case, the answer might have been influenced by an attachment preference. The existence of different possible sources of errors might at least partly explain the relatively high error rates even in the native speaker controls, which were also reported for similar experiments (e.g. Felser et al. 2003, Fernández 2003).

The preference for low disambiguated sentences in the comprehension questions also became manifest in participants' reading times. All three groups evinced a clear low attachment preference, both in the of-conditions and in the with-conditions.[60] While the native speakers' universal low attachment preference and all groups' low attachment preference in the with-condition are in line with previous findings, the learners' preference in the of-condition is rather unexpected and requires further discussion. The advanced learners' lack of a clear preference and the intermediate learners' high attachment preference in the off-line study were attributed to L1 transfer effects. It is surprising that no such effects were found in the on-line task because L1 influence would be expected especially in situations of time pressure as in the on-line task. There are several explanations for this finding. First, it is possible that transfer effects only occur in

[60] Interestingly, the effects of the disambiguating region were replicated in the wrap-up region for all groups but the intermediate learners. This might be related to reading speed differences. Since the intermediate learners took comparatively long to read the disambiguating word, they had probably integrated the information when reaching the wrap-up region. The other groups were faster, so that the effects on the disambiguating word spilled over to the following region.

off-line processing and that the learners' on-line attachment preferences are the result of a recency strategy. Second, the German high attachment preference might not be a robust effect, with the consequence that the learners' L1 transfer in the present study does not necessarily lead to a high attachment preference. Third, a bias in the materials or in the method of the present study might be responsible for the results. These three possibilities are discussed in turn in the following paragraphs.

It has been suggested that cross-linguistic differences only appear off-line and thus do not influence on-line processing at all (see Chapter 2.2.3.). Such an account is in line with the findings reported by Fernández (2003), who claims that cross-linguistic differences only appear off-line and are caused by non-structural factors. However, as discussed in Chapter 3.2.4.1, such an account would contradict studies that found an on-line high attachment preference with various languages, methods and materials.

An alternative explanation for the present findings would be that the high attachment preference in German is not as robust as generally assumed. Hence, L1 transfer would not necessarily lead to a high attachment preference in the learners of the present study. Support for this explanation comes from Augurtzky (2006), who did not find any clear preferences for German genitive sentences, neither in an off-line questionnaire nor in on-line SPR and ERP studies. However, a robust high attachment preference was found in on-line studies by Hemforth and colleagues and by the off-line studies reported in Chapter 5.1. Taken together these findings are indicative of a tendency to attach high in German which is weakened by some factor in Augurtzky's experiments, for example dialectal variation (see the discussion in Chapter 5.1.). In the light of these findings, it is improbable that the learners' clear low attachment preference in the present on-line study can be attributed to L1 transfer.

Another possibility is that the present study was biased towards low attachment. It is improbable that such a bias was inherent in the materials because they were carefully chosen, constructed and evaluated. Hence, it is more plausible that the way of presentation was the decisive factor. While the present study and the SPR task reported by Augurtzky (2006) used a word-by-word presentation, all other comparable studies presented the sentences either chunk-wise or, in the case of eye-tracking or ERP studies, as a whole. It is conceivable that the word-by-word presentation induces higher WM load and thus causes a recency preference, especially in the German ESL learners. Such an account would be in line with findings reported by Frenck-Mestre & Pynte (1997, see Chapter 3.2.3.) for a different construction. Furthermore, a word-by-word SPR study on

RC attachment preferences with Japanese ESL learners (Omaki 2005) found a weak low attachment preference.[61] The lack of any preference reported by Felser et al. (2003) might thus be related to the chunk-wise presentation in this experiment. Furthermore, processing load in the Felser et al. (2003) study might have been reduced because the participants had completed the off-line questionnaire before the SPR task and thus were already used to the type of experimental sentences. Another finding supporting this explanation is the discrepancy between the WM span experiments by Swets et al. (2007) and Traxler (2007) (see also Chapter 3.1.3.). Whereas in the former off-line study, the high span individuals showed the strongest low attachment preference, the on-line eye-tracking study by Traxler (2007) found exactly the contrary, i.e. the low span readers had the strongest low attachment preference. Under the assumption that WM demands of the specific task influence participants' processing strategies, it is conceivable that the on-line task forced the low-span participants to employ a recency strategy while this was not the case in the off-line task.

Although the low attachment preference found for the ESL learners in the present study prima facie contradicts the SSH, which predicts no preference for these constructions, it is possible that learners who do not process deep syntactic structures employ a default strategy. Such a default might be a preference for the most recently processed element, especially if the construction is long and difficult and thus imposes high processing load. Such an account would also explain the discrepancy between the present off-line and the on-line experiment. Whereas in the on-line task time pressure forces participants to use a strategy which reduces processing load, the off-line task allows them to reflect upon their decisions. Hence, they choose the option that is preferred in their L1.

Although this explanation seems plausible, it is difficult to test experimentally. One possibility to investigate this issue would be to translate the experimental sentences into German and to run the study with native speakers of German to determine whether the high attachment preference reported by Hemforth et al. (1998) can be replicated with the present materials. However, a one-to-one translation of the English materials into German is impossible because German reflexive pronouns are neutral with regard to number and gender and thus cannot serve as disambiguating elements. Furthermore, the

[61] The fact that this difference did not reach statistical significance in Omaki's study might be related to the additional manipulation of the length of the RCs. Interestingly, the low attachment preference was only found in the long condition, which indicates that the results are to some extent influenced by WM load imposed by the respective constructions.

German relative pronoun is marked for gender, so that the sentence would already be disambiguated on the relative pronoun. Thus, the results of such a study would not be fully comparable to those with the English materials. A more promising possibility to further investigate the findings of the present study would be to replicate the study with chunk-wise instead of word-by-word presentation. The reading of larger elements might reduce WM load and thus the necessity for a recency strategy. If the learners' low attachment preference in the present study was caused by the way of presentation, the low attachment preference should be weaker when larger elements are presented. The native speakers, however, are expected to show the same low attachment preference as before. The next chapter presents an experiment to test these predictions.

5.5 Experiment 3: On-line RC Attachment II

The aim of Experiment 3 is to further investigate L2 learners' on-line RC attachment preferences. As discussed in Chapter 5.4.5. it is possible that the way of presentation and the resulting processing load manipulate L2 learners' RC attachment preferences. The present study tests this possibility by replicating Experiment 2 with a chunk-wise presentation.

While the native speakers are still expected to show a low attachment preference in the present study, there are different possible predictions for the learners. Firstly, it is conceivable that they show a low attachment preference as in Experiment 2. Such a result could either be attributed to the mastery of native-like processing or to a preference to attach to the most recently processed element in order to reduce processing load. Secondly, the learners might show a high attachment preference as in Experiment 1, which might be attributed to L1 transfer. Such a finding would imply that the low attachment preference in Experiment 2 was a strategy to reduce processing load in the more demanding task.

Since the present study presents the sentences in the same chunks as the study reported by Felser et al. (2003, see Chapter 3.2.4.2.), the finding of a high or a low attachment preference in the present study would seriously challenge the results of Felser et al. and thus the SSH. If the present study does not find a clear attachment preference, this would be in line with the SSH. But, like the results of Felser et al., the individual participants' results would have to be analysed because the reason for the absence of a significant preference might be that some learners prefer high attachment while others prefer low attachment.

The present study thus permits to further investigate the questions that were raised by Experiment 2 and to test the predictions of different L2 processing theories. Hence, the study addresses the following research questions:

A) Are ESL learners influenced by their L1 or by an L2-specific strategy when processing RC attachment ambiguities?

B) To what extent are L2 learners' RC attachment preferences influenced by the way of presentation?

5.5.1 Participants

Participants were a group of intermediate ESL learners (n = 20) and a native speaker control group (n = 20)[62] who had not taken part in any of the other reported studies. The biographical data of the learners are summarised in Table 5-5-1:[63]

Group		Age	Years of learning English	Months in English-speaking countries	Grammar
Intermediate learners (n = 20)	**mean**	21.5	10.9	___	77.4
	SD	(1.2)	(1.4)		(6.6)
	range	19-24	7-14		65-82

Table 5-5-1: Biographical data of the learners in Experiment 3

5.5.2. Materials and Method

Materials and method were the same as in Experiment 2 (see Chapter 5.4.), except that the sentences in Experiment 3 were not presented word-by-word but chunk-wise.

(55) a.	The student	had liked	the secretary of the professor	who	had killed	herself	in the office.
b.	The student	had liked	the secretary of the professor	who	had killed	himself	in the office.
c.	The student	had liked	the professor with the secretary	who	had killed	himself	in the office.
d.	The student	had liked	the professor with the secretary	who	had killed	herself	in the office.
Region	1	2	3	4	5	6	7

The chunks were the same as in the Felser et al. (2003) study so that the two studies were directly comparable. Sentences and regions were matched for word and syllable number (see Appendix B3).

5.5.3 Results

As in Experiment 2, the results of the comprehension questions and the reading times are presented in turn. The accuracy rates for each of the three groups are summarised in the following table:[64]

[62] Half of them were native speakers of British English, the other half were speakers of American English. Their mean age was 28.5 years (range = 20-38, SD = 6.1), and 80% of them were female.

[63] There was a significant correlation between years of learning and grammar performance ($r(20) = 0.473$, $p \leq 0.035$).

	of high	*of* low	*with* high	*with* low	Total
Intermed. learners	85.0 (22.1)	80.0 (17.4)	87.5 (17.2)	87.5 (22.2)	85.3 (14.5)
Native speakers	90.0 (14.9)	81.3 (19.7)	80.0 (17.4)	91.3 (14.7)	85.6 (11.1)

Table 5-5-2: Mean accuracy rates and (standard deviations) in the comprehension questions in Exp. 3

The total percentages of correct answers for the experimental items were submitted to a one-factor ANOVA with Group as a between-subjects variable. This analysis did not yield a significant effect ($F(1,39) = 0.006$, $p \leq 0.936$), which suggests that both groups were approximately equally accurate in the comprehension questions. To analyse the accuracy data more closely, the four conditions were then considered separately. A mixed three-factor ANOVA with Preposition (of vs. with) and Attachment (high vs. low) as within-subjects variables and Group as a between-subjects variable did not yield any significant main effects (Attachment: $F(1,38) = 0.048$), $p \leq 0.828$, Preposition: $F(1,38) = 1.264$, $p \leq 0.271$, Group: $F(1,38) = 0.027$, $p \leq 0.871$). But there was a significant interaction of Attachment x Preposition ($F(2,38) = 6.507$, $p \leq 0.015$), which suggests that high-disambiguated sentences were judged more successfully than low-disambiguated sentences only in the of-conditions. None of the interactions with Group were significant (Preposition x Group: $F(1,38) = 1.264$, $p \leq 0.271$; Attachment x Group: $F(1,38) = 0.433$, $p \leq 0.515$). This finding is interesting in comparison to Experiment 2, which found a marginally significant Attachment x Group interaction that was attributed to a low attachment preference in the learners. In Experiment 3, in contrast, the learners preferred high attachment in the of-conditions. In order to further analyse the difference between the experiments, the results of the comprehension questions were compared for the two groups that had taken part in both experiments, i.e. the intermediate learners and the native speakers. The results are summarised in Table 5-5-3: [65]

		of high	*of* low	*with* high	*with* low
Intermediate learners	**Exp.2**	89.2 (17.0)	92.5 (11.7)	79.2 (22.8)	95.0 (10.2)
	Exp.3	85.0 (22.1)	80.0 (17.4)	87.5 (17.2)	87.5 (22.2)
Native speakers	**Exp.2**	89.2 (17.0)	87.5 (18.3)	84.2 (23.2)	86.7 (21.5)
	Exp.3	90.0 (14.9)	81.3 (19.7)	80.0 (17.4)	91.3 (14.7)

Table 5-5-3: Accuracy rates and (standard deviations) in Experiments 2 and 3

Since the participants were not the same across the two experiments, the factor Experiment was treated as a between-subjects variable in the three-factor ANOVAs with

[64] Participants' total accuracy rates for the filler items were at ceiling level: 96.0 % (SD = 4.2) for the intermediate learners and 97.6 % (SD = 2.5) for the native speakers.

[65] Since comparisons of the total results were not statistically significant (Group: $F(1,99) = 0.015$, $p \leq 0.902$, Experiment: $F(1,99) = 1.988$, $p \leq 0.162$, Group x Experiment: $F(1,99) = 0.072$, $p \leq 0.790$), the results of the four conditions were analysed to further explore possible differences across studies.

Preposition and Attachment as within-subjects factors on the results. For the intermediate learners these analyses yielded no significant main effects, but a significant Attachment x Experiment interaction ($F(1,48) = 6.365$, $p \leq 0.015$). The interactions of Preposition x Experiment ($F(1,48) = 3.317$, $p \leq 0.075$) and Preposition x Attachment ($F(1,48) = 3.413$, $p \leq 0.071$) approached significance. These interactions indicate that the learners' results were more strongly influenced by Preposition and Attachment than in Experiment 3. For the native speakers, in contrast, only the interaction of Preposition x Attachment ($F(1,48) = 4.435$, $p \leq 0.040$) was significant, which indicates that the native speakers' results did not vary significantly between the two experiments.

The results of the comprehension questions indicate that the learners did not have more difficulties with the comprehension questions than the native speakers. Both groups showed only small differences between the conditions. This result implies that all participants paid attention to the task and were sufficiently advanced to comprehend the experimental sentences. In contrast to the findings from Experiment 2, where the learners' performance in the comprehension questions was more strongly affected by recency considerations than the native speakers' performance, such a tendency did not show in the present experiment. This finding indicates that the chunk-wise presentation did not require a learner-specific recency strategy. Further evidence is provided by the response times for the comprehension questions, which are illustrated in Table 5-5-4:[66]

	***of* high**	***of* low**	***with* high**	***with* low**
Intermed. learners	2158.98 (529.29)	2375.62 (611.54)	2492.29 (872.48)	2123.88 (449.74)
Native speakers	2156.83 (595.72)	1939.43 (551.59)	2380.59 (873.03)	2069.01 (429.09)

Table 5-5-4: Response times in ms and (standard deviations) for correctly answered questions in Exp. 3

Mixed two-factor ANOVAs with Attachment (high vs. low) as a within-subjects variable and Group as a between-subjects variable were run on the results for the of-condition and the with-condition in turn. In the of-condition no significant main effects were found, neither for Attachment ($F \leq 1$), nor for Group ($F(1,38) = 1.767$, $p \leq 0.192$), but there was a significant interaction between these factors ($F(1,38) = 6.193$, $p \leq 0.017$). In the with-condition there was a significant effect for Attachment ($F(1,38) = 9.743$, $p \leq 0.003$), but not for Group ($F \leq 1$), and no significant interaction ($F \leq 1$). To analyse these findings more closely, separate repeated-measures ANOVAs with the factors Preposition and Attachment were conducted for each group, respec-

[66]The response times only include correctly answered questions. Outliers were excluded from the analysis following the standard procedure (see Chapter 4.5.) The procedure affected 3.9% of the intermediate learners', 3.5% of the advanced learners' and 3.8 % of the native speakers' data.

tively with subjects and with items as a random variable. For the intermediate learners, these analyses neither yielded a significant main effect for Attachment ($F_1 \leq 1$; $F_2 \leq 1$), nor for Preposition ($F_1 \leq 1$; $F_2 \leq 1$), but a significant interaction ($F_1(1,19) = 9.897$, $p \leq 0.005$; $F_2(1,7) = 15.309$, $p \leq 0.006$). In contrast, the analyses of the native speakers' results yielded a significant effect for Attachment ($F_1(1,19) = 9.587$, $p \leq 0.006$; $F_2(1,7) = 6.731$, $p \leq 0.036$). There was neither a significant effect for Preposition ($F_1(1,19) = 1.438$, $p \leq 0.246$; $F_2(1,7) = 2.747$, $p \leq 0.141$), nor a significant interaction ($F_1 \leq 1$; $F_2 \leq 1$). These results confirm the impression that both groups preferred low attachment in the with-condition, while in the of-condition the learners tended to prefer high attachment. Although these tendencies are in line with the accuracy results, it must be considered that long response times are not necessarily the result of difficulties with a particular sentence but may be influenced by various other factors such as temporary lapses of concentration. Therefore, the following paragraphs focus on the third measure of this task, i.e. the word-by-word reading times, in order to test whether the tendencies found for the comprehension questions can be confirmed by a more immediate measure of participants' processing decisions.

Preceding the analyses of the reading times, outliers were eliminated following the same procedure as in Experiment 2. Since no participant had answered less than 80% of all questions correctly, no participants were excluded because of this factor. All sentences for which the comprehension question had been answered incorrectly were excluded. This procedure led to the exclusion of 15.0% of the learners' and 9.2% of the native speakers' data. The screening for data points that deviated by more than 2.5 standard deviations from the mean for each region of every sentence led to the additional deletion of 0.5% of the learners' and 1.1% of the native speakers' data. Thus, a total of 15.5% of the learners' and 10.3% of the native speakers' data were excluded from further analyses. The resulting mean reading times and standard deviations per group and per region in the four experimental conditions are summarised in Appendix E2.

As in Experiment 2, reading times on the disambiguating reflexive pronoun and the wrap-up region were submitted to separate mixed three-factor ANOVAs with Preposition (of vs. with) and Attachment (high vs. low) as within-subject variables and Group as a between-subjects variable. For the disambiguating region, this analysis yielded a significant effect for Attachment ($F(1,38) = 8.758$, $p \leq 0.005$), but not for Preposition ($F \leq 1$). No significant main effect was found for Group ($F(1,38) = 2.292$, $p \leq 0.139$), but some interactions reached significance: Group x Attachment ($F(1,38) = 4.539$,

$p \leq 0.004$), Attachment x Preposition ($F(1,38) = 8.429$, $p \leq 0.006$), and Attachment x Preposition x Group ($F(1,38) = 4.254$, $p \leq 0.046$). These interactions make the results difficult to interpret and require further analyses of the individual factors. The analysis of the wrap-up region only found a significant effect for Attachment ($F(1,38) = 9.743$, $p \leq 0.003$). There were no significant main effects for Preposition ($F \leq 1$) or Group ($F \leq 1$), and no significant interactions: Attachment x Group ($F(1,38) = 1.846$, $p \leq 0.182$), Preposition x Group ($F \leq 1$), Attachment x Preposition x Group ($F \leq 1$).

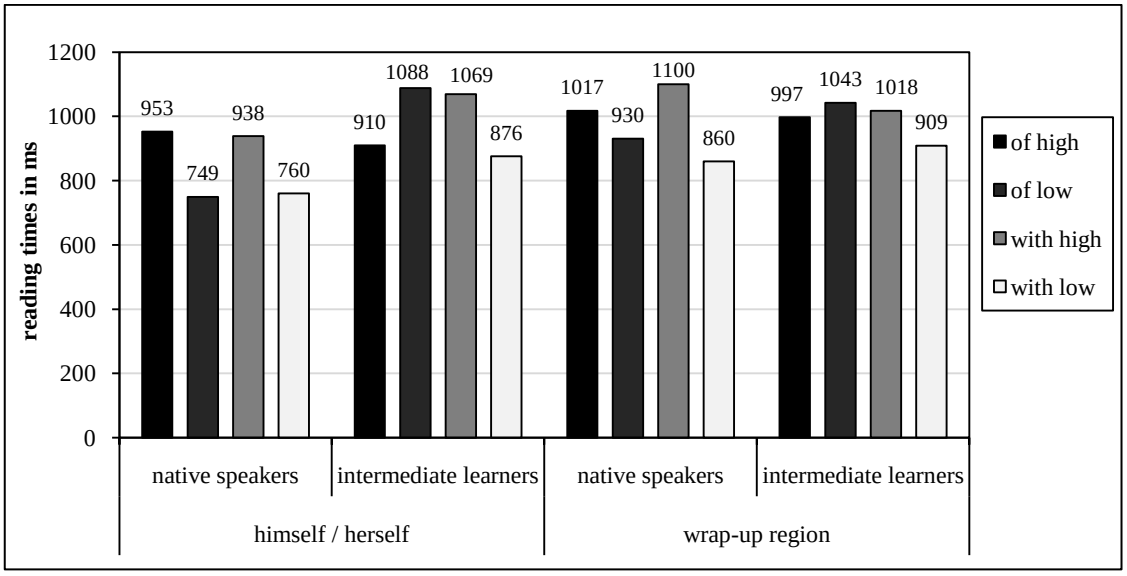

Figure 5-5-1: Reading times per group and per region in Experiment 3

In order to further explore the interactions, separate two-factor ANOVAs with Preposition and Attachment as within-subject factors were run for each group, respectively with subjects and items as random variables. For the native speakers, the analysis for the disambiguating region yielded a significant effect for Attachment ($F_1(1,19) = 29.036$, $p \leq 0.0001$; $F_2(1,7) = 6.707$, $p \leq 0.036$), but not for Preposition ($F_1 \leq 1$; $F_2 \leq 1$) and no significant interaction ($F_1 \leq 1$; $F_2 \leq 1$).[67] The analyses for the wrap-up region also yielded a significant effect for Attachment ($F_1(1,19) = 11.182$, $p \leq 0.003$; $F_2(1,7) = 8.690$, $p \leq 0.021$) but not for Preposition ($F_1 \leq 1$; $F_2 \leq 1$). As in the disambiguating region, there was no significant interaction between the two factors ($F_1(1,19) = 2.285$, $p \leq 0.147$; $F_2 \leq 1$).[68] These findings confirm the impression that the native speakers read the low attachment conditions faster than the high attachment conditions.

[67] An additional ANOVA with Condition as a within-subjects and Variety as a between-subjects factor found no significant effect for Variety ($F(1,18) = 3.120$, $p \leq 0.094$) and no significant interaction ($F(3,54) = 1.091$, $p \leq 0.338$), which indicates that the variety did not affect reading times in this region.

[68] As in the disambiguating region, there was no significant main effect for Variety ($F(1,18) = 0.378$, $p \leq 0.547$) and no significant interaction of Condition x Variety ($F(3,54) = 0.504$, $p \leq 0.653$), which indicates that the variety did not significantly influence the reading times of this region.

The analysis of the intermediate learners' reading times on the disambiguating word did not yield any significant main effects, neither for Attachment ($F_1 \leq 1$; $F_2 \leq 1$) nor Preposition ($F_1 \leq 1$; $F_2 \leq 1$), but a significant interaction ($F_1(1,19) = 9.684$, $p \leq 0.006$; $F_2(1,7) = 8.339$, $p \leq 0.023$). The analysis for the wrap-up region did not yield any significant effects either, neither for Attachment ($F_1(1,19) = 1.490$, $p \leq 0.250$; $F_2 \leq 1$), nor for Preposition ($F_1 \leq 1$; $F_2 \leq 1$). Moreover, no significant interaction was found ($F_1(1,19) = 1.269$, $p \leq 0.274$; $F_2(1,7) = 1.860$, $p \leq 0.751$). The interaction on the disambiguating word indicates that attachment preferences are influenced by the linking preposition. These results confirm the impression that the learners' reading times were shorter for forced high attachment in the *of*-condition, but shorter for forced low attachment in the *with*-condition. Such a high attachment preference in the of-condition contradicts the findings from Experiment 2, where the learners preferred low attachment in both conditions. Hence, it would be interesting to directly compare the results of the two experiments, at least on the disambiguating reflexive pronoun, which was identical in both experiments. Therefore, separate three-factor ANOVAs with Attachment and Preposition as within-subjects factors and Experiment as a between subjects factor were run on the results of the native speakers and the intermediate learners in turn. The results are summarised in Table 5-5-5:

	Intermediate learners	**Native speakers**
Attachment	$F(1,48) = 4.200$, $p \leq 0.046$	$F(1,48) = 17.331$, $p \leq 0.0001$
Preposition	$F(1,48) = 0.223$, $p \leq 0.639$	$F(1,48) = 0.193$, $p \leq 0.662$
Experiment	$F(1,48) = 40.039$, $p \leq 0.0001$	$F(1,48) = 0.751$, $p \leq 0.391$
Attachment x Preposition	$F(1,48) = 7.115$, $p \leq 0.010$	$F(1,48) = 0.437$, $p \leq 0.512$
Attachment x Experiment	$F(1,48) = 0.961$, $p \leq 0.332$	$F(1,48) = 0.531$, $p \leq 0.470$
Preposition x Experiment	$F(1,48) = 0.009$, $p \leq 0.924$	$F(1,48) = 0.064$, $p \leq 0.801$
Att. x Prep. x Experiment	$F(1,48) = 15.076$, $p \leq 0.0001$	$F(1,48) = 0.001$, $p \leq 0.996$

Table 5-5-5: Results of the Attachment x Preposition x Experiment ANOVAs on the disambiguating reflexive pronouns in Experiments 2 and 3.

As Table 5-5-5 indicates, only the learners showed a significant main effect for the factor Experiment and a significant Attachment x Preposition x Experiment interaction. Hence, this group's reading times differed significantly between Experiments 2 and 3, in contrast to the native speakers' results.

5.5.4 Discussion

Experiment 3 found a clear difference between learners' and native speakers' reading times of the disambiguating element. The native speakers preferred low attachment in both conditions, while the intermediate learners preferred low attachment in the *with*-

condition and high attachment in the *of-condition*. This result is not in line with Experiment 2 and comparable previous studies (see Felser et al. 2003, and 3.2.4.2).

In comparison to Experiment 2, the chunk-wise presentation in the present study was expected to elicit different patterns, either because of a possibly prosodic bias or a difference in processing load induced by the way of presentation. Since the native speakers in Experiment 3 still found the same low attachment preference as Experiment 2 and various other experiments with comparable materials in English (see Chapter 2.), it is unlikely that the chunk-wise presentation introduced a general bias towards high attachment in the present study. Therefore, the intermediate learners' high attachment preference in the *of*-conditions in this study is attributable to transfer from the learners' L1, as in the off-line study (see Chapter 5.3.). In Experiment 2, the comparatively high processing load induced by the word-by-word presentation seems to have suppressed the L1 influence and triggered attachment of the RC to the most recently encountered DP. The influence of task demands on L1 transfer in L2 sentence processing is in line with other studies which found transfer effects only in off-line but not in on-line tasks (e.g. Roberts, Gullberg & Indefrey 2008).

In the light of the present findings, the reason that Felser et al. (2003) found no preference in their German ESL learners might be the learners' comparatively high proficiency. Hence, at least some of them might have adopted a native-like low attachment preference, while others transferred the German high attachment preference to their L2. Another factor that might have played a role is the number disambiguation used by Felser and colleagues. This type of disambiguation was found to introduce a bias towards low attachment (see Chapter 5.4.2.) which might have competed with the high attachment preference induced by L1 transfer. In order to further investigate this possibility, it would be helpful to replicate the present study with learners at a similar proficiency level as the learners in the Felser et al. (2003) study.

5.6 Summary

Taken together, the results of the studies on German ESL learners' RC attachment preferences provide interesting results. In the off-line study, the native speaker control group preferred low attachment in both conditions, which is in line with previous findings. The advanced learners also showed a clear low attachment preference in the *with*-condition, but no preference in the *of*-condition. This difference is in line with findings reported by Felser et al. (2003) with German ESL learners at a similar proficiency level. The intermediate learners, in contrast, preferred high attachment in the *of*-condition,

which contradicts the SSH. Although additional off-line experiments on monolingual speakers of German found that the German high attachment preference is less stable than formerly assumed, the German learners' high attachment tendency can be attributed to L1 transfer. The finding that the high attachment preference was insignificant in the advanced learners indicates that L1 transfer subsides with increasing proficiency.

Unexpectedly, this pattern was not replicated in the on-line word-by-word SPR task, in which all three groups took comparatively longer to process sentences with forced high disambiguation, i.e. they preferred low attachment in both conditions. Since it is unlikely that learners show a native-like preference only in on-line but not in off-line processing, the learners' low attachment preference in Experiment 2 might result from a strategy to attach the RC to the most recently processed element to reduce processing load. In order to further investigate this possibility, a second SPR task was conducted in which sentences were not presented word-by-word but chunk-wise in order to reduce processing load. While the native speakers again preferred low attachment, the intermediate learners showed a significant high attachment preference in this task. This finding suggests that learners resort to L1 transfer in this on-line task, while in situations of increased processing load they employ a recency strategy. It is difficult to determine whether the recency preference in Experiment 2 is learner-specific because the native speaker control group showed a consistent low attachment preference in all three experiments. While the intermediate learners were found to resort to transfer more often than the advanced learners, the present results cannot answer the questions whether learners make use of transfer from early on and whether it is possible for highly proficient learners to acquire native-like processing strategies. In order to resolve these questions and to obtain a clearer picture of the development of L2 processing, further research with learners at various proficiency levels is necessary.

Although the experiments reported here did not explicitly investigate the underlying reasons for participants' attachment preferences and the predictions of specific monolingual processing models, they permit certain conclusions. The cross-linguistic differences between RC attachment preferences in English and German were confirmed, even though the German monolinguals' preference was found to be less stable than expected. The influence of the properties of the preposition linking the two potential attachment sites suggests that lexical factors such as thematic properties play an important role in sentence processing. While a purely syntactic processing model faces some difficulties explaining such phenomena, most of the processing models presented in Chapter 2.

have been adapted to account for the present findings. The facts that there are some differences between English and German and that German ESL learners resort to some extent to the preferences of their L1 are not in line with the assumption of a language-universal parser.

It is interesting that none of the results from the different tasks are in line with the predictions of the SSH, i.e. a lack of attachment preferences as it was found for example by Felser et al. (2003) and Papadopoulou & Clahsen (2003, 2006) (see Chapter 3.2.4.2.). In contrast, the present results suggest that learners transfer strategies from their L1 to their L2, unless increased processing load forces them to employ a recency strategy. For the learners tested by Felser et al. (2003), who were more proficient than the learners tested in Experiment 3, this result indicates that their lack of a clear preference might partly result from transfer, which biases them towards high attachment, and partly from native-like processing, which biases them towards low attachment. The results of the experiments reported by Papadopouolou (2006) and Papadopoulou & Clahsen (2003), in which L2 learners of Greek did not show any preference although their respective L1s and their L2 were biased towards high attachment, are not necessarily caused by shallow structure processing, either. Instead, learners could simply have difficulties integrating various sources of syntactic and pragmatic information, or, in the case of Greek as an L2, complex morphological information (Papadopoulou 2006: 187).

Taken together, the results of the present experiments contradict the absence of any preference predicted by the SSH. Instead, learners seem to be influenced by their L1 to some extent. In situations of increased processing load, they employ a recency strategy. While it is conceivable that both the recency strategy and the transfer effects observed in the present studies are default strategies assumed by the learners in the absence of deep structural representations to help them to process the sentences in question, this would considerably weaken the predictions of the SSH and would make it difficult to empirically test the hypothesis.

To conclude, the studies on German ESL learners' RC attachment preferences provided interesting insights into transfer. However , on the basis of these studies it is difficult to draw conclusions on the employment of L2-specific strategies because L1 transfer factor is a confounding variable. For this reason, the following experiments focus on ESL learners' attachment preferences in structures that do not underlie cross-linguistic differences, so that any differences observed between learners and native speakers can be interpreted with regard to L2-specific processing strategies.

6 Experiments on PP attachment

As discussed in the previous chapter, studies on L2 learners' RC attachment preferences are difficult to interpret because various factors might influence the learners' processing. In order to find out more about L2 learners' processing strategies, it might be helpful to exclude some of these influencing factors in an experiment. PP attachment ambiguities (see example (1), here repeated as (62)) provide such an opportunity because they do not underlie cross-linguistic variation but elicit a general VP attachment preference (see Chapters 2.2.1. and 3.2.3.).

(62) The man saw the boy with the binoculars.

Therefore, the learners' PP attachment preferences might provide some evidence in support of or against the employment of learner-specific strategies such as a recency strategy (e.g. Frenck-Mestre & Pynte 1997 and the discussion in Chapter 3.2.3.) or shallow syntactic processing (see Clahsen & Felser 2006b and the discussion in Chapter 3.2.4.2.). Since this ambiguity has rarely been studied with learners (see Frenck-Mestre & Pynte 1997), the present studies might provide new insights into L2 processing.

6.1 Experiment 4: On-line PP attachment preferences

Experiment 4 investigates ESL learners' on-line PP attachment preferences with disambiguated versions of sentences like (62) above, manipulating argument status of the verb (transitive vs. ditransitive) and forced disambiguation (high vs. low). The experiment addresses the following research questions:

A) Do ESL learners' on-line PP attachment preferences differ from native speakers?
B) Do the learners employ L2-specific strategies when processing PP attachment ambiguities?
C) Are there any proficiency effects on ESL learners' processing PP attachment ambiguities?

In the on-line SPR task with forced VP- or DP disambiguation, faster reading times in the disambiguating region are expected to reveal participants' attachment preferences. Although a general VP attachment preference was reported for the structures in question, processing models differ in their predictions for the native speakers' processing of the experimental sentences. As discussed in Chapter 2.2.1. the Garden Path Model assumes that DP attachment is more costly because it involves more nodes in the tree structure than VP attachment. Therefore, according to such a phrase-structure-driven parsing model, VP attachment should be generally preferred. Parameterised parsing

models (see Chapter 2.3.) predict the same effect. According to Parametrized Head Attachment (see Chapter 2.3.2.) VP attachment should be preferred because the PP is the most recently encountered theta-role assigner. In the Recency & Predicate Proximity Model (see Chapter 2.3.3.), the principle of predicate proximity would require attachment to the predicate-VP. While all these models predict a general VP attachment preference, regardless of lexical and pragmatic factors, lexically-driven parsing models (see Chapter 2.4.2.) predict an influence of the respective VP argument structure. Therefore, VP attachment might be preferred for ditransitive verbs only, because for these verbs the reader should anticipate two arguments. If the native speakers show different preferences in the VP and the DP conditions, this would indicate that their processing decisions are to some extent influenced by verb argument structure. If, however, a general VP attachment preference is found, argument structure cannot be the only factor that influences the native speakers' attachment preferences.

The predictions for the ESL learners as compared to the native speakers are more heterogeneous. As for the native speakers, whether or not the learners show a stronger VP attachment preference in the ditransitive than in the transitive condition indicates to what extent their processing decisions are guided by verb subcategorisation. The SSH predicts a high sensitivity to argument structure information and thus a high reliance on this factor in the present study. For the transitive condition, the predictions of the SSH are less clear. It is possible that the learners have no clear attachment preference in this condition because they do not make any commitment in the absence of syntactic information to guide them. But it is also possible that the learners prefer VP attachment in this condition, considering that the SSH predicts that learners are sensitive to verb argument structure and discourse factors. Since the VP is both the most recently encountered theta assigner and the most salient discourse entity, the learners' attachment preferences might be guided by one of these factors. But it is also possible that the learners show a stronger preference for DP attachment than the native speakers and thus employ a recency strategy. Finally, it is conceivable that the learners have the same attachment preferences as the native speakers. Since the structures in question require the same processing mechanisms in the learners' L1 and L2, such a pattern would not provide any information about the learners' state of L2 acquisition. Nevertheless, such a result would be instructive in that it would show that the learners do not employ an L2-specific processing strategy.

The following chapters describe the design of the present experiment before the results are presented and then discussed with regard to the predictions outlined in the present chapter.

6.1.1 Participants

Since Experiments 2 and 4 were combined in such a way that the experimental items of one study constituted at the same time the filler items for the other, the participants in the present experiment were the same as in Experiment 2 (see Chapter 5.4.1).

6.1.2 Development of Materials

Similar as in the French materials of Frenck-Mestre & Pynte (1997, see Chapter 3.2.3.), verb subcategorisation and attachment site were manipulated. Accordingly, sentences of the structure DP – VP – DP – PP – DP were constructed in such a way that the verb was either transitive or ditransitive and the PP could either be attached to the VP or to its argument DP. Sentences of this type can only be disambiguated via pragmatic information. This has been achieved in two ways in previous studies. First, disambiguation could be implemented by a second PP following the first, causing a garden-path effect if the second PP forces a reinterpretation. This can be illustrated by the following example from Boland & Boehn-Jernigan (1998):

(56) John gave a letter to his son to a friend a month ago.

The second possibility is that the meaning of the PP determines where it is attached, as illustrated in the following example:

(57) John read a letter to a friend / about a friend.

Since in (57) the sentence is directly disambiguated on the PP, a real garden-path effect is unlikely to occur. However, previous studies found attachment effects with this type of disambiguation (e.g. Clifton, Speer & Abney 1991; Schütze & Gibson 1999). The main reason why the second way of disambiguation was chosen for the present study is that it permits the direct comparison of transitive and ditransitive sentences, which is not possible with the first way of disambiguation. Since the transitive sentences have significant implications for the research questions, the experimental sentences had to allow this manipulation.

The verb argument status posed another difficulty for the construction of the experimental materials because the distinction between arguments and adjuncts and the subcategorisation frames of certain verbs are not always clear. Schütze & Gibson (1999) developed a list of criteria to distinguish between arguments and adjuncts. These criteria were used to choose the verbs for the present study. In the experimental sentences with

ditransitive verbs the PPs had to meet at least one of these criteria, whereas the PPs in the sentences with transitive verbs did not meet any of these criteria. Still, it must be noted that the distinction between arguments and adjuncts is not simple and binary: "... we assume that argumenthood is not an all-or nothing phenomenon, but that it occurs in degrees." (Schütze & Gibson 1999: 428). Since the argument or adjunct status of the critical PPs is crucial in the present study, native speaker judgements were collected as a further indication of verb argument status. Ditransitive constructions were only included in the actual study if the VP disambiguation was judged as being significantly more natural than the DP disambiguation. Transitive constructions were only included if the judgements for the two disambiguations did not differ significantly. Twelve sentences of the structure DP – VP – DP – PP – DP were included in the norming study. The verbs were manipulated in such a way that twelve ditransitive and twelve transitive sentences were created. Each of these sentences could be disambiguated either towards high (VP) or low (DP) attachment via the meaning of the PPs so that there was a total of 48 sentences of four types: ditransitive verbs disambiguated for VP attachment as in (58a), ditransitive verbs disambiguated for DP attachment as in (58b), transitive verbs disambiguated for VP attachment as in (59a), and transitive verbs disambiguated for DP attachment as in (59b):

(58) a. The journalist sent the report to the magazine.
b. The journalist sent the report about the murder.

(59) a. The journalist read the report in the evening.
b. The journalist read the report about the murder.

Two lists were created so that each participant saw the ditransitive sentence in one disambiguation and the corresponding transitive sentence in the opposite disambiguation. Ten native speakers of English judged the naturalness and ten others the plausibility of the sentences on a scale from one (very natural / plausible) to five (very unnatural / implausible) (see Chapter 4.4.). The results are illustrated in Table 6-1-1.

	ditrans. high	**ditrans. low**	**trans. high**	**trans. low**
Naturalness	1.94 (1.09)	2.88 (0.88)	2.25 (0.81)	3.54 (0.80)
Plausibility	2.25 (1.14)	2.48 (0.95)	2.28 (1.35)	2.41 (0.85)

Table 6-1-1: Naturalness and plausibility ratings for Experiment 4

Repeated-measures ANOVAs with the factors Verb Subcategorisation and Attachment were calculated on these results. The analysis of the naturalness judgements yielded a significant effect for Attachment ($F_1(1,9) = 10.619$, $p \leq 0.005$; $F_2(1,11) = 30.197$, $p \leq 0.0001$). There was no significant effect for Verb Subcategorisation

($F_1(1,9) = 2.474$, $p \leq 0.151$; $F_2(1,11) = 0.892$, $p \leq 0.365$), but a significant interaction ($F_1(1,9) = 14.298$, $p \leq 0.002$; $F_2(1,11) = 6.322$, $p \leq 0.029$).[69] The clear VP attachment preference in the ditransitive condition confirms that the verbs chosen for this condition can be considered as ditransitive verbs. Although the transitive verbs also yielded a VP attachment preference in this task, the interaction of Attachment x Verb Subcategorisation was significant. This interaction indicates that the effect was comparatively stronger in the ditransitive conditions. Thus, participants distinguished between the ditransitive and the transitive verbs in this task. The effects were also significant in the items analysis, which indicates that the observed differences are not related to lexical preferences of individual lexical items but rather to the experimental manipulations.

