Goals in Discourse

From Actions to Rhetorical Relations

Frank Zickenheiner

Danksagung

Als ich angefangen habe, mich mit rhetorischen Relationen zu beschäftigen, hatte ich noch kaum eine Ahnung von Linguistik und linguistischen Frage- und Problemstellungen. Auch waren für mich die Themenbereiche Semantik und Pragmatik und insbesondere der Bereich den man mit "Diskursstruktur" überschreiben könnte, zunächst neu und fremd. Als ich mich dann in diese, für mich neuen Themen eingearbeitet habe war ich schnell davon fasziniert, wie facettenreich und vielschichtig sie erscheinen und von wie vielen verschiedenen Blickwinkeln man sie aus betrachten kann: das selbe Phänomen kann von einer fast schon rein logisch/mathematischen Perspektive, über einer philosophischen, einer cognitiven, einer computerwissenschaftlichen, bis hin zu einer neurowissenschaftlichen Perspektive betrachtet werden und erscheint dabei jedesmal in einem anderen Licht. Die vorliegende Arbeit ist von vielen dieser Zugänge beeinflusst und ranchiert zwischen ihnen in dem Sinne, als dass sie mit Bausteinen aus der Philosophie (insbesondere der Sprechakt- und der Handlungstheorie), der künstlichen Intelligenz und der Cognitionsforschung arbeitet, aber eben auch mit formalen Methoden, wie eben Formalisierungen von Grundannahmen. Es wird außerdem versucht innerhalb dieser Formalisierungen eine stringente Beweisführung zu realisieren.

Wie schon angedeutet, hatte ich bevor ich mit der vorliegenden Arbeit begonnen habe, im wesentlichen keine Ahnung von Linguistik im Allgemeinen, und keine Ahnung von linguistischer Pragmatik im Speziellen. Dass sich dieser Umstand ändern konnte, verdanke ich im wesentlichen meiner Betreuerin Katja Jasinskaja, die das Wagnis auf sich genommen hat, jemanden wie mich, also eine bis Dato gänzlich fachfremde Person als Mitarbeiter in ihr Projekt aufzunehmen. Dabei verstand sie es nicht nur, an den richtigen Stellen meine Arbeit engmaschig zu überblicken, sondern auch die Zügel an den anderen Stellen so locker zu lassen, dass ich Kapazitäten hatte mich innerhalb der Linguistik jenseits meines eigentlichen Fachbereiches weiter zu bilden. So wurde ich beispielsweise immer wieder dazu ermutigt, meinen Horizont durch den Besuch von Sommerschulen und Workshops zu erweitern und mich dadurch fachlich stärker zu etablieren. Mein größter Dank gilt daher an dieser Stelle meiner Betreuerin Katja Jasinskaja!

Ein weitere Dank gilt Klaus von Heusinger, der sowohl die Zweitbetreuung meiner Arbeit übernommen hat und in seiner Funktion als Sprecher des Sonderforschunsbereichs 1252 *Prominenz in Sprache*, in dem ich mitarbeiten durfte, meine Arbeit erst möglich gemacht hat. Neben den vielen Kollegen aus der Kölner Linguistik und insbesondere aus dem SFB 1252, möchte ich insbesondere noch Stefan Hinterwimmer als Drittgutachter meiner Arbeit erwähnen und danken, sowie die Betreuer unseres Doktorandenkolloqiums, Petra Schumacher, Doris Mücke und Marco Garcia, die es tatsächlich geschafft haben, diese Kolloqium nicht nur zu einer sehr hilfreichen, sondern auch zu einer sehr kurzweiligen Veranstaltung werden zu lassen. Auch ihnen an dieser Stelle, sowie allen Kollegen, die ich hier nicht aufführen konnte, mein großer Dank.

Zu guter Letzt will ich noch Christian Reinecke danken, der in der letzten Phase vor der Abgabe dieser Arbeit, längere Passagen Korrektur gelesen hat und mich erheblich bei LaTeX-Problemen und Fragen des Textsatzes unterstützt hat. Außerdem danke ich natürlich meinen Eltern, denn die haben ja schließlich nicht unwesentlich dazu beigetragen, dass dieses ganze Projekt realisiert werden konnte.

27. Dezember 2021 F.Z.

Diese Arbeit wurde von Frank Christian Zickenheiner an der Universität zu Köln angefertigt. Sie ist ein Projekt zur Erlangung des Doktorgrades der philosophischen Fakultät an der Universität zu Köln. Als solche wurde sie angenommen und bewertet.

Contents

List of Figures

List of Principles

List of Symbols

$\cap$	set intersection
$\cup$	set union
$\subset$	subset relation
$x \setminus y$	set difference
$\in$	element relation
$>>$	presupposes
$\exists$	existential quantifier
$\forall$	universal quantifier
$\wedge$	conjunction
$\vee$	disjunction
$\uplus$	conjoint of DRSs
$\neg$	negation
$\Diamond$	modal possibility operator
BEL	believe operator
DES	desire operator
INT	intention operator
MOD	general mode operator
Att	attitude description set
$\oplus$	conjoint of mental states
$A \twoheadrightarrow B$	A causes B
$\Rightarrow$	implication
$\Leftrightarrow$	equivalence
$:\Leftrightarrow$	equivalence by definition
$f : x \rightarrow y$	f maps from set x to y
$\mathrm{CG}_{A,B}$	the common ground of A and B
$\prec$	temporally preceeds

Chapter 1

Introduction

Consider the following pair of sentences:

(1.1) a. What are you doing tonight?
b. I wouldn't mind having a beer.

These two utterances form a coherent discourse, even though they do not seem to have much in common in the sense that they do not share any semantic content. However, a competent language user is able to establish a certain relation between these two utterances and, hence, is able to turn them into a meaningful discourse. The question is now the following: How does a competent language user construct a relation between these utterances which turns them into a coherent discourse?

In (1.1), the two utterances seem to be somehow related by a certain goal of the speaker. In the case of this example, one may assume that the speaker's goal was actually to have a drink with the hearer. Given this assumption, both utterances in (1.1) can be understood as means for achieving this goal which finally turns these two utterances into a coherent discourse. This means that both utterances belong to a a plan for achieving a particular goal and the fact that these utterances belong to the same plan makes the discourse in (1.1) coherent.

Example (1.1