In order to examine the sentences for lexical or pragmatic biases, the plausibility ratings were analysed in the same way. The two-factor ANOVAs neither yielded a significant effect for Verb Subcategorisation ($F_1 \leq 1$; $F_2 \leq 1$), nor for Attachment ($F_1(1,9) = 9.000$, $p \leq 2.489$; $F_2(1,11) = 1.746$, $p \leq 0.107$) and no significant interaction between these factors ($F_1 \leq 1$; $F_2 \leq 1$).[70] Despite a tendency to prefer DP attachment in the transitive conditions, the lack of a significant difference indicates that the sentences did not differ substantially in terms of real world plausibility. Hence, this factor is unlikely to influence participants' performance in the actual on-line study.

6.1.3 Procedure

From the sentences created for the norming task, eight were chosen for the on-line SPR study. In order to make sure that the disambiguating element was not the final element of the sentence, a supplementary wrap-up region was added to all sentences. As in the norming tasks, there was a transitive and a ditransitive version of each sentence, which could either be disambiguated towards VP or DP attachment:

(60)	a.	The journalist	sent	the report	to the magazine	once again.
	b.	The journalist	sent	the report	about the murder	once again.
(61)	a.	The journalist	read	the report	in the evening	once again.
	b.	The journalist	read	the report	about the murder	once again.
Region		1	2	3	4	5

[69] This was confirmed by separate one-factor ANOVAs for the ditransitive and the transitive conditions. A significant difference was found for the sentences with ditransitive verbs ($F_1(1,9) = 14.000$, $p \leq 0.002$; $F_2(1,11) = 7.370$, $p \leq 0.002$) and for the sentences with transitive verbs ($F_1(1,9) = 28.655$, $p \leq 0.0001$; $F_2(1,11) = 31.354$, $p \leq 0.0001$).

[70] This was confirmed by separate one-factor ANOVAs for the ditransitive and the transitive conditions. These analyses did not find a significant difference between VP and DP attachment, neither for the ditransitive ($F(1,9) = 2.507$, $p \leq 0.148$) nor the transitive verbs ($F(1,9) = 1.272$, $p \leq 0.289$). In the analyses by items, there was no significant difference between VP and DP attachment, either: ($F(1,11) = 2.617$, $p \leq 0.134$ for the ditransitive and $F(1,11) = 3.716$, $p \leq 0.080$ for the transitive verbs.

In order to make the sentences within a condition comparable, the sentences and the regions for analysis within the sentences were matched for word and syllable number. The experimental sentences as well as a table representing the word and character counts per region can be found in Appendix C1. Two sets of 16 experimental sentences were created in such a way that each participant saw the transitive version in one part and the ditransitive version in the other part of the experiment. One of each of these two versions of the sentences was disambiguated towards VP attachment and the other version towards DP attachment. In addition to the 16 experimental sentences, each participant read 16 sentences from another study (i.e. Experiment 2) and 24 additional filler sentences. The sentences were randomised in such a way that no two sentences of the same study directly followed each another. The procedure was the same as for Experiment 2 (see Chapter 5.4.).

6.1.4 Results

The SPR task provided three measures, namely the answers to the comprehension questions, the time participants took to answer these questions, and the reading times for each word. The results of these three measures are presented in turn. The accuracy rates for the comprehension questions that followed the experimental items are illustrated in the following table (see Table 5-4-3 for the total accuracy rates including fillers).

	ditrans. VP	**ditrans. DP**	**trans. VP**	**trans. DP**	**total**
Intermed. learners	93.8 (12.9)	91.1 (12.2)	98.2 (6.6)	98.2 (6.6)	94.4 (6.1)
Advanced learners	91.7 (12.0)	90.8 (15.4)	97.5 (7.6)	95.0 (12.1)	94.0 (7.2)
Native speakers	97.5 (7.6)	90.0 (15.5)	98.3 (6.3)	96.7 (8.6)	95.7 (5.7)

Table 6-1-2: Percentages of correctly answered comprehension questions in Exp. 4

Overall, the native speakers were slightly more accurate than the learners. This difference however did not reach statistical significance in a one-factor ANOVA with Group as a between-subjects variable ($F(2,87) = 0.633$, $p \leq 0.543$). In order to explore whether the experimental condition had an effect on participants' success in the comprehension questions, the results were also analysed by experimental conditions. A mixed three-factor ANOVA with Verb Subcategorisation and Attachment as within-subjects variables and Group as a between-subjects variable was calculated. This analysis yielded significant effects for Verb Subcategorisation ($F_1(1,87) = 19.687$, $p \leq 0.0001$; $F_2(1,21) = 5.200$, $p \leq 0.033$) and Attachment ($F_1(1,87) = 8.003$, $p \leq 0.006$; $F_2(1,21) = 4.274$, $p \leq 0.045$), but not for Group ($F_1 \leq 1$; $F_2 \leq 1$). Since none of the interactions were significant, participants were slightly less successful with ditransitive than with transitive verbs, and slightly more successful with VP attachment than with DP

attachment.[71] Although the VP attachment preference seemed to be stronger for the ditransitive verbs, the tendency was not confirmed by the statistical analyses, in that no significant interaction was found between Verb Subcategorisation and Attachment. The finding that there were neither a main effect for Group nor any significant interactions with this factor is in line with the total results. Hence the learners did not have more difficulties with any of the experimental conditions than the native speakers.

In order to further explore the results of each group, separate two-factor ANOVAs with Verb Subcategorisation and Attachment as within-subjects factors were conducted for each of the three groups, respectively with subjects and items as random variables. For the intermediate learners, there was a significant effect for Verb Subcategorisation in the analysis by subjects but not in the analysis by items ($F_1(1,29) = 11.956$, $p \leq 0.002$; $F_2(1,29) = 1.770$, $p \leq 0.225$). No significant effect for Attachment ($F_1 \leq 1$; $F_2 \leq 1$) and no significant interaction between the factors ($F_1 \leq 1$; $F_2 \leq 1$) were found. Thus, the intermediate learners were slightly less successful with ditransitive than with transitive verbs, but showed no significant difference between VP and DP attachment.[72] The pattern was similar for the advanced learners, who also evinced a significant effect for Verb Subcategorisation only in the analysis by subjects ($F_1(1,29) = 5.118$, $p \leq 0.031$; $F_2(1,29) = 1.895$, $p \leq 0.211$). There was neither a significant effect for Attachment ($F_1(1,29) = 1.000$, $p \leq 0.326$; $F_2 \leq 1$), nor a significant interaction ($F_1 \leq 1$; $F_2 \leq 1$).[73] The native speakers' pattern was slightly different. Not only did the effect for Verb Subcategorisation reach significance, at least in the analysis by subjects ($F_1(1,29) = 3.857$, $p \leq 0.059$; $F_2(1,29) = 1.576$, $p \leq 0.250$), but there was also a significant effect for Attachment ($F_1(1,29) = 10.666$, $p \leq 0.003$; $F_2(1,29) = 6.686$, $p \leq 0.036$). As for the other groups, no significant interaction was found ($F_1(1,29) = 2.217$, $p \leq 0.147$; $F_2(1,29) = 2.046$, $p \leq 0.196$). The findings support the impression that the native speakers' comprehension results in the VP attachment conditions were better than in the DP

[71] This impression is confirmed by an additional ANOVA with the four-level factor Condition as a within-subjects factor and Group as a between-subjects factor. This analysis yielded a significant effect for Condition ($F(3,261) = 9.842$, $p \leq 0.001$), no significant effect for Group ($F \leq 1$), and no significant interaction ($F \leq 1$). Planned comparisons show that only the differences between the ditransitive-VP and the ditransitive-DP conditions ($F(1,87) = 4.810$, $p \leq 0.031$) and between the ditransitive-VP and the transitive-VP conditions ($F(1,87) = 10.023$, $p \leq 0.002$) were statistically significant.

[72] This impression was supported by an ANOVA with the four-level within-subjects factor Condition, which yielded a significant main effect ($F_1(3,81) = 3.446$, $p \leq 0.020$; $F_2(3,21) = 1.066$, $p \leq 0.385$). Only the planned comparison between ditransitive-DP and transitive-DP conditions turned out significant in the analysis by subjects ($F_1(1,29) = 6.353$, $p \leq 0.018$; $F_2(1,7) = 1.339$, $p \leq 0.285$).

[73] In the ANOVA with the four-level within-subjects factor Condition only the main effect approached significance in the analysis by subjects ($F_1(3,87) = 2.248$, $p \leq 0.088$; $F_2(3,21) = 0.891$, $p \leq 0.426$).

attachment conditions, especially for ditransitive verbs.[74] None of the learner groups showed sensitivity to this distinction.

However, it has to be noted that the observed effects were relatively small and all groups performed at over 90% levels in all conditions. Thus, it can be concluded that the learners did not show more difficulties with the comprehension questions than the native speakers, and the condition only had a small influence on the accuracy rates. This finding indicates that all learners paid attention to the task and were sufficiently proficient to comprehend the experimental sentences, even in the word-by-word presentation which requires keeping comparatively many words in mind until the end of the sentences in order to successfully answer the comprehension questions.

Table 6-1-3 summarises the response times for the correctly answered comprehension questions:[75]

	ditransitive VP	**ditransitive DP**	**transitive VP**	**transitive DP**
Intermed. learners	2574.91 (737.80)	2686.74 (643.63)	2226.74 (584.59)	2388.39 (796.78)
Advanced learners	2408.21 (509.32)	2765.86 (733.08)	2429.55 (644.53)	2508.03 (677.17)
Native speakers	2030.24 (600.62)	2156.4 (628.64)	1863.98 (474.83)	1816.14 (446.29)

Table 6-1-3: Response times in ms and (standard deviations) for correctly answered comprehension questions in Exp. 4

A mixed three-factor ANOVA with Verb Subcategorisation (transitive vs. ditransitive) and Attachment (high vs. low) as within-subjects variables and Group as a between-subjects variable yielded significant effects for Attachment ($F(1,87) = 13.046$, $p \leq 0.001$), Verb Subcategorisation ($F(1,87) = 22.955$, $p \leq 0.0001$) and Group ($F(2,87) = 12.109$, $p \leq 0.0001$). There were also significant interactions between Verb Subcategorisation and Attachment ($F(2,87) = 6.142$, $p \leq 0.015$), and the interaction of Verb Subcategorisation x Attachment x Group approached significance ($F(2,87) = 2.514$, $p \leq 0.087$). There was neither a significant interaction of Group x Verb Subcategorisation ($F_1(2,87) = 1.560$, $p \leq 0.216$) nor of Group x Attachment ($F(2,87) = 0.303$, $p \leq 0.739$).

[74] This tendency also appeared in the one-factor ANOVA with the four-level factor Condition, which found a significant main effect in the analysis by subjects ($F_1(3,87) = 4.610$, $p \leq 0.015$; $F_2(3,21) = 2.617$, $p \leq 0.216$). Planned comparisons yielded a significant difference between ditransitive-VP and ditransitive-DP conditions ($F_1(1,29) = 7.602$, $p \leq 0.010$; $F_2(1,7) = 4.769$, $p \leq 0.065$).

[75] Outliers, i.e. values that were 2.5 standard deviations below or above the mean for each group and each item, were also excluded from the analysis. This procedure affected 3.7% of the intermediate learners' data, 3.3 % of the advanced learners' data, and 3.6 % of the native speakers' data.

In order to analyse these findings more closely, separate repeated-measures ANOVAs with the factors Verb Subcategorisation and Attachment were run on the results for each group, respectively with subjects and items as a random variable.

	Intermediate learners	**Advanced learners**	**Native speakers**
Attachment	$F_1(1,29) = 2.661$, $p \leq 0.114$ $F_2(1,7) = 1.235$, $p \leq 0.303$	$F_1(1,29) = 5.584$, $p \leq 0.025$ $F_2 \leq 1$	$F_1(1,29) = 5.732$, $p \leq 0.023$ $F_2(1,7) = 1.225$, $p \leq 0.035$
Subcat	$F_1(1,29) = 16.987$, $p \leq 0.0001$ $F_2(1,7) = 7.276$, $p \leq 0.031$	$F_1(1,29) = 1.579$, $p \leq 0.219$ $F_2 \leq 1$	$F_1(1,29) = 10.806$, $p \leq 0.003$ $F_2(1,7) = 3.920$, $p \leq 0.088$
Att x Subcat	$F_1 \leq 1$ $F_2 \leq 1$	$F_1(1,29) = 3.894$, $p \leq 0.058$ $F_2 \leq 1$	$F_1(1,29) = 7.065$, $p \leq 0.013$ $F_2(1,7) = 10.181$, $p \leq 0.015$

Table 6-1-4: Verb Subcategorisation x Attachment ANOVAs on the response time data in Exp. 4

For the intermediate learners' results, the analyses yielded only a significant effect for Verb Subcategorisation, which indicates that response times were longer for the ditransitive than for the transitive conditions. Although Table 6-1-3 suggests that the intermediate learners, like the other two groups, took longer to read the DP disambiguated than the VP disambiguated sentences, the effect did not reach statistical significance. The same analysis of the advanced learners' data, in contrast, yielded a significant effect for Attachment, at least in the analysis by subjects, but no significant effect for Verb Subcategorisation. The interaction between the factors was marginally significant in the analysis by subjects and highly significant in the analysis by items. For the native speakers, there was also a significant effect for Attachment, but only in the analysis by items. This group also showed a significant effect for Verb Subcategorisation in the analysis by subjects, and a significant interaction. The interaction indicates that the native speakers' VP attachment preference was stronger in the ditransitive than in the transitive condition.

Although the response time results are in line with the accuracy results in that both measures point to a preference for the VP disambiguated sentences, it must be noted that long response times do not necessarily result from processing difficulties but might be influenced by other factors such as a temporary lapse of concentration. This would explain the relatively high standard deviations and the finding that the tendencies were not always statistically significant. Furthermore, the effects reached statistical significance more often in the analyses by items than in the analyses by subjects, which is indicative of interpersonal variation. The following paragraphs therefore have a closer look at the third measure in the present study, i.e. the reading times.

Preceding the analyses of the reading times, outliers were excluded following the same procedure as in Experiment 2. Since no participant had answered less than 80% of

the questions correctly, no participants were excluded because of this factor. The results of two of the 32 intermediate learners had to be excluded from further analyses because their mean reading time by word exceeded the mean of this group by more than 3.0 standard deviations. No participants from any of the other two groups were excluded because of this criterion. In a next step, all sentences for which the comprehension question had been answered incorrectly were excluded from further analyses. This procedure affected 4.2% of the intermediate learners' data, 5.8% of the advanced learners' data, and 4.2% of the native speakers' data. The screening for outliers, i.e. data points which were 2.5 standard deviations above or below the mean for each region of every sentence, led to the deletion of 3.1% of the intermediate learners' data, 2.5% of the advanced learners' data, and 1.0% of the native speakers' data. Thus, a total of 7.3% of the intermediate learners' data, 8.3% of the advanced learners' data, and 5.2% of the native speakers' data were excluded from further analyses. The resulting reading times per group and per region in the four experimental conditions are summarised in Appendix E3. The three groups' reading times of the PP region are illustrated in Figure 6-1-1:

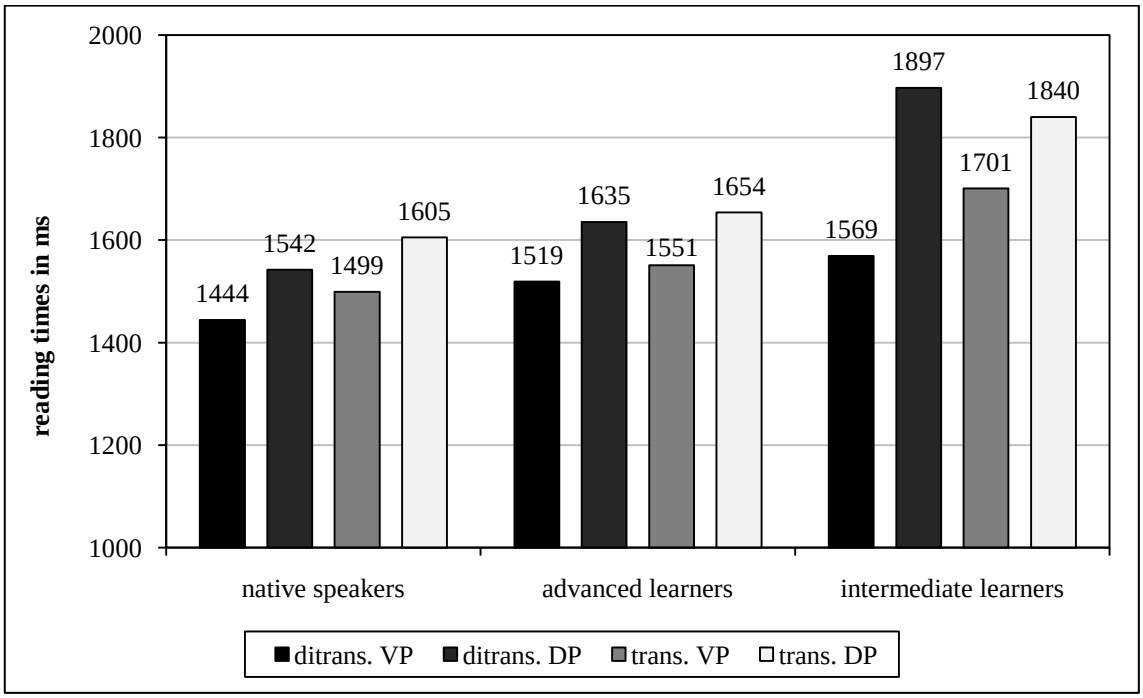

Figure 6-1-1: Reading times in the PP region by Group and Condition in Exp. 4

Mixed three-factor ANOVAs with Verb Subcategorisation and Attachment as within-subject variables and Group as a between-subjects variable were calculated for the disambiguating PP region and the wrap-up region in turn. The results are summarised in the following table:

	disambiguating region	wrap-up region
Attachment	$F_1(1,87) = 56.220$, $p \leq 0.0001$	$F_1 \leq 1$
	$F_2(1,21) = 28.948$, $p \leq 0.0001$	$F_2(1,21) = 3.100$, $p \leq 0.625$
Subcategorisation	$F_1(1,87) = 10.954$, $p \leq 0.0001$	$F_1(1,87) = 2.284$, $p \leq 0.137$
	$F_2(1,21) = 9.326$, $p \leq 0.042$	$F_2(1,21) = 1.172$, $p \leq 0.291$
Group	$F_1(1,87) = 2.217$, $p \leq 0.115$	$F_1 \leq 1$
	$F_2(1,21) = 2.333$, $p \leq 0.077$	$F_2 \leq 1$
Attachment x Group	$F_1 \leq 1$	$F_1(2,87) = 8.206$, $p \leq 0.001$
	$F_2 \leq 1$	$F_2 \leq 1$
Attachment x Subcat	$F_1 \leq 1$	$F_1 \leq 1$
	$F_2 \leq 1$	$F_2(1,21) = 12.822$, $p \leq 0.002$
Subcat x Att x Group	$F_1(2,87) = 3.503$, $p \leq 0.034$	$F_1 \leq 1$
	$F_2(2,21) = 6.116$, $p \leq 0.022$	$F_2 \leq 1$

Table 6-1-5: Prep x Att x Group ANOVA on the RTs of the disambiguating and wrap-up region in Exp. 4

In order to further explore the complex interactions found in the omnibus ANOVAs, separate two-factor ANOVAs with Verb Subcategorisation and Attachment as within-subjects factors were conducted for each group, respectively with subjects and items as a random variable. The results are summarised in the following table:

	Native speakers	Advanced learners	Intermediate learners
PP-region			
Attach-ment	$F_1(1,29) = 23.158$, $p \leq 0.0001$	$F_1(1,29) = 12.619$, $p \leq 0.001$	$F_1(1,29) = 21.400$, $p \leq 0.0001$
	$F_2(1,7) = 19.647$, $p \leq 0.003$	$F_2(1,7) = 8.272$, $p \leq 0.021$	$F_2(1,7) = 6.433$, $p \leq 0.039$
Subcat	$F_1(1,29) = 3.909$, $p \leq 0.058$	$F_1(1,29) = 2.689$, $p \leq 0.112$	$F_1(1,29) = 4.580$, $p \leq 0.041$
	$F_2(1,7) = 1.243$, $p \leq 0.302$	$F_2 \leq 1$	$F_2(1,7) = 1.910$, $p \leq 0.209$
Att x Subcat	$F_1 \leq 1$	$F_1 \leq 1$	$F_1(1,29) = 4.249$, $p \leq 0.048$
	$F_2(2,7) = 2.814$, $p \leq 0.137$	$F_2 \leq 1$	$F_2(1,7) = 2.719$, $p \leq 0.134$
Wrap-up region			
Attach-ment	$F_1(1,29) = 9.523$, $p \leq 0.008$	$F_1(1,29) = 1.869$, $p \leq 0.181$	$F_1(1,29) = 2.108$, $p \leq 0.151$
	$F_2(1,7) = 13.282$, $p \leq 0.008$	$F_2(1,7) = 3.948$, $p \leq 0.087$	$F_2(1,7) = 1.574$, $p \leq 0.250$
Subcat	$F_1(1,29) = 2.311$, $p \leq 0.139$	$F_1 \leq 1$	$F_1(1,29) = 0.368$, $p \leq 0.549$
	$F_2(1,7) = 0.314$, $p \leq 0.593$	$F_2 \leq 1$	$F_2(1,7) = 0.875$, $p \leq 0.381$
Att x Subcat	$F_1(1,29) = 2.620$, $p \leq 0.116$	$F_1 \leq 1$	$F_1(1,29) = 1.564$, $p \leq 0.229$
	$F_2(1,7) = 2.234$, $p \leq 0.179$	$F_2(1,7) = 1.175$, $p \leq 0.314$	$F_2(1,7) = 2.205$, $p \leq 0.110$

Table 6-1-6: Verb Subcategorisation x Attachment ANOVAs on the reading time data in Exp. 4

The native speakers' results confirm the impression that the ditransitive conditions were read faster than the transitive conditions and that VP attachment conditions were read faster than DP attachment conditions.[76] In the wrap-up region, the only significant fac-

[76]This impression is further supported by an ANOVA with the four-level within-subjects factor Condition. For the PP region, this analysis yielded a significant main effect ($F_1(3,87) = 7.990$, $p \leq 0.0001$; $F_2(3,21) = 6.122$, $p \leq 0.004$). Planned pair-wise comparisons found a significant difference between the transitive-VP and transitive-DP conditions ($F_1(1,29) = 11.555$, $p \leq 0.002$; $F_2(1,7) = 10.573$, $p \leq 0.014$). In the wrap-up region, there was also a significant main effect ($F_1(3,87) = 5.739$, $p \leq 0.001$; $F_2(3,21) = 4.374$, $p \leq 0.047$). Planned comparisons found a significant effect between the ditransitive-VP and the ditransitive-DP conditions ($F_1(1,29) = 9.692$, $p \leq 0.004$; $F_2(1,7) = 7.062$, $p \leq 0.033$).

tor was Attachment. This finding implies that, as in the PP region, native speakers took longer to read the DP than the VP attachment conditions. As Figure 6-1-1 indicates, the advanced learners showed a similar pattern. For this group, there was a significant effect for Attachment in the PP region, indicating that the VP disambiguations were read significantly faster than the DP disambiguations.[77] There was neither a significant effect for Subcategorisation nor a significant interaction. Thus, in contrast to the native speakers, the advanced learners did not take longer to process the transitive conditions than the ditransitive conditions. In the wrap-up region, no significant effects were found. Figure 6-1-1 suggests similar results for the two learner groups, although the intermediate learners' reading times were generally higher. Although none of the effects reached significance in the analysis by items, there was a tendency for Verb Subcategorisation and Attachment to interact in the analysis by subjects, which indicates that this group's VP attachment preference in the transitive condition was not as strong as in the ditransitive condition.[78] Like the advanced learners, the intermediate learners did not show any significant effects in the wrap-up region.

6.1.5 Discussion

The present section discusses the results, i.e. participants' accuracy rates and response times for the comprehension questions and the word-by-word reading times, with regard to the research questions and the predictions formulated in the beginning of this chapter. Methodological issues that are important for the interpretation of the data are also addressed.

The first point to note about the comprehension questions is that the accuracy rates were generally high but the response times showed some variation within the groups. One reason for the high accuracy rates might be that the experimental sentences were comparatively short and simple and the questions relatively easy to answer. But this finding also indicates that participants understood the sentences and were paying attention to the task. The findings from the comprehension questions and the reading times

[77] An additional ANOVA with the four-level within-subjects factor Condition yielded a significant main effect ($F_1(3,87) = 8.247$, $p \leq 0.0001$; $F_2(3,21) = 3.566$, $p \leq 0.031$). Planned paired comparisons found statistically significant differences between the ditransitive-VP and ditransitive-DP conditions ($F_1(1,28) = 5.438$, $p \leq 0.027$; $F_2(1,7) = 5.925$, $p \leq 0.045$) and between the transitive-VP and transitive-DP conditions ($F_1(1,28) = 18.736$, $p \leq 0.0001$; $F_2(1,7) = 10.107$, $p \leq 0.016$).

[78] To investigate this further an ANOVA with the four-level within-subjects factor Condition was run on the results. This analysis found a significant main effect ($F_1(3,87) = 9.724$, $p \leq 0.0001$; $F_2(3,21) = 5.169$, $p \leq 0.008$). Planned paired comparisons found statistically significant differences between the ditransitive-VP and ditransitive-DP conditions ($F_1(1,29) = 22.206$, $p \leq 0.0001$; $F_2(1,7) = 13.151$, $p \leq 0.008$) and between the transitive-VP and transitive-DP conditions ($F_1(1,29) = 5.814$, $p \leq 0.022$; $F_2(1,7) = 10.180$, $p \leq 0.015$). Moreover, there was a significant difference between the ditransitive-VP and transitive-VP conditions ($F_1(1,29) = 5.353$, $p \leq 0.028$; $F_2(1,7) = 5.546$, $p \leq 0.051$).

converge in that VP attachment was found to be preferred in both conditions and all three groups. But there is a discrepancy between the comprehension results and the reaction time data for the verb subcategorisation manipulation. Both the response times and the reading times point to an advantage of the ditransitive over the transitive condition, which however did not reach statistical significance for all groups. In the comprehension questions, in contrast, there was an advantage of the transitive over the ditransitive conditions. Although the comprehension results thus suggest that the transitive conditions were easier to process, a look at the reaction time data shows that the effect is actually reversed, i.e. the transitive sentences were harder to process and thus took participants longer to read. This processing difficulty, in turn, might have led to more attentive reading and to slightly better results in the comprehension questions. The results thus indicate that comprehension questions following the sentences in a SPR task do not necessarily provide a complete picture of participants' reading difficulties.

The SPR patterns have to be interpreted with similar care as the results of the comprehension questions. One crucial factor in this respect is the specification of the target regions. Since in the present study the sentences were disambiguated in the PP region rather than on an isolated word, the reading times for the three words in this region were accumulated for the analyses, thus yielding a relatively long region of interest. The subsequent wrap-up region was also analysed in order to discern possible spill-over or wrap-up effects. Such effects were only found in the native speakers, whose patterns from the disambiguating region were replicated in the wrap-up region. This finding might be related to the native speakers' generally short reading times. The learners' reading times were comparatively longer, which indicates that they processed all the relevant information in the PP region, while the native speakers were so fast on this region that some of the processing was only accomplished in the following region. Since all groups showed similar effects in the disambiguating PP region, this region is the focus of the following discussion.

The native speakers' VP attachment preference in the ditransitive condition is in line with previous findings (see Chapter 2.2.1.) and can be explained by principle based and lexically based processing models (see Chapter 2.). The native speakers' VP attachment preference in the transitive condition, in which the PP was not an obligatory argument of the verb, indicates that verb subcategorisation is not the only factor that guides participants' attachment decisions. This finding is not in line with a purely lexically based processing model. Although the finding converges with the results from other studies on

monolingual PP attachment preferences, it contrasts with the results reported by Frenck-Mestre & Pynte (1997) (see Chapter 3.2.3.). These authors found no preference or even a slight DP attachment preference for French native speakers in the transitive condition. It is possible that this finding results from a cross-linguistic difference between French on the one hand and English and German on the other hand. But such an explanation would require a theoretical explanation and additional evidence. Another possible explanation is that the VP attachment preference in the present study is triggered by the absence of referential context. Since the present sentences only contain one object-DP, modification of this DP is pragmatically unnecessary. However, such an account is unlikely because none of the reported monolingual studies included such contexts for the experimental sentences. Instead, the DP attachment preference in the Frenck-Mestre & Pynte (1997) study might be triggered by the lexical factors. A closer look at these materials reveals that many of the DP-attached PPs are close to idiomatic expressions (e.g. salle de séjour 'living room', train de nuit 'night train', or femme de chambre 'chamber maid'). Therefore, at least some of these sentences might be biased towards DP attachment. Since a VP attachment preference was found in many other studies which were not affected by such a bias, it seems to be a more robust finding.

As outlined in the beginning of this chapter, theoretical approaches predict different patterns for the ESL learners as compared to the native speakers in the present study. It is conceivable that the learners show (a) no attachment preference, (b) a stronger VP attachment preference than the native speakers, (c) a stronger DP attachment preference than the native speakers, or (d) similar patterns as the native speakers. In the following, these possibilities and their implications for L2 processing theories are considered in turn in the light of the present experimental findings.

The SSH would predict no attachment preference for the learners, similar as for RC attachment. This prediction is based on the assumption that learners who are not guided by structural processing principles prefer to make no commitment and hence to show no attachment preference (see Chapter 3.2.4.2.). Since both learner groups had a clear VP attachment preference in both conditions of the present study, the ESL learners' results do not directly support the SSH.

The second possibility is that the learners show a stronger VP attachment preference than the native speakers. If this was the case in the ditransitive condition only, it would indicate that the learners rely more strongly on verb subcategorisation than the native speakers. Such a pattern in both conditions or only in the transitive condition would

suggest that the learners over-rely on a principle like Predicate Proximity (Gibson et al. 1996; see Chapter 2.3.3.) or on attachment to the most recent theta-assigner. Since these strategies are lexical or pragmatic rather than syntactic, such a finding would be in line with the SSH, which predicts that learners rely on lexical and pragmatic rather than syntactic factors. Although no such tendency was found in the advanced learners, the intermediate learners showed a stronger VP attachment preference in the ditransitive than in the transitive condition. It is debatable, however, whether this tendency is related to a VP attachment preference. The intermediate learners' reading times are comparatively long for the DP disambiguation of the ditransitive condition, which suggests that these learners have difficulties interpreting their dispreferred disambiguation, possibly resulting from difficulties to recover from misanalyses. To conclude, the data do not provide clear evidence for a learner-specific VP attachment preference.

Third, it is conceivable that the learners have a stronger DP attachment preference than the native speakers. Such a tendency was reported by Frenck-Mestre & Pynte (1997, see Chapter 3.2.3.) and interpreted as an indication of an L2-specific recency strategy. But in the present study the learners evinced the opposite pattern: Like the English native speaker control group, both the advanced and the intermediate learners preferred attachment to the VP. This finding strongly suggests that the learners did not employ a recency strategy.[79]

Fourth, it was suggested that the learners show the same processing patterns as the native speaker control group. This is what was found in the present study. In order to determine the reason for the learners' processing preferences in the present task, it would be necessary to know more about the factors that determine the native speakers' attachment preferences. Despite the extensive research on monolinguals' PP attachment preferences (see Chapter 2.1.) it is still unresolved whether the VP attachment preference is caused by syntactic principles such as Minimal Attachment or Predicate Proximity or by lexical or pragmatic factors. The interpretation of the learners' preferences in the present study depends on the explanation of the preferences in L1 processing. Assuming that native speakers are guided by syntactic principles, the learners' native-like preference in the present study indicates that they make native-like use of syntactic principles. But it is also conceivable that both native speakers and learners are guided by discourse principles, or the native speakers employ syntactic principles while the

[79] This finding does not necessarily imply that learners never employ such a strategy. It is conceivable that learners resort to such a strategy in constructions that impose higher processing load than the sentences of the present study. Furthermore, the eye-tracking measure used by Frenck-Mestre & Pynte might be more sensitive to early and weak preferences than the present SPR task.

learners rely on discourse principles. The latter possibility is predicted by the SSH. The following paragraphs discuss the implications of the results for the research questions.

The only systematic difference between learners and native speakers in this on-line reading task was one of general reading speed, which suggests a lack of processing automaticity or an insecurity on the side of the learners. The absence of significant qualitative differences between the learners' and the native speakers' patterns suggests that the learners employ similar processing mechanisms as the native speakers.

The two L2-specific strategies in the focus of the present thesis are shallow structure processing and an L2-specific recency preference. With regard to the latter strategy, the present study did not provide any supportive evidence. Shallow structure processing, however, cannot be excluded on the basis of the present data. Although the learners neither evinced an absence of any preference nor a stronger VP attachment preference than the native speakers as would be predicted by the SSH, it is possible that the learners' native-like processing patterns are in fact a result of lexical and pragmatic rather than structural processing strategies.

With regard to possible proficiency effects, the intermediate learners only differed from the advanced learners in their longer reading times for all conditions. The difference was especially noticeable for the dispreferred disambiguations, which indicates that the intermediate learners' processing is less automatic and more insecure and they have more difficulties to recover from misanalyses compared to the other groups. In contrast to the advanced group, the intermediate learners showed a stronger VP attachment preference in the ditransitive than in the transitive conditions. Although this finding could indicate that the intermediate learners are more sensitive to the argument structure violation in the ditransitive-DP condition, it is also possible that these learners have difficulties interpreting their dispreferred disambiguation, possibly resulting from difficulties to recover from misanalyses.

Although the present study is instructive with regard to L2-specific processing strategies, several questions remain. Since the present study found highly native-like preferences in the ESL learners, further research on the mechanisms that guide L1 processing of these structures is required. Furthermore, it would be helpful to study L2 learners' processing of related structures that are less controversial in L1 processing in order to find out if there are systematic differences between L1 and L2 processing. Such a follow-up study is also expedient because the present study had some methodological shortcomings, in spite of the careful choice and development of materials. First, the dis-

ambiguating PP region is relatively long, which makes it difficult to determine where exactly the disambiguation takes place and where to expect processing difficulties. Hence, it is not sure whether the present study measured initial processing decisions. Second, the sentence type examined here does not allow for an ambiguous control condition. Such a control condition would be especially instructive because the SSH predicts that learners prefer ambiguous structures which do not force them to make any commitments, over disambiguated structures. Third, despite the careful choice and development of the materials, the distinction between the argument and adjunct status of the critical PPs is not entirely clear. Fourth, it is possible that the argument structure differences in the ditransitive conditions also biased participants in the transitive conditions. Thus it might be interesting to investigate structures in which all verbs have the same argument status. This is the aim of the following two studies.

6.2 Research questions and expectations for Experiments 5 and 6

Experiments 5 and 6 aim to further explore ESL learners' PP attachment preferences. Since Experiment 4 found that ESL learners were sensitive to verb argument structure, the present studies attempt to exclude this factor by using materials in which all PPs are attached to VPs of the same argument status. Experimental sentences are of the nature DP1 – VP1 – DP2 – that – DP1 –VP2 – PP, where DP1 is one single referent and the PP can either attach to VP1 or to VP2. A further advantage of these structures in comparison to those used in Experiment 4 is that they can be left ambiguous. Therefore, participants' performance in off-line and on-line tasks can be compared. Such a comparison has provided interesting results in the studies on RC attachment (see Chapter 3.2.4.). The present studies test whether similar insights can be gained from PP attachment preferences. Hence, it is conceivable that the learners' use of non-native like strategies increases in on-line tasks as compared to off-line tasks. In the present on-line task, three conditions, i.e. ambiguous sentences, sentences disambiguated towards high attachment and sentences disambiguated towards low attachment, can be directly compared. In studies on other ambiguities with monolinguals (e.g. Swets et al. 2008; Traxler, Pickering & Clifton 1998; Van Gompel et al. 2005), it was found that participants were faster in ambiguous than in disambiguated structures. Hence, monolinguals do not seem to have any difficulties with ambiguous constructions. There are two possible explanations for such a finding: (a) participants interpret the sentences along the lines of their preferred interpretation; or (b) they create a 'good enough' representation (Christianson et al. 2001), i.e. they do not commit to a specific interpretation. The present studies test

whether the learners show a similar pattern or whether they have difficulties pursuing the two interpretations of an ambiguous sentence at the same time.

In the present study direct transfer of attachment preferences from the learners' L1 is improbable. The direct German counterpart to the English construction cannot be ambiguous because German word order rules do not permit a low attachment reading of a structure like (62a). The following examples illustrate the German literal translation (62b) and the high (62c, 62d) and low (62e) disambiguated equivalents:

(62) a. Peter copied the document that he had hidden from Julia.
b. Peter kopierte das Dokument dass er hatte versteckt von / vor Julia.
c. Peter kopierte das Dokument dass er versteckt hatte von Julia.
d. Peter kopierte das Dokument von Julia dass er versteckt hatte.
e. Peter kopierte das Dokument dass er vor Julia versteckt hatte.

As the examples show, the direct German counterpart to the English construction is disambiguated towards high attachment (62c). This construction is unnatural because high attachment would rather be expressed as in (62d). It is possible that the German ESL learners in the present study directly translate the English construction into German and assume that English has an alternative construction to express low attachment as German. However, such an outcome is improbable, at least in the relatively proficient learners who took part in the present study. These learners are assumed to be sufficiently familiar with English word order to realise that in English a construction as (62e) is ungrammatical. Hence, transfer is not predicted to substantially influence the learners in the present study.

The target structure has, to the author's knowledge, only been investigated by one previous study with Spanish monolinguals (Igoa et al. 1998, see Chapter 2.3.2.). The study found a low attachment preference with Spanish materials, which was clearer in the off-line than in the on-line task. In the on-line SPR task, participants read sentences like (63) that were pragmatically disambiguated towards high (63a) or low (63b) attachment or were left ambiguous (63c).

(63) a. Raul vendió el libro que tenía subrayado a su amigo.
b. Raul examinó el libro que había robado a su amigo.
c. Raul vendió el libro que tenía subrayado a su amigo.

On the disambiguating PP, reading times turned out to be shortest for low disambiguated and ambiguous sentences and longest for sentences disambiguated towards high attachment. The authors take this low attachment preference as evidence for Late Closure. Additional corpus analyses indicated that the low attachment preference was not

determined by the second verbs (VP2) being more likely to take an extra argument than those located at a higher position (VP1), as lexical models would assume. Hence, the authors argue that the data are in line with exposure-based models such as Tuning, but not with lexically driven constraint-based models.

Given that a structural principle like Late Closure or Recency guides participants' attachment preferences in the Spanish experiment, participants in the present studies are expected to also adhere to such a language universal principle. Therefore, a low attachment preference is predicted for the English native speakers in the present studies. The predictions for the learners are less straightforward. It is improbable that they transfer the preference from their L1 to their L2 because the construction is different in the two languages. The learners are more likely to employ a recency strategy. Such a strategy might either be native-like or even stronger than in the native speakers. If the German ESL learners employ a shallow processing strategy, they are not expected to be guided by a structural principle such as Late Closure. Accordingly, they should not evince a clear attachment preference in the off-line task in which they have to choose one of two interpretations. In the on-line task they would be predicted to prefer the ambiguous over the disambiguated structures. Given these heterogeneous predictions, Experiments 5 and 6 address the following research questions:

A) Do ESL learners' off-line and on-line attachment preferences for PPs differ from those of English native speakers?
B) Are the learners' preferences different for off-line and on-line tasks?
C) Do the learners employ learner-specific parsing strategies when processing PP attachment ambiguities?

6.3 Experiment 5: Off-line PP Attachments

6.3.1 Participants

Participants were from the same pool as in the previously reported studies, but the participants had not taken part in any of the studies reported here. Hence, they could not have developed any strategies for processing the experimental sentences and were unaware of the purpose of the study.

The control group consisted of 30 native speakers of English.[80] As in the previous studies, the two learner groups differed with regard to the time of learning English and

[80] Half of them were speakers of British English and the other half were speakers of American English. They were living in English-speaking countries at the time of testing and reported that they had not lived in non-English-speaking countries for more than two years. Their mean age was 28.4 years (range = 18-60, SD = 17.7) and 66.7% of them were female.

the time spent in English-speaking countries. As an additional proficiency measure, the learners self-rated their English proficiency on a scale from one (very low) to ten (very high).[81] The learners' biographical data are summarised in Table 6-3-1:

Group		Age	Years of learning English	Months in English-speaking countries	Proficiency (scale 1-10)
Intermediate learners (n = 30)	**mean**	21.9	11.7	—	6.6
	SD	(2.0)	(2.1)		(1.0)
	range	20-27	8-17		5-9
Advanced learners (n = 30)	**mean**	26.5	16.1	10.3	8.0
	SD	(4.9)	(4.9)	(5.8)	(0.6)
	range	23-49	12-39	3-26	7-9

Table 6-3-1: Biographical data of the learner groups in Experiment 5

6.3.2 Materials and Method

The experimental sentences were of the type DP1 – VP1 – DP2 – that – DP1 –VP2 – PP, where DP1 is one single referent and the PP can either attach to VP1 or to VP2. Both verbs were ditransitive, i.e. they fulfilled at least one of the criteria for ditransitivity proposed by Schütze & Gibson (1999, see Chapter 6.1.3.). For the norming study, 16 sentences of this type were created. In eight of these sentences the two ditransitive verbs took the same preposition so that attachment of the PP was ambiguous (64a). For the other eight sentences, ditransitive verbs with corresponding prepositions were chosen so that the sentences were disambiguated either towards high (64b) or low attachment (64c) when the preposition was encountered. Hence, only the subcategorisation frame of one of the two verbs was satisfied.

(64) a. Janet concealed the present that she had borrowed from Peter.
b. Lily offered the CD that she had stolen from John.
c. Lily offered the CD that she had sold to John.

The sentences were constructed in such a way that the two verbs within each sentence were approximately matched for frequency (see Chapter 4.3.). Two lists were created so that each participant saw a sentence either disambiguated towards high or low attachment. Each list thus contained four ambiguous sentences, four high and four low disambiguated sentences. For the norming study, the 12 experimental sentences per list were intermingled with 16 sentences of unrelated constructions (see Chapter 7.3.1.). Ten English native speakers (five per list) judged the naturalness and ten others the real-

[81] Although self-ratings have been criticised for their unreliability, they are widely employed (e.g. Fernández 2003) and were found to correlate with other proficiency measures (Grosjean 1982). In order to assure their validity for the present study, correlations with the other two proficiency measures were calculated. There was a significant correlation of the self-ratings with years of learning ($r(60) = 0.422$, $p \leq 0.001$) and months spent in English-speaking countries ($r(60) = 0.433$, $p \leq 0.001$).

world plausibility of the sentences on a scale from one (very natural / plausible) to five (very unnatural / implausible). The results are summarised in the following table:

	ambiguous	**high attachment**	**low attachment**
Naturalness	2.45 (0.76)	4.21 (0.94)	2.83 (0.71)
Plausibility	2.38 (0.65)	2.87 (0.98)	2.26 (0.59)

Table 6-3-2: Naturalness and plausibility ratings for Experiment 5

As Table 6-3-2 shows, high-disambiguated sentences were judged as being more unnatural than ambiguous and low-disambiguated sentences. To test this observation, the naturalness judgements were submitted to a one-factor repeated-measures ANOVA with subjects as a random variable. This analysis yielded a significant main effect ($F(2,18) = 9.872$, $p \leq 0.0001$). Planned comparisons confirmed the impression that high attachment was judged as being significantly less natural than the other two conditions (high vs. low attachment: $F(1,9) = 10.782$, $p \leq 0.009$; high attachment vs. ambiguous: $F(1,9) = 10.782$, $p \leq 0.009$); low attachment vs. ambiguous: $F \leq 1$). This result provides a first impression of English native speakers' off-line preferences, which, as expected, tend towards low attachment. Ambiguous sentences are judged even more positively than low disambiguated sentences. The relatively low standard deviations in all conditions indicate that participants' judgements of the sentences were quite homogeneous.

The purpose of the plausibility judgements was to ensure that participants' judgements were not biased by the pragmatic plausibility of the sentences. As in the naturalness rating task, the high disambiguated sentences were less accepted than ambiguous and low disambiguated sentences in the plausibility task, although the differences were comparatively small. The one-factor repeated-measures ANOVA on the results did not yield a significant main effect ($F_1(2,18) = 2.095$, $p \leq 0.152$).[82] The differences between conditions indicate that participants' plausibility judgements are influenced to some extent by the grammaticality or naturalness of the sentences, which is in line with the previously reported studies. Nevertheless, these differences were relatively small and insignificant, which indicates that the experimental sentences did not differ substantially in their real-world plausibility across conditions. Hence, this factor should not influence participants' attachment preferences in the off-line questionnaire.

In the off-line questionnaire, the eight ambiguous sentences from the norming study were listed with the two possible interpretations, as illustrated in (65):

[82] Since identical items (for different lists) were only used for the sentences disambiguated towards high or low attachment of the PP, an ANOVA with items as a random variable could only be run to compare these two types of sentences. This analysis did not reach statistical significance either ($F_2(1,7) = 4.421$, $p \leq 0.074$).

(65) Peter copied the term paper that he had hidden from Julia.

□ Peter copied the term paper from Julia.

□ Peter had hidden the term paper from Julia.

In half of the items the high-disambiguated interpretation preceded the low-disambiguated interpretation; in the other half the order was reversed. The sentences were intermingled with eight ambiguous sentences from Experiment 7 (see Chapter 7.3.) and eight unrelated ambiguous sentences (see Appendix C2 for a list of experimental items). Participants were instructed to choose their preferred interpretation as spontaneously as possible, but, as in the other off-line tasks, there was no time limit.

6.3.3 Results

The three groups' off-line results are summarised in the following table:[83]

	% high attachment	% low attachment
Native speakers	25.6 (17.1)	74.4 (17.1)
Advanced learners	32.8 (22.8)	67.2 (22.8)
Intermediate learners	35.0 (18.4)	65.0 (18.4)

Table 6-3-3: Mean attachment choices and (standard deviations) by group in Exp. 5

These results were submitted to a mixed two-factor ANOVA with Attachment as a within-subjects variable and Group as a between-subjects variable. The analysis yielded a significant effect for Attachment ($F(1,87) = 71.647$, $p \leq 0.0001$), but no significant effect for Group ($F \leq 1$) and no significant interaction ($F \leq 1$). This finding indicates that the three groups performed similarly. However, the comparatively high standard deviations suggest some inter-individual differences. Since such differences turned out to be interesting in the off-line RC attachment study (see Chapter 5.3.), they are also considered here. The native speakers' low attachment preference was relatively consistent. Only two of the 30 native speakers (i.e. 6.7%) showed a high attachment preference and two others no preference. For the advanced learners the picture was similar. Four of them (i.e. 13.3%) preferred high attachment, and another four did not show any preference. Of the 30 intermediate learners, five (i.e. 16.7%) preferred high attachment, and six (i.e. 20%) did not show any preference. These individual results indicate that most of the learners showed the same low attachment preference as the native speakers, but the learners' decisions are more heterogeneous. Still proficiency, at least as it was measured in the present task, is not a reliable predictor of participants' attachment choices because neither for the intermediate nor for the advanced learners were there

[83]The results of the filler items in this task will not be considered here because there is no correct or wrong answer to these completely ambiguous sentences. The results for the eight filler items that were at the same time the experimental sentences for Experiment 7 are reported in Chapter 7.3.3..

any significant correlations between participants' low attachment choices and their time of learning English (intermed. learners: $r(29) = -0.041$, $p \leq 0.838$; adv. learners: $r(29) = 0.073$, $p \leq 0.706$), the time they had spent in English-speaking countries (adv. learners: $r(29) = -0.329$, $p \leq 0.081$) or their self-rated proficiency (intermed. learners: $r(29) = -0.197$, $p \leq 0.325$; adv. learners: $r(29) = 0.137$, $p \leq 0.478$).

6.3.4 Discussion

The only previous study with similar constructions is reported by Igoa et al. (1998) for Spanish (see Chapter 2.3.2.). These authors found a low attachment preference with this type of materials, which they take as evidence for the Late Closure Principle. The English native speakers in the present study showed an even stronger preference for low attachment. The results provide further support of the Late Closure explanation, but as Igoa et al. (1998) admit, the low attachment preference would also be in line with the Recency & Predicate Proximity Model (see Chapter 2.3.3.). The ESL learners in the present study showed a similar low attachment preference. However, this preference was not as strong as in the native speaker control group, and the individual participants' results revealed that some of the learners preferred high attachment for these constructions. Despite the absence of significant correlations with proficiency measures such as grammar performance, years of learning or time abroad, the tendency to attach high or to make no commitment was stronger in the intermediate learners than in the advanced learners.

Given that a structural principle like Late Closure or Recency is responsible for the native speakers' low attachment preference, the learner's results indicate that they are also sensitive to this factor and show highly native-like processing mechanisms. Such an interpretation is in line with the learners' native-like attachment preferences in Experiment 4 (see Chapter 6.1.). Hence, the results of the present study cannot be attributed to learner-specific strategies. As outlined in Chapter 6.2. an L2-specific recency strategy would predict a stronger low attachment preference for the learners than for the native speakers. In contrast, the effect in the present study was reversed, in that the low attachment preference was strongest in the native speakers and weakest in the intermediate ESL learners. The present results are not in line with the SSH, either, because the SSH would predict no preferences in the learners.

In spite of this piece of evidence against L2-specific processing strategies, the question remains why the learners' low attachment preference is less clear-cut and more heterogeneous as compared to the native speakers. The difference could simply be re-

lated to the learners' general insecurity in the task, which might prompt at least some of them to choose high attachment as a default. It is also conceivable that the learners are more insecure than the native speakers about the argument structure requirements of the verbs in the experimental sentences. Hence, they might sometimes respect the requirements of the first verb rather than those of the second verb, although both verbs subcategorise for a PP complement. Another possible explanation would be based on cross-linguistic influence. As outlined in Chapter 6.2. the direct German counterpart to the English construction is disambiguated towards high attachment, which might bias the learners' initial preference. If this is the case, the learners' tendency to attach high should be stronger in the on-line SPR task, which is supposed to reflect more immediate preferences than an off-line task. But it is also conceivable that the learners show a stronger tendency to attach low under the time pressure and processing load of the on-line task. Such a tendency would support an explanation based on learner-specific strategies and would be in line with the results of the studies on RC attachment (see Chapter 5.). If the learners show no preferences or prefer the ambiguous over the disambiguated structures, in contrast, this would be in line with the SSH. The next chapter addresses these issues by presenting an on-line SPR study of the same construction in order to compare the learners' off-line and on-line processing.

6.4 Experiment 6: On-line PP attachments II

Experiment 6 is the on-line version of Experiment 5 and thus addresses the same research questions (see Chapter 6.2.). The predictions for the two studies are also similar. It is conceivable, however, that the higher processing load of the on-line task affects the learners' processing. Furthermore, the present study allows for a direct comparison between ambiguous sentences and disambiguated sentences.

If learners employ native-like processing strategies in the present task, they are expected to prefer low attachment. The learners' employment of an L2-specific recency strategy would entail a stronger low attachment preference than in the native speakers. If learners employ a shallow syntactic processing strategy, they should be more comfortable with ambiguous than with disambiguated sentences. But if L1 influence is responsible for the learners' slight tendency to attach high in the off-line task, this tendency is expected to be stronger in the on-line task in which participants have less time to consciously reflect their processing decisions.

6.4.1 Participants

The participants were from the same pool as in the previously reported studies (see Chapter 4.1.). A native speaker control group[84] and two groups of German university students of English (see Table 6-4-1) participated in the study.

Group		Age	Years of learning English	Months in English-speaking countries	Grammar
Intermediate learners (n = 30)	**mean**	21.6	11.07	___	80.27
	SD	2.03	2.30		5.67
	range	19-24	8-15		65-82
Advanced learners (n = 30)	**mean**	25.12	15.75	12.11	90.84
	SD	2.63	3.68	9.14	4.32
	range	23-28	13-25	4-36	85-100

Table 6-4-1: Biographical data of the learner groups in Experiment 6

6.4.2 Materials and method

The materials were of the same structure as those used in the off-line task described in the previous chapter. The ambiguous (66a), high (66b) and low (66c) disambiguated conditions are illustrated in the following examples.

(66) a.	Janet	concealed	the present	that	she had borrowed	from Peter	once again.
b.	Lily	offered	the CD	that	she had stolen	to John	once again.
c.	Lily	offered	the CD	that	she had stolen	from John	once again.
Region	1	2	3	4	5	6	7

The norming study that preceded the actual experiment was the same as reported in Experiment 5, including both ambiguous and disambiguated elements. The results are thus the same as reported in Chapter 6.3.2, here repeated in Table 6-4-2:

	high attachment	low attachment
Naturalness	4.21 (0.94)	2.83 (0.71)
Plausibility	2.87 (0.98)	2.26 (0.59)

Table 6-4-2: Naturalness and plausibility judgements for Experiment 6

Since the on-line study used disambiguated sentences, it was important that the same sentence was equally natural and plausible in both disambiguations. Therefore, two additional one-factor repeated-measures ANOVAs, respectively with subjects and with items as a random variable, were calculated to compare the sentences disambiguated for high and for low attachment. The analyses yielded a significant main effect by subjects and by items ($F_1(2,18) = 9.872$, $p \leq 0.0001$; $F_2(1,7) = 11.262$, $p \leq 0.012$). This result suggests that the judgements for each condition were relatively consistent and the varia-

[84] Half of these 30 participants spoke British English; the other half spoke American English. Their mean age was 24.0 years (range = 20-36, SD = 5.0) and 60% of them were female.

tion between conditions was considerably higher than within conditions. If the obtained statistical effect had been related to the particular verbs in the sentences, no significant effect would have been found in the item-based analysis.

The same procedure was applied to the plausibility judgements. As Table 6-4-2 indicates, the sentences disambiguated towards high attachment were less accepted than ambiguous and low-disambiguated sentences. To test this observation, the results were submitted to a one-factor repeated-measures ANOVA, which yielded no significant main effects ($F_1(1,9) = 2.512$, $p \leq 0.147$; $F_2(1,7) = 4.421$, $p \leq 0.074$). As in the previous studies, participants' plausibility judgements thus seem to be influenced to some extent by the grammaticality or naturalness of the sentences. Nevertheless, the comparatively small standard deviations within conditions indicate that sentences within a condition did not differ substantially with regard to this factor. Hence, plausibility should not influence participants' attachment preferences in the on-line SPR task.

Six ambiguous and six disambiguated sentences from the norming task were included in the on-line study. The sentences were closely matched for phonological word number and character lengths (see Appendix C3). Moreover, the verbs within each sentence were matched for frequency (see Chapter 4.3.). There was only one list so that the disambiguated sentences appeared in both high and low disambiguation. Hence, each participant read two versions of a sentence. This procedure has the advantage that the conditions can be compared without inter-individual differences affecting the outcomes. However, participants might be affected by reading the same sentence twice during the task. Consequently, there should be some distance between two versions of the same item and a sufficient number of filler and distracter items. The task was divided into three parts of equal length such that in each part every participant saw the six experimental items in different disambiguations. The three versions of an item were distributed across the three parts. Each participant saw two ambiguous, two high-disambiguated and two low-disambiguated experimental items in every part. The parts were divided by a short break.

The 18 experimental sentences (see Appendix C3) were intermingled with 18 sentences of another study (see Chapter 7.4.) and 24 independent fillers of various structures and lengths. The order of presentation was pseudorandomised for each participant in such a way that no two experimental sentences directly followed one another. Since the items for Experiment 8 were all questions, half of the filler items were also questions. Moreover, half of all filler sentences were ungrammatical in order to approxi-

mately match the number of grammatical and ungrammatical sentences in the experiment. However, the division into grammatical and ungrammatical sentences was not straightforward for the sentences of the present study, considering that only the sentences disambiguated towards high attachment were clearly judged as being unnatural in the norming task. For the other sentences the judgements were less clear and the means were slightly above "3", i.e. the middle value on the scale. Since grammaticality here was not an all-or-nothing phenomenon, it was not possible to exactly match the number of grammatical and ungrammatical sentences. In order to provide at least an approximate balance, half of the 12 declarative filler sentences were grammatical and the other half were ungrammatical. Altogether, each participant read 60 sentences in the word-by-word SPR task. Each sentence was followed by a timed grammaticality judgement. Participants completed all experimental tasks in the same order (see Chapter 4.2.).

6.4.3 Results

The present chapter first analyses the results of the grammaticality judgement task, which provides a first indicator of possible processing difficulties. Then the discussion focuses on the word-by-word reading times.

Preceding the analysis of the grammaticality judgements for the experimental items, the judgements of the filler items are considered. They provide a first indicator whether participants were able to process the sentences on-line and whether they were paying attention to the task. The results for the filler items are summarised in Table 6-4-3:

	Intermediate learners	Advanced learners	Native speakers
Questions	96.39 (5.66)	98.61 (3.16)	96.94 (4.63)
Declaratives	87.5 (11.32)	91.94 (6.74)	95.0 (6.42)
Total fillers	92.08 (7.45)	95.14 (3.29)	95.97 (4.02)

Table 6-4-3: Accuracy rates and (standard deviations) in the judgements of the filler items in Exp. 6

Table 6-4-3 indicates that the intermediate learners were slightly less successful than the other groups with the total fillers and showed a higher within-group variation, whereas the two other groups were approximately on the same level. An ANOVA with Group as a between-subjects variable for all sentences yielded a significant main effect ($F(2,89) = 4.573$, $p \leq 0.013$). Tukey post-hoc tests found a significant difference between the intermediate learners and the native speakers ($p \leq 0.014$) while the difference between the intermediate and advanced learners approached significance ($p \leq 0.067$). No significant difference was found between the advanced learners and the native speakers ($p \leq 0.812$). These patterns suggest that the intermediate learners were not as successful as the other groups in their grammaticality judgements. But their judgements

were still sufficiently successful to show that they were able to process complex sentences in this word-by-word reading task.

Table 6-4-3 suggests that the performance of all three groups differed between the questions and the declaratives, the latter being judged less successfully. In order to further investigate this effect, the results of these two conditions were submitted to a mixed two-factor ANOVA with Condition as a within-subjects variable and Group as a between-subjects variable.[85] The analysis yielded significant main effects for Group ($F(2,87) = 5.040$, $p \leq 0.008$) and Condition ($F(1,87) = 40.801$, $p \leq 0.0001$). The interaction was also significant ($F(2,87) = 5.027$, $p \leq 0.009$), which indicates that the effect for Condition was significant in the learners but not in the native speakers. Analysing possible reasons for the learners' slight difficulties with the declaratives is beyond the scope of the present thesis. It is more relevant for the purpose of the present studies that participants were comparatively successful judging the ungrammatical questions. This finding implies that they mastered the complex constructions and task demands and did not have a bias to accept ungrammatical sentences. The judgements of the experimental sentences are summarised in Table 6-4-4:[86]

	Intermediate learners	**Advanced learners**	**Native speakers**
Ambiguous	73.33 (26.03)	82.78 (17.77)	66.67 (23.16)
High	45.00 (23.63)	61.67 (28.08)	60.00 (23.41)
Low	69.44 (21.12)	71.67 (18.65)	70.00 (19.28)
Total	62.59 (19.56)	72.04 (14.03)	65.56 (17.16)

Table 6-4-4: Mean percentages and (standard deviations) of sentences judged grammatical in Exp. 6

As a first step, the total judgements for the three experimental structures were submitted to a one-factor ANOVA with Group as a between-subjects variable. This analysis did not yield a significant effect ($F(1,89) = 2.367$, $p \leq 0.100$). Hence, the three groups did not differ substantially in their judgements.

With regard to the judgements of each of the three structures, Table 6-4-4 indicates that all participants accepted the high disambiguated sentences less often than the other two structures. The difference was clearer in the learners than in the native speakers. Interestingly, all groups accepted approximately 70% of the low disambiguated sentences, whereas the judgements of the other two structures differed considerably between groups. In order to further investigate these effects, the results for the three conditions were submitted to a mixed two-factor ANOVA with the three-level factor Condi-

[85] Since the items differed across the two conditions, items could not be treated as a random variable.

[86] Since the grammaticality of the experimental items is difficult to determine, the number of sentences that were accepted by the participants rather than the number of correct answers is considered here.

tion as a within-subjects variable and Group as a between-subjects variable. This analysis yielded a significant main effect for Condition ($F(2,174) = 27.078$, $p \leq 0.0001$). Pair wise comparisons found significant differences between high and ambiguous conditions ($F(1,87) = 38.577$, $p \leq 0.0001$) as well as between high and low disambiguated conditions ($F(1,87) = 29.035$, $p \leq 0.0001$), but not between ambiguous and low disambiguated conditions ($F(1,87) = 3.052$, $p \leq 0.084$). The main effect for Group did not reach statistical significance ($F(2,87) = 2.423$, $p \leq 0.095$), but the interaction was significant ($F(4,348) = 3.766$, $p \leq 0.007$), confirming that the three groups' reading time patterns differed across conditions.

The possible reasons for the interaction were further explored by separate one-factor ANOVAs with the three-level factor Condition on the results of the three groups, respectively with subjects and items as a random variable. For the native speakers, the effect approached but did not reach statistical significance ($F_1(2,58) = 2.719$, $p \leq 0.074$; $F_2(2,10) = 3.321$, $p \leq 0.079$), which indicates that this group judged the three structures in a similar way. This was different for the advanced learners, for whose results the analysis yielded a significant main effect ($F_1(2,58) = 7.770$, $p \leq 0.001$; $F_2(2,10) = 21.664$, $p \leq 0.0001$). Pair-wise comparisons indicate that the difference between high and ambiguous conditions ($F_1(1,29) = 11.843$, $p \leq 0.0001$; $F_2(1,5) = 53.223$, $p \leq 0.001$) was significant, whereas the difference between ambiguous and low conditions ($F_1(1,29) = 6.591$, $p \leq 0.016$; $F_2 \leq 1$) was only significant in the analysis by subjects. The difference between high and low conditions did not reach statistical significance ($F_1(1,29) = 3.361$, $p \leq 0.077$; $F_2 \leq 1$). These results confirm the impression that this group clearly preferred the ambiguous structures over the structures disambiguated towards high or low attachment. For the intermediate learners, there was also a significant main effect ($F_1(2,58) = 25.605$, $p \leq 0.0001$; $F_2(2,10) = 6.623$, $p \leq 0.023$). Pair-wise comparisons showed that the difference between high and ambiguous conditions was significant ($F_1(1,29) = 31.311$, $p \leq 0.0001$; $F_2(1,5) = 15.223$, $p \leq 0.018$), whereas the difference between high and low conditions was significant only in the analysis by subjects ($F_1(1,29) = 35.002$, $p \leq 0.0001$; $F_2(1,5) = 4.261$, $p \leq 0.108$). The difference between ambiguous and low conditions was insignificant ($F_1(1,29) = 1.203$, $p \leq 0.282$; $F_2(1,5) = 1.054$, $p \leq 0.363$). Thus, in contrast to the advanced learners, the intermediate group had a tendency to judge the high disambiguated structures as being ungrammatical more often than the other two constructions.

The question arises whether the apparent difficulty with the high disambiguated sentences was also reflected in the response times for the grammaticality judgements. In contrast to the response times for the comprehension questions that followed the sentences in Experiments 2 to 4, the response times for the grammaticality judgements in the present study (see Table 6-4-5) include the judgements for all sentences. Sentences that were judged incorrect were not excluded from the analyses for two reasons. First, the experimental sentences cannot be definitely classified as being grammatical or ungrammatical, taking into account that although they contained argument structure violations, they were accepted by most of the English native speaker controls. Second, even if a participant judges a sentence as being ungrammatical, this does not necessarily imply that she did not understand and process the respective sentence successfully. Outliers, i.e. values deviating from the mean for each group and each item by more than 2.5 standard deviations, were excluded from the analysis. This procedure affected 4.1% of the intermediate learners' data, 4.6% of the advanced learners' data, and 3.3% of the native speakers' data. The resulting judgement times are summarised in Table 6-4-5:

	Intermediate learners	**Advanced learners**	**Native speakers**
ambiguous	1584.83 (612.98)	1298.16 (392.42)	1190.08 (303.16)
high	1826.05 (831.62)	1367.19 (784.99)	1381.63 (671.77)
low	1443.98 (470.09)	1191.97 (308.13)	1215.48 (374.90)

Table 6-4-5: Response times in ms and (standard deviations) for the grammaticality judgments in Exp. 6

A mixed two-factor ANOVA with Condition as a within-subjects variable and Group as a between-subjects variable yielded significant effects for Condition ($F(2,174) = 6.684$, $p \leq 0.002$) and Group ($F(2,87) = 7.011$, $p \leq 0.002$), but no significant interaction ($F(4,348) = 0.608$, $p \leq 0.658$). To analyse this finding more closely, separate repeated-measures ANOVAs with Condition were conducted for each group, respectively with subjects and with items as a random variable. For the intermediate learners, the by subjects analysis yielded a significant main effect ($F_1(2,58) = 5.092$, $p \leq 0.009$; $F_2(2,10) = 2.869$, $p \leq 0.104$). Planned pair-wise comparisons by subjects found a significant difference between high and low disambiguations ($F_1(1,29) = 8.341$, $p \leq 0.007$; $F_2(1,5) = 4.074$, $p \leq 0.100$) and the difference between high and ambiguous sentences approached significance ($F_1(1,29) = 3.370$, $p \leq 0.077$; $F_2(1,5) = 3.005$, $p \leq 0.144$). In contrast, the difference between ambiguous and low disambiguated conditions was not statistically significant ($F_1(1,29) = 2.151$, $p \leq 0.153$; $F_2 \leq 1$). The findings indicate that these learners took longer to judge the high disambiguated sentences than any of the other two conditions. Although Table 6-4-5 suggests a similar tendency for the other

groups, neither the advanced learners ($F_1 \leq 1$; $F_2 \leq 1$) nor the native speakers ($F_1(2,58) = 2.099$, $p \leq 0.132$; $F_2(2,10) = 1.351$, $p \leq 0.303$) showed a significant main effect for Condition.

The following paragraphs focus on the word-by-word reading times. Preceding the analyses, it must be considered which data are included in the analyses. In contrast to other experiments using similar experimental paradigms (e.g. Juffs 1998a,b; Juffs & Harrington 1995, 1996, and the discussion in Chapter 4.2.), sentences that were judged as incorrect were not excluded from the reading time analyses. The deletion of outliers affected 3.5% of the intermediate learners' data, 3.4% of the advanced learners' data, and 2.9% of the native speakers' data. A summary of the resulting reading times per group in all regions is provided in Appendix E4.

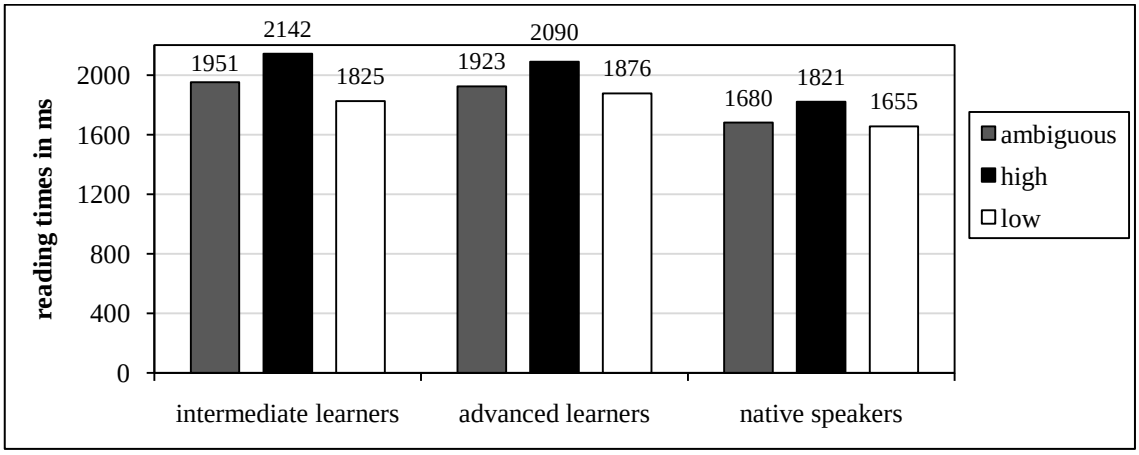

Figure 6-4-1: Participants' reading times in the PP-region in Exp. 6

In a first step, a mixed three-factor ANOVA with Group as a between-subjects variable and Condition as a within-subjects variable was run on the results for the regions of interest, namely the PP-region and the wrap-up region. For the PP-region, this analysis yielded significant effects for Condition ($F(2,174) = 20.672$, $p \leq 0.0001$) and Group ($F(2,87) = 4.617$, $p \leq 0.012$), but no significant interaction ($F(4,348) = 0.787$, $p \leq 0.535$). The Tukey post-hoc test found significant differences between intermediate learners and native speakers ($p \leq 0.048$), as well as between advanced learners and native speakers ($p \leq 0.017$), but not between the two learner groups ($p \leq 0.911$). For the wrap-up region, no significant effect for Condition was found ($F(2,174) = 0.186$, $p \leq 0.830$) and the effect for Group approached significance ($F(2,174) = 2.590$, $p \leq 0.081$). No significant interaction of the two factors was found, either ($F(4,348) = 0.718$, $p \leq 0.580$). This result indicates that the effect from the PP region did not spill over to the following wrap-up region. Hence, the wrap-up region is not considered in the following analyses.

Although no significant interaction with Group was found, Figure 6-4-1 suggests some differences between the three groups. Therefore, the results of the individual groups were analysed separately in order to better understand the specific reading time patterns. The native speakers seemed to take longer for the high disambiguated sentences than for any of the other two structures. This impression was confirmed by ANOVAs with Condition respectively as a within-subjects and a within-items variable, which yielded a significant main effect ($F_1(2,58) = 4.306$, $p \leq 0.018$; $F_2(2,10) = 4.514$, $p \leq 0.040$). Pair-wise comparisons found that only the differences between high and ambiguous ($F_1(1,29) = 5.780$, $p \leq 0.023$; $F_2(1,5) = 4.129$, $p \leq 0.098$) and high and low disambiguated sentences ($F_1(1,29) = 5.952$, $p \leq 0.021$; $F_2(1,5) = 6.021$, $p \leq 0.058$) were significant in the analysis by subjects. In contrast, the difference between ambiguous and low disambiguation conditions did not yield a significant effect ($F_1 \leq 1$; $F_2 \leq 1$).[87]

The advanced learners' reading time pattern (see Figure 6-4-1) was similar. This impression was confirmed by ANOVAs with Condition respectively as a within-subjects and a within-items variable. For the PP-region, these analyses yielded a significant main effect ($F_1(2,58) = 14.286$, $p \leq 0.0001$; $F_2(2,10) = 6.450$, $p \leq 0.016$). Pair-wise comparisons found significant differences between ambiguous and high disambiguated structures ($F_1(1,29) = 16.108$, $p \leq 0.0001$; $F_2(1,5) = 4.944$, $p \leq 0.077$) and the two disambiguated structures ($F_1(1,29) = 19.704$, $p \leq 0.0001$; $F_2(1,5) = 14.158$, $p \leq 0.013$), but not between ambiguous and low disambiguated structures ($F_1 \leq 1$; $F_2 \leq 1$).

As Figure 6-4-1 indicates, the intermediate learners' reading times seemed to differ not only between the high disambiguated condition and each of the other two conditions, but also between ambiguous and low disambiguated conditions. However, this impression was not confirmed by the statistical analyses, which replicated the pattern of the other two groups. The ANOVAs with Condition as within-subjects and within-items variable found a significant main effect ($F_1(2,58) = 5.217$, $p \leq 0.008$; $F_2(2,10) = 10.544$, $p \leq 0.003$). Pair-wise comparisons found that the difference between ambiguous and high disambiguated sentences approached significance ($F_1(1,29) = 3.777$, $p \leq 0.062$; $F_2(1,5) = 5.463$, $p \leq 0.067$). The difference between the two disambiguated structures was highly significant ($F_1(1,29) = 9.204$, $p \leq 0.005$; $F_2(1,5) = 24.241$, $p \leq 0.004$), but no significant effect was found between ambiguous and low disambiguated structures ($F_1(1,29) = 1.637$, $p \leq 0.221$; $F_2(1,5) = 4.322$, $p \leq 0.092$).

[87] Although reading times in the wrap-up region also seemed to be higher for the high disambiguated structures than for any of the two other two structures, this effect did not reach statistical significance ($F_1(2,58) = 1.427$, $p \leq 0.248$; $F_2(2,10) = 0.814$, $p \leq 0.470$).

Following Igoa et al. (1998), the present analyses focused on the combined reading times for the entire disambiguating region. However, the English sentence is disambiguated at the preposition itself, depending on whether it is of or from. Accordingly, it is possible that early processing effects already show upon the preposition. This possibility is investigated in the following paragraphs. Preceding the analyses, outliers were excluded using the procedure described in Chapter 4.5. Thereby 3.5% of the intermediate learners' data, 4.3% of the advanced learners' data, and 2.8% of the native speakers' data were excluded. The resulting reading times are summarised in Figure 6-4-2:

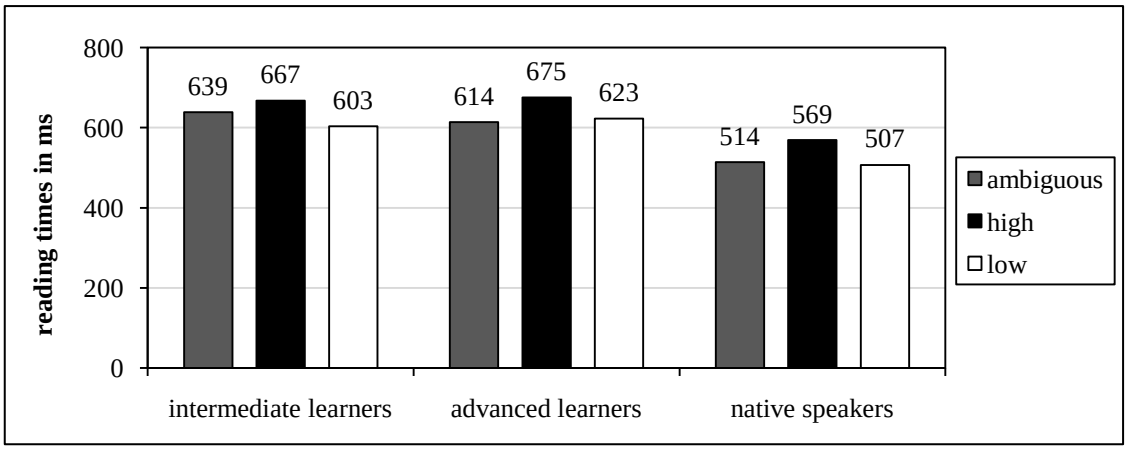

Figure 6-4-2: Reading times on the disambiguating preposition in Exp. 6

The reading times on the preposition were analysed in a similar way as those of the PP-region. First, a mixed two-factor ANOVA with Group as a between-subjects variable and Condition as a within-subjects variable was run on the results, yielding significant effects for Condition ($F(2,174) = 26.226$, $p \leq 0.0001$) and Group ($F(2,87) = 4.586$, $p \leq 0.013$), but no significant interaction ($F \leq 1$). The Tukey post-hoc test found significant differences between intermediate learners and native speakers ($p \leq 0.034$), as well as between advanced learners and native speakers ($p \leq 0.023$), but not between the two learner groups ($p \leq 0.983$).

In a next step, the results of the individual groups were analysed separately. As Figure 6-4-2 indicates, the native speakers' pattern was similar as for the entire PP-region. This impression was confirmed by ANOVAs with Condition as a within-subjects and as a within-items variable, which yielded a significant main effect ($F_1(2,58) = 8.800$, $p \leq 0.001$; $F_2(2,10) = 5.161$, $p \leq 0.029$) for this group. Pair-wise comparisons showed that only the differences between high and ambiguous ($F_1(1,29) = 10.609$, $p \leq 0.003$; $F_2(1,5) = 7.408$, $p \leq 0.042$) and between high and low disambiguated conditions ($F_1(1,29) = 10.492$, $p \leq 0.003$; $F_2(1,5) = 5.925$, $p \leq 0.051$) were significant. In contrast, the difference between ambiguous and low disambiguated conditions did not yield a

significant effect ($F_1 \leq 1$; $F_2 \leq 1$). For the advanced learners, the analyses also yielded a significant main effect ($F_1(2,58) = 10.078$, $p \leq 0.0001$; $F_2(2,10) = 5.203$, $p \leq 0.028$). As pair-wise comparisons showed, the differences between ambiguous and high disambiguated sentences ($F_1(1,29) = 14.307$, $p \leq 0.001$; $F_2(1,5) = 4.834$, $p \leq 0.079$) and between the two disambiguated structures ($F_1(1,29) = 23.657$, $p \leq 0.0001$; $F_2(1,5) = 6.384$, $p \leq 0.053$) were significant, at least in the analysis by subjects. The difference between ambiguous and low disambiguated structures did not reach statistical significance ($F_1 \leq 1$; $F_2 \leq 1$). Figure 6-4-2 shows that the intermediate learners differed from the other two groups in that they took longer to read the ambiguous than the low-disambiguated structures. The respective analyses yielded a significant main effect ($F_1(2,58) = 5.477$, $p \leq 0.007$; $F_2(2,10) = 6.772$, $p \leq 0.014$). Pair-wise comparisons found significant differences between ambiguous and high disambiguated ($F_1(1,29) = 6.043$, $p \leq 0.002$; $F_2(1,5) = 14.655$, $p \leq 0.012$) and between ambiguous and low disambiguated sentences ($F_1(1,29) = 7.657$, $p \leq 0.010$; $F_2(1,5) = 8.547$, $p \leq 0.033$), but not between the two disambiguated structures ($F_1 \leq 1$; $F_2 \leq 1$). Taken together, the results lead to the conclusion that the patterns for the prepositions were highly similar to those for the entire PP regions. All three groups read the critical regions of the ambiguous and the low disambiguated structures approximately equally fast and faster than the same regions in the high disambiguated structures. The following chapter discusses the theoretical implications of these findings.

6.4.4 Discussion

Although the measures provided by the task, i.e. the grammaticality judgements and the word-by-word reading times, cannot be directly compared, the question arises whether participants showed the same tendencies in these tasks. In order to address this question, the results from both measures are discussed in turn in the following.

The tendency across groups and structures to judge the sentences grammatical is unexpected because all sentences contained a violation of one of the verbs' subcategorisation requirements. This result does not necessarily imply that participants were insensitive to these requirements, but it indicates that the violations were not strong enough to make a sentence ungrammatical. The native speakers' tendency to accept the sentences is especially noteworthy because it contradicts findings from previous studies using the same type of on-line task (Juffs 1998a,b; Juffs & Harrington 1995, 1996), in which the native speakers tended to reject the (grammatical) garden-path sentences presented to them. The present results thus suggest that the previous findings do not result from the

native speakers' oversensitivity but rather from the sentence types in the respective studies. Hence, it is possible that the high acceptance rates of the present sentences can be related to an incomplete processing route as it has been proposed in Ferreira, Bailey & Ferraro's (2002) notion of "good enough" representations and Sanford & Sturt's (2002) underspecification account. Similar as in Ferreira et al.'s (2002) garden-path sentences, participants in the present study might have associated one argument with two verbs at the same time. Such an account would be in line with the SSH, which predicts that learners rely more strongly on incomplete syntactic representations than native speakers. However, the pattern found in the present study was not in line with this prediction. Instead, the intermediate learners rejected the sentences more often than the other two groups, possibly because they were more insecure than the other groups in the comparatively demanding task and accordingly processed the sentences more carefully and consciously. This finding indicates that these learners rely less strongly on incomplete representations than the other groups, in contrast to the predictions of the SSH.

However, such a conclusion might be too strong on the basis of the total results only, considering that the three groups differed substantially across experimental structures. While there were hardly any differences between groups in the low disambiguated condition, the intermediate learners accepted the high disambiguated sentences less often than the other two groups, which implies that they were sensitive to this manipulation. In the ambiguous sentences, in contrast, both learner groups but especially the advanced learners accepted the sentences more often than the native speakers. Assuming that the rejections of sentences were related to argument structure violations, the learners' pattern for the ambiguous structures indicates that they did not always perceive the violations in these structures. But there are two arguments against such an explanation. First, it is implausible that the advanced learners were less sensitive to argument structure than the intermediate learners. Second, all learners were comparatively successful in the disambiguated conditions. Therefore, the learners' high acceptance rate for the ambiguous structures is more likely related to 'good-enough' representations.

It is possible, however, that the subcategorisation requirements for some of the verbs were particularly strong and hence the learners did not perceive the argument structure violation. But such an explanation is unlikely for two reasons. First, the native speakers did not show such a tendency. Second, all sentences used the same verbs, although in different constellations. So if the learners had misclassified the argument structure requirements of any of the verbs, this would have shown in their high acceptance of the

high- or low-disambiguated sentences. Nevertheless, in order to fully clarify this point, an additional off-line grammaticality judgement task was conducted with ten intermediate learners, ten advanced learners and ten native speakers who had not taken part in any of the other studies of the present thesis. In this task, participants judged the grammaticality of simple sentences containing the experimental verbs on a scale from one (very natural) to five (very unnatural). Two lists were created so that each participant saw a verb either in its grammatical (67a) or in its ungrammatical version (67b).

(67) a. Susan offered the book to Mary.

b. Susan offered the book.

Table 6-4-6 presents the results for each verb combination that was used in the ambiguous sentences of Experiment 6.

Verbs	Verb 1		Verb 2		Verb x Gram. ANOVA
	3-arg.	2-arg.	3-arg.	2-arg.	
Interm. learners	1.4	2.9	1.4	3.0	Verb: $F(1,9) = 0.086$, $p \leq 0.772$ Gram: $F(1,9) = 83.620$, $p \leq 0.0001$ Verb x Gram: $F(1,9) = 0.536$, $p \leq 0.472$
Advanced Learners	1.4	2.8	1.5	2.9	Verb: $F(1,9) = 0.125$, $p \leq 0.561$ Gram: $F(1,9) = 45.307$, $p \leq 0.0001$ Verb x Gram: $F(1,9) = 0.392$, $p \leq 0.447$
Native speakers	1.5	2.9	1.5	2.9	Verb: $F(1,9) = 0.172$, $p \leq 0.684$ Gram: $F(1,9) = 43.124$, $p \leq 0.0001$ Verb x Gram: $F(1,9) = 0.020$, $p \leq 0.889$

Table 6-4-6: Mean ratings and statistical effects for the control task for Exp. 6

The findings of a significant effect for Grammaticality but no significant effect for Verb and no significant interaction between the factors indicate that the learners were relatively consistent in their grammaticality judgements for the verbs and were sensitive to the respective verbs' subcategorisation requirements.[88]

Accordingly, the learners' high acceptance of the ambiguous structures in this task requires an alternative explanation. In Chapter 6.2. it was suggested that a preference of ambiguous over disambiguated structures can have two reasons. Either participants interpret the sentences in line with their preferred interpretation, or they create a 'good enough' representation (Christianson et al. 2001), not deciding upon a specific interpretation. The first explanation is not supported by the present results for the following reason: If participants interpreted the sentences in line with their favourite interpretation, they should accept them as often as the other structures. Instead, the learners ac-

[88] A mixed three-factor ANOVA with Group as a between-subjects variable and Verb and Grammaticality as within-subjects variables did not yield a significant effect for Group ($F(1,27) = 0.073$, $p \leq 0.788$) and no significant interactions of Group x Verb ($F \leq 1$) and Group x Grammaticality ($F \leq 1$). Hence, the groups did not differ significantly in their judgements.

cepted particularly the ambiguous structures, which indicates that they took the PP as attaching to both verbs so that both verbs' subcategorisation frames would be satisfied.

The finding that such a pattern was particularly frequent in the learners is in line with the SSH, which predicts that learners under-use structural principles and create 'good enough' representations more often than native speakers. However, this effect did not show in the learners' reading time data. The advanced learners' reading times were approximately equal for ambiguous and low disambiguated sentences. The intermediate learners' reading times were even higher for the ambiguous than for the low disambiguated constructions, although this difference did not yield statistical significance. This finding suggests that the learners used a structural principle like Late Closure. Moreover, the learners did not seem to interpret the ambiguous structures in the same way as the low disambiguated structures. Instead, the learners' higher acceptance rates for the ambiguous structures in the grammaticality judgements suggest that they were aware of both interpretations of these structures and hence found them more acceptable than the high disambiguated structures that violate Late Closure. Although prima facie the learners' patterns are not in line with the SSH, it is possible that the learners' low attachment preference is a default chosen in the absence of structural information to guide their attachment preferences. If this was the case, however, the SSH would be extremely difficult to test, considering that any attachment preference found in the learners could be interpreted as a default.

Given that the structures under investigation were directly disambiguated by the preposition, the present study allowed for a comparison of intermediate processing decisions on the preposition with less intermediate effects on the entire PP-region. Interestingly, this comparison yielded highly similar results for all three groups, which indicates that the low attachment preference was immediate and participants understood the disambiguating function of the preposition as soon as it was encountered.

6.5 Summary

The rationale behind the three studies on PP attachment was to investigate ESL learners' attachment preferences in cases that were not subject to cross-linguistic differences between the learners' L1 and L2 so that no transfer effects were to be expected.

The first of these studies found a native-like high attachment preference both in the intermediate and advanced learners. The strong high attachment preference in the ditransitive-verb condition is in line with previous L2 processing studies that found native-like sensitivity to verb argument structure in L2 learners (Altarriba, Kambe, Pollat-

sek & Rayner 2001; Frenck-Mestre & Pynte 1997; Juffs 1998a,b; Juffs & Harrington 1995, 1996). Like the native speaker controls the learners in Experiment 4 preferred VP attachment of the PP, even if the PP was not required by the verb's argument structure. This finding provides evidence against the learner-specific recency strategy that was suggested in previous studies (Frenck-Mestre & Pynte 1997 and Experiment 2 of the present thesis), at least for these comparatively short and simple sentences.

Experiments 5 and 6 investigated a more complex PP attachment ambiguity in which a learner-specific recency strategy would be more likely. Furthermore, they allowed for a direct comparison of off-line and on-line processing of the same structure, which turned out to be similar. But it has to be noted that the learners' low attachment preference was not as strong as that of the native-speakers in the off-line task, while there were no significant differences between the groups in the on-line SPR task. This finding would be in line with a processing capacity theory claiming that a recency strategy would be employed in on-line rather than in off-line processing. Such a strategy is not necessarily L2-specific, considering that the native speakers also showed a low attachment preference in Experiments 5 and 6. Therefore, the learners' low attachment preference implies that they make use of Late Closure or Recency in the same way as native speakers.

Experiment 6 found no evidence for processing differences between ambiguous and low disambiguated structures. This finding suggests that the learners did not prefer ambiguous over disambiguated sentences, as would be predicted by the SSH. Moreover, the learners did not have considerable difficulties computing the two interpretations in parallel. It is possible that the learners simply did not construct two interpretations. But such an explanation is improbable because the learners accepted these structures considerably more often than the other two structures. This finding implies that they associated the PP to the two verbs at the same time. Thus, they were at least subconsciously aware of the two possible interpretations.

Taken together, the results of the six studies reported so far provide some insights into ESL learners' off-line and on-line syntactic ambiguity resolution. Both intermediate and the advanced learners showed highly native-like processing with regard to verb argument structure (Experiment 4), but also with regard to structural processing principles such as Late Closure (Experiments 5 and 6). In tasks that involve increased processing load, the learners were found to resort to Late Closure or Recency (Experiment 2).

When processing structures that underlie cross-linguistic variation, the learners showed evidence of transfer effects (Experiments 1 and 3).

But one important issue could not be fully resolved on the basis of these studies: the validity of the SSH (Clahsen & Felser 2006b; see the discussion in Chapter 3.1.2.). The prediction of the SSH that L2 learners should show no attachment preferences was not confirmed by the previous studies. However, closer analyses indicated that some of the results might still be compatible with the SSH. One example is the learners' sensitivity to verb argument structure that was found in Experiment 4 (see Chapter 6.1.). Furthermore, it is conceivable that learners do not construct deep syntactic structures but nevertheless show default attachment preferences which are guided by non-syntactic factors. Such default strategies could be a recency strategy in complex constructions like the RCs in Experiment 2 or the PPs in Experiment 6, a reliance to discourse principles and verb argument structure in Experiment 4, or L1 transfer as in Experiments 1 and 3. Although the predictions of the SSH were not fully confirmed by any of the present studies, the counterevidence was inconclusive, either, because both L1 transfer and a learner-specific recency strategy might be default strategies employed by the learners in the absence of deep structural information to guide their processing decisions. Furthermore, the SSH does not make clear predictions for the syntactic structures investigated in the studies. Therefore, the structures that are tested in the next chapters involve syntactic long-distance movement, which is predicted to be non native-like according to the SSH. Furthermore, Experiments 7 and 8 allow for a comparison of learners' off-line and on-line processing of these structures, which might differ according to the SSH.

7 Attachment preferences in adjunct questions

In order to obtain clearer results with regard to the SSH, the present studies investigate structures which provide a better testing ground for syntactic processing, i.e. structures that involve syntactic movement traces. While previous L2 processing studies on such constructions obtained inconclusive results (see Chapter 3.2.5.), the present studies examine a so far unexplored structure that involves syntactic movement and at the same time tests learners' attachment preferences. The ambiguous version of this construction is illustrated in (68).

(68) When did the watchman think that the robber had entered the building?

In this question the attachment of the adjunct when is ambiguous. It could either be extracted from the matrix clause and thus modify the verb think, or from the embedded clause and thus modify the verb enter. In the latter reading, a generative analysis would assume an intermediate trace in the CP-specifier position preceding the complementiser that, as illustrated in the following example:

(69) When$_i$ did the watchman think t_i that the robber had entered the building t_i?

According to this analysis, the intermediate trace should show an effect in psycholinguistic studies when participants assume the long movement reading. In comparison to previously investigated structures that involve intermediate traces (see Chapter 3.2.5.), these constructions have the advantage that the moved element is not an argument but an adjunct of the verb. In the case of argument extraction, any observed trace effect might be purely lexical, induced by the integration of the operator into the verb's argument structure (see Marinis et al. 2005). Since this is not the case for the adjunct questions in the present study, any delay on the SpecCP, which is expected to show on the complementiser that, can be attributed to the intermediate trace. The study by Marinis et al. (2005) has been criticised for the reason that a possible trace effect in the learners might not have been discovered because it spilled over to the region following the complementiser, where it was covered up by the comparative length of that region (see Dekydtspotter, Schwartz & Sprouse 2006, and the discussion in Chapter 3.2.5.). The present study addresses such criticism by using a word-by-word SPR paradigm and by carefully analysing all the critical regions separately.

While Experiment 7 tested participants' off-line attachment decisions for ambiguous sentences, Experiment 8 examined the on-line processing of both ambiguous and disambiguated sentences. Preceding the presentation of the experiments, the corresponding structures in German, i.e. the learners' L1, are considered in order to better understand

the possible processing difficulties German ESL learners might face when reading these structures in English.

7.1. Ambiguous adjunct questions in German

The previously reported studies showed that cross-linguistic influences have to be considered in all L2 processing studies, even if the constructions under investigation are similar in the learners' L1 and L2. Therefore, the present chapter presents a control study that tested German native speakers' attachment preferences for the German counterpart to the construction investigated here. An illustrative example is provided below:

(70) Wann hat der Wachmann gedacht dass der Räuber das Gebäude betreten hat?
When has the watchman thought that the robber the building entered has?
'When did the watchman think that the robber had entered the building?'

Although the construction is similar in the two languages, the use of the complementiser dass / that seems to be slightly different. In German it is quite natural to leave out the complementiser and to raise the verb into the head position of the CP of the embedded sentence, as illustrated in (71):

(71) Wann hat der Wachmann gedacht hat der Räuber das Gebäude betreten?
When has the watchman thought has the robber the building entered?
'When did the watchman think had the robber entered the building?'

The corresponding English construction seems to be unacceptable (see Chapter 7.3.1. for a naturalness judgement task with native speakers of English). The German sentences, however, are perfectly acceptable, and, interestingly, they elicit the interpretation of when as an adjunct to the embedded verb. This impression of the author, who is a native speaker of German, was tested by an off-line task in which 30 native speakers of German read sentences such as (70) and (71) with their two possible interpretations and indicated their preferred interpretation.[89] The results of the study are summarised in the following table:

	with dass		**verb raising without dass**	
	non-local attachment	**local attachment**	**non-local attachment**	**local attachment**
Mean (SD)	33.31 (31.56)	66.64 (31.57)	89.43 (14.85)	10.54 (14.82)

Table 7-1-1: German native speakers' attachment preferences in German ambiguous adjunct questions

[89]The sentences were close translations of the English materials. There were six sentences with the complementiser *that*, six sentences without the complementiser and verb raising to the CP-head, and 16 ambiguous filler sentences of different constructions. All participants were German university students. Their mean age was 24 (SD = 5.6, range = 18-31), and 76.7% of them were female.

The results suggest that participants' preferences differ across the two conditions, in that they prefer extraction of the adjunct from the matrix clause in the first and extraction from the embedded clause in the second condition. This observation was confirmed by an ANOVA with the factors Condition (dass vs. without-dass) and Attachment (local vs. non-local), which yielded no significant effect for Condition ($F \leq 1$), but a significant effect for Attachment ($F(1,29) = 10.906$, $p \leq 0.003$) and a highly significant interaction ($F(1,29) = 93.905$, $p \leq 0.0001$).

To conclude, the direct German counterpart of the English construction elicits a local attachment preference because there is an alternative in German that is biased towards the non-local reading.

7.2 Predictions and research questions

If the integration of syntactic operators with their governing categories at the trace position causes increased processing load (Gibson & Warren 2004), the English native speakers are expected to prefer local over non-local attachment in the present study. This effect might be stronger in the on-line task than in the off-line task. For the learners, who are assumed to have limited processing capacity, the local attachment preference is expected to be even stronger. Furthermore, it is possible that the learners' local attachment preference is reinforced by L1 influence, because the directly corresponding German construction elicits a local attachment preference (see Chapter 7.1.). The German ESL learners might not be aware that, unlike in their L1, there is no alternative construction in English that is biased towards non-local attachment.

According to the SSH, only the native speakers should show an intermediate trace effect in the on-line task, i.e. longer reading times in the SpecCP-position when non-local attachment is forced. The learners, in contrast, are not predicted to show an intermediate trace effect because they should be unable to construct deep syntactic structures that require such traces. Moreover, the SSH predicts that learners process ambiguous sentences faster than disambiguated sentences because the former do not force them to make any commitment. It is also conceivable, however, that the on-line results for the ambiguous sentences resemble those of Experiment 6 (see Chapter 6.4.), which also compared a non-local attachment condition, a local attachment condition and an ambiguous condition. Experiment 6 found that reading times were similar for ambiguous and locally disambiguated sentences and faster for these two structures than for the sentences disambiguated towards non-local attachment. This pattern was interpreted as evidence that the learners neither preferred these structures over the locally disambigu-

ated structures, as it would be predicted by the SSH, nor did they have particular difficulties with the two interpretations.[90] In the present study, both learners and native speakers might interpret the ambiguous structures like their preferred disambiguation, i.e. local attachment. Accordingly, reading times should be similar for these two constructions, while they are expected to be elevated for the structures that are disambiguated towards non-local attachment.

To sum up, the learners are expected to show a local attachment preference in the present studies. This finding might be related to shallow structure processing, to L1 influence, or to a general preference for the structurally simpler construction because of processing capacity restrictions. In all these cases, the local attachment preference is expected to be stronger in the intermediate learners than in the advanced learners and stronger in the on-line task as compared to the off-line task. The present studies address the following research questions:

A) How do ESL learners process ambiguous adjunct questions off-line and on-line?
B) To what extent are these preferences influenced by the learners' proficiency?
C) Are ESL learners influenced by their L1 or by an L2-specific strategy when processing ambiguous adjunct questions?

7.3 Experiment 7: Off-line attachment preferences

7.3.1 Participants and procedure

The participants were the same as in Experiment 5 (see Chapter 6.3.1.). The materials were adjunct questions like (68), here repeated as (72).

(72) When did the watchman think that the robber had entered the building?

For the norming study, eight sentences of the same type were created. While this construction is ambiguous, the alternative construction in (73a) corresponds to the German construction in (71) that has a stronger tendency towards non-local attachment. The manipulation of the presence or absence of the complementiser was included into the norming task in order to test whether the tendency towards non-local attachment found in German (see Chapter 7.1.) is similar in English. Furthermore, an alternative construction without the complementiser but without verb raising (73b) was included in the norming study.

(73) a. When did the watchman think had the lady stolen the painting?
b. When did the watchman think the lady had stolen the painting?

[90] While it is possible that the learners simply did not consider two interpretations, this is improbable because the learners accepted these structures considerably more often than the other two structures. This indicates that they took the PP to fulfil the argument structure of the two verbs in parallel.

For the norming task, two lists were created so that each participant saw a sentence in only one of the following constructions: (a) with the complementiser *that*, (b) without the complementiser and with inversion, or (c) without the complementiser and without inversion. Since the on-line Experiment 8 required disambiguated sentences, these were included in the norming studies as well. Sentences were disambiguated via pragmatic information on the auxiliary, as illustrated in (74):

(74) How could / did the sergeant think that the lady had stolen the painting?

The disambiguated sentences were integrated into the two lists in such a way that each participant saw a sentence either disambiguated towards local or non-local attachment. Each list thus contained four ambiguous sentences with the complementiser that, four without that, four sentences with high and four with low disambiguation. For the norming study, the 16 experimental sentences per list were intermingled with 12 sentences from Experiment 5 (see Chapter 6.3.). Ten native speakers of English (five per list) judged the naturalness and ten others judged the real-world plausibility of the sentences on a scale from one (very natural / plausible) to five (very unnatural / implausible).

Although all types of items were included in the norming study, only the results for the ambiguous sentences are reported here because these were used in the present off-line study (for a discussion of the disambiguated sentences, see Chapter 7.4.1.). Table 7-3-1 provides an overview of the naturalness and plausibility judgements for the ambiguous items.

	ambiguous	**raising without *that***	**non-raising without *that***
Naturalness	2.15 (0.97)	4.68 (0.49)	1.97 (1.03)
Plausibility	1.75 (0.71)	4.13 (1.33)	1.89 (0.91)

Table 7-3-1: Naturalness and plausibility judgements for Experiment 7

The table shows that the ambiguous sentences with verb raising were judged as being considerably more unnatural than the other two structures. This impression was statistically confirmed by a three-level repeated measures ANOVA, which yielded highly significant main effects both with subjects and with items as a random variable ($F_1(1,9) = 2.809$, $p \leq 0.0001$; $F_2(1,7) = 70.422$, $p \leq 0.0001$). Since the verb raising construction is obviously unacceptable in English, it was neither included in the off-line questionnaire nor in the on-line SPR task, but it will be considered in the further discussion, especially concerning the corresponding German construction.

As Table 7-3-1 indicates, the plausibility judgements showed a similar pattern as the naturalness judgements. The raising construction received considerably more negative

judgements than the other two constructions. The difference was highly significant in the one-factor ANOVA, both with subjects and with items as random variables ($F_1(1,9) = 42.638$, $p \leq 0.0001$; $F_2(1,7) = 53.708$, $p \leq 0.0001$). As in the previous studies, participants' plausibility judgements were influenced to some extent by the grammaticality or naturalness of the sentences. Nevertheless, the differences within conditions are relatively small and insignificant, which indicates that the experimental sentences do not differ considerably in their real-world plausibility. Hence, this factor is not expected to influence participants' attachment preferences in the off-line questionnaire.

In the off-line questionnaire, the eight ambiguous questions from the norming study were listed with their two possible interpretations, as illustrated in (75).

(75) When did the detective say that the thief had stolen the painting from the gallery?

□ The detective said that the thief had stolen the painting at midnight.

□ The detective said in the morning that the thief had stolen the painting.

In half of the experimental items the non-local reading preceded the local reading; in the other half the order was reversed. The sentences were intermingled with eight ambiguous sentences from Experiment 6 (see Chapter 6.4.) and eight unrelated ambiguous sentences (see Appendix D1 for a list of experimental items). Participants were instructed to choose their preferred interpretation as spontaneously as possible, but, as in the other off-line tasks, there was no time limit.

7.3.2 Results

The three groups' results of the off-line questionnaire are summarised in Table 7-3-2:[91]

	% non-local attachment	% local attachment
Native speakers	79.58 (18.12)	20.42 (18.12)
Advanced learners	57.33 (27.56)	42.67 (27.65)
Intermediate learners	47.91 (27.48)	52.08 (27.48)

Table 7-3-2: Mean off-line attachment choices and (standard deviations) in Exp. 7

A two-factor ANOVA with Attachment as a within-subjects variable and Group as a between-subjects variable yielded a significant effect for Attachment ($F(1,87) = 18.742$, $p \leq 0.0001$), no significant effect for Group ($F \leq 1$), and a highly significant interaction ($F(2,87) = 12,384$, $p \leq 0.0001$), which indicates that attachment preferences differed between the groups. In order to further explore the interaction, separate one-factor ANOVAs were run on the results of each group. For the native speakers, there was a

[91] The results of the filler items in this task will not be considered here. As all items were completely ambiguous, there is no correct or wrong answer. The results for the eight filler items that were at the same time the experimental sentences for Experiment 5 are reported in Chapter 6.3.3..

significant difference between local and non-local attachment ($F_1(1,29) = 65.002$, $p \leq 0.0001$; $F_2(1,7) = 8.972$, $p \leq 0.020$). For the two learner groups, in contrast, the analyses did not yield any significant effects (advanced learners: $F_1(1,29) = 2.036$, $p \leq 0.165$; $F_2(1,7) = 1.703$, $p \leq 0.233$, the intermediate learners: $F_1 \leq 1$; $F_2 \leq 1$).

Although none of the effects in the learner groups reached statistical significance, Table 7-3-2 suggests that the advanced learners showed a small but insignificant advantage for non-local attachment, while the intermediate learners did not show any preference at all. In Experiment 1, it turned out to be interesting to investigate whether the absence of a preference is related to a general indeterminacy in all participants or rather to some participants having a high and some a low attachment preference. A similar pattern was found in the present study. Of the 30 intermediate learners, only 3 (i.e. 10%) preferred long and short movement in half of the sentences. 13 intermediate learners (i.e. 43.3%) clearly preferred short movement, and 14 (i.e. 46.7%) preferred long movement in most of the sentences. The advanced learners' pattern was similar in that 4 of them (i.e. 13.3%) did not show any preference, while 18 (i.e. 60%) preferred long movement and 8 (i.e. 26.7%) preferred short movement in most of the cases. The pattern suggests that the more proficient the learners are, the more they prefer attachment to the embedded clause. This possibility was tested by conducting a correlation analysis of the learners' embedded-clause attachment choices and their proficiency as measured by the years they had been learning English, the time they had spent in English-speaking countries, and their self-rated proficiency. There were no significant correlations of long movement choices, neither with years of learning English ($r(60) = 0.205$, $p \leq 0.119$), nor with months spent in English-speaking countries ($r(60) = 0.167$, $p \leq 0.207$), nor with self-rated proficiency ($r(60) = 0.193$, $p \leq 0.142$).

7.3.3 Discussion

The native speakers' clear preference for long movement in the present study is unexpected because there is evidence from monolingual processing studies that shorter wh-movement is preferred over longer movement (Crain & Fodor 1985; Frazier 1987; Gibson 1998; Stowe 1986). It is possible that the long movement preference in the present off-line study was triggered by the lack of contexts for the sentences. For some reason it might have seemed more plausible to the participants that an event had taken place "last week" than "yesterday". But such an account cannot fully explain the present results since a follow-up study with contexts that provided equally probable non-local and local

attachment interpretations (Rah in prep.) found the same long movement preference in a group of English native speakers (n = 30). An example sentence is given in (76):

(76) John's girlfriend Mary was very fond of travelling. Yesterday at the party, he told his friends that she had travelled to China by train last summer.
When did he say that his girlfriend had travelled to China?
□ Yesterday
□ Last summer

The results for all three groups are summarised in the following table:

	% non-local attachment	% local attachment
Native speakers	96.4 (7.1)	3.6 (7.1)
Intermediate learners	38.9 (36.7)	61.1 (36.7)

Table 7-3-3: Off-line attachment preferences and (standard deviations) in a follow-up study with context (Rah in prep)

As Table 7-3-3 shows, the native speakers' preference for extraction from the embedded clause in this follow-up study was even stronger than in the study without contexts, while the learners' preferences were highly similar in the two studies. This finding indicates that the results of Experiment 7 are not only caused by the absence of contexts for the sentences.[92] To conclude, the English native speaker control group in the present study did not prefer the local attachment of the wh-element, in contrast to the two learner groups. And even these learner groups did not show a clear preference for local attachment as it might have been expected. Although no correlation of local attachment choices with the proficiency was found, the tendency to attach the wh-operator to the embedded clause was stronger in the advanced than in the intermediate learners.

This tendency towards local attachment in the learners was not strong, and the individual results even reveal that in both groups non-local attachment was preferred more often than local attachment. Nevertheless, the tendency towards non-local attachment was considerably weaker in the learners than in the native speakers. This effect could be explained by L1 transfer, considering that the direct German equivalent of the English construction that was tested in the present study was found to elicit a preference for local attachment in German native speakers. Another possible explanation for the learners' results would be in terms of shallow structure processing. According to such an account, the learners do not have access to the syntactic representations necessary for long movement and therefore choose the structurally closer construction, leading to a

[92]The long movement preference could not have been induced by the complementiser *that*, either. In a control study with sentences without the complementiser (see (73b) above), a different group of 20 English native speakers chose the long-movement reading in 90.1% of all cases (SD = 13.46).

local attachment of the wh-element. It is also possible, however, that the learners have access to the necessary structural interpretations, but simply do not have the processing capacity that is necessary to keep the wh-operator in mind over a long distance. In order to determine which of these three accounts is most probable, an on-line SPR task was conducted with similar materials and participants. This task might reveal whether or not learners make native-like use of deep syntactic structures in on-line processing.

7.4 Experiment 8: On-line attachment preferences

The present study tests the on-line processing of the same constructions as Experiment 7. Hence, the experimental sentences are similar, but in addition to the ambiguous sentences in Experiment 7, the present study also includes sentences disambiguated towards long or short movement.

7.4.1 Participants and procedure

Since Experiments 6 and 8 were combined in such a way that experimental items of one study were at the same time filler items for the other, the participants in the present study were the same as described in Chapter 6.4.1.

The sentences for this study were pragmatically disambiguated towards long (77a) or short movement (77b) or they were left ambiguous (77c). These three conditions are illustrated in the following examples:

(77) a.	How	did	the watchman	think	that	the robber	had entered	the building	?
b.	How	could	the watchman	think	that	the robber	had entered	the building	?
c.	When	did	the watchman	think	that	the robber	had entered	the building	?
Region	1	2	3	4	5	6	7	8	9

The norming study was the same as for Experiment 5 (see Chapter 6.3.). Since the results of the ambiguous sentences have already been reported (see Chapter 7.3.1.), only the judgements for the disambiguated sentences are considered here. These results are summarised in the following table:

	local attachment	**non-local attachment**
Naturalness	2.15 (0.87)	2.28 (0.68)
Plausibility	1.18 (1.26)	2.0 (0.5)

Table 7-4-1: Naturalness and plausibility judgements for Experiment 8

The table indicates that the judgements for sentences disambiguated towards non-local attachment were judged slightly worse than sentences disambiguated towards local attachment. This impression was statistically confirmed by a two-level repeated measures

ANOVA which did not yield any significant effects ($F_1 \leq 1$; $F_2(1,7) = 2.391$, $p \leq 0.174$). Thus, the two types of sentences did not differ significantly in terms of naturalness.

The plausibility judgements showed a similar tendency. The difference between local and non-local attachment was highly significant in the one-factor ANOVAs with subjects and with items as random variables ($F_1 \leq 1$; $F_2 \leq 1$). As in the previous studies, the plausibility judgements seem to be influenced to some extent by the grammaticality or naturalness of the sentences. Nevertheless, the differences between conditions are relatively small and insignificant, which indicates that the experimental sentences do not differ considerably in their real-world plausibility. Hence, this factor should not influence participants' attachment preferences in the on-line task.

Six of the ambiguous sentences from the norming study were included in the on-line experiment. Some of them had to be slightly modified so that they would be equally possible with non-local and local disambiguation (see Appendix D2 for a complete list of experimental items). Furthermore, all versions had to be compatible with regard to phonological words and character length per region (see Appendix D2). Hence the study included six sentences, each in three versions: the ambiguous version and the two versions disambiguated for embedded- and for matrix-clause attachment, respectively.

Since Experiments 6 and 8 were combined, the 18 experimental sentences for the present study were intermingled with 18 sentences of Experiment 6 (see Chapter 6.4.) and 24 independent fillers of various structures and lengths. Half of the filler items were ungrammatical wh-questions, which contained extractions from DP-islands as in (78a), from adjunct islands as in (78b), and from PPs as in (78c):

(78) a. Who does Sam deny the story that he kissed?
b. Who did you meet Tom after you saw?
c. Who did a book about interest you?

A detailed description of the other fillers and of the experimental procedure is provided in Chapter 6.4..

7.4.2 Results

Like the combined SPR and grammaticality judgement task reported in Experiment 6 (see Chapter 6.4.), the present study provides three dependent measures, namely participants' grammaticality judgements, the time they took to make these judgements, and their reading times on each word. The analyses of these measures are presented in turn in the following.

Since Experiments 6 and 8 were combined, participants' judgements for the filler items were the same as for Experiment 6 (see Table 6-4-3, here repeated as 7-4-2).

	Intermediate learners	Advanced learners	Native speakers
Questions	96.39 (5.66)	98.61 (3.16)	96.94 (4.63)
Declaratives	87.5 (11.32)	91.94 (6.74)	95.00 (6.42)
Total fillers	92.08 (7.45)	95.14 (3.29)	95.97 (4.02)

Table 7-4-2: Accuracy rates and (standard deviations) in the judgements of the filler items in Exp. 8

Detailed analyses of these results can be found in Chapter 6.4.3.. The most important finding for the present study is that the learners' overall judgements were comparatively successful. Hence, they were able to process complex sentences in this word-by-word reading task. Furthermore, it is important that the learners judged the ungrammatical questions comparatively successfully, which indicates that they mastered the complex constructions and were not biased to accept ungrammatical sentences. The grammaticality judgements for the experimental sentences are summarised in the following table:

	Intermediate learners	Advanced learners	Native speakers
Ambiguous	85.56 (21.32)	91.67 (12.18)	92.78 (9.47)
Short movement	88.89 (15.37)	90.56 (10.43)	90.56 (16.19)
Long movement	58.33 (33.55)	68.33 (24.11)	83.33 (20.99)
Total	77.59 (23.41)	83.52 (15.58)	88.89 (15.55)

Table 7-4-3: Accuracy rates and (standard deviations) in the judgements of the exp. items in Exp. 8

The total judgements for the three experimental structures were submitted to a one-factor ANOVA with Group as a between-subjects variable. This analysis yielded a significant main effect ($F(2,89) = 5.144$, $p \leq 0.008$). Tukey post-hoc tests only found a significant difference between intermediate learners and native speakers ($p \leq 0.001$). Neither the difference between the two learner groups ($p \leq 0.220$) nor between advanced learners and native speakers ($p \leq 0.287$) turned out significant. Table 7-4-3 indicates that all groups judged the non-locally disambiguated sentences ungrammatical more often than the other two structures. Judgements for the ambiguous and the locally disambiguated structures were similar in the native speakers and the advanced learners. The intermediate learners, however, judged the ambiguous sentences ungrammatical slightly more often than the locally disambiguated sentences. In order to statistically test these observations, the results were submitted to a mixed two-factor ANOVA with Condition as a within-subjects variable and Group as a between-subjects variable. This analysis yielded significant main effects for Condition ($F(2,174) = 40.862$, $p \leq 0.0001$) and Group ($F(2,87) = 5.144$, $p \leq 0.008$). The interaction was also significant

($F(4,348) = 4.014$, $p \leq 0.007$), confirming that the differences between the three conditions varied across groups.

The reasons for this interaction were further explored by separate one-factor ANOVAs with the three-level factor Condition on the results of each group, respectively with subjects and items as a random variable. For the native speakers, there was a significant main effect ($F_1(2,58) = 3.268$, $p \leq 0.045$; $F_2(2,10) = 9.186$, $p \leq 0.005$). Additional analyses for pair-wise contrasts found a significant difference between ambiguous and non-locally disambiguated sentences ($F_1(1,5) = 5.439$, $p \leq 0.027$; $F_2(1,29) = 16.236$, $p \leq 0.010$). The difference between the two disambiguated structures only reached significance in the analysis by items ($F_1(1,29) = 2.951$, $p \leq 0.095$; $F_2(1,5) = 9.494$, $p \leq 0.027$), and the difference between ambiguous and locally disambiguated structures was not statistically significant ($F_1 \leq 1$; $F_2(1,29) = 1.000$, $p \leq 0.363$). The advanced learners also showed a significant main effect for Condition ($F_1(2,58) = 23.300$, $p \leq 0.0001$; $F_2(2,10) = 53.291$, $p \leq 0.0001$). Pair-wise comparisons yielded significant differences between non-locally disambiguated and ambiguous conditions ($F_1(1,29) = 27.863$, $p \leq 0.0001$; $F_2(1,5) = 105.000$, $p \leq 0.0001$) and between the two disambiguated conditions ($F_1(1,29) = 24.681$, $p \leq 0.0001$; $F_2(1,5) = 76.923$, $p \leq 0.0001$). The difference between ambiguous and locally disambiguated conditions ($F_1 \leq 1$; $F_2 \leq 1$) did not reach statistical significance. The results thus confirm the impression that this group judged the non-locally disambiguated sentences as being ungrammatical more often than any of the other two structures. For the intermediate learners there was also a significant main effect ($F_1(2,58) = 19.416$, $p \leq 0.0001$; $F_2(2,10) = 33.974$, $p \leq 0.001$). Pair-wise comparisons found significant differences between non-locally disambiguated and ambiguous conditions ($F_1(1,29) = 22.111$, $p \leq 0.0001$; $F_2(1,5) = 22.385$, $p \leq 0.005$) and between the two disambiguated conditions ($F_1(1,29) = 23.936$, $p \leq 0.0001$; $F_2(1,5) = 63.774$, $p \leq 0.0001$). The difference between ambiguous and locally disambiguated conditions only reached statistical significance in the analysis by items ($F_1 \leq 1$; $F_2(1,5) = 7.105$, $p \leq 0.045$). Thus, in contrast to the advanced learners, the intermediate group showed a tendency to judge the ambiguous sentences as being ungrammatical slightly more often than the locally disambiguated structures.

The question arises whether the difficulty with the non-locally resolved sentences was also reflected in the judgement times (see Table 7-4-4). As in Experiment 6, sentences that were judged incorrect were not excluded from the initial analyses (see Chap-

ter 6.4.4). The screening for outliers led to the exclusion of 4.3% of the intermediate learners', 3.9 % of the advanced learners' and 3.5 % of the native speakers' data.

	Intermediate learners	Advanced learners	Native speakers
ambiguous	1224.73 (491.28)	1041.58 (425.52)	887.85 (331.21)
short movement	1436.64 (779.41)	1152.52 (455.91)	907.59 (343.78)
long movement	1872.45 (1097.51)	1824.85 (1061.02)	1249.27 (589.92)

Table 7-4-4: Response times in ms and (standard deviations) for the comprehension questions in Exp. 8

A mixed two-factor ANOVA with Condition as a within-subjects variable and Group as a between-subjects variable yielded a significant effect for Condition ($F(2,174) = 32.884$, $p \leq 0.0001$), and for Group ($F(2,87) = 6.885$, $p \leq 0.002$), but no significant interaction ($F(4,348) = 1{,}537$, $p \leq 0.193$). To analyse the findings more closely, separate repeated measures ANOVAs with the factor Condition were conducted for each group, respectively with subjects and with items as a random variable. For the intermediate learners, these analyses yielded a significant main effect ($F_1(2,58) = 8.614$, $p \leq 0.001$; $F_2(2,10) = 12.245$, $p \leq 0.002$). Planned pair-wise comparisons found significant differences the two disambiguated conditions ($F_1(1,29) = 6.586$, $p \leq 0.016$; $F_2(1,5) = 13.919$, $p \leq 0.014$) and between non-locally disambiguated and ambiguous conditions ($F_1(1,29) = 11.945$, $p \leq 0.002$; $F_2(1,5) = 15.188$, $p \leq 0.011$). The difference between ambiguous and locally disambiguated conditions only approached significance in the analysis by subjects ($F_1(1,29) = 3.741$, $p \leq 0.063$; $F_2(1,5) = 2.378$, $p \leq 0.184$). The results suggest that these learners had more difficulties to judge the long-movement conditions than any of the two other conditions. The advanced learners showed a similar pattern. There was a significant main effect ($F_1(2,58) = 16.934$, $p \leq 0.0001$; $F_2(2,10) = 12.454$, $p \leq 0.002$). Planned pair-wise comparisons found significant differences between the two disambiguated structures ($F_1(1,29) = 19.394$, $p \leq 0.0001$; $F_2(1,5) = 9.905$, $p \leq 0.025$) and between non-locally disambiguated and ambiguous sentences ($F_1(1,29) = 19.290$, $p \leq 0.0001$; $F_2(1,5) = 25.769$, $p \leq 0.004$), but not between ambiguous and locally disambiguated conditions ($F_1(1,29) = 1.451$, $p \leq 0.238$; $F_2(1,5) = 0.980$, $p \leq 0.368$). Like the two learner groups, the native speakers showed a significant main effect for Condition, which however only approached significance in the analysis by items ($F_1(2,58) = 9.688$, $p \leq 0.004$; $F_2(2,10) = 3.646$, $p \leq 0.065$). Planned comparisons yielded significant differences between the two disambiguated structures ($F_1(1,29) = 13.369$, $p \leq 0.001$; $F_2(1,5) = 3.185$, $p \leq 0.134$) and between non-locally disambiguated and ambiguous sentences ($F_1(1,29) = 9.763$, $p \leq 0.004$; $F_2(1,5) = 4.712$, $p \leq 0.082$), but not between ambiguous and locally disambiguated conditions ($F_1 \leq 1$; $F_2 \leq 1$).

The following paragraphs analyse the on-line reading times. The data were submitted to the same treatment as in Experiment 6 (see Chapter 6.4.4.). Sentences that were judged incorrect were not excluded from the initial reading times analyses. The standard procedure of excluding outliers lead to the deletion of 3.0% of the intermediate learners' data, 3.6% of the advanced learners' data, and 2.7% of the native speakers' data (see Appendix E5 for the resulting reading times for all groups in all regions).

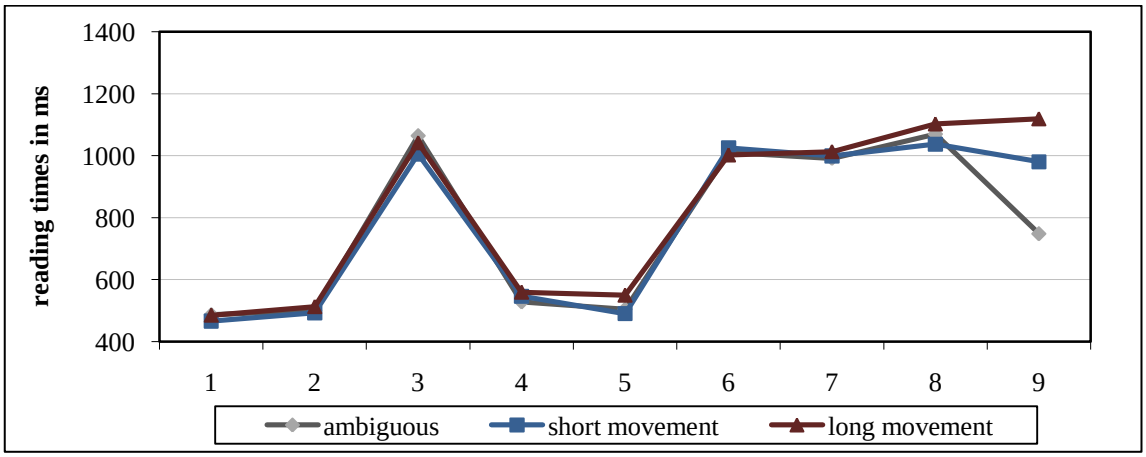

Figure 7-4-1: Native speakers' RTs per region in Exp. 8

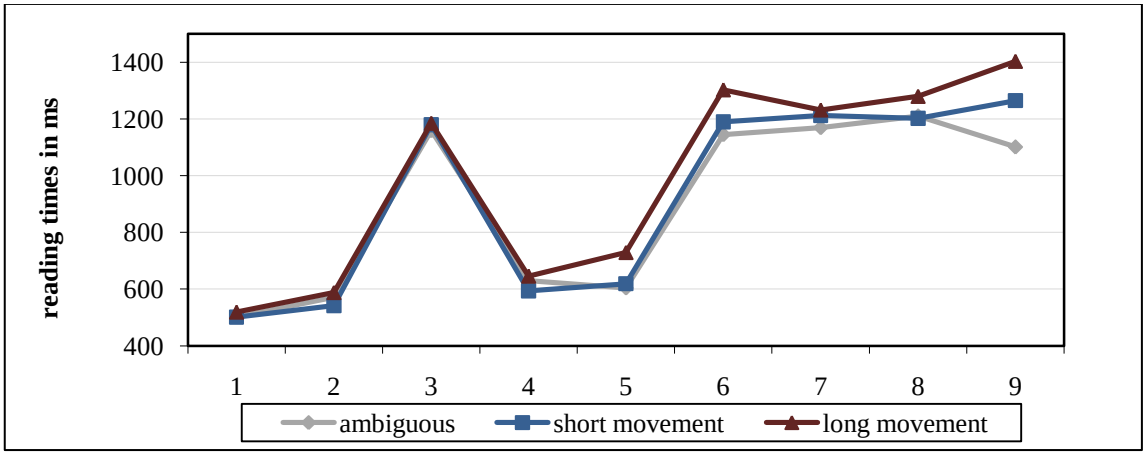

Figure 7-4-2: Advanced learners' RTs per region in Exp. 8

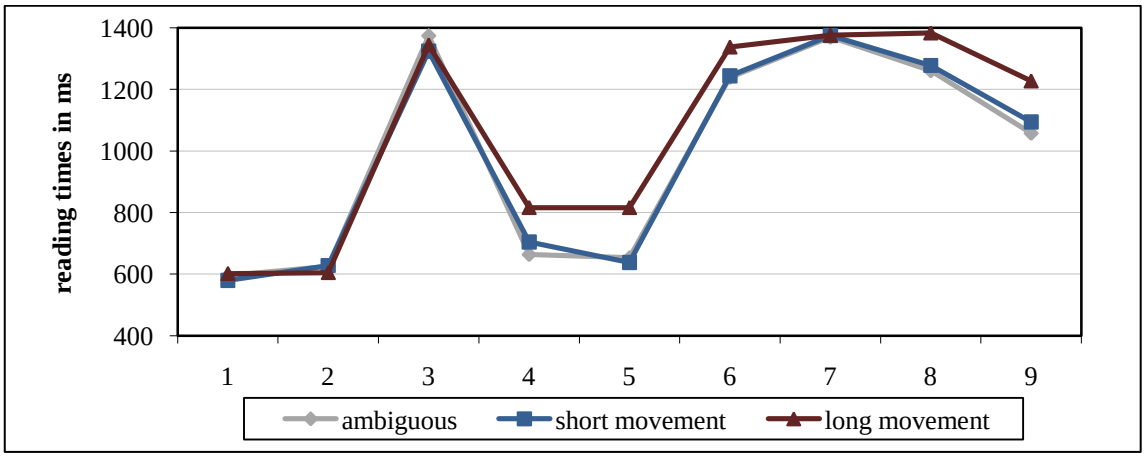

Figure 7-4-3: Intermediate learners' RTs per region in Experiment 8

As a first data analysis, a mixed three-factor ANOVA with Group as between-subjects variable and Condition as within-subjects variable was run on the results of regions 4, 5, 6, 8, and 9 with the following results:

Region	Condition	Group	Condition x Group
4	$F(2,174) = 7.364$, $p \leq 0.001$	$F(2,87) = 4.400$, $p \leq 0.015$	$F(4,174) = 4.334$, $p \leq 0.002$
5	$F(2,174) = 28.759$, $p \leq 0.0001$	$F(2,87) = 7.858$, $p \leq 0.001$	$F(4,174) = 2.857$, $p \leq 0.032$
6	$F(2,174) = 4.219$, $p \leq 0.016$	$F(2,87) = 7.004$, $p \leq 0.002$	$F(4,174) = 3.365$, $p \leq 0.011$
8	$F(2,174) = 29.152$, $p \leq 0.0001$	$F(2,87) = 5.531$, $p \leq 0.006$	$F(4,174) = 0.169$, $p \leq 0.945$
9	$F(2,174) = 27.546$, $p \leq 0.0001$	$F(2,87) = 2.754$, $p \leq 0.070$	$F(4,174) = 2.810$, $p \leq 0.033$

Table 7-4-5: Mixed Condition x Group ANOVAs for the RTs per region in Exp. 8

Since in all regions but region 8 there was a significant interaction between the two factors, separate one-factor analyses for each group were run in order to analyse the findings more closely. The results of each group are summarised in Figures 7-4-1 to 7-4-3:

The results of the ANOVA with Conditon as a within-subjects and as a within-items variable that was run on the native speakers' results are summarised in Table 7-4-6. Pair-wise comparisons were conducted in order to further compare the three conditions.

Region	Condition	long vs. short movement	long movement vs. ambiguous	short movement vs. ambiguous
4	$F_1(2,58) = 1.237$, $p \leq 0.298$ $F_2(2,10) = 1.116$, $p \leq 0.365$	$p_1 \leq 0.515$ $p_2 \leq 0.645$	$p_1 \leq 0.135$ $p_2 \leq 0.115$	$p_1 \leq 0.877$ $p_2 \leq 0.104$
5	$F_1(2,58) = 9.587$, $p \leq 0.0001$ $F_2(2,10) = 9.544$, $p \leq 0.005$	$p_1 \leq 0.0001$ $p_2 \leq 0.004$	$p_1 \leq 0.005$ $p_2 \leq 0.019$	$p_1 \leq 0.313$ $p_2 \leq 0.428$
6	$F_1 \leq 1$, $F_2 \leq 1$			
8	$F_1(2,58) = 12.595$, $p \leq 0.0001$ $F_2(2,10) = 8.911$, $p \leq 0.006$	$p_1 \leq 0.001$ $p_2 \leq 0.003$	$p_1 \leq 0.001$ $p_2 \leq 0.036$	$p_1 \leq 0.111$ $p_2 \leq 0.324$
9	$F_1(2,58) = 10.594$, $p \leq 0.0001$ $F_2(2,10) = 10.160$, $p \leq 0.004$	$p_1 \leq 0.045$ $p_2 \leq 0.040$	$p_1 \leq 0.001$ $p_2 \leq 0.008$	$p_1 \leq 0.003$ $p_2 \leq 0.083$

Table 7-4-6: ANOVAs for the native speakers' RTs in Exp. 8

Table 7-4-6 shows a significant main effect in regions 5, where the filler-gap effect is expected, and in regions 8 and 9, where the integration of the operator with its governing category is expected. Tests for pair-wise contrasts showed that in these regions only the differences between the two disambiguated structures and between long-movement and ambiguous structures were significant. The difference between short-movement and ambiguous conditions did not reach statistical significance in any of these regions. In order to ensure that the results were not biased by the sentences that were judged ungrammatical, a second analysis was run on the native speakers' reading time data with-

out the sentences that were judged ungrammatical. These analyses included 88.9% of the entire data set. The results are summarised in Table 7-4-7:

Region	Condition	long vs. short movement	long movement vs. ambiguous	short movement vs. ambiguous
4	$F_1(2,58) = 1.080$, $p \leq 0.348$ $F_2(2,10) = 0.098$, $p \leq 0.456$	$p_1 \leq 0.472$ $p_2 \leq 0.784$	$p_1 \leq 0.243$ $p_2 \leq 0.203$	$p_1 \leq 0.181$ $p_2 \leq 0.156$
5	$F_1(2,58) = 11.869$, $p \leq 0.0001$ $F_2(2,10) = 8.972$, $p \leq 0.006$	$p_1 \leq 0.0001$ $p_2 \leq 0.005$	$p_1 \leq 0.004$ $p_2 \leq 0.023$	$p_1 \leq 0.173$ $p_2 \leq 0.257$
6	$F_1 \leq 1$, $F_2 \leq 1$			
8	$F_1(2,58) = 12.814$, $p \leq 0.0001$ $F_2(2,10) = 9.129$, $p \leq 0.005$	$p_1 \leq 0.001$ $p_2 \leq 0.004$	$p_1 \leq 0.0001$ $p_2 \leq 0.028$	$p_1 \leq 0.329$ $p_2 \leq 0.272$
9	$F_1(2,58) = 11.113$, $p \leq 0.0001$ $F_2(2,10) = 9.939$, $p \leq 0.009$	$p_1 \leq 0.042$ $p_2 \leq 0.036$	$p_1 \leq 0.007$ $p_2 \leq 0.039$	$p_1 \leq 0.0001$ $p_2 \leq 0.002$

Table 7-4-7: ANOVAs for the native speakers' RTs of the sentences judged grammatical in Exp. 8

Table 7-4-7 shows similar effects as the complete data set, in that long movement elicited significantly higher reading times than the other conditions in regions 5, 8 and 9.

The advanced learners' pattern resembled that of the native speakers, although their reading times were generally higher. The results of the ANOVA with Conditon as a within-subjects and as a within-items variable and the planned comparisons for each pair of conditions on the advanced learners' results are summarised in Table 7-4-8.

Region	Condition	long vs. short movement	long movement vs. ambiguous	short movement vs. ambiguous
4	$F_1(2,58) = 1.819$, $p \leq 0.171$ $F_2(2,10) = 1.816$, $p \leq 0.212$	$p_1 \leq 0.092$ $p_2 \leq 0.070$	$p_1 \leq 0.690$ $p_2 \leq 0.543$	$p_1 \leq 0.113$ $p_2 \leq 0.347$
5	$F_1(2,58) = 10.274$, $p \leq 0.0001$ $F_2(2,10) = 6.633$, $p \leq 0.015$	$p_1 \leq 0.0001$ $p_2 \leq 0.073$	$p_1 \leq 0.001$ $p_2 \leq 0.011$	$p_1 \leq 0.715$ $p_2 \leq 0.629$
6	$F_1(2,58) = 7.889$, $p \leq 0.002$ $F_2(2,10) = 7.806$, $p \leq 0.009$	$p_1 \leq 0.010$ $p_2 \leq 0.006$	$p_1 \leq 0.002$ $p_2 \leq 0.043$	$p_1 \leq 0.506$ $p_2 \leq 0.919$
8	$F_1(2,58) = 10.541$, $p \leq 0.0001$ $F_2(2,10) = 5.336$, $p \leq 0.026$	$p_1 \leq 0.0001$ $p_2 \leq 0.015$	$p_1 \leq 0.010$ $p_2 \leq 0.034$	$p_1 \leq 0.122$ $p_2 \leq 0.806$
9	$F_1(2,58) = 16.262$, $p \leq 0.0001$ $F_2(2,10) = 7.541$, $p \leq 0.014$	$p_1 \leq 0.0001$ $p_2 \leq 0.019$	$p_1 \leq 0.0001$ $p_2 \leq 0.038$	$p_1 \leq 0.858$ $p_2 \leq 0.314$

Table 7-4-8: ANOVAs for the advanced learners' RTs in Exp. 8

As Table 7-4-8 indicates, the advanced learners' pattern resembled that of the native speakers. Significant main effects were found in region 5 and in regions 8 and 9. But the advanced learners also showed a significant main effect in region 6, which might have spilled over from region 5. Tests for pair-wise contrasts showed that in regions 5, 6, 8, and 9, only the differences between the two disambiguated structures and between long-movement and ambiguous sentences were significant. The difference between short-movement and ambiguous conditions did not reach statistical significance. As for the native speakers, a second analysis was run on the advanced learners' results without the

sentences that were judged ungrammatical. These analyses included 83.5% of the entire data set. Table 7-4-9 shows that the results resemble those for the complete data set:

Region	Condition	long vs. short movement	long movement vs. ambiguous	short movement vs. ambiguous
4	$F_1(2,58) = 1.225$, $p \leq 0.301$ $F_2(2,10) = 1.923$, $p \leq 0.278$	$p_1 \leq 0.385$ $p_2 \leq 0.294$	$p_1 \leq 0.468$ $p_2 \leq 0.622$	$p_1 \leq 0.179$ $p_2 \leq 0.511$
5	$F_1(2,58) = 5.561$, $p \leq 0.006$ $F_2(2,10) = 7.624$, $p \leq 0.011$	$p_1 \leq 0.016$ $p_2 \leq 0.043$	$p_1 \leq 0.001$ $p_2 \leq 0.013$	$p_1 \leq 0.687$ $p_2 \leq 0.494$
6	$F_1(2,58) = 6.218$, $p \leq 0.004$ $F_2(2,10) = 10.408$, $p \leq 0.008$	$p_1 \leq 0.026$ $p_2 \leq 0.018$	$p_1 \leq 0.003$ $p_2 \leq 0.039$	$p_1 \leq 0.483$ $p_2 \leq 0.856$
8	$F_1(2,58) = 8.595$, $p \leq 0.001$ $F_2(2,10) = 6.884$, $p \leq 0.027$	$p_1 \leq 0.001$ $p_2 \leq 0.018$	$p_1 \leq 0.020$ $p_2 \leq 0.035$	$p_1 \leq 0.211$ $p_2 \leq 0.686$
9	$F_1(2,58) = 8.660$, $p \leq 0.0021$ $F_2(2,10) = 6.135$, $p \leq 0.026$	$p_1 \leq 0.0017$ $p_2 \leq 0.023$	$p_1 \leq 0.001$ $p_2 \leq 0.041$	$p_1 \leq 0.799$ $p_2 \leq 0.396$

Table 7-4-9: ANOVAs for the advanced learners' RTs of the sentences judged grammatical in Exp. 8

The intermediate learners showed a similar pattern as the other groups, except that their reading times were highest. However, there were some small but notable differences to the other groups, especially in region 4. Table 7-4-10 summarises the results of the ANOVA with Conditon and the planned comparisons for each pair of conditions on the intermediate learners' results.

Region	Condition	long vs. short movement	long movement vs. ambiguous	short movement vs. ambiguous
4	$F_1(2,58) = 10.222$, $p \leq 0.0001$ $F_2(2,10) = 35.519$, $p \leq 0.0001$	$p_1 \leq 0.001$ $p_2 \leq 0.002$	$p_1 \leq 0.001$ $p_2 \leq 0.0001$	$p_1 \leq 0.278$ $p_2 \leq 0.225$
5	$F_1(2,58) = 12.205$, $p \leq 0.0001$ $F_2(2,10) = 15.109$, $p \leq 0.001$	$p_1 \leq 0.001$ $p_2 \leq 0.015$	$p_1 \leq 0.002$ $p_2 \leq 0.007$	$p_1 \leq 0.091$ $p_2 \leq 0.107$
6	$F_1(2,58) = 1.365$, $p \leq 0.263$ $F_2(2,10) = 1.876$, $p \leq 0.203$	$p_1 \leq 0.130$ $p_2 \leq 0.755$	$p_1 \leq 0.104$ $p_2 \leq 0.184$	$p_1 \leq 0.888$ $p_2 \leq 0.163$
8	$F_1(2,58) = 7.494$, $p \leq 0.001$ $F_2(2,10) = 11.154$, $p \leq 0.003$	$p_1 \leq 0.002$ $p_2 \leq 0.016$	$p_1 \leq 0.006$ $p_2 \leq 0.002$	$p_1 \leq 0.330$ $p_2 \leq 0.812$
9	$F_1(2,58) = 3.552$, $p \leq 0.035$ $F_2(2,10) = 7.107$, $p \leq 0.012$	$p_1 \leq 0.044$ $p_2 \leq 0.041$	$p_1 \leq 0.036$ $p_2 \leq 0.013$	$p_1 \leq 0.764$ $p_2 \leq 0.511$

Table 7-4-10: ANOVAs for the intermediate learners' RTs in Exp. 8

The pattern for intermediate learners turned out to be similar to the two other groups in that there were also significant main effects for Condition in regions 5, as well as in regions 8 and 9. In contrast to the advanced learners, the intermediate learners did not show a significant main effect in region 6, although the descriptive analysis indicates that long movement elicited increased reading times in comparison to the other two conditions. The intermediate learners do not show such a spill-over effect from region 5. This finding might be a result of their generally longer reading times, which indicate that they only moved on when they had fully integrated the new materials into the sentence. An interesting difference to the results of the other two groups is that the inter-

mediate learners showed a significant main effect in region 4, i.e. on the verb of the matrix clause. Tests for pair-wise contrasts showed that in regions 4, 5, 6, 8, and 9, only the differences between the two disambiguated sentences and between long-movement and ambiguous sentences were significant, whereas the difference between short-movement and ambiguous conditions did not reach statistical significance. As for the other two groups, a second analysis of the reading time data without the sentences that were judged ungrammatical was conducted, including 77.6% of the entire data set. The results are summarised in Table 7-4-11.

Region	Condition	long vs. short movement	long movement vs. ambiguous	short movement vs. ambiguous
4	$F_1(2,58) = 6.623$, $p \leq 0.003$ $F_2(2,10) = 15.222$, $p \leq 0.0001$	$p_1 \leq 0.0018$ $p_2 \leq 0.012$	$p_1 \leq 0.003$ $p_2 \leq 0.001$	$p_1 \leq 0.255$ $p_2 \leq 0.341$
5	$F_1(2,58) = 7.809$, $p \leq 0.001$ $F_2(2,10) = 12.775$, $p \leq 0.001$	$p_1 \leq 0.005$ $p_2 \leq 0.017$	$p_1 \leq 0.003$ $p_2 \leq 0.004$	$p_1 \leq 0.374$ $p_2 \leq 0.144$
6	$F_1 \leq 1$, $F_2 \leq 1$			
8	$F_1(2,58) = 6.197$, $p \leq 0.004$ $F_2(2,10) = 10.676$, $p \leq 0.032$	$p_1 \leq 0.023$ $p_2 \leq 0.019$	$p_1 \leq 0.042$ $p_2 \leq 0.034$	$p_1 \leq 0.171$ $p_2 \leq 0.356$
9	$F_1(2,58) = 2.991$, $p \leq 0.043$ $F_2(2,10) = 7.762$, $p \leq 0.052$	$p_1 \leq 0.047$ $p_2 \leq 0.054$	$p_1 \leq 0.039$ $p_2 \leq 0.028$	$p_1 \leq 0.832$ $p_2 \leq 0.741$

Table 7-4-11: ANOVAs for the intermed. learners' RTs for the sentences judged grammatical in Exp. 8

To sum up, all analyses of the results excluding the sentences judged ungrammatical yielded similar results as the entire data set. This finding indicates that the observed effects are not biased by the sentences judged ungrammatical. The following chapter discusses the implications of the observed tendencies.

7.4.3 Discussion

This chapter discusses the results of the grammaticality judgements and the reading times in turn, especially with regard to the predictions made in Chapter 7.1..

The number of correct judgements for the ungrammatical filler sentences was unexpectedly high for all participants, also in comparison to the results of Juffs & Harrington (1995) and Juffs (2005), who used ungrammatical fillers of similar constructions. Hence, all participants were able to understand the sentences in this relatively demanding task and did not overaccept ungrammatical sentences. Since the ungrammatical questions all included subjacency-violations, i.e. extractions of a wh-element from an island,[93] the learners' successful judgements of these questions indicate that they mas-

[93] Subjacency is a structural principle introduced by Chomsky (1973) to explain restrictions in long-distance movement. According to this principle, a moved element must not cross more than one bounding node in a single movement, whereby bounding nodes are assumed to be IPs and DPs, at least in English. For an overview of subjacency studies with L2 learners, see for example Hawkins (2001, Chapter 7).

tered this abstract grammatical principle. This finding is particularly important because it indicates that learners are as sensitive to syntactic barriers to long movement as native speakers.

The grammaticality judgements of the experimental sentences confirm the first impression from the control sentences, i.e. that all groups were able to judge the sentences almost equally successfully. However, all three groups but especially the intermediate learners judged the long-movement sentences as being ungrammatical more often than the two other structures, and standard deviations were highest for this condition. One explanation for this result might be that some participants preferred to interpret this condition like the short movement condition. Since this interpretation is pragmatically infelicitous, it might have led participants to judge the sentence as being ungrammatical. The intermediate group moreover showed a tendency to judge the ambiguous sentences as being ungrammatical slightly more often than the short-movement sentences, while the other two groups did not show any difference between these conditions. Hence, these learners seemed to be aware of the ambiguity and did not simply consider the matrix-clause attachment interpretation of the ambiguous structures. Such a conclusion is further supported by the significant correlation between the intermediate learners' judgements of ambiguous and long-movement sentences ($r(30) = 0.402$, $p \leq 0.028$). This finding indicates that when participants understood the long-movement interpretation, they also accepted the ambiguous conditions more easily. There was no such correlation for the advanced learners ($r(30) = 0.245$, $p \leq 0.193$), nor for the native speakers $r(30) = 0.096$, $p \leq 0.613$). Hence, both groups rejected the long-movement reading comparatively often but did not have any difficulties with the ambiguous condition.

The finding that none of the three groups accepted the ambiguous sentences significantly more often than the sentences disambiguated towards short movement is replicated in the judgement times. This result is not in line with previous studies on various structures (e.g. Swets et al. 2008; Traxler, Pickering & Clifton 1998; Van Gompel et al. 2005) which found a preference for ambiguous over disambiguated sentences in monolingual speakers of English. Furthermore, the result is not in line with the SSH, which would predict that especially the learners prefer ambiguous over disambiguated structures because the former do not force them to commit to one structural analysis.

Although participants' grammaticality judgements were timed, the response latencies cannot be considered as on-line data because participants had already read the whole sentence before making their judgements. The SPR data might be more reliable

in this respect. A look at the word-by-word reading times indicates that, similar as in the other SPR tasks reported so far, the intermediate learners' general reading times were highest, followed by the advanced learners and the native speakers. Apart from this difference, the three groups' qualitative patterns were similar in most regions, which indicates that the learners employed highly native-like processing strategies. In order to interpret the findings more closely, the following paragraphs consider the three groups' results for the regions of interest in turn.

In region 4, i.e. the main verb of the matrix sentence, no particular effects were expected, except a slight advantage of the long-movement condition because a short-movement interpretation might cause a delay when the moved element is integrated at its base position. However, none of the groups showed such a tendency. The intermediate learners even evinced the opposite effect, i.e. their reading times were significantly higher for long movement than for the other two conditions. A possible explanation is that the learners were aware at this point of the sentence that their preferred interpretation, i.e. short movement, is pragmatically infelicitous. This might cause a delay here, especially considering that the two disambiguated constructions only differ with respect to the auxiliary (i.e. could for local attachment and did for non-local attachment). Therefore, the learners might have tried to interpret both constructions in terms of short movement initially, which might have led to a surprise effect and ensuing reanalysis when the main verb was encountered. Under this assumption, the absence of such an effect in the native speakers and advanced learners indicates that these groups were more successful than the intermediate learners when reanalysing the sentence.

Alternatively, it is possible that the intermediate learners' longer reading times for the long-movement condition simply spilled over to region 5, so that the effect found in that region is not necessarily a filler-gap effect. But such an explanation is unlikely because the advanced learners and the native speakers showed a similar effect in region 5, but no effect in region 4. Thus, both learner groups show an intermediate trace effect in region 5 in the long-movement condition. Since such an effect would not be expected if the learners were unable to process deep syntactic structures, the finding suggests that they perform the full syntactic movement operation in a similar way as the native speakers.

The findings are supported by the results from region 8, the original position of the moved element, where the integration of the operator with its governing category is expected. In this region, all groups took longer in the long-movement condition than in the

short-movement and the ambiguous conditions. This result supports the proposed analysis of region 5, i.e. that both learner groups show native-like syntactic processing. It is possible that the filler integration effect in region 8 is a semantic effect that does not necessarily require deep structural processing. Such an explanation would be in line with the SSH. However, this explanation is unlikely because the moved element is not an argument of the verb of the embedded clause. Furthermore, the intermediate gap effect found in region 5 cannot be explained by direct lexical association between the antecedent and the trace. Hence, the present results can be interpreted as evidence against the SSH.

The results for region 9 are difficult to interpret because it only contains the question mark. It is possible that this region does not only reflect sentence wrap-up effects. Instead, participants might already think about the grammaticality of the sentence at this point. Therefore, the longer reading times for long movement in this region might not only reflect filler integration or a sentence wrap-up effect, but rather participants' insecurity about the grammaticality of the sentences disambiguated towards long movement. Similar considerations apply for the faster reading times of the ambiguous in comparison to the short-movement sentences in this region. The effect does not necessarily imply that ambiguous sentences caused less processing difficulty. Instead, the grammaticality of the ambiguous sentences might have been easier to determine because the first word differed, whereas the disambiguated conditions only differed in the auxiliary did or could.

In more general terms, the obtained effects show that all participants were sensitive to the experimental manipulations because they were able to distinguish between long and short movement, both in their grammaticality judgements and in their reading times. This result is especially important because of the disambiguation via pragmatic information. There was no clearly defined disambiguating word as in classical garden-path sentences, where a possible effect can be clearly identified. But while all three groups distinguished between long and short movement on the one hand and long movement and ambiguous structures on the other hand, reading times did not significantly differ between short movement and ambiguous structures. This finding is in line with the results of the grammaticality judgements and indicates that participants preferred the short movement interpretation of the ambiguous sentences. Like the results of the grammaticality judgements, these findings contradict the SSH, which would predict a preference of ambiguous over disambiguated structures for the learners.

Interestingly, none of the groups showed any filler-integration effects in the ambiguous condition, although most of the native speakers and at least some of the learners preferred long movement in the off-line task. Although the native speakers displayed a slight effect in this direction in the on-line task, such that in region 8 the ambiguous condition yielded higher reading times than the short-movement condition, this effect did not reach statistical significance and no intermediate gap effect was found in region 5. The native speakers' apparent short-movement preference for the ambiguous constructions in this task is in line with previous studies which found a preference for short over long distance movements in monolingual sentence processing (see Chapter 3.2.5.). The finding that the learners also preferred the short movement interpretation of the ambiguous sentences thus indicates that they prefer a similar recency interpretation in these structures as the native speakers. Although it is possible that this is a transfer effect, considering the German off-line preference for short movement (see Chapter 7.1.), such an explanation is improbable because of the difference between the learners' lack of any preference in the off-line task (Chapter 7.3.) and the preference for short movement in the on-line task. The discrepancy between off-line and on-line processing rather suggests that participants chose the structurally simplest interpretation under the higher processing demands of an on-line task.

7.5 Summary

Taken together, the results from the off-line and on-line studies on attachment preferences in adjunct questions provide interesting insights into the mechanisms that guide L1 and L2 processing and attachment preferences. The native speakers' preferences were found to differ for off-line and on-line processing, which shows that even native speakers resort to a recency preference in sentences which are difficult to process. This pattern might be replicated by the learner groups. It is also possible that it is not the complexity of the sentences which causes the effect but rather the word-by-word presentation in the SPR task. Therefore, it would be interesting to replicate the study with an eye-tracking paradigm in which participants see a sentence as a whole.

The learners' lack of any preference in the off-line task could be attributed to a recency preference, shallow structure processing or L1 influence. But the results from the on-line task contradict the SSH in that they suggest that the learners process deep syntactic structures in a similar way as native speakers. Therefore, the learners are influenced by their L1 or by an L2-specific recency strategy. In order to resolve this question, a parallel study with learners whose L1 shows a similar off-line preference as Eng-

lish would be necessary. If these learners show the same patterns as the learners of the present study, this would suggest that transfer is not the decisive factor here. In the light of the data from the on-line study, it is more likely that the learners' bias towards short movement in the off-line task is a processing load effect because the long movement preference disappears in the on-line task, even in the native speakers.

With regard to proficiency influences on learners' processing, it can be concluded that both learner groups showed highly similar tendencies in the off-line and on-line task. In the off-line task, the advanced learners showed a slight tendency towards a native-like long-movement preference. This tendency indicates that increased proficiency leads to more native-like processing strategies, although it remains open whether it is L1 transfer or a learner-specific recency strategy that subsides in the advanced learners. In the on-line task, both learner groups showed highly similar preferences, the only difference being that the intermediate learners had slightly more difficulties recovering from initial misanalysis than the advanced learners, although they were able to recover and to process the entire sentences in the same way as the other groups.

8 Conclusions

The aim of the present chapter is to give an overall interpretation of the findings of the studies presented here with regard to the main research questions. The following general discussion considers some implications for general processing theories (chapter 8.1), and for L2 learners' processing (chapter 8.2). After some directions for future research (chapter 8.3), chapter 8.4 closes with some concluding remarks.

8.1 Implications for first language processing

Although the studies presented here were not designed for testing and evaluating the predictions of the monolingual processing models (see chapter 2.1), the results permit some preliminary conclusions about the nature of the sentence processing mechanism, in particular about the following issues: (1.) the language universality of the parser, (2.) the role of syntax in relation to other modules of the grammar, and (3.) the time course of processing decisions.

With regard to the language universality of the human processing mechanism, the results from the RC attachment studies are of particular interest. The present study focuses on English sentences and does not include an on-line experiment with German monolinguals, so that the on-line processing in the two languages cannot be directly compared. Nevertheless, the present studies provide some evidence that English and German RC attachment preferences differ and the German learners transferred the German high attachment preference when processing English ambiguities. Interestingly, such cross-linguistic differences were not only found in RC attachment but also in the wh-questions that were investigated in Experiments 7 and 8. The latter difference might be related to the existence of an alternative construction in German, such that one construction was biased towards non-local and one towards local attachment. A similar explanation might be applicable for the RC attachment ambiguity. Such a prediction would be in line with an account that relates RC attachment preferences to the presence or absence of an alternative Saxon genitive construction in a specific language. But since this account could not explain the data from all previously investigated languages (see chapter 2.1.1), the present evidence is only tentative and further research on a variety of languages is necessary to fully resolve this issue.

Another implication of the present findings is that the observed L1 transfer in the learners is incompatible with a strictly language universal processing mechanism. Hence, the assumption that there are universal principles and principles of cross-linguistically varying strength is more in line with the data. The present results do not

allow any firm conclusions about the reasons for such cross-linguistic variation. The finding that particularly the less proficient learners transfer preferences from their first to their L2 suggests that L2 exposure plays a certain role. However, this finding does not necessarily support a frequency based processing model for the following reasons: First, it is also conceivable that learners transfer structural principles rather than frequency information from their first to their L2. Second, L2 proficiency rather than the frequency of a certain structure might be responsible for the subsiding of transfer in the advanced learners compared to the intermediate learners.

As regards the role of syntax relative to other modules of the grammar, the present studies provide evidence for the influence of non-syntactic factors on participants' processing decisions. Both in English and German the properties of the preposition linking the two potential attachment sites was found to influence attachment preferences, which suggests that lexical factors such as thematic properties play an important role in processing decisions. Whereas a purely syntactic processing model faces some difficulties explaining such phenomena, most of the processing models presented in chapter 2 have been adapted in such a way as to account for the present findings. The influence of the theta-properties of verbs was investigated in Experiment 4 by comparing PP attachment preferences for ditransitive verbs where the PP was required by the verb's argument structure and for transitive verbs where the PP was an adjunct. Since a VP attachment preference was found for both types of verbs, verb subcategorisation cannot be the only factor that determines attachment preferences. This finding therefore argues against purely lexical processing models.

The comparison of off-line and on-line processing of the same materials permits a distinction between initial and final attachment preferences and thus an insight into the time course of processing decisions. The only sentences for which the English native speaker control groups showed distinct attachment preferences in off-line and on-line processing were the wh-questions investigated in Experiments 7 and 8. Only in the on-line study did these participants adopt a local attachment preference, possibly to facilitate the processing of these comparatively complex structures. No such difference was found between off-line and on-line processing for any of the other experimental structures. Hence, there is no evidence for a systematic difference between initial and final processing decisions as postulated by the Late Closure Theory (see chapter 2.1.1). But it must be noted that self-paced reading as an on-line measure is possibly not sensitive

enough to capture immediate processing decisions. An additional measure such as eye-tracking might provide further insights in this respect.

Such a more fine-grained measure would also be helpful to resolve the question whether processing is serial, i.e. whether the processor opts for one potential analysis, running the risk of getting wrong, or whether it considers several analyses in parallel, despite the risk of a processing overload. In the present on-line studies all groups showed surprise effects when a sentence was disambiguated towards their dispreferred interpretation. This finding indicates that the participants had not considered all possible interpretations in the same way. However, the distinction between recent serial and parallel processing models is more complex, considering that research focuses on the question "whether or not there are some circumstances in which multiple structural interpretations are retained." (Gibson and Pearlmutter, 2000: 231). The experimental evaluation of this possibility requires sophisticated experimental manipulations (see Gibson and Pearlmutter, 2000, for an overview). Although a comparison of serial versus parallel processing would also be interesting for L2 processing research, especially considering that L2 learners might have more difficulties than native speakers with computing several possible interpretations in parallel, it is beyond the scope of the present study.

A related question is whether participants sometimes resort to a reduced processing route, even in their native language (e.g. Christianson et al., 2001; Ferreira, 2003; Ferreira, Ferraro and Bailey, 2002; Ferreira and Patson, 2007; Townsend and Bever, 2001). Although this possibility was not explicitly tested in the present studies, Experiments 4 and 6 included a fully ambiguous condition, which should be preferred over the two disambiguated structures if participants rely on 'good enough' representations. Although such a tendency was found in the grammaticality judgements of Experiment 4, it was neither replicated in the judgement times nor in the reading times, where the native speakers preferred the low disambiguated sentences. In Experiment 6, participants showed an equally strong preference for ambiguous and non-locally disambiguated structures. This finding suggests that they tended to interpret the ambiguous construction like their preferred interpretation instead of leaving the interpretation unresolved. Such a finding is not in line with the assumption of reduced processing in monolinguals.

Taken together, the results of the present studies could make some contributions to unresolved issues in L1 processing, but further systematic research is necessary in order to gain more insights into this broad and interesting field. Such insights would also be helpful for a theory of L2 processing, considering that at least some L1 processing

mechanisms were found to be transferred to the learners' L2. The implications of the reported studies for theories of L2 processing are discussed in the following chapter.

8.2 Implications for second language processing

The present experiments investigated L2 learners' processing of various syntactic structures. Therefore, each of the studies addressed the same broad research questions from a slightly different perspective. The main research question focused on possible differences between first and L2 processing. As the summary of the main results chapter 8.1 shows, some differences between L1 and L2 processing were found. In this chapter I discuss possible factors that influence these differences, namely L1 transfer, L2-specific processing, cognitive resource limitations and proficiency effects, in the light of the present experimental findings.

The issue of L1 transfer was explicitly addressed in the studies on RC attachment preferences (see chapter 3) because previous studies found cross-linguistic variation for these ambiguities. Although additional off-line studies with monolingual speakers of German (see chapter 3.1) indicated that the German high attachment preference is less stable than formerly assumed, the German learners showed a tendency for high attachment, at least in the off-line study and the self-paced reading task with chunk-wise presentation. This tendency is indicative of L1 transfer. But since it is controversially debated in the literature which factors are responsible for the cross-linguistic influences, it is yet unclear which specific features of the L1 are transferred to the learners' L2 in processing.

Interestingly, the advanced learners showed a weaker high attachment preference in the off-line study than the intermediate learners. Moreover, the analyses of the individual participants' data showed that some of the advanced learners preferred high attachment and others preferred low attachment and only a few of these learners had no preference. These findings suggest that the observed transfer effects subside with increasing proficiency. Further research with less proficient learners on the one hand and more advanced learners on the other hand is required to investigate how exactly L2 processing develops, whether learners make use of transfer from early on and whether it is possible for highly proficient learners to acquire native-like processing.

These issues are related to the question whether transfer is a mechanism inherent in L2 processing or whether learners only resort to transfer when facing specific processing difficulties as suggested for example by Frenck-Mestre (2005). Such processing difficulties could either be related to defective access to grammatical representations or

to processing capacity limitations. These two possibilities are discussed in turn in the following paragraphs.

Since the Shallow Structure Hypothesis assumes that learners do not construct deep syntactic representations, it predicts an overreliance on lexical and pragmatic information and an under-specification of syntactic representations. For the RC attachment ambiguity the Shallow Structure Hypothesis predicts no significant preferences in the learners. This prediction is not confirmed by Experiments 1, 2 and 3, where the intermediate learners showed clear preferences in all three tasks. And although the advanced learners did not show a clear attachment preference in the off-line task, analyses of the individual participants' data showed that some of these advanced learners preferred high attachment and some preferred low attachment, whereas only a few of them had no preference. These findings indicate that the learners make use of L1 transfer, which gradually gives way to a native-like low attachment preference with increasing proficiency.

The studies on PP attachment (chapter 4) found native-like attachment preferences in the learners. This result is not in line with the Shallow Structure Hypothesis, which predicts that learners do not make any commitment in these structures. The learners' readiness to choose a syntactic interpretation showed particularly clearly in Experiment 4. Here the learners, like the native speakers, did not prefer ambiguous over disambiguated structures, which indicates that they had a structural preference. Nevertheless, it has to be noted that the learners' attachment preferences in these studies might be a default mechanism to avoid structural processing. Furthermore, the patterns observed in these studies could be explained by a preference to attach to the most recently encountered theta-assigner or discourse entity. Since the Shallow Structure Hypothesis predicts that learners are sensitive to such factors, the results do not contradict the Shallow Structure Hypothesis.

The only direct evidence against the Shallow Structure Hypothesis comes from the on-line attachment preferences in wh-questions that were examined in Experiment 6. In this on-line study all three groups showed highly similar effects, which indicates that only in the non-local attachment condition all three groups assumed an intermediate gap in the head position of the embedded CP and show a filler-integration effect at the base position of the movement-chain. This effect cannot be explained by direct lexical association of the trace and the operator because the moved element was not an argument but an adjunct of the verb. Hence, the results suggest that the learners processed deep

syntactic structures in the same way as native speakers, contrary to the predictions of the Shallow Structure Hypothesis.

The present findings also contradict the Declarative-Procedural Model (see Ullmann, 2005, and the discussion in chapter 2.3.4), which predicts that L2 syntactic processing is determined by declarative knowledge whereas L1 processing also accesses procedural knowledge. Hence, learners should not have access to rule based automatic processing in the same way as native speakers. Especially the results from Experiment 8 provide evidence against this model because the filler-activation effects found in the learners in this experiment can only result from the application of rule based grammatical knowledge. However, unlike the Shallow Structure Hypothesis, the Declarative-Procedural Model assumes that the development of native-like processing is possible with increasing exposure and proficiency (Ullman, 2005). Hence, it is possible that the learners of the present study were sufficiently proficient to have some access to the procedural memory system.

Although the present data do not support the proposal that learners are unable to process deep syntactic information, as suggested by the Shallow Structure Hypothesis and the Declarative-Procedural Model, it has to be noted that the learners' reliance on structural information might depend on the demands of the respective on-line task. As outlined in chapter 2.4.5, there seems to be a discrepancy between tasks that are combined with a metalinguistic task such as grammaticality judgements on the one hand and reading tasks that are combined with comprehension questions on the other hand. Whereas studies of the former type found native-like processing of wh-dependencies (e.g. Juffs, 2005; Williams, Möbius and Kim, 2001), studies of the latter type found evidence against deep structural processing for these structures (e.g. Felser and Roberts, 2007; Marinis et al., 2005). The present study fits into this pattern in that it found native-like processing of filler-gap dependencies in a self-paced reading task combined with grammaticality judgements. These findings suggest that metalinguistic tasks such as grammaticality jugdgements raise the learners' syntactic awareness. In comprehension tasks, in contrast, the learners focus on the contents of the sentences and might even show shallow syntactic processing. Further research with different types of tasks is necessary in order to further explore this possibility.

With regard to a L2-specific recency preference, the present studies did not find any supporting evidence. The only exception to this is Experiment 5, where the learners had no clear preference whereas the native speakers preferred non-local attachment. How-

ever, this effect could also be related to L1 influence. Furthermore, in the other studies the learners showed either native-like local attachment preferences, as in Experiment 4, or a native-like non-local attachment preference, as in Experiment 3. The latter finding shows that the learners were able to attach high if the sentence required it. In Experiment the chunk-wise reading task of Experiment 2, the learners even showed a high attachment preference where the native speakers preferred low attachment. The only result that is compatible with a recency preference was the difference between off-line and on-line processing found in Experiments 2 vs. 1 and 8 vs. 7. For these constructions there was a stronger tendency towards local attachment in on-line than in off-line processing. This discrepancy might result from the increased processing load of the word-by-word reading tasks. It has to be noted, however, that the native speaker control groups showed comparable effects in these studies, which indicates that the recency preference is not L2-specific. Taken together, these results suggest that local attachment in situations of increased processing load is a general rather than a second-language specific principle.

It is of course possible that L2-specific processing is found in tasks of yet increased processing load, with more fine-grained measures and / or in less proficient learners, but neither shallow structure processing nor a second-language specific recency preference are necessary to explain the learners' processing preferences in the present studies. According to principles of explanatory parsimony, the account which makes the fewest possible assumptions is to be preferred. Thus, the null hypothesis should be that L2 processing follows the same mechanisms as L1 processing until the opposite is proven wrong.

One finding common to all the reported on-line studies is that learners had higher reading times than native speakers. The tendency was stronger in the intermediate than in the advanced learners and is in line with findings from previous studies (see chapter 2.4). The relatively slower L2 processing can have several reasons. Learners might have restricted lexical or grammatical access, which might be a consequence of L1 interference or limited WM capacity. It is also possible that learners are generally more insecure, especially in experimental tasks which they might have considered as a test situation, in spite of being explicitly told that this was not the case. As a result of this insecurity in the comprehension questions or grammaticality judgements following each sentence, the learners might have read the sentences extremely carefully, also considering that they had no possibility to go back and to reanalyse the sentences. Moreover, L2

learners generally approach language input more consciously and more carefully than native speakers. These factors have to be taken into account when comparing learners' and native speakers' on-line data because the learners' reading might be more conscious and less influenced by immediate processing decisions than native speakers' reading. The same is true for the comprehension questions and grammaticality judgements, where native speakers might make more mistakes relative to their actual competence because they are less anxious to make mistakes.

Apart from the learners' higher general reading times in comparison to the native speakers, none of the present studies found any direct effects of cognitive resource limitations in the learners that were not found in the native speakers. Although there were more local attachments in on-line than in off-line processing in Experiments 2 vs. 1 and 8 vs. 7, which might result from the increased processing load of the word-by-word reading tasks, these effects also appeared in the native speakers. This finding suggests that it is a natural rather than a second- language specific mechanism to resort to a recency preference in situations of increased processing load.

Another effect that is possibly related to processing capacity restrictions is the learners' comparative difficulty to recover from misanalyses. Evidence for these difficulties comes from the strong garden-path effects for the learners' dispreferred interpretations in the on-line studies on RC and PP attachment. Furthermore, in the on-line reading of adjunct questions, only the intermediate learners had difficulties with the non-local attachment interpretation at an early point in the sentence. These unexpected difficulties might have been caused by the need to reanalyse the structures in the way that imposes higher processing load.

It is also conceivable that processing capacity restrictions have an indirect influence on L2 processing, in that they lead to transfer effects. But such an account would contradict the finding from the word-by-word reading task in Experiment 2, where the transfer effect normally observed for these constructions was found to give way to a recency interpretation that was less demanding in terms of processing capacity. Hence, transfer is more likely to occur when participants have time to think about their attachment decisions.

Taken together, the findings of the present studies indicate that L2 processing is influenced by a shortage of automaticity and processing capacity in comparison to L1 processing. This conclusion is in line with the Fundamental Identity Hypothesis:

> "There are no fundamental differences between non-native and native grammatical representation or processing architecture forced by a critical period. Differences, if found, relate to factors characterizing L2 acquisition independently of a critical period, e.g. L1 transfer or performance factors, such as computational limitations, etc." (Hopp, 2006: 348)

The experimental findings furthermore imply that the influence of performance factors on L2 processing are not severe, at least in the learners tested in the present studies. Moreover, even native speakers are to some extent influenced by processing capacity limitations. Hence, the learners' limited processing resources do not necessarily lead to non-native like processing.

The previous discussion has shown that non native-like processing, which might be caused by L1 transfer or processing capacity limitations, are generally expected to subside with increasing proficiency. The questions when, how fast and under which circumstances the differences subside are still the subject of controversial debates. The present study addressed this issue by testing learners at two proficiency levels on various syntactic processing tasks. Since participants from both groups had started learning English at approximately the same age, i.e. at around eleven years, the age of first exposure can be excluded as an influencing factor. But it remains open whether differences between these learner groups are related to differences in times of exposure or whether the immersion period the more advanced learners had experienced was the decisive factor in changing participants' processing behaviour.

Since most processing studies with advanced learners report non native-like processing even in highly advanced and immersed learners, it is not surprising that the present studies found evidence for non-native processing in the intermediate and in the advanced learners, who were not immersed in an English-speaking country at the time of testing. Nevertheless, as outlined in the previous chapters, the differences between learners and native speakers in the present studies were not fundamental, in that the learners did not show evidence of L2-specific processing. This finding indicates that even non-immersed learners are able to employ native-like processing mechanisms.

The main difference between intermediate and advanced learners was in speed rather than in qualitatively different processing. The advanced learners' general reaction times were slightly faster, which indicates that their processing was comparatively more confident and more automatic. The garden-path effects for their dispreferred disambiguations were also weaker, which suggests that the advanced learners had less difficulty to recover from misanalyses. Further studies with yet more advanced learners are required to investigate whether L2 processing can possibly become native-like.

To conclude, even the non-immersed intermediate learners did not show evidence of L2-specific processing which would be indicative of a fundamental difference between first and L2 processing. Nevertheless, the learners still showed some non native-like processing patterns, such as L1 transfer as well as reduced processing automaticity and recovery from misanalyses. These effects were attenuated in the advanced in comparison to the intermediate learners, which indicates that increased proficiency and / or exposure may lead to more native-like L2 processing.

8.3 Future research

The last few decades have seen an increasing interest in sentence processing. Combined with modern experimental techniques such as self-paced reading, eye-tracking and ERPs, this interest has triggered an extensive body of research and the development of various theoretical processing models. Although these models can explain a variety of experimental data, especially the cross-linguistic differences found for the RC attachment ambiguity and the resulting debate on the language universality of the parser are still controversially discussed in the literature. Moreover, the exact working mechanisms of the parser are still far from fully understood. Hence, it is controversially debated (a) whether syntactic information is available in advance of or simultaneously with other information, (b) whether the parser works in a probabilistic or deterministic way, (c) whether alternative interpretations are considered in a serial or in a parallel way. Therefore, further systematic research is necessary in order to gain more insights into this broad and interesting field. Such research must include various languages and a range of materials and methods. Clearer results on general processing theories would also be helpful for a theory of L2 processing, considering that at least some processing mechanisms are transferred from the learners' L1 to their L2.

The present studies could provide some interesting insights into the mechanisms that underlie L2 processing. But it remains to be seen whether the results can be replicated and generalised to other grammatical structures, or to languages and language combinations other than those that have been examined so far. Furthermore, more data on the processing of other learner groups at various proficiency levels and in various learning situations are required to gain a better understanding of the development of L2 processing. In this respect, it would also be interesting to compare early and late language learners or natural versus instructed acquisition. Such research with various groups of learners can help to develop not only a property theory but also a transition theory (Gregg, 1996, White, 2003: 151) of L2 processing. In order to gain further in-

sights into the time course of attachment decisions, the reaction-time methods used in the present studies have to be supplemented with other psycholinguistic methods, such as eye-tracking, event-related potentials and imaging techniques, in order to furnish converging evidence on L2 processing performance.

The present study focuses on the differences between first and L2 processing and on the question to what extent learners make use of L2-specific processing. The logical next step would be to explore possible reasons for such differences. The present study provides evidence that L2 processing is not fundamentally different from L1 processing, in that learners do not show evidence of L2-specific mechanisms such as shallow syntactic processing or a recency preference, and that the differences between first and L2 processing observed in these studies are related to a relative lack of processing automaticity or processing capacity. In order to explore the first of these possibilities, the learners' exposure to the L2 could be manipulated experimentally. The second possibility could be explored by considering WM as an additional variable. Furthermore, the question arises whether native speakers' processing changes under additional processing load. Conversely, L2 learners' performance under increased task demands should suffer to a greater extent than native speakers' performance under the same demands. In order to investigate this issue, processing load could be systematically manipulated via sentence length, or by varying the presentation rate in speeded grammaticality judgements and thus increasing time pressure. This line of research is important, considering that a further understanding of the potential reasons for L2 learners' reduced processing speed and accuracy in comparison to native speakers also has significant pedagogical implications. Hence, a better understanding of the mechanisms that underlie L2 processing could lead to improved instruction methods as well as the development of tools to distinguish normal from impaired L2 development.

8.4 Concluding remarks

The aim of the eight studies reported here was to obtain a better understanding of L2 off-line and on-line processing by investigating attachment preferences of two groups of German learners of English at different proficiency levels for four syntactic ambiguities. The main research questions for the studies were to what extent first and L2 processing differ and which factors are responsible for these differences.

The learners turned out to be influenced to some extent by L1 transfer, which subsided with increasing proficiency. Apart from these transfer effects, there was no clear evidence for second-language specific processing such as shallow structure processing or a

preference to attach to the most recently encountered element in any of the examined structures and tasks. The finding indicates that learners are able to employ native-like structural processing mechanisms. Such a result is noteworthy because the intermediate learners tested in the studies had learned their L2 exclusively in a school context without immersion. Therefore, the present findings indicate that even non-immersed learners at a certain proficiency level do not employ fundamentally different processing mechanisms in on-line tasks as compared to native speakers.

First and L2 processing were guided by fundamentally similar mechanisms, but even the highly proficient learners' processing patterns were found to deviate from the native speakers in processing speed and the ability to recover from misanalyses. Hence, L2 processing is influenced by a shortage of automaticity and processing capacity in comparison to L1 processing. However, the consequences of these restrictions are not severe and attenuate with increasing proficiency. Moreover, it depends on the demands of the specific tasks and materials whether the learners' limited processing resources lead to non-native like processing.

Although further research is still required, the present studies show that behavioural off-line and on-line studies of syntactic ambiguity resolution provide interesting insights into the mechanisms that underlie L2 syntactic processing.

References

Abdelghani, H. & Fodor, J.D. (1999). *Low attachment of relative clauses in Arabic*. Poster presented a AMLaP (Arachitectures and Mechanisms of Language Processing) 1999, Edinburgh, UK.

Abney, S.P. (1987). *The English Noun Phrase in its Sentential Aspect*. PhD dissertation, MIT, Cambridge, MA.

Abney, S.P. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research* 18,1, 129-144.

Abutalebi, J., Cappa, S. & Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. *Bilingualism: Language and Cognition* 4, 179-190.

Acheson, D. J., Wells, J. B. & MacDonald, M. C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods* 40, 278-289.

Altarriba, J., Kambe, G., Pollatsek, A., & Rayner, K. (2001). Semantic codes are not used in integrating information across eye fixations in reading: Evidence from fluent Spanish-English bilinguals. *Perception and Psychophysics* 63, 875-890.

Altmann, G.M. & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition* 30, 191-238.

Aoshima, S., Phillips, C. & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language* 51, 23-54.

Augurtzky, P. (2006). *Attaching Relative Clauses in German: The Role of Implicit and Explicit Prosody in Sentence Processing*. PhD dissertation. Leipzig: Max-Planck-Institut für Kognitions- und Neurowissenschaften, MPI Series in Human Cognitive and Brain Sciences 77.

Augurzky, P., Alter, K. & Pechmann, T. (2004). The influence of prosody on reading - an ERP study on relative clause attachment. In: *Proceedings of the Second CamLing Postgraduate Conference on Language Research*. Cambridge: University of Cambridge.

Baccino, T., De Vincenzi, M. & Job, R. (2000). Cross-linguistic studies of the late closure strategy: French and Italian. In: M. De Vincenzi & V. Lombardo (Eds.), *Cross-Linguistic Perspectives on Language Processing*. Dordrecht: Kluwer, 89-118.

Baddeley, A. (2007). *Working Memory, Thought, and Action*. Oxford: OUP.

Baddeley, A.D., Gathercole, S.E. & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review 105*, 158-173.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In: J.D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing*. Dordrecht: Kluwer, 1-46.

Balthazar, D. & Kister, L. (1995). *Corrélation entre determination et selection d'un anaphorisé dans une structure N de N*. Paper presented at the Seminar 'Anaphore et Référence'. Nancy (CRIN), September 20-22.

Bard, E., Robertson, D. & Sorace, A. (1996) Magnitude estimation of linguistic acceptability. *Language* 72,1, 32-68.

Bates, E. & MacWhinney, B. (1987). Competition, variation, and language learning. In: B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum 157-193.

Bates, E. & MacWhinney, B. (1989). Functionalism and the Competition Model. In: B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing*. Cambridge: CUP, 3-73.

Berwick, R.C. & Weinberg, A.S. (1984). *The grammatical basis of linguistic performance*. Cambridge, MA: MIT Press.

Bever, T. (1970). The cognitive basis for linguistic structures. In: J.R. Hayes (Ed.), *Cognition and the Development of Language*. New York: Wiley, 278-362.

Bever, T.G. & Sanz, M. (1997). Empty Categories Access their Antecedents During Comprehension: Unaccusatives in Spanish. *Linguistic Inquiry* 28,1, 69-91.

Bhatia, T.K. & Ritchie, W.C. (1998). *The Handbook of Bilingualism*. Malden: Blackwell.

Bialystok, E. & Miller, B. (1999). The problem of age in second language acquisition: Influences from language, structure and task. *Bilingualism: Language and Cognition* 2, 127-145.

Blackwell, A. & Bates, E. (1995). Inducing agrammatic profiles in normals: Evidence for the selective vulnerability of morphology under cognitive resource limitations. Journal *of Cognitive Neuroscience* 7, 228-257.

Bock, J. K. & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology* 23, 45-93.

Boland, J. E. (1997). Resolving syntactic category ambiguities in discourse context: Probabilistic and discourse constraints. *Journal of Memory and Language* 36, 588-615.

Boland, J.E. & Boehm-Jernigan, H. (1998). Lexical constraints and prepositional attachment. *Journal of Memory and Language* 39, 684-719.

Bornkessel, I., Fiebach, C. & Friederici, A. D. (2004). On the cost of syntactic ambiguity in human language comprehension: An individual differences approach. *Cognitive Brain Research* 21, 11-21.

Bradley, D., Fernández, E.M. & Lovrić, N. (2003). *Overt prosody in the RC-attachment construction: Elicitation protocols*. Poster presented at the 9th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP), Glasgow, UK.

Brennan, S., Friedman, M. & Pollard, C. (1987). A centering approach to pronouns. In: *Proceedings of the 25th Annual Meeting of the ACL*, 155-162.

Britt, M.A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language* 33, 251-283.

Brysbaert, M. & Mitchell, C. (1996). Modifier attachment in sentence parsing: evidence from Dutch. *The Quarterly Journal of Experimental Psychology* A, 49,3, 664-695.

Caplan, D. & Waters, G. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences* 22, 77-126.

Carreiras, M. (1992). Estrategias de análisis sintáctico en el procesamiento de frases: Cierre temprano versus cierre último [Strategies for syntatic analysis in sentence processing: Early closure vs. Late closure]. *Cognitiva* 4, 3-27.

Carreiras, M., Betancort, M. & Meseguer, E. (2001). *Relative clause attachment in Spanish: Do readers use different strategies when disambiguating by gender and number?* Poster presented at the 14th Annual Conference on Human Sentence Processing, Philadelphia, AP.

Carreiras, M. & Clifton, C. (1993). Relative clause interpretation preferences in Spanish and English. *Language and Speech* 36, 353-372.

Carreiras, M. & Clifton, C. (1999). Another word on parsing relative clauses: Eyetracking evidence from Spanish and English. *Memory and Cognition* 27, 826-833.

Carreiras, M., Gillon-Dowens, M., Barber, H. & Betancourt, M. (2004). *Do proficient second language learners agree like natives?* Paper presented at the ESF Workshop 'Language Processing in First and Second Language Learners', University of Essex.

Carroll, S.E. (2001). *Input and Evidence: The Raw Material of SLA*. Amsterdam: Benjamins.

Carroll, S.E. (2004). Acquisition by processing theory: A theory of everything? *Bilingualism: Language and Cognition* 7: 23-25.

Carstens, V. (2000). Concord in Minimalist Theory. *Linguistic Inquiry* 31, 319-355.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1973). Conditions on Transformations. In: S.R. Anderson & P. Kiparsky (Eds.), *Festschrift for Morris Halle*. New York: Holt. 232-286.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. (1986a). *Barriers*. Cambridge, MA: MIT Press.

Chomsky, N. (1986b). *Knowledge of Language: its Nature, Origin, and Use*. London: Praeger.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge: CUP.

Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.

Christianson, K., Hollingworth, A., Halliwell, J. & Ferreria, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology* 42, 368-407.

Christianson, K. & Slattery, T. (2005). *Lingering effects of garden paths: Evidence for "good enough" processing.* Unpublished manuscript, University of Illinois at Urbana-Champaign.

Christianson, K., Williams, C., Zacks, R. & Ferreira, F. (2006). Younger and older adults' good enough interpretations of garden-path sentences. *Discourse Processes* 42, 205-238.

Clahsen, H. & Featherston, S. (1999). Antecedent priming at trace positions: Evidence from German scrambling. *Journal of Psycholinguistic Research* 28,4, 415-437.

Clahsen, H. & Felser, C. (2006a). Continuity and shallow structures in language processing. *Applied Psycholinguistics* 27, 107-126.

Clahsen, H. & Felser, C. (2006b). Grammatical processing in language learners. *Applied Psycholinguistics* 27, 3-42.

Clahsen, H. & Felser, C. (2006c). How native-like is non-native language processing? *Trends in Cognitive Sciences* 10, 564-570.

Clifton, C., Speer, S. & Abney, S.P. (1991). Parsing arguments: phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language* 30, 251-271.

Coles, M. G. H. & Rugg, M.D. (1995): Event-related brain potentials: An introduction. In: M.D. Rugg & M.G. H.Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition*. New York: OUP, 1-26.

Constantino, G., Oria-Meriono, M.D., Heydel, M. & Sainz, J.S. (1999). *Resolving attachment ambiguity: Evidence from Spanish, German and English sentence production*. Poster presented at the IV Symposium on Psycholinguistics, Miraflores de la Sierra, Madrid.

Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O. & Engle, R.W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review* 12, 769-786.

Corley, M. & Corley, S. (1995). *Cross-linguistic and inter-linguistic evidence for the use of statistics in human sentence processing*. Unpublished manuscript, University of Exeter, Exeter, England, and University of Edinburgh, Edinburgh, Scotland.

Corley, M. & Cuthbert, M. (1996). *Individual differences in modifier attachments: Experience-based factors*. Paper presented at the 2nd Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP), Torino, Italy.

Cowart W. (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgements*. Thousand Oaks, CA: Sage.

Crain, S. & Fodor, J.D. (1985). Rules and constraints in sentence processing. *Proceedings of the North Eastern Linguistic Society* 15, 87-104.

Crain, S. & Steedman, M. (1985). On not being led up the garden path: the use of context by the psychological syntax processor. In: D.R. Dowty, L. Karttunen & A. Zwicky (Eds.), *Natural language parsing*. Cambridge: CUP.

Crocker, M. (1996). *Computational Psycholinguistics: An Interdisciplinary Approach to the Study of Language*, Dordrecht: Kluwer.

Crocker, M.W., Pickering, M. & Clifton, C. (2000). *Architectures and Mechanisms for Language Processing*. Cambridge: CUP.

Cuetos, F. & Mitchell, D. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition* 30, 73-105.

Cuetos, F., Mitchell, D. & Corley, M. (1996). Parsing in different languages. In: M. Carreiras, J. Garcia-Albea & N. Sebastian-Galles (Eds.), *Language processing in Spanish*. Mahwah, NJ: Erlbaum, 145-187.

Curtiss, S. (1988). Abnormal language acquisition and the modularity of language. In: F.J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey. Volume II Linguistic Theory: Extensions and Implications*. Cambridge: CUP, 96-116.

Daneman, M. & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19, 450-466.

De Vincenzi, M. (1991). *Syntactic parsing strategies in Italian*. Dordrecht: Kluwer.

De Vincenzi, M. & Job, R. (1993). Some observations on the universality of the late-closure strategy. *Journal of Psycholinguistic Research* 22, 189-206.

De Vincenzi, M. & Job, R. (1995). An investigation of Late Closure: The role of syntax, thematic structure, and pragmatics in initial and final interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21, 1303–21.

Deevy, P.L. (2000). Agreement checking in comprehension: evidence from relative clauses. *Journal of Psycholinguistic Research* 29, 1, 69-79.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22, 499-534.

DeKeyser, R. (2001). Automaticity and automatization. In: P. Robinson (Ed), *Cognition and Second Language Instruction*. New York: CUP, 125-151.

DeKeyser, R. (2003). Implicit and explicit learning. In: M. H. Long & C. Doughty (Eds.), *The Handbook of Second Language Acquisition*. Oxford: Blackwell, 313-348.

Dekydtspotter, L., Kim, B., Kim, H., Wang, Y.-T., Kim, H.-K. & Lee, J. (2008). Intermediate traces and anaphora resolution in the processing of English as a second language. In: H. Chan, H. Jacob & E. Kapia, *BUCLD 32 Proceedings*. Somerville, MA: Cascadilla Press, 84-95.

Dekydtspotter, L. & Renaud, C. (2007). On Intermedieate Traces in English-French Grammar and Sentence Processing. In: H. Caunt-Nulton et al. (Eds.), *BUCLD 31 Proceedings*. Somerville, MA: Cascadilla Press, 160-171.

Dekydtspotter, L., Schwartz, B.D. & Sprouse, R.A. (2006). The Comparative fallacy in L2 processing research. In: M. Grantham O'Brien et al. (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)*. Somerville, MA: Cascadilla Proceedings, 33-40.

Desmet, T., Brysbaert, M. & De Baecke, C. (2002). The correspondence between sentence production and corpus frequencies in modifier attachment. *The Quarterly Journal of Experimental Psychology* 55A,3, 879-896.

Dick, F., Bates, E., Wulfeck, B., Utman, J. A., Dronkers, N. & Gernsbacher, M.A. (2001). Language deficits, localization, and grammar: Evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychological Review* 108, 759-788.

Dijkstra, A.F.J. & Haverkort, J.M.W. (2004). APT: Costs and benefits of a hybrid model. *Bilingualism: Language and Cognition* 7,1, 26-28.

Duden. (2005). *Duden: Grammatik der deutschen Gegenwartssprache*. Mannheim: Bibliographisches Institut.

Dussias, P.E. (2001). Sentence Parsing in Fluent Spanish-English Bilinguals. In: J.L. Nichol (Ed.), *One mind, two languages. Bilingual Sentence Processing*. Malden, MA: Blackwell.

Dussias, P.E. (2003). Syntactic ambiguity resolution in L2 learners. Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition* 25, 529-557.

Dussias, P.E. & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition* 10, 101-116.

Edmonds, A.C., Killam, J. & Liljestrand, A. (2006). *Prosody and the production of structurally ambiguous phrases*. University of Illinois at Urbana-Champaign: UIUC Working Papers.

Ehrlich, K., Fernández, E., Fodor, J.D., Stenshoel, E. & Vinereanu, M. (1999). *Low attachment of relative clauses: New data from Swedish, Norwegian and Romanian*. Poster presented at the 12th Annual CUNY Conference on Human Sentence Processing, New York, March 18-20.

Ellis, N. (2001). Memory for language. In: P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge / New York: CUP, 33-68.

Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24, 143-188.

Emmorey, K., Bellugi, U., Friederici, A. & Horn, P. (1995). Effects of age of acquisition on grammatical sensitivity - Evidence from online and off-line tasks. *Applied Psycholinguistics* 16, 1-23.

Engle, R.W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science* 11, 11-15.

Epstein S. D., Flynn, S. & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences* 1, 677-758.

Epstein S.D., Flynn, S. & Martohardjono, G. (1998). The strong continuity hypothesis: some evidence concerning functional categories in adult L2 acquisition. In: S. Flynn, G. Martohardjono & W. O'Neil (Eds.), *The Generative Study of Second Language Acquisition*. Mahwah, NJ: Erlbaum, 61-77.

Fabbro, F. (2001). The bilingual brain: cerebral representation of languages. *Brain and Language* 79, 211-222.

Featherston, S. (2008). Thermometer judgements as linguistic evidence. In: C. Riehl & A. Rothe (Eds.), *Was ist linguistische Evidenz?* Aachen: Shaker, 69-89.

Felser, C. & Roberts, L. (2004). *Plausibility and recovery from garden paths in second language sentence processing*. Poster presented at AMLaP, Aix-en-Provence.

Felser, C. & Roberts, L. (2007). Processing wh-dependencies in a second language: a cross-modal priming study. *Second Language Research* 23, 9-36.

Felser, C., Roberts, L., Marinis, T. & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics* 24, 453-489.

Felser, C., Sato, M. & Bertenshaw, N. (to appear). The on-line application of Binding Principle A in English as a second language. *Bilingualism: Language and Cognition*.

Fender, M. (2001). A review of L1 and L2/ESL word integration skills and the nature of L2/ESL word integration development involved in lower-level text processing. *Language Learning* 51,2, 319-396.

Fernández, E. (1999). Processing strategies in second language acquisition: some preliminary results. In: E. Klein (Ed.), *The development of second language grammars: A generative approach*. Amsterdam: Benjamins, 217-239.

Fernández, E. (2003). *Bilingual Sentence Processing*. Amsterdam: Benjamins.

Fernández, E., Fodor, J. D., de Almeida, R. G., Bradley, D. & Quinn, D. (2003). *Relative clause attachment in Canadian French: Prosodic boundary or F0-matching?* Poster presented at the 16th Annual CUNY Conference on Human Sentence Processing, Cambridge, MA, March 27-29.

Fernández, E. & Sainz, J. (2004). *Early preferences in RC attachment in Spanish: Two methods, and disambiguation by number agreement*. Poster presented at AMLaP 2004, Aix-en-Provence, France.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology* 47, 164-203.

Ferreira, F. & Clifton, C. E. (1986). The independence of syntactic processing. *Journal of Memory and Language* 25, 348-368.

Ferreira, F., Ferraro, V. & Bailey, K. G. D. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science* 11, 11-15.

Ferreira, F. & Henderson, J.M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16, 555-569.

Ferreira, F. & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass* 1,1-2, 71-83.

Fiebach, C.J., Schlesewski, M. & Friederici, A.D. (2002). Separating syntactic memory costs and syntacti integration costs during parsing: the processing of German wh-questions. *Journal of Memory and Language* 47, 250-272.

Finger, I. & Zimmer, M. (2000). *Relative clause attachment preference in Brazilian Portuguese.* Unpublished Manuscript.

Flege, J., Yeni-Komshian, G. & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language* 41, 78-104.

Flynn S. (1996). A parameter-setting approach to second language acquisition. In:W. Ritchie & T. Bhatia (Eds.), *Handbook of language acquisition*. San Diago: Academic Press, 121-158.

Flynn S. & Martohardjono, G. (1994). Mapping from the initial state to the final state: the separation of universal principles and language specific principles. In: B. Lust, M. Suñer & J. Whitman (Eds.), *Syntactic theory and first language acquisition: cross-linguistic perspectives. Vol. 1: Heads, projections and learnability*. Hillsdale, NJ: Erlbaum, 319-335.

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, J.A., Bever, T.G. & Garrett, M. (1974). *The Psychology of Language*. New York: Mac Graw Hill.

Fodor, J.D. (1998). Learning to parse? *Journal of Psycholinguistic Research* 27, 285-319.

Fodor, J.D. (2000). Garden path re-analysis: attach (anyway) and revision as last resort. In: M. De Vincenzi & V. Lombardo (Eds.), *Cross-Linguistic Perspectives on Language Processing*. Dordrecht: Kluwer, 21-61.

Fodor, J.D. (2002). Psycholinguistics cannot escape prosody. In: *Proceedings of the Speech Prosody* 2002 Conference, Aix-en-Provence, France.

Fodor, J.D. & Inoue, A. (1994). The diagnosis und cure of garden paths. *Journal of Psycholinguistic Research* 23, 407-434.

Fodor, J. D. & Inoue, A. (1998). Attach Anyway. In: J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing.* Dordrecht: Kluwer, 101-141.

Francis, W.S. (1999). Cognitive integration of language and memory in bilinguals: Semantic representation. *Psychological Bulletin* 125, 193-222.

Francis, W.S., Romo, L.F. & Gelman, R. (2002). Syntactic structure, grammatical accuracy, and content in second-language writing. In: R. Heredia & J. Altarriba (Eds.), *Bilingual Sentence Processing*. Amsterdam: Elsevier, 317-337.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Unpublished PhD dissertation, University of Connecticut, Storrs.

Frazier, L. (1987). Theories of sentence processing. In: Garfield, J. (Ed.), *Modularity in knowledge representation and natural language understanding*. Cambridge, MA: MIT Press, 291-307.

Frazier, L. (1990). Parsing modifiers: special purpose routines in the human sentence processing mechanism? In: D.A. Balota, G. B. Flores D'Arcais & K. Rayner (Eds.), *Comprehension Processes in Reading*. Hillsdale: Erlbaum, 303-330.

Frazier, L & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.

Frazier, L. & Clifton, C. (1997). *Construal: overview, motivation, and some new evidence*. Journal of Psycholinguistic Research 26,3, 277-297.

Frazier, L. & Fodor, J.D. (1978). The sausage machine: a new tow-stage parsing model. *Cognition* 6, 291-325.

Frazier, L. & Rayner, K. (1982). Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14, 178-210.

Frazier, L. & Vonk, W. (1997). Modifier attachment. In: J.P. de Ruiter & D. Wilkins (Eds.), *Max-Planck-Institut für Psycholinguistik: Annual Report 1996*. Nijmegen, NL: Max-Planck-Institut für Psycholinguistik.

Frenck-Mestre, C. (1997). Examining second language reading: An on-line look. In: A. Sorace (Ed.), *Proceedings of the GALA 1997 conference on language acquisition*. Edinburgh: University of Edinburgh, 474-478.

Frenck-Mestre, C. (2002). An on-line look at sentence processing in a second language. In: R. Herrida & J. Altarriba (Eds.), *Bilingual sentence processing*. Amsterdam: Elsevier.

Frenck-Mestre, C. (2004). Ambiguities and anomalies: What can eye-movements and event-related potentials reveal about second language sentence processing? In: J.F. Kroll & A. De Groot (Eds.), *Tutorials in Bilingualism*. Mahwah, NJ, 268-284.

Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: a review of methodologies and experimental findings. *Second Language Research* 21, 175-198.

Frenck-Mestre, C. & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *Quarterly Journal of Experimental Psychology* 50A, 119-148.

Frenck-Mestre, C. & Pynte, J. (2000). Resolving syntactic ambiguities: cross-linguistic differences? In: M. De Vincenzi & V. Lombardo (Eds.), *Cross-Linguistic Perspectives on Language Processing*. Dordrecht: Kluwer, 119-148.

Friederici, A. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences* 6, 78-84.

Friederici, A. D., Steinhauer, K., Mecklinger, A. & Meyer, M. (1998). Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biological Psychology* 47, 193-221.

Friedman, N.F. & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. Behavior Research Methods 37, 581-590.

Garnsey, S. M., Pearlmutter, N. J., Myers, E. & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language* 37, 58-93.

García-Orza, J., Fraga, I., Tejido, M. & Acuña, J.C. (2000). *High attachment preferences in Galician relative clauses: Preliminary data*. Poster presented at AMLaP (Architectures and Mechanisms of Language Processing) 2000, Leiden, NL, September 20-23.

Garrod, S.C. & Sanford, A.J. (1985). On the Real-time Character of Interpretation during Reading. *Language and Cognitive Processes* 1, 43-61.

Gathercole, S.E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics* 27,4, 513-613.

Gazdar, G., Klein, E., Pullum, G. & Sag, I. (1985). *Generalized Phrase Structure Grammar*. Oxford: Blackwell.

Genesee, F. & Nicoladis, E. (2006). Bilingual first language acquisition. In:E. Hoff & M. Shatz (Eds.), *Handbook of Language Development*. Oxford: Blackwell, 324-342.

Gibson, E. (1992). Review article: On the adequacy of the Competition Model. *Language* 68, 812-830.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68, 1-76.

Gibson, E. & Pearlmutter, N. (2000). Distinguishing serial and parallel parsing. J*ournal of Psycholinguistic Research* 29, 231-240.

Gibson, E., Pearlmutter, N., Caneco-Gonzaliez, E. & Hickok, G. (1996). Recency preference in human sentence processing mechanism. *Cognition* 59, 23-59.

Gibson, E., Pearlmutter, N.J. & Torrens, V. (1998). Recency and lexical preferences in Spanish. *Memory & Cognition* 27, 603-611.

Gibson, E. & Schütze, C. (1999). Disambiguation *preferences in noun phrase conjunction do not mirror corpus frequency.* Journal of Memory and Language 40, 263-279.

Gibson, E., Schütze, C.T. & Salomon, A. (1996). The relationship between the frequency and the processing complexity of linguistic structure. *Journal of Psycholinguistic Research* 25, 59-92.

Gibson, E. & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax* 7, 55-78.

Gibson, E. & Wexler, K. (1994). Triggers. *Linguistic Inquiry* 25,3, 407-454.

Gilboy, E., & Sopena, J.-M. (1996). Segmentation effects in the processing of complex NPs with relative clauses. In: M. Carreiras, J.E. García-Albea & N. Sebastián Gallés (Eds.), *Language Processing in Spanish*. Mahwah, NJ: Erlbaum, 191-206.

Gilboy, E., Sopena, M., Clifton, C. & Frazier, L. (1995). Argument structure and association preferences in Spanish and English compound NPs. *Cognition* 54, 131-67.

Golestani, N., Alario, F.-X., Meriaux, S., Le Bihan, D., Dehaene, S. & Pallier, C. (2006). Syntax production in bilinguals. *Neuropsychologia* 44,4, 1029-1040.

Gordon, P. C., Hendrick, R. & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language* 51,1, 97-114.

Gorrel, P. (1995). Syntax and Parsing. Cambridge: CUP.

Green, D. W. (1986). Control, activation and resource: A framework and a model for the control of speech in bilinguals. *Brain and Language* 27, 210-223.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* 1, 67-81.

Gregg, K. (1996). The logical and developmental problem of second language acquisition. In: W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition*. San Diego, CA: Academic Press, 49-81.

Gregg, K. R. (2003). The state of emergentism in second language acquisition. *Second Language Research* 19,2, 95-128.

Grice, H.P. (1975). Logic and conversation. In: P. Cole & J.L. Morgan (Eds.), *Syntax and semantics. Vol. 3: Speech acts*. New York: Academic Press, 41-58.

Grosjean, F. (1982). *Life with Two Languages. An Introduction to Bilingualism*. Cambridge, MA: Harvard University Press.

Groß, Rebecca (2002). "*Someone who shot the servant of the actress who was on the balcony." Zur Verarbeitung von Relativsätzen im Deutschen und Englischen*. Unpublished master's thesis. Universität Düsseldorf.

Grosz, B., Joshi, A. & Weinstein, S. (1995). Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics* 2,21, 203-225.

Gunter, T.C., Friederici, A.D. & Schriefers, H.J. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience* 12,4, 556-568.

Hahne, A. (2001).What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research* 30, 251-266.

Hahne, A. & Friederici, A. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition* 4, 123-141.

Håkansson, G., Pienemann, M. & Sayehli, S. (2002). Transfer and typological proximity in the context of L2 processing. *Second Language Research* 18, 250-273.

Harrington, M. (1987). Processing transfer*: language specific strategies as a source of interlanguage variation. Applied* Psycholinguistics 8, 351-378.

Harrington, M. (2004). Input processing as a theory of processing input. In: B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary*. Mahwah, NJ: Erlbaum, 79-92.

Harrington, M. & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition* 14, 25-38.

Hasegawa, M. (1990). Comments on Mazuka and Lust's paper. In: L. Frazier (Ed.), Language processing and language acquisition. Dordrecht: Kluwer, 207-223.

Hasegawa, M., Carpenter P.A. & Just, M.A. (2002). An fMRI study of bilingual sentence comprehension and workload. *Neuroimage* 15, 647-660.

Hawkins, R. (2001). *Second language syntax*. Oxford: Blackwell.

Hemforth, B., Konieczny, L. & Scheepers, C. (2000a). Modifier attachment: relative clauses and coordinations. In: B. Hemforth (Ed.), *German sentence processing*. Dordrecht: Kluwer, 161-186.

Hemforth, B., Konieczny, L. & Scheepers, C. (2000b). Syntactic attachment and anophor resolution: the tow sides of relative clause attachment. In: M.W. Crocker, M. Pickering & C. Clifton (Eds.), *Architectures and Mechanisms for Language Processing*. Cambridge: CUP, 259-281.

Hemforth, B., Konienczny, L., Scheepers, C. & Strube, G. (1992). SOUL-Processing: Semantik-orientierten Prinzipien menschlicher Sprachverarbeitung. In: G. Görz (Ed.). *Konvens 92. 1. Konferenz "Verarbeitung natürlicher Srpache", Nürnberg, 7.-9. Oktober 1992*. Berlin et al.: Springer, 197-208.

Hemforth, B., Konieczny, L, Selig, H. & Walter, M. (2000). Case matching and relative clause attachment. *Journal of Psycholinguistic Research* 29,1, 81-88.

Hemforth, B., Konieczny, L, Scheepers, C. & Strube, G. (1998). Syntactic ambiguity resolution in German. In: D. Hillert (Ed.), *Sentence processing: a cross-linguistic perspective*. New York: Academic Press, 294-311.

Henstra, J.-A. (1996). *On the Parsing of Syntactically Ambiguous Sentences: Coordination and Relative Clause Attachment*. Unpublished PhD dissertation, University of Sussex, UK.

Hernandez, A.E., Bates, E.A. & Avila, L.X. (1994). On-line sentence interpretation in Spanish-English bilinguals: What does it mean to be "in between". *Applied Psycholinguistics* 15, 417-446.

Hernandez, A.E., Bates, E.A. & Avila, L.X. (1996). Processing Across the Language Boundary: A Cross-Modal Priming Study of Spanish-English Bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22,4, 846-864.

Hirose, Y. (1999). *Resolving Reanalysis Ambiguity in Japanese Relative Clauses*. PhD dissertation, CUNY.

Hocking, I. (2003). *Resources and Parsing*. PhD dissertation. University of Exeter, UK.

Hoover, M.L. & Dwivedi, V.D. (1998). Sentence processing by skilled bilinguals. *Language Learning* 48, 1-29.

Hopp, H. (2006). Ultimate attainment at the interfaces in second language acquisition: grammar and processing. PhD dissertation. Groningen University.

Huynh, H. & Feldt, L. S. (1976). Estimation of the box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics* 1, 69-82.

Igoa, J.M., Carreiras, M. & Meseguer, E. (1998). A study on late closure in Spanish: Principle-ground vs. frequency-based accounts of attachment preferences. *Quarterly Journal of Experimental Psychology* 51A,3, 561-592.

Illes, J., Francis, W. S., Desmond, J. E., Gabrieli, J. D. E., Glover, G. H., Poldrack, R., Lee, C. J. & Wagner, A. D. (1999). Convergent cortical representation of semantic processing in bilinguals. *Brain and Language* 70, 347 63.

Indefrey, P. (2006). A meta-analysis of hemodynamic studies on frist and second language processing: Which suggested differences can we trust and what do they mean? Language Learning 56,1, 279-304.

Inoue, A & Fodor, J.D. (1995). Information-paced parsing of Japanese, Japanese Sentence Processing, In: R. Mazuka & N.Nagai (Eds.), *Japanese sentence processing*. Hillsdale: Erlbaum, 9-64.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics* 25, 603-634.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning* 57,1, 1-33.

Johnson, J.S. & Newport, E.L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21, 60-99.

Juffs, A. (1998a). Main verb vs. reduced relative clause ambiguity resolution in L2 sentence processing. *Language Learning* 48, 107-147.

Juffs, A. (1998b). Some effect of first language argument structure and morphosyntax on second language processing. *Second Language Research* 14, 406-424.

Juffs, Alan (2001). Psycholinguistically oriented second language research. *Annual Review of Applied Linguistics* 21, 207-223.

Juffs, Alan (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society* 102, 199-226.

Juffs, A. (2005). The influence of first language on the processing of wh-movement in English as a second language. *Second Language Research* 21, 121-151.

Juffs, Alan (2006). Processing reduced relative vs. main verb ambiguity in English as a Second Language: a replication study with working memory. In: R. Slabakova (Ed.), *Inquiries in Linguistic Development in Honor of Lydia White*. Amsterdam: Benjamins, 213-232.

Juffs, A. & Harrington, M. (1995). Parsing effects in second language sentence processing: subject and object asymmetries in wh-extraction. *Studies in Second Language Acquisition* 17, 483-516.

Juffs, Alan & Harrington, M. (1996). Garden Path Sentences and Error Data in Second Language Sentence Processing. *Language* Learning 46, 283-326.

Juliano, C. & Tanenhaus, M.A. (1994). A constraint-based lexicalist account of the subject/object attachment preference. *Journal of Psycholinguistic Research* 23, 459-471.

Jun, S.-A. (2003). Prosodic Phrasing and Attachment Preferences. *Journal of Psycholinguistic Research* 32, 219-249.

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. (1992). A Capacity Theory of Comprehension: Individual Differences in Working Memory. *Psychological Review* 99, 122-149.

Just, M.A., Carpenter, P.A. & Wolley, J.D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111, 228-238.

Just, M. A. & Varma, S. (2002). A hybrid architecture for working memory: Reply to MacDonald and Christiansen (2002). *Psychological Review* 109, 55-65.

Kaan E (2002). Investigating the effects of distance and number interference in agreement processing: An ERP study. *Journal of Psycholinguistic Research* 31,2, 165-193.

Kamide, Y. & Mitchell, D.C. (1997). Relative clause attachment: Nondeterminism in Japanese parsing. *Journal of Psycholinguistic Research* 26,2, 247-254.

Kazanina, N., Lau, Lieberman, M., Yoshida, M. & Phillips, C. (2007). The Effect of Syntactic Constraints on the Processing of Backwards Anaphora. *Journal of Memory and Language* 56, 384-409.

Kilborn, K. (1989). Sentence processing in a second language: The timing of transfer. *Language and Speech* 32, 1-23.

Kilborn, K. (1992). On-line integration of grammatical information in a second language. In R. Harris (Ed.), *Cognitive Processing in Bilinguals*. Amsterdam: Elsevier, 337-350.

Kilborn, K. & Cooreman, A. (1987). Sentence interpretation strategies in adult Dutch-English bilinguals. *Applied Psycholinguistics* 8, 415-431.

Kim, K.H.S., Relkin, N.R., Lee, K.M. & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature* 388, 171-174.

Kimball, J. (1973). Seven principles of surface structure parsing. *Cognition* 2, 15-47.

Klein, E. (1999). Just parsing through: notes on the state of L2 processing research today. In: E. Klein & G. Martohardjono (Eds.), *The development of second language grammars: a generative approach*. Amsterdam: Benjamins, 197-216.

Konieczny, L. (1996). *Human sentence processing: a semantics-oriented parsing approach*. PhD dissertation. Universität Freiburg: IIG-Berichte 3/96.

Konieczny, L. & Hemforth, B. (2000). Modifier attachment in German: relative clauses and prepositional phrases. In: A. Kennedy, R. Radach. D. Heller & J. Pynte (Eds.), *Reading as a Perceptual Process*. Amsterdam: Elsevier, 517-527.

Konieczny, L., Hemforth, B., Scheepers, C. & Strube, G. (1997). The role of lexical heads in parsing: evidence from German. *Language and Cognitive Processes* 12,2-3, 307-348.

Kroll, J. (2006). *Reading & speaking words in two languages: A problem in representation & control*. Paper presented at the The Neurocognition of Second Language: A CBBC Workshop, Georgetown University.

Kroll, J.F. & De Groot, A. (1997). Lexical and Conceptual Memory in the Bilngual: Mapping Form to Meaning in two Languages. In: J.F. Kroll & A. De Groot (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Mahwah, NJ: Erlbaum, 169-200.

La Pointe, L.B. & Engle, R.W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, Cognition* 16, 1118-1133.

Lenneberg, E. (1967). *Biological Foundations of Language*. New York: Wiley.

Liljestrand Fultz, A. (2007). Prosody in syntactic diambiguation in English-French interlanguage. In: H. Caunt-Nulton et al. (Eds.), *BUCLD 31 Proceedings*. Somerville, MA: Cascadilla Press, 394-405.

Liu, H., Bates, E. & Li, P. (1992): Sentence interpretation in bilingual speakers of English and Chinese. *Applied Psycholinguistics* 13, 451-484.

Long, M. H. (2005). Problems with supposed counter-evidence to the Critical Period Hypothesis. *International Review of Applied Linguistics* 43, 287-317.

Lourenço-Gomes, M.C. (2002). *The length of the relative clause as a factor for testing the IPH in Brazilian Portuguese: A preliminary study*. Unpublished manuscript.

Love, T., Maas, E. & Swinney, D. (2003). The influence of language exposure on syntactic language processing. *Experimental Psychology* 50,3, 204-216.

Lovrić, N. (2002). *It's the prosody that matters in Croatian*. Poster presented at the Fifteenth Annual CUNY Conference on Human Sentence Processing, New York.

Lovrić, N., Bradley, D. & Fodor, J.D. (2000). *RC attachment in Coratian with and without preposition*. Poster presented at AMLaP (Architectures and Mechanisms of Language Processing) 2000, Leiden, NL.

Lovrić, N. & Fodor, J.D. (2000). *Relative clause attachment in sentence parsing: Evidence from Croatian*. Poster presented at the 13th Annual CUNY Conference on Human Sentence Processing, La Jolla, CA.

MacDonald, M.C. (1989) Priming effects from gaps to antecedents. *Language and Cognitive Processes* 4, 35-56.

MacDonald, M.C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes* 9, 157-201.

MacDonald, M. C. & Christiansen, M. H. (2002). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review* 109, 35-54.

MacDonald, M.C., Pearlmutter, N.J. & Seidenberg, M.S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review* 101, 676-703.

McRae, K., Spivey-Knowlton, M.J. & Tanenhaus, M.K. (1998). Modeling thematic fit (and other constraints) within an integration competition framework. *Journal of Memory and Language* 38, 283-312.

MacWhinney, B. (1997). Second language acquisition and the competition model. In: J.F. Kroll & A. De Groot (Eds.), *Tutorials in Bilingualism*. Mahwah, NJ: Erlbaum 113-144.

Maia, M., Costa, A., Fernández, E. & Lourenço-Gomes, M. do C. (2004). The comprehension of ambiguous relative clauses in Brazilian and European Portuguese: A contrastive study. Revista da ABRALIN, 3,1-2, 11-39.

Mandell, P. (1999). On the reliability of grammaticality judgement tests in second language acquisition research. *Second Language Research* 15, 73-99.

Marinis, T. (2003). Psycholinguistic techniques in second language acquisition research. *Second Language Research* 19, 144-161.

Marinis, T., Roberts, L., Felser, C. & Clahsen, H. (2005). Gaps in second language processing. *Studies in Second Language Acquisition* 27, 53-78.

Marslen-Wilson, W.D. & Tyler, L.K (1987). Against modularity. In: J.L.Garfield (Ed.), *Modularity in Knowledge Representation and Natural Language Understanding*. Cambridge, MA: MIT Press.

Mayberry, R. I. & Fischer, S. D. (1989). Looking through phonological shape to lexical meaning: The bottleneck of nonnative sign language processing. *Memory and Cognition* 17, 740-754.

Mayberry, R. I. & Lock, E. (2003). Age constraints on first versus second language acquisition: Evidence for linguistic plasticity and epigenesis. *Brain and Language* 87, 369-384.

Mazuka, R. & Lust, B. (1990). On parameter setting and parsing: predictions for cross-linguistic differences in adult and child processing. In: L. Frazier (Ed.), *Language processing and language acquisition*. Dordrecht: Kluwer, 163-205.

McDonald, J.L. (1987). Sentence interpretation in bilingual speakers of English and Dutch. *Applied Psycholinguistics* 8, 379-414.

McDonald, J.L. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics* 21, 395-423.

McDonald, J.L. (2006). Beyond the critical period: processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language* 55, 381-401.

McElree, B. & Bever, T.G. (1989). The psychological reality of linguistically defined gaps. *Journal of Psycholinguistic Research* 18, 21-35.

McElree, B., Jia, G. X. & Litvak, A. (2000). The time-course of conceptual processing in three bilingual populations. *Journal of Memory and Language* 42, 229-254.

McKoon, G., Allbritton, D.W. & Ratcliff, R. (1996). Sentential context effects on lexical decisions with cross-model and all-visual procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22, 1494-1497.

Mendelsohn, A. & Pearlmutter, N. (1999). *Individual differences in relative clause attachment preferences*. Poster presented at the Annual Meeting of the Psychonomics Society, Los Angeles, CA.

Meng, M. & Bader, M. (2000). Ungrammaticality detection and garden-path strength: Evidence for serial Parsing. *Language and Cognitive Processes* 15, 615-666.

Mitchell, D.C. (1994). Sentence Parsing. In: M. Gernsbacher (Ed.), *Handbook of Psycholinguistics*. New York: Academic Press, 375-409.

Mitchell, D.C. & Brysbaert, M. (1998). Challenges to recent theories of crosslinguistic variation in parsing: evidence from Dutch. In: D. Hillert (Ed.), *Sentence processing: a cross-linguistic perspective*. New York: Academic Press, 313-344.

Mitchell, D.C., Brysbaert, M., Grondelaers, S. & Swanepoel, P. (2000). Modifier attachment in Dutch: Testing aspects of Construal Theory. In: A. Kennedy, R. Radach, D. Heller & J. Pynte (Eds.), *Reading as a perceptual process*. Oxford: Elsevier, 493-516.

Mitchell, D. & Cuetos, F. (1991). The origins of parsing strategies. In: C. Smith (Ed.), *Current issues in natural language processing*. Austin, TX: University of Austin, Center for Cognitive Science, 1-12.

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 (non-lexical) statistical records. *Journal of Psycholinguistic Research* 24,6, 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.

Miyake, A. & Friedman, N. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In: A. F. Healy & L. E. Bourne (Eds.), *Foreign Language Learning: Psycholinguistic Studies on Training and Retention*. Mahwah, NJ: Erlbaum, 339-364.

Miyamoto, E.T. (1998). *Relative clause attachment in Brazilian Portuguese*. Unpublished manuscript, MIT, Cambridge, MA.

Miyamoto, E.T. (2005). In Processamento da Linguagem. *Pelotas: Educat*, 71-90.

Miyamoto, E. T., Nakamura, M. & Takahashi, S. (2004). *Word order and prosody in the attachment of relative clauses in Japanese*. Poster presented at the 17th Annual CUNY Conference on Human Sentence Processing, University of Maryland, College Park, MD.

Müller, J.L. (2005). Electrophysiological correlates of second language processing. *Second Language Research* 21, 152-174.

Murphy, V. (1997). The effect of modality on a grammaticality judgment task. *Second Language Research* 13, 1, 34-65.

Murray, W.S., Rahman, A. & Heydel, M. (2000). *Modifier attachment preferences across languages and across structures*. Paper presented at the 6th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP), Leiden, NL, September 21-23.

Myers-Scotton, C. (1993). *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: Clarendon Press.

Nakano, Y., Felser, C. & Clahsen, H. (2002). Antecedent priming at trace positions in Japanese long-distance scrambling. *Journal of Psycholinguistic Research* 31, 531-571.

Newport, E.L., Bavelier, D. & Neville, H.J. (2001). Critical thinking about critical periods: Perspectives on a critical period for language acquisition. In: E. Dupoux (Ed.), *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler*. Cambridge, MA: MIT Press, 481-502.

Nichol, J., Fodor, J.D. & Swinney, D. (1994). Using Cross-Modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Leaming, Memory, and Cognition* 20,5, 1229-1238.

Nichol, J., Forster, K.I. & Veres, C. (1997). Subject-verb agreement processes in comprehension. *Journal of Memory and Language* 36, 569-587.

Nichol, J. & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* 18, 5-20.

Nowak, A. (2000). *On relative clause attachment in Polish: Evidence for Late Closure and against case matching*. Poster presented at the 13th Annual CUNY Conference on Human Sentence Processing, La Jolla, CA.

Odlin, T. (1989). *Cross-linguistic Influence in Language Learning*. Cambridge: CUP, 1989.

Odlin, T. (2003). Cross-linguistic Influence. In: C. Doughty & M. Long, (Eds.), *Handbook on Second Language Acquisition*. Oxford: Blackwell, 436-486.

Ojima, S., Nakata, H. & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience* 17, 1212-1228.

Omaki, A. (2005). *Working memory and relative clause attachment in first and second language processing*. Unpublished MA thesis, University of Hawaii.

Omaki, A. & Ariji, K. (2005). Testing and attesting the use of structural information in L2 sentence processing. In: L. Dekydtspotter et al. (Eds.), *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference (GASLA 2004)*. Somerville, MA: Cascadilla Press, 205-218.

Opitz, B. & Friederici, A.D. (2003). Interactions of the hippocampal system and the prefrontal cortex in learning language like-rules. *NeuroImage* 19, 1730-1737.

Oria-Merino, L., Costantino, G.V., Heydel, M. & Sainz, J.S. (2000). *Cross-linguistic differences in relative clause attachment resolution: Tuning vs. Construal*. Poster presented at the 13th Annual CUNY Conference on Human Sentence Processing, La Jolla, CA.

Oria-Merino, L. & Sainz, J.S. (1998). *Producing agreement in Spanish: The role of morphological, syntactic and referential factors*. Poster presented at AMLaP (Architectures and Mechanisms of Language Processing) 1998, Freiburg.

Oria-Merino, L. & Sainz, J.S. (1999). *Processing subject-verb agreement in Spanish: Referential factors and word order*. Poster presented at the IV Symposium on Psycholinguistics, Miraflores de la Sierra, Madrid.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C. & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning* 56,1, 199-230.

Oxford University Computing Services (2004). *The British National Corpus*. (Simple Search of BNC World) http://sara.natcorp.ox.ac.uk/lookup.html

Oxford University Language Centre. *Oxford Placement Test*. <http://www.lang.ox.ac.uk/tests/english_placement.html> (last viewed 20/07/07)

Pallier, C. (2007). Critical periods in language acquisition and language attrition. In: B. Köpke, M. S. Schmid, M. Keijzer & S. Dostert (Eds.), *Language Attrition: Theoretical perspectives*. Amsterdam: Benjamins.

Papadopoulou, D. (2006). *Cross-linguistic variation in sentence processing*. Amsterdam: Springer.

Papadopoulou, D. & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing. A study of relative clause attachment in Greek. *Studies in Second Language Acquisition* 24, 501-528.

Papadopoulou, D. & Clahsen, H. (2006). Ambiguity resolution in sentence processing: the role of lexical and contextual information. *Journal of Linguistics* 42, 109-138.

Papadopoulou, D. & Tsimpli, I. (2004). *Agreement and Case* Cues in Sentence Processing: Evidence from Subject / Object Ambiguities in Greek. Poster presented at AMLAP 2004, Aix-en-Provence.

Papagno, C. & Velar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology* 48A, 98-107.

Paradis, M. (1997). The cognitive neuropsychology of bilingualism. In: A. de Groot & J. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Hillsdale, NJ: Erlbaum, 331-354.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: Benjamins.

Paradis, J. & Genesee, F. (1996). Syntactic acquisition in bilingual children: Autonomous or interdependent? *Studies in Second Language Acquisition* 18, 1-25.

Pearlmutter, N. (2000). Linear versus hierarchical agreement feature processing in comprehension. *Journal of Psycholinguistic Research* 29, 89-98.

Pearlmutter, N., Garnsey, S.M. & Bock, K. (1999). Agreement Processes in Sentence Comprehension. *Journal of Memory and Language* 41, 427-456.

Penfield, W. & Roberts, L. (1959). *Speech and Brain Mechanisms*. Princeton, NJ: Princeton University Press.

Perani, D. & Abutalebi, J. (2005) The neural basis of first and second language processing. *Current Opinions in Neurobiology* 15, 202-206.

Phillips, C. (1996). *Order and Structure*. PhD dissertation. MIT Working Papers in Linguistics.

Phillips, C. (2006). The real-time status of island phenomena. *Language* 82,4, 795-823.

Pickering, M. (1999). Sentence comprehension. In: S. Garrod & M. Pickering (Eds.), *Language processing.* Hove: Psychology Press, 123-153.

Pickering, M. & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes* 6, 229-259.

Pienemann, M. (1998). *Language processing and second language development*. Amsterdam: Benjamins.

Poulisse, N. (1997). Language production in bilinguals. In: A. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*. Mahwah, NJ: Erlbaum, 201-224.

Prévost, P. & White, L. (2000). Missing surface inflection or impairment in second language acquisition? *Second Language Research* 16,2, 103-133.

Pritchett, D. (1991). Head position and parsing ambiguity. *Journal of Psycholinguistic Research* 20, 251-270.

Pynte, J. (1998). The time course of attachment decisions: evidence from French. In: D. Hillert (Ed.), *Sentence processing: A crosslinguistic perspective*. San Diego: Academic Press, 227-245.

Pynte, J. & Colonna, S. (2001). Competition between primary and non-primary relations during sentence comprehension. *Journal of psycholinguistic research* 30, 569-599.

Quinn, D., Abdelghani, H. & Fodor, J.D. (2000). *More evidence of implicit prosody in reading: French and Arabic relative clauses*. Poster presented at the 13th Annual CUNY Conference on Human Sentence Processing, La Jolla, CA.

Radford, A. (2004). *Minimalist* Syntax. Cambridge: CUP.

Rah, A. & Adone, D. (to appear*). Processing of the Reduced Relative vs. Main Verb Ambiguity in L2 Learners at Different Proficiency Levels. Studies in Second Language Acquisition*. Studies in Second Language Acquisition.

Rah, A. (2009). *L2 Learners' Processing of Ambiguous Wh-Questions. Unpublished Manuscript*. University of Cologne.

Rayner, K., Kambe, G. & Duffy, S.A. (2000). The effect of clause wrap-up on eye movements during reading. *Quarterly Journal of Experimental Psychology* 53A, 1061-1080.

Rizzi, L. (1990). *Relativized Minimality*. Cambridge, MA: MIT Press.

Roberts, L., Gullberg, M. & Indefrey, P. (2008). On-line pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition* 30,3, 333-357.

Roberts, L., Marinis, T., Felser, C. & Clahsen, H. (2006). Antecedent priming at trace positions in children's sentence processing. *Journal of Psycholinguistic Research* 35, 175-188.

Rodriguez, G. (2007). *Parsing cataphoric pronouns in a second language*. Presentation held on the Second Language Research Forum: University of Illinois at Urbana-Champaign.

Rodriguez, G. (2008). *Second Language Processing: Is it fundamentally different?* PhD dissertation. University of Pittsburgh.

Rohde, D. (2001). *Linger*. Retrieved October 24, 2007, from http://tedlab.mit.edu/~dr/Linger/

Rossi, S., Gugler, M.F., Friederici, A.D. & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: evidence from event-related brain potentials. *Journal of Cognitive Neuroscience* 18,12, 2030-2048.

Rüschemeyer S.-A., Fiebach, C., Kempe, V. & Friederici, A.D. (2005). Processing lexical semantic and syntactic information in first and second language: fMRI evidence from German and Russian. *Human brain mapping* 25,2, 266-86.

Rüschemeyer, S.A., Zysset, S., & Friederici, A.D. (2006). Native and non-native reading of sentences: An fMRI experiment. *Neuroimage* 31,1, 354-365.

Sag, I. A. & Fodor, J. D. (1994). Extraction Without Traces. In: *Proceedings of the Thirteenth Annual Meeting of the West Coast Conference on Formal Linguistics*. Stanford: CSLI Publications, 365-384.

Sanford, A.J. & Garrod, S.C. (1981). *Understanding Written Language: Explorations of Comprehension Beyond the Sentence*. Chichester: Wiley.

Sanford, A. & Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Sciences* 6, 382-386.

Sasaki, Y. (1994). Paths of processing strategy transfers in learning Japanese and English as foreign languages. *Studies in Second Language Acquisition* 16,1, 43-72.

Sato, M. & Felser, C. (2006). Sensitivity to semantic and morphosyntactic violations in L2 sentence processing: Evidence from speeded grammaticality judgements. *Essex Research Reports in Linguistics* 51, 1-39.

Sauerland, U. & Gibson, E. (1998). *Case matching in relative clause attachment*. Paper presented at the 11th Annual CUNY Conference on Human Sentence Processing, New Brunswick, NJ.

Schachter, J., Tyson, A. F. & Diffley, F. J. (1976). Learner intuition of grammaticality. *Language Learning* 26, 67-76.

Scherag, A., Demuth, L., Rösler, F., Neville, H. & Röder, B. (2004). The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. *Cognition* 93, B97-B108.

Schütze, C.T. & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language* 40, 409-431.

Schwartz, B. D. & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research* 12, 40-72.

Segalowitz, N. (2003). Automaticity and second languages. In: M. H. Long & C. Doughty (Eds.), *The Handbook of Second Language Acquisition*. Oxford: Blackwell, 382-408.

Segalowitz, N. & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In: J. Kroll & A. de Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches*. Oxford: OUP, 371-388.

Sekerina, I. (1997). The Late Closure Principle vs. the Balance Principle: Evidence from on-line processing of ambiguous Russian Sentences. In: P. Costa (Ed.), *The Proceedings of the Second European Conference on Formal Description of Slavic Languages*. University of Potsdam, Germany.

Sekerina, I. (2004). *Cross-linguistic variation in gender use as a parsing constraint: Dutch vs. Russian*. Paper presented at the 13th Annual Meeting Formal Description of Slavic Languages, 13 FASL, Columbia, February 24-29.

Sekerina, I., Fernández, E. & Petrova, K. (2004). Relative clause attachment in Bulgarian. O. Arnaudova, W. Browne, M. L. Rivero & D. Stojanović (Eds.), *Formal Approaches to Slavic Linguistics: The Ottawa Meeting 2003*. Ann Arbor, MI, 375-394.

Service, E., Simola, M., Metsanheimo, O. & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology* 14,3, 383-408.

Shaked, A., Bradley, D. & Fernández, E.M. (2004). *Relative clause Attachment in Hebrew: Free versus construct state nominals*. Poster presented at the 10th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP), Aix-en-Provence, France.

Singleton, D. (2005). The Critical Period Hypothesis: A coat of many colours. *International Review of Applied Linguistics* 43, 269-285.

Slabakova, R. (2000). Transfer Revisited: The L2 Acquisition of Telicity Marking in English by Spanish and Bulgarian Native Speakers. *Linguistics* 38, 4, 739-770.

Sorace, A. (2006). Possible manifestations of shallow processing in advanced second language speakers. *Applied Psycholinguistics* 27, 88-91.

Spivey-Knowlton, M. & Sedivy, M. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition* 55, 227-267.

Stevens, S.S. (1946). On the theory of scales of measurement. *Science* 103, 677-680.

Stevenson, B. (1993). *Ungrammaticality and stages in sentence processing*. PhD dissertation, Monash University.

Stowe, L.A. (1986). Parsing wh-constructions: evidence for on-line gap location. *Language and Cognitive Processes* 1, 227-245.

Stowe, L.A. & Sabourin, L. (2005). Imaging the processing of a second language: Effects of maturation and proficiency on the neural processes involved. *International Review of Applied Linguistics* 43, 329-353.

Swets, B., Desmet, T., Hambrick, D.Z. & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experiment Psychology: General* 136, 64-81.

Swets, B., Desmet, T., Clifton, C. & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition* 36,1, 201-216.

Swinney, D. (1979). Lexical access during sentence comprehension: (Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior* 18, 645-659.

Swinney, D. & Hakes, D. (1976). Effects of prior context upon lexical access during sentence comprehension. *Journal of Verbal Learning and Verbal Behavior* l5, 68l-689.

Swinney, D., Ford, M., Frauenfelder, U. & Bresnan, J. (1988). On the temporal course of gap-filling and antecedent assignment during sentence comprehension. In: B. Grosz, R. Kaplan, M. Macken & I. Sag (Eds.), *Language Structure and Processing*. Stanford, CA: CSLI.

Taraban, R. & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language* 27, 597-632.

Thornton, R., Gil, M. & MacDonald, M.C. (1998). Accounting for crosslinguistic variation: a constraint-based perspective. In: D. Hillert (Ed.), *Syntax and Semantics, Vol. 31: A crosslinguistic perspective*. San Deigo, CA: Academic Press, 211-225.

Thornton, R., MacDonald, M.C. & Gil, M. (1999). Pragmatic Constraints on the Interpretation of Complex Noun Phrases in Spanish and English. *Journal of Experimental Psychology: Learning, Memory and Cognition* 25, 1347-1365.

Townsend, D. J., & Bever, T. G. (2001). Sentence comprehension: The integration of habits and rules. Cambridge, MA: MIT Press.

Traxler, M.J. (2007). Working memory contributions to relative clause attachment processing: A hierarchical linear modelling analysis. *Memory & Cognition* 35,5, 1107-21.

Traxler, M.J., & Pickering, M.J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35, 454-475.

Traxler, M.J., Pickering, M.J. & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language* 39, 558-592.

Traxler, M.J., Pickering, M., Clifton, C. & van Gompel, R. (2000). Architectures and mechanisms for sentence processing: is syntactic parsing a form of lexical ambiguity resolution? In: M. De Vincenzi & V. Lombardo (Eds.), *Cross-Linguistic Perspectives on Language Processing*. Dordrecht: Kluwer, 149-174.

Truscott, J. & M. Sharwood Smith (2004). Acquisition by processing: a modular perspective on language development. *Bilingualism: Language and Cognition* 7,2, 1-20.

Trueswell, J.K. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language* 35, 566-585.

Trueswell, J.C. & Tanenhaus, M.K. (1994). Toward a lexicalist framework for constraint based syntactic ambiguity resolution. In: C. Clifton, L. Frazier & K. Rayner (Eds.), *Perspectives on Sentence Processing*. Hillsdale, NJ: LEA Press, 155-179.

Trueswell, J. C., Tanenhaus, M. K. & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language* 33, 285-318.

Tsimpli, I.M. (2003). I kataktisi tu genous stin elliniki os defteri glossa (The acquisition of gender in Greek as a second language). In: A. Anastasiadi-Simeonidi, A. Ralli & D. Chila-Markopoulou (Eds.), *To Genos (Gender)*. Athens: Pataki, 168-189.

Tyler, L. K. & Marslen-Wilson, W. D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning of Verbal Behavior* 16, 683-692.

Ullman, M. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition* 4, 105-122.

Ullman, M. (2004). Contributions of memory circuits to language: the declarative/procedural model. Cognition 92, 231-270.

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In: C. Sanz (Ed.), *Mind and Context in Adult Second Language Acquisition*. Washington, D.C.: Georgetown, 141-178.

Van Berkum, J.J.A., Brown, C.M. & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language* 41,2, 147-182.

Van Gompel, R., Pickering, M., Pearson, J. & Liversedge, S.P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language* 52, 284-307.

Vasishth, S., Agnihotri, R.K., Fernández, E. & Bhatt, R. (2004). *Relative clause attachment in Hindi: Effects of RC length and RC placemen*t. Poster presented at the 10th Annual Conference on Architectures and mechanisms for Language Processing (AMLaP), Aix-en-Provence, France.

Vasishth, S., Agnihotri, R.K, Fernández, E. & Bhatt, R. (2005). *Noun modification preferences in Hindi*. Proceedings of Construction of Knowledge Conference, Vidya Bhawan Society, Udaipur.

Vassiere, J. (1983). Language-independent prosodic features. In: A. Cutler & D.R. Ladd (Eds.), *Prosody: Models and measurement*. New York: Springer.

Verhoeven, L. (1992). Assessment of bilingual proficiency. In: L. Verhoeven & J. de Jong (Eds.), *The construct of language proficiency*. Amsterdam: Benjamins.

Vigliocco, G., Butterworth, B. & Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition* 61,3, 261-298.

Vigliocco, G., Hartsuiker, R., Jarema, G. & Kolk, H. (1996). One or more labels on the bottles? Notional concord in Dutch and French. *Language and Cognitive Processes* 11, 407-442.

Vigliocco, G. & Nicol, J. (1998). Separating hierarchical relations and word order in language production: is proximity concord syntactic or linear? *Cognition* 68, 13-29.

Vos, S. H., Gunter, T. C., Schriefers, H. & Friederici, A. D. (2001). Syntactic parsing and working memory: The effects of syntactic complexity, reading span, and concurrent load. *Language and Cognitive Processes* 16, 65-103.

Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Psycholinguistics* 25, 315-339.

Wartenburger, I., Heekeren, H.R., Abutalebi, J., Cappa, S.F., Villringer, A. & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron* 37, 159-170.

Waters, G.S. & Caplan, D. (1996a). The capacity theory of sentence comprehension: critique of Just and Carpenter (1992). *Psychological Review* 103, 761-772.

Waters, G.S. & Kaplan, D. (1996b). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology* 49A,1, 51-79.

Waters, G.S. & Caplan, D. (2004). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, and Computers* 35, 550-564.

Waters, G. S., Caplan, D. & Rochon, E. (1995). Processing capacity and sentence comprehension in patients with Alzheimer's disease. *Cognitive Neuropsychology* 12, 1-30.

Waters, G., Caplan, D. & Yampolsky, S. (2003). On-line syntactic processing under concurrent memory load. *Psychonomic Bulletin and Review* 10, 88-95.

Weissenborn, J. & Höhle, B. (2001). *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition*. Amsterdam: Benjamins.

Weskott, T. & Fanselow, G. (2008). *Different measures of linguistic acceptability – Not so different at all?* Paper presented at Linguistic Evidence, University of Tübingen.

Werker, J. T. & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology* 46,3, 233-251.

White, L. (1987). Against comprehensible input: the Input Hypothesis and the development of second-language competence. *Applied Linguistics* 8, 95-110.

White, L. (2003). *Second language acquisition and Universal Grammar*. Cambridge: CUP.

Wijnen, F. (1998). *Dutch relative clause attachment in tow- and three-site contexts*. Poster , 11th Annual CUNY Conference on Human Sentence Processing, New Brunswick, NJ.

Wijnen, F. (2004). The implicit prosody of jabberwocky and the relative clause attachment riddle. In:H. Quené & V. van Heuven (Eds.), *On Speech and Language. Studies for Sieb G. Nooteboom*. Utrecht: Landelijke Onderzoeksschool Taalwetenschap, 169-178.

Williams, J.N. (2006). Incremental interpretation in second language sentence processing. *Language and Cognition* 9,1, 71-88.

Williams, J.N., Möbius, P. & Kim, C. (2001). Native and non-native processing of English wh-questions: Parsing strategies and plausibility constraints. Applied Psycholinguistics 22, 509-540.

Wolf, F., Gibson, E. & Desmet, T. (2004.) Discourse coherence and pronoun resolution. *Language and Cognitive Processes* 19, 665-675.

Year, J. (2003). Sentence processing within the Competition Model. *TESOL & Applied Linguistics* 3,1.

Yuan, B. (1998): Interpretation of binding and orientation of the Chinese reflexive ziji by English and Japanese speakers. *Second Language Research* 14, 324-340.

Zagar, D., Pynte, J. & Rativeau, S. (1997). Evidence for early closure attachment on first-pass reading times in French. *Quarterly Journal of Experimental Psychology* 50A, 421-438.

Appendix

A. General Materials

A1. Language Profile (German)

Fragen zur Person:

Name:

Geschlecht:

Alter:

Muttersprache (L1):

Zweitsprache (L2):

Andere Sprachen:

Zweitspracherwerb und –kontakt:

Wie haben Sie ihre Zweitsprache (L2) erworben? (Familie, Schule, Austauschprogramm, Reisen, Arbeit):

Seit wie vielen Jahren lernen Sie insgesamt Englisch?

Seit wie vielen Jahren studieren Sie Englisch?

Wie lange und wann haben Sie in englischsprachigen Ländern gelebt?

Gegenwärtiger Sprachgebrauch (Prozent der Zeit in jeder Sprache):

	L1	L2
Freizeit		
Beruf/Studium		
Gesamt		

Wie würden Sie Ihre englische Sprachkompetenz auf einer Skala von 1 = sehr niedrig bis 10 = sehr hoch einschätzen?
1 2 3 4 5 6 7 8 9 10

A2. Language Profile (English)

Name:

Age:

Sex:

First language:

Second language(s):

A3. Information for Participants in on-line experiments

Dear participant,
The experimental series "Attachment Preferences in German ESL learners" is part of the research project "Sentence Processing in Second Language Learners", which is conducted by Ms. Anne Rah in the context of her dissertation at the Chair of Applied English Linguistics at the University of Cologne.
Experimental data are collected through questionnaires and reading time measurements. Participation in all experiments is voluntary.
Collected data will only be used for publication in the context of the above mentioned research project and will not be passed on to third parties. Participants are anonymous in all publications.
Please confirm with your signature that you are aware of the above mentioned facts and that you agree to take part in the study.
Thank you very much for your participation!

A4. Information for Participants in off-line experiments

Dear participant,
The experimental series "Attachment Preferences in German ESL learners" is part of the research project "Sentence Processing in Second Language Learners", which is conducted by Ms. Anne Rah in the context of her dissertation at the Chair of Applied English Linguistics at the University of Cologne.
Experimental data are collected through questionnaires and reading time measurements. Participation in all experiments is voluntary.
Collected data will only be used for publication in the context of the above mentioned research project and will not be passed on to third parties. Participants are anonymous in all publications.
By sending the filled in questionnaire back to us, you confirm that you are aware of the above mentioned facts and that you agree to take part in the study.
Thank you very much for your participation!

B. Materials for the experiments on RC-attachment

B1. German off-line questionnaire

Instruction:

Im folgenden Fragebogen werden Ihnen zunächst einige persönliche Fragen gestellt. Die Daten sind selbstverständlich anonym und dienen nur der statistischen Auswertung der Studie. Im Anschluss lesen Sie einige ambige Sätze, d.h., Sätze mit mindestens zwei verschiedenen Bedeutungen. Lesen Sie die Sätze bitte aufmerksam durch und entscheiden Sie möglichst spontan, welche der beiden Antwortalternativen für Sie am natürlichsten klingt (Kommas und andere Satzzeichen wurden in allen Sätzen bewusst weggelassen).

Sentences:

Adjunct-attachment:

1. Wann sagte Peter dass seine Freundin das Buch gelesen hat?
   a. Er sagte letzte Woche dass seine Freundin das Buch gelesen hat.
   b. Er sagte dass seine Freundin das Buch letzte Woche gelesen hat.
2. Wann sagte Lisa dass Tom das Auto repariert hat?
   a. Sie sagte dass Tom das Auto gestern repariert hat.
   b. Sie sagte gestern dass Tom das Auto repariert hat.
3. Wann dachte Michael dass das Auto eine Reparatur braucht?
   a. Er dachte sofort dass das Auto eine Reparatur braucht.
   b. Er dachte dass das Auto sofort eine Reparatur braucht.
4. Wann dachte der Chauffeur dass die Dame den Brief versteckt hatte?
   a. Er dachte dass die Dame den Brief sofort versteckt hatte.
   b. Er dachte sofort dass die Dame den Brief versteckt hatte.

5. Wann behauptete der Detektiv dass der Dieb das Bild gestohlen hat?
   a. Er behauptete gestern Nacht dass der Dieb das Bild gestohlen hat.
   b. Er behauptete dass der Dieb das Bild gestern Nacht gestohlen hat.
6. Wann dachten die Kinder dass der Weihnachtsmann die Geschenke bringt?
   a. Sie dachten dass der Weihnachtsmann die Geschenke am Heiligabend bringt.
   b. Sie dachten dass der W. die Geschenke bringt als sie jünger waren.
7. Wann sagte die Mutter habe sie die Pizzas gekauft?
   a. Sie sagte sie habe die Pizzas gestern gekauft.
   b. Sie sagte gestern sie habe die Pizzas gestern gekauft.
8. Wann sagte der Lehrer habe die Schülerin die Antworten abgeschrieben?
   a. Er sagte dass die Schülerin die Antworten in der Pause abgeschrieben habe.
   b. Er sagte in der Pause dass die Schülerin die Antworten abgeschrieben habe.
9. Wann behauptete der Wachmann habe der Dieb das Gebäude betreten?
   a. Er dachte in der Nacht der Dieb das Gebäude betreten habe.
   b. Er dachte dass der Dieb das Gebäude in der Nacht betreten habe.
10. Wann sagte der Mann hätten seine Nachbarn ihn gestört?
    a. Er sagte dass seine Nachbarn ihn spät am Abend gestört hätten.
    b. Er sagte spät am Abend dass seine Nachbarn ihn gestört hätten.
11. Wann sagten die Kinder habe ihr Vater die Tickets verkauft?
    a. Sie sagten letzte Woche dass ihr Vater die Tickets verkauft habe.
    b. Sie sagten dass ihr Vater die Tickets letzte Woche verkauft habe.
12. Wann sagte der Tourist habe der Fremdenführer sich verlaufen?
    a. Er sagte dass der Fremdenführer sich am letzten Tag verlaufen habe.
    b. Er sagte am letzten Tag dass der Fremdenführer sich verlaufen habe.

PP-attachment:

1. Der Mann sah den Jungen mit dem Fernglas.
   a. Der Mann hatte ein Fernglas.
   b. Der Junge hatte ein Fernglas.
2. Die Frau garnierte den Kuchen mit den Schokoladenstreuseln.
   a. Die Frau garnierte den Schokostreuselkuchen.
   b. Die Frau tat Schokostreusel auf den Kuchen.
3. Der Lehrer bestrafte die Schülerin mit den schlechtesten Noten.
   a. Der Lehrer bestrafte die Schülerin die die schlechtesten Noten hatte.
   b. Der Lehrer gab der Schülerin die schlechtesten Noten um sie zu bestrafen.
4. Der Junge dekorierte den Raum mit den Girlanden.
   a. Der Junge dekorierte den Raum der voller Girlanden war.
   b. Der Junge dekorierte den Raum mit Hilfe von Girlanden.
5. Das Mädchen schlug den Jungen mit dem Stock.
   a. Das Mädchen schlug den Jungen der einen Stock hatte.
   b. Das Mädchen benutzte den Stock um den Jungen zu schlagen.
6. Die Lehrerin belohnte den Schüler mit den besten Noten.
   a. Die Lehrerin belohnte den Schüler der die besten Noten hatte.
   b. Die Lehrerin gab dem Schüler die besten Noten um ihn zu belohnen.
7. Das Mädchen entzückte die Frau mit dem schönen Kleid.
   a. Das Mädchen entzückte die Frau dadurch dass es ein schönes Kleid trug.
   b. Das Mädchen entzückte die Frau die ein schönes Kleid trug.
8. Die Frau erwürgte den Mann mit der gelben Krawatte.
   a. Die Frau erwürgte den Mann der eine gelbe Krawatte trug.
   b. Die Frau benutzte die gelbe Krawatte um den Mann zu erwürgen.

RC-attachment:

1. Der Student mochte den Assistenten des Professors der bei dem Überfall ums Leben kam.
   a. Der Assistent kam ums Leben.
   b. Der Professor kam ums Leben.
2. Der Arzt untersuchte den Vater des Jungen der bei dem Unfall schwer verletzt wurde.
   a. Der Junge wurde verletzt.
   b. Der Vater wurde verletzt.
3. Der Reporter ignorierte die Cousine der Schauspielerin die um ein Interview bat.

a. Die Cousine bat um ein Interview.
b. Die Schauspielerin bat um ein Interview.

4. Die Sekretärin fand die Tochter der Managerin die die Party früher verlassen hatte.
   a. Die Managerin hatte die Party früher verlassen.
   b. Die Tochter hatte die Party früher verlassen.
5. Der Pilot bemerkte die Nichte von der Stewardess die eine Postkarte schrieb.
   a. Die Nichte schrieb eine Postkarte.
   b. Die Stewardess schrieb eine Postkarte.
6. Das Hausmädchen sah den Fahrer von dem Millionär der neben der Garage wartete.
   a. Der Millionär wartete neben der Garage.
   b. Der Fahrer wartete neben der Garage.
7. Der Richter kannte den Anwalt von dem Industriellen der einen Prozess verloren hatte.
   a. Der Anwalt hatte einen wichtigen Prozess verloren.
   b. Der Industriell hatte einen wichtigen Prozess verloren.
8. Der Mitarbeiter beriet die Kollegin von der Professorin die bei den Studenten sehr beliebt war.
   a. Die Professorin war sehr beliebt.
   b. Die Kollegin war sehr beliebt.

B2. English off-line questionnaire

Please read the following sentences and decide which of the alternative answers seems most plausible to you.

1. a. The student had liked the secretary of the professor who was killed in the robbery.
   □ the secretary was killed.
   □ the professor was killed.
   b. The student had liked the professor with the secretary who was killed in the robbery.
   □ the secretary was killed.
   □ the professor was killed.
2. a. The doctor examined the mother of the small boy who was badly injured in the accident.
   □ the nurse was injured.
   □ the boy was injured.
   b. The doctor examined the small boy with the mother who was badly injured in the accident.
   □ the mother was injured.
   □ the boy was injured.
3. a. The photographer ignored the daughter of the manager who was very impolite and arrogant.
   □ the daughter was arrogant.
   □ the manager was arrogant.
   b. The photographer ignored the manager with the daughter who was very impolite and arrogant.
   □ the daughter was arrogant.
   □ the manager was arrogant.
4. a. The secretary found the mistress of the baron who had left the party very early.
   □ the mistress had left the party.
   □ the baron had left the party.
   b. The secretary found the baron with the mistress who had left the party very early.
   □ the mistress had left the party.
   □ the baron had left the party.
5. a. The housemaid saw the driver of the lady who was waiting in front of the garage.
   □ the lady was waiting.
   □ the driver was waiting.
   b. The housemaid saw the lady with the driver who was waiting in front of the garage.
   □ the lady was waiting.
   □ the driver was waiting.
6. a. The pilot smiled at the nephew of the stewardess who was writing a postcard in the cockpit.
   □ the stewardess was writing a postcard.
   □ the nephew was writing a postcard.

b. The pilot smiled at the stewardess with the nephew who was writing a postcard in the cockpit.
□ the stewardess was writing a postcard.
□ the nephew was writing a postcard.

7. a. The judge knew the solicitor of the baroness who had lost an important trial recently.
□ the baroness had lost a trial.
□ the solicitor had lost a trial.

b. The judge knew the baroness with the solicitor who had lost an important trial recently.
□ the baroness had lost a trial.
□ the solicitor had lost a trial.

8. a. The little girl envied the lover of the princess who was looking at the wonderful castle.
□ the princess was looking at the castle.
□ the lover was looking at the castle.

b. The little girl envied the princess with the lover who was looking at the wonderful castle.
□ the princess was looking at the castle.
□ the lover was looking at the castle.

B3. On-line self-paced reading tasks (Experiments 2+3)

Instruction experiment 2:

Welcome. In this experiment, you'll be reading some sentences on the computer screen. When you press the space bar, the first word will appear. With every press of the space bar, a new word will appear and the last word will disappear again. You should try to read as naturally as possible, making sure that you understand what you read.
When you finish reading the last word of a sentence, press the space bar again, and you will see a question about the sentence you just read. To answer the question, press the "F" key for YES or the "J" key for NO. Try to answer as quickly and as accurately as possible. You will be told if your answer was incorrect. You should take this as an indication to read more slowly and carefully. After the question, you will see a fixation mark again and you can start with the next sentence by pressing the space bar.
You can take breaks as you need them, but please try to do so before you've started reading a new sentence, when the screen just shows a fixation mark.
That's all there is to it - so relax and press the space bar to go on :-)

Instruction experiment 3:

In this experiment, you'll be reading some sentences on the computer screen. When you press the space bar, the first segment of the sentence will appear. With every press of the space bar, a new segment will appear and the last segment will disappear again. You should try to read as naturally as possible, making sure that you understand what you read.
After each sentence, you will see a question about that sentence. To answer the question, press the key for YES or the one for NO. Try to answer as quickly and as accurately as possible. You will be told if your answer was incorrect. You should take this as an indication to read more slowly and carefully. After the question, you will see a plus again and you can start with the next sentence by pressing the space bar.
You can take breaks as you need them, but please try to do so before you've started reading a new sentence, when the screen just shows a plus.
That's all there is to it - so relax and press the space bar to go on :-)

Sentences:

1. a. The student had liked / the secretary of the professor / who / had killed / herself / in the office.
   b. The student had liked / the secretary of the professor / who / had killed / himself / in the office.
   c. The student had liked / the professor with the secretary /who / had killed / himself / in the office.
   d. The student had liked / the professor with the secretary / who / had killed / herself / in the office.

2. a. The doctor examined / the mother of the boy / who / had injured / herself / with the knife.

b. The doctor examined / the mother of the boy / who / had injured / himself / with the knife.
c. The doctor examined / the boy with the mother / who / had injured / himself / with the knife.
d. The doctor examined / the boy with the mother / who / had injured / herself / with the knife.

3. a. The photographer ignored / the daughter of the manager / who / introduced / herself / to the actor.
b. The photographer ignored / the daughter of the manager / who / introduced / himself / to the actor.
c. The photographer ignored / the m. with the d. / who / introduced / himself / to the actor.
d. The photographer ignored / the m. with the d. / who / introduced / herself / to the actor.

4. a. The secretary found / the mistress of the baron / who / had excused / herself / from the party.
b. The secretary found / the mistress of the baron / who / had excused / himself / from the party.
c. The secretary found / the baron with the mistress / who / had excused / himself / from the party.
d. The secretary found / the baron with the mistress / who had excused / herself / from the party.

5. a. The housemaid saw / the driver of the lady / who / admired / himself / in the mirror.
b. The housemaid saw / the driver of the lady / who / admired / herself / in the mirror.
c. The housemaid saw / the lady with the driver / who / admired / herself / in the mirror.
d. The housemaid saw / the lady with the driver / who / admired / himself / in the mirror.

6. a. The pilot smiled / at the nephew of the stewardess / who / was writing / himself /a postcard.
b. The pilot smiled / at the nephew of the stewardess / who / was writing / herself / a postcard.
c. The pilot smiled / at the stewardess with the nephew / who / was writing / herself / a postcard.
d. The pilot smiled / at the stewardess with the nephew / who / was writing / himself / a postcard.

7. a. The judge knew / the solicitor of the baroness / who / tried to drown / himself / in the river.
b. The judge knew / the solicitor of the baroness / who tried to drown / herself / in the river.
c. The judge knew / the baroness with the solicitor / who tried to drown / herself / in the river.
d. The judge knew / the baroness with the solicitor / who tried to drown / himself /in the river.

8. a. The little girl envied / the lover of the princess / who / was amusing / himself / in the castle.
b. The little girl envied / the lover of the princess / who / was amusing / herself / in the castle.
c. The little girl envied / the princess with the lover / who / was amusing / herself / in the castle.
d. The little girl envied / the princess with the lover / who / was amusing / himself / in the castle.

Length counts for Experiment 2

word number per condition		**DP**	**VP**	**DP**	**DP**	**rel. pron.**	**VP**	**refl. pron.**	**VP**	**Total**
of high	**Mean**	2.1	1.3	2	3	1	1.9	1	1.1	**13.4**
	(SD)	(0.4)	(0.5)	(0)	(0)	(0)	(0.6)	(0)	(0.4)	**(1.1)**
of low	**Mean**	2.1	1.3	2	3	1	1.9	1	1.1	**13.4**
	(SD)	(0.4)	(0.5)	(0)	(0)	(0)	(0.6)	(0)	(0.4)	**(1.1)**
with high	**Mean**	2.1	1.3	2	3	1	1.9	1	1.1	**13.4**
	(SD)	(0.4)	(0.5)	(0)	(0)	(0)	(0.6)	(0)	(0.4)	**(1.1)**
with low	**Mean**	2.1	1.3	2	3	1	1.9	1	1.1	**13.4**
	(SD)	(0.4)	(0.5)	(0)	(0)	(0)	(0.6)	(0)	(0.4)	**(1.1)**

character number per condition		**DP**	**VP**	**DP**	**DP**	**rel-pron**	**VP**	**refl-pron**	**VP**	**Total**
of high	**Mean**	10.9	6.1	10.1	11.8	3	9.8	7	10.8	**69.4**
	(SD)	(2.5)	(1.9)	(1.6)	(2.5)	(0)	(1.4)	(0)	(1.3)	**(4.2)**
of low	**Mean**	10.9	6.1	10.1	11.8	3	9.8	7	10.8	**69.4**
	(SD)	(2.5)	(1.9)	(1.6)	(2.5)	(0)	(1.4)	(0)	(1.3)	**(4.2)**
with high	**Mean**	10.9	6.1	9.8	13.9	3	9.8	7	10.8	**71.1**
	(SD)	(2.5)	(1.9)	(2.5)	(1.4)	(0)	(1.4)	(0)	(1.3)	**(4.0)**
with low	**Mean**	10.9	6.1	9.8	13.9	3	9.8	7	10.8	**71.1**
	(SD)	(2.5)	(1.9)	(2.5)	(1.4)	(0)	(1.4)	(0)	(1.3)	**(4.0)**
ANOVA across conditions				**p = 0.697**	**p = 0.065**					**p ≤ 0.0001**

B4. Grammaticality judgement task

Instruction:

Please read the following sentences attentively and mark whether you consider them as grammatical English sentences.

Sentences:

Grammatical
of-f.-f.
1. The student had liked the secretary of the woman who had killed herself in the office.
2. The girl saw the sister of the lady who was admiring herself in the mirror.
of-m.-m.
1. The passenger smiled at the brother of the pilot who was writing himself a postcard.
2. The doctor contacted the father of the young man who was talking to himself in the garden.
with-f.-f.
1. The photographer ignored the duchess with the daughter who introduced herself to the actor.
2. The cameraman adored the actress with the sister who liked to show herself in public.
with-m.-m.
1. The gardener disliked the rich man with the son who bankrupted himself in the casino.
2. The journalist interviewed the politician with the advisor who criticised himself far too often.

Ungrammatical:
of-f.-m.
1. The woman knew the niece of the actress who was admiring himself a bit too much.
2. The teacher liked the grandmother of the schoolgirl who often found himself in a lot of trouble.
of-m.-f.
1. The policeman arrested the nephew of the salesman who had allowed herself a long holiday.
2. The judge recognised the solicitor of the baron who tried to drown herself in the river.
with-f.-m.
1. The doctor examined the small girl with the mother who had injured himself with a knife.
2. The secretary found the lady with the housemaid who had excused herself from the party.

C. Materials for the experiments on PP-attachment

C1. Experiment 4 (on-line)

Instruction: As Experiment 2 (see Appendix B3.)

Sentences:

1	a.	The guard concealed the weapon from the criminal yesterday.
	b.	The guard concealed the weapon of his new colleague yesterday.
2	a.	The author mailed the story to the editor once again.
	b.	The author mailed the story with the improvements once again.
3	a.	The police informed the guards about the danger this morning.
	b.	The police informed the guards of the old castle this morning.
4	a.	The journalist sent the report to the magazine right away.
	b.	The journalist sent the report about the murder right away.
5	a.	The robber hid the jewels from the policemen yesterday.
	b.	The robber hid the jewels of the millionaire yesterday.
6	a.	The lady showed the necklace to her new neighbour once again.
	b.	The lady showed the necklace with the diamond once again.
7	a.	The salesman offered the apples to the little girl this morning.
	b.	The salesman offered the apples with the red patches this morning.
8	a.	The soldier handed the weapon to his new comrade right away.

b. The soldier handed the weapon of his new comrade right away.
9 a. The guard fired the weapon on the training ground yesterday.
b. The guard fired the weapon of his new colleague yesterday.
10 a. The author wrote the story within two weeks once again.
b. The author wrote the story with the happy end once again.
11 a. The police interviewed the guards during the morning once again.
b. The police interviewed the guards of the old castle this morning.
12 a. The journalist read the report in the evening once again.
b. The journalist read the report about the murder right away.
13 a. The robber sold the jewels on the black market yesterday.
b. The robber sold the jewels of the millionaire yesterday.
14 a. The lady lost the necklace in the afternoon once again.
b. The lady lost the necklace with the diamond once again.
15 a. The salesman ate the apple with great appetite this morning.
b. The salesman ate the apple with the red patches right away.
16 a. The soldier destroyed the weapon in the fierce battle yesterday.
b. The soldier destroyed the weapon of his new comrade yesterday.

Length counts

syllable number per condition		**DP**	**VP**	**DP**	**PP**	**wrap-up**	**Total**
ditrans. high	Mean	9.5	6	9.75	14.5	9.5	**49.3**
	(SD)	(1.9)	(2.4)	(1.5)	(1.9)	(0.9)	**(1.7)**
ditrans. low	Mean	9.5	6	9.75	15.1	9.5	**49.9**
	(SD)	(1.9)	(2.4)	(1.5)	(1.2)	(0.9)	**(3.4)**
monotrans. high	Mean	9.5	6	9.75	15.6	9.5	**50.3**
	(SD)	(1.9)	(2.4)	(1.5)	(2.2)	(0.9)	**(3.4)**
monotrans. low	Mean	9.5	6	9.75	15.1	9.5	**49.9**
	(SD)	(1.9)	(2.4)	(1.5)	(1.2)	(0.9)	**(3.4)**
ANOVA across conditions					**p = 0.373**		**p = 0.651**

word number per condition		**DP**	**VP**	**DP**	**PP**	**wrap-up**	**Total**
ditrans. high	Mean	2	1	2	3.4	1.8	**10.1**
	(SD)	(0)	(0)	(0)	(0.5)	(0.4)	**(0.8)**
ditrans. low	Mean	2	1	2	3.6	1.8	**10.4**
	(SD)	(0)	(0)	(0)	(0.5)	(0.5)	**(0.5)**
monotrans. high	Mean	2	1	2	3.3	1.6	**10**
	(SD)	(0)	(0)	(0)	(0.5)	(0.5)	**(0.5)**
monotrans. low	Mean	2	1	2	3.5	1.6	**10.1**
	(SD)	(0)	(0)	(0)	(0.5)	(0.5)	**(0.4)**
ANOVA across conditions					**p = 0.656**	**p = 0.351**	**p = 0.505**

C2. Experiment 5 (off-line)

Instruction: The following sentences are ambiguous, which means that they have two possible interpretations, depending on how you read them. Please read the sentences carefully and decide which of the two interpretations you prefer (which is easier for you to understand or which sounds more natural).

Sentences:

1. Peter copied the term paper that he had hidden from Julia.
   a. Peter copied the term paper from Julia.
   b. Peter had hidden the term paper from Julia.
2. Lily offered the book that she had sold to John.

a. Lily offered the book to John.
b. Lily had sold the book to John.

3. Tom received the necklace that he bought from Mary.
   a. Tom bought the necklace from Mary.
   b. Tom received the necklace from Mary.
4. Lisa sent the book that she had shown to Bill.
   a. Lisa sent the book to Bill.
   b. Lisa had shown the book to Bill.
5. Jeff recommended the report that he had promised to the magazine.
   a. Jeff recommended the report to the magazine.
   b. Jeff had promised the report to the magazine.
6. Janet concealed the cell phone that she had borrowed from her boyfriend.
   a. Janet concealed the cell phone from her boyfriend.
   b. Janet had borrowed the cell phone from her boyfriend.
7. Mike withdrew the money that he had stolen from the bank.
   a. Mike withdrew his money from the bank.
   b. Mike had stolen the money from the bank.
8. Carrie handed the money that she had transferred to the detective.
   a. Carrie handed the money to the detective.
   b. Carrie had transferred the money to the detective.

C3. Experiment 6 (on-line)

Instruction:

In this experiment, you'll be reading sentences word by word. With every press of the space bar, a new word will appear and the last word will disappear again. You should try to read as naturally as possible, making sure that you understand what you read.
When you finish reading the last word of a sentence, press the space bar again, and you will be asked whether the sentence you have read was a possible sentence in English. Concentrate on how you feel about these sentences. For example, you might feel that a) and b) sound possible in English, whereas c) and d) do not.
a) Mary is likely to read the book.
b) It seems that John is late.
c) John seems that he is late.
d) Mary offers the book.
You can take breaks as you need them, but please try to do so before you've started reading a new sentence, when the screen just shows a plus.
That's all there is to it - so relax and press the space bar to go on:-)

Sentences:

Ambiguous:
1. Lily offered the CD that she had sold to her Mum yesterday.
2. Tom received the book that he had bought from his friend yesterday.
3. Jerry announced the report that he had promised to his boss yesterday.
4. Janet concealed the present that she had borrowed from her friend yesterday.
5. Mike removed the money that he had stolen from the bank yesterday.
6. Carrie handed the papers that she had announced to her boss yesterday.

High:
1. Julia had sold the book that he was hiding to her friend yesterday.
2. John offered the CD that he had bought to his friend yesterday.
3. Carrie handed the money that she had borrowed to her Dad yesterday.
4. Mary received the letter that she had shown from her Mum yesterday.
5. Janet concealed the present that she had promised from her friend yesterday.
6. Mike withdrew the money that he had transferred from the bank yesterday.

Low:

1. Peter had sold the book that he was hiding from his friend yesterday.
2. Lily offered the CD that she had bought from her Mum yesterday.
3. Carrie handed the money that she had borrowed from her Dad yesterday.
4. Tom received the letter that he had shown to his friend yesterday.
5. Janet concealed the present that she had promised to her friend yesterday.
6. Mike withdrew the money that he had transferred to the bank yesterday.

Length counts

word number per condition		DP1	VP1	DP2	that	$Pron_{DP1}$	VP2	PP	wrap-up	Total
ambiguous	**Mean**	1	1	2	1	1	2	3	1	12
	(SD)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
high	**Mean**	1	1	2	1	1	2	3	1	12
	(SD)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
low	**Mean**	1	1	2	1	1	2	3	1	12
	(SD)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)

character number per condition		DP1	VP1	DP2	that	$Pron_{DP1}$	VP2	PP	wrap-up	Total
ambiguous	**Mean**	4.5	7.7	8	4	2.5	9.8	10.2	9	55.7
	(SD)	(1.0)	(1.2)	(1.8)	(0)	(0.5)	(1.8)	(2.5)	(0)	(5.5)
high	**Mean**	4.7	7.5	7.8	4	2.5	10.3	10.7	9	56.5
	(SD)	(0.8)	(1.0)	(1.7)	(0)	(0.5)	(2.1)	(1.6)	(0)	(4.7)
low	**Mean**	4.5	7.5	7.8	4	2.5	10.3	10.7	9	56.3
	(SD)	(1.0)	(1.0)	(1.7)	(0)	(0.5)	(2.1)	(1.3)	(0)	(3.7)

D. Materials for the experiments on adjunct-attachment

D1. Experiment 7 (off-line)

Instruction:

The following sentences are ambiguous, i.e. they have two possible interpretations, depending on how you read them. Please read the sentences carefully and decide which of the two interpretations you prefer (which is easier for you to understand or which sounds more natural).

Sentences:

1. When did Peter say that his girlfriend read the book on wild horses?
   a. Peter said so last week.
   b. Mary read the book last week.
2. When did Lisa say that Tom repaired the engine of the car?
   a. Lisa said so yesterday.
   b. Tom did that yesterday.
3. When did Michael think that the car needed a repair at the garage?
   a. He thought so when he first heard the funny noises.
   b. The car needed a repair very soon.
4. When did the driver think that the lady had hidden a letter in her pocket?
   a. He thought so when he saw her.
   b. She had hidden the letter when she received it.
5. When did the detective find out that the thief had stolen the painting from the gallery?
   a. The detective found out during the night.
   b. The thief had stolen the painting in the afternoon.
6. When did the teacher say that the girl copied the answers from her neighbour?
   a. He said so last week.
   b. The girl had copied the answers during the break.

7. When did the mother say that she bought the ice-cream from the shop?
   a. She bought it yesterday.
   b. She said so yesterday.
8. When did the children think that Santa Clause brought the presents for the family?
   a. They thought he had brought them at night.
   b. They thought that when they were younger.

D2. Experiment 8 (on-line)

Instruction: As Experiment 6 (see Appendix C3.)

Sentences:

Ambiguous:
When did the watchman think that the robber had entered the building ?
When did the sergeant think that the robber had stolen the painting ?
When did the teacher say that the student had copied the answers ?
When did the old man claim that his neighbours had kept him awake ?
When did the children say that their father had sold their tickets ?
When did the tourist claim that the guide had lost the way back ?

High:
How could the watchman think that the robber had entered the building ?
How could the sergeant think that the lady had stolen the painting ?
How could the teacher say that the student had copied the answers ?
How could the old man claim that his neighbours had kept him awake ?
How could the children say that their father had sold their tickets ?
How could the tourist claim that the guide had lost the way back ?

Low:
How did the watchman think that the robber had entered the building ?
How did the sergeant think that the robber had stolen the painting ?
How did the teacher say that the student had copied the answers ?
How did the old man claim that his neighbours had kept him awake ?
How did the children say that their father had sold their tickets ?
How did the tourist claim that the guide had lost the way back ?

Length counts

word number per condition		**CP**	**DP1**	**VP1**	**that**	**DP2**	**VP2**	**DP**	**?**	**Total**
ambi-guous	**Mean (SD)**	2 (0)	2 (0)	1 (0)	1 (0)	2 (0)	2 (0)	2 (0)	_	12 (0)
high	**Mean (SD)**	2 (0)	2 (0)	1 (0)	1 (0)	2 (0)	2 (0)	2 (0)	_	12 (0)
low	**Mean (SD)**	2 (0)	2 (0)	1 (0)	1 (0)	2 (0)	2 (0)	2 (0)	_	12 (0)

character number per condition		**CP**	**DP1**	**VP1**	**that**	**DP2**	**VP2**	**DP**	**?**	**Total**
ambi-guous	**Mean (SD)**	7 (0)	10.3 (0.8)	4.3 (1.0)	4 (0)	10.0 (1.8)	8.2 (1.3)	10.3 (1.4)	1 (0)	54.2 (2.2)
high	**Mean (SD)**	8 (0)	10.3 (0.8)	4.3 (1.0)	4 (0)	10.0 (1.8)	8.2 (1.3)	10.3 (1.4)	1 (0)	55.2 (2.2)
low	**Mean (SD)**	6 (0)	10.3 (0.8)	4.3 (1.0)	4 (0)	10.0 (1.8)	8.2 (1.3)	10.3 (1.4)	1 (0)	53.2 (2.2)

E. Results of the self-paced reading tasks

E1. Experiment 2

	1	2	3	4	5	6	7	8
Native speakers								
of high	985.08	535.17	966.52	1582.29	559.04	953.17	782.28	1683.79
	(140.66)	(108.59)	(72.17)	(170.12)	(60.73)	(226.53)	(170.27)	(155.19)
of low	1071.11	571.76	956.63	1611.36	586.08	920.027	636.83	1460.44
	(110.34)	(109.02)	(45.53)	(54.94)	(26.95)	(190.82)	(41.28)	(137.39)
with high	1061.42	600.21	1029.84	1585.07	675.92	1066.68	859.66	1909.31
	(112.87)	(179.11)	(110.09)	(142.46)	(80.55)	(221.56)	(206.01)	(190.55)
with low	985.16	604.64	980.66	1597.55	593.29	1024.45	576.69	1544.94
	(140.81)	(130.39)	(131.37)	(89.37)	(52.66)	(251.44)	(38.87)	(102.77)
Advanced learners								
of high	1091.99	611.29	947.16	1495.78	584.86	961.19	666.92	1570.77
	(376.78)	(242.67)	(58.33)	(192.83)	(121.43)	(54.16)	(126.04)	(337.31)
of low	1028.87	533.54	1059.48	1526.95	585.38	959.18	623.06	1464.15
	(37.62)	(67.81)	(197.81)	(128.80)	(123.65)	(488.25)	(108.69)	(173.10)
with high	1037.19	502.73	924.63	1354.21	569.02	824.7	750.11	1724.37
	(198.93)	(64.57)	(77.23)	(90.89)	(117.79)	(227.62)	(95.99)	(154.72)
with low	962.05	579.54	1023.48	1566.65	550.31	1012.88	560.38	1541.44
	(75.63)	(161.89)	(7.33)	(133.49)	(109.04)	(416.98)	(22.10)	(219.06)
Intermediate learners								
of high	1106.24	543.61	1071.02	1555.86	514.39	969.96	711.37	1569.24
	(71.12)	(36.88)	(49.64)	(60.33)	(17.59)	(241.66)	(66.61)	(97.68)
of low	1107.94	663.54	1114.13	1639.48	546.46	1020.09	577.856	1517.84
	(122.99)	(95.63)	(86.79)	(88.24)	(29.909)	(185.65)	(49.29)	(124.96)
with high	1078.27	641.21	1079.44	1605.53	591.049	1035.11	659.61	1530.22
	(72.65)	(123.82)	(90.50)	(67.03)	(41.17)	(234.83)	(24.20)	(181.34)
with low	1103.89	644.79	1026.17	1473.04	555.13	889.87	591.53	1416.94
	(152.57)	(67.79)	(51.89)	(93.18)	(79.82)	(130.26)	(36.25)	(136.63)

E2. Experiment 3

	Segment					
Condition	**1**	**2**	**3**	**4**	**5**	**6**
Native speakers						
of high	868.91	1480.60	793.78	757.84	952.69	1016.96
	(160.12)	(300.40)	(41.80)	(53.49)	(192.54)	(152.97)
of low	965.08	1495.78	737.15	740.91	749.22	930.31
	(149.69)	(485.62)	(89.54)	(119.27)	(65.34)	(157.01)
with high	946.54	1702.44	724.66	750.34	938.21	1099.60
	(207.35)	(351.79)	(138.35)	(109.98)	(180.75)	(203.54)
with low	839.19	1429.54	669.73	760.62	759.79	859.59
	(165.92)	(216.97)	(27.42)	(130.73)	(59.79)	(143.54)
Intermediate learners						
of high	1169.68	2078.29	782.80	888.51	909.52	997.17
	(264.34)	(385.74)	(137.89)	(187.55)	(172.32)	(135.82)
of low	1243.09	1885.94	891.18	837.32	1088.06	1042.67
	(201.434)	(504.59)	(195.37)	(129.25)	(140.75)	(114.22)
with high	1229.16	2076.31	863.43	894.75	1068.87	1017.72
	(226.24)	(556.56)	(117.61)	(178.92)	(219.92)	(253.46)
with low	1217.59	2089.24	773.27	884.90	875.77	908.55
	(284.37)	(336.88)	(96.08)	(151.45)	(149.20)	(124.82)

E3. Experiment 4

	Segment				
Condition	**1**	**2**	**3**	**4**	**5**
Native speakers					
ditrans.-high	877.28	501.99	947.42	1444.42	891.68
	(73.32)	(56.22)	(100.46)	(154.21)	(147.08)
ditrans.-low	948.70	528.16	972.66	1542.86	961.55
	(122.06)	(67.40)	(152.72)	(227.14)	(240.05)
monotrans.-high	920.59	504.61	984.99	1499.43	853.79
	(50.52)	(19.67)	(35.19)	(144.01)	(88.87)
monotrans.-low	968.46	500.56	951.23	1605.19	893.15
	(45.72)	(22.72)	(50.92)	(230.60)	(141.93)
Advanced learners					
ditrans.-high	925.13	522.90	1031.38	1518.72	924.46
	(11.07)	(108.73)	(91.79)	(70.87)	(95.28)
ditrans.-low	944.33	509.70	1031.59	1635.18	881.44
	(38.86)	(41.68)	(26.95)	(199.35)	(291.52)
monotrans.-high	978.66	524.16	997.43	1550.55	894.27
	(38.15)	(53.86)	(43.25)	(140.09)	(96.57)
monotrans.-low	966.07	500.24	940.64	1654.49	878.38
	(112.053)	(50.23)	(46.08)	(224.74)	(174.25)
Intermediate learners					
ditrans.-high	982.96	522.05	1021.43	1569.38	803.14
	(71.13)	(34.35)	(82.89)	(191.97)	(339.78)
ditrans.-low	1014.33	590.58	1075.01	1896.59	762.32
	(94.01)	(106.37)	(101.45)	(376.67)	(348.54)
monotrans.-high	1036.96	553.12	1055.28	1700.69	951.44
	(22.83)	(24.061)	(50.34)	(295.49)	(235.68)
monotrans.-low	1034.09	559.34	1065.52	1840.41	885.65
	(76.85)	(20.48)	(34.74)	(304.09)	(267.98)

E4. Experiment 6

	Segment						
Condition	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Native speakers							
ambiguous	551.74	599.81	1085.41	561.08	1639.33	1680.36	1163.45
	(70.92)	(52.69)	(54.87)	(60.67)	(85.13)	(36.59)	(231.34)
high	529.51	576.9	1092.75	554.99	1558.08	1820.57	1353.94
	(55.21)	(25.41)	(64.27)	(64.04)	(114.83)	(252.14)	(295.83)
low	518.95	590.82	1119.57	558.34	1585.31	1655.27	1163.81
	(26.31)	(51.52)	(42.26)	(46.40)	(61.37)	(118.19)	(198.89)
Advanced learners							
ambiguous	558.15	662.63	1244.59	728.42	1877.53	1923.39	1516.85
	(46.96)	(72.50)	(64.88)	(120.39)	(80.26)	(220.25)	(158.24)
high	544.31	640.68	1174.45	593.33	1761.20	2089.88	1512.20
	(45.68)	(65.78)	(62.89)	(45.20)	(130.54)	(173.53)	(376.15)
low	561.86	744.38	1293.13	694.67	2025.55	1875.2	1544.39
	(61.41)	(92.36)	(100.76)	(81.31)	(230.89)	(131.34)	(406.79)
Intermediate learners							
ambiguous	643.78	763.93	1387.99	731.09	1979.18	1951.37	1316.39
	(78.35)	(103.09)	(71.40)	(74.33)	(102.43)	(170.98)	(160.65)
high	607.34	811.75	1321.15	666.52	2014.97	2142.33	1333.42
	(34.30)	(116.38)	(94.24)	(97.28)	(190.14)	(100.37)	(338.02)
low	565.54	824.92	1338.43	701.07	2095.26	1824.75	1335.74
	(39.44)	(199.87)	(96.12)	(76.28)	(55.57)	(106.35)	(309.67)

E5. Experiment 8

	Segment								
Condition	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Native speakers									
ambiguous	486.92	500.57	1064.44	528.26	504.78	1010.86	991.47	1069.69	748.16
	(49.53)	(28.95)	(72.05)	(32.09)	(38.01)	(65.43)	(42.53)	(106.41)	(57.13)
low	466.35	492.62	1005.67	546.19	490.29	1025.09	994.09	1031.68	980.27
	(24.39)	(19.49)	(44.23)	(27.99)	(17.68)	(28.63)	(70.90)	(80.97)	(262.01)
high	485.15	512.72	1041.76	559.32	549.69	1002.43	1013.11	1102.52	1119.17
	(37.49)	(13.36)	(61.33)	(47.59)	(18.85)	(31.57)	(36.62)	(88.94)	(220.37)
Advanced learners									
ambiguous	502.57	571.64	1158.83	629.33	604.45	1144.88	1169.48	1210.44	1101.26
	(35.73)	(66.18)	(88.76)	(55.50)	(53.49)	(67.25)	(56.13)	(75.10)	(184.02)
low	500.85	541.49	1179.19	593.87	619.04	1189.77	1212.01	1202.23	1264.59
	(35.55)	(19.51)	(59.78)	(71.83)	(50.47)	(158.39)	(70.35)	(93.80)	(481.24)
high	519.45	588.29	1184.89	646.22	729.37	1302.73	1231.61	1279.71	1403.19
	(46.29)	(35.99)	(104.31)	(44.70)	(97.35)	(152.09)	(101.21)	(68.21)	(247.84)
Intermediate learners									
ambiguous	584.58	624.97	1374.31	676.74	672.31	1238.85	1369.19	1285.34	1082.61
	(95.72)	(26.98)	(180.32)	(66.12)	(115.70)	(86.77)	(100.95)	(91.44)	(188.92)
low	579.59	627.71	1323.71	703.91	637.67	1344.11	1375.18	1303.96	1094.04
	(31.37)	(47.07)	(62.68)	(70.91)	(48.63)	(71.06)	(108.53)	(65.16)	(247.46)
high	601.66	604.71	1343.86	808.07	807.14	1324.51	1363.10	1354.24	1217.25
	(74.83)	(35.95)	(85.53)	(64.63)	(113.28)	(98.35)	(138.79)	(75.73)	(198.48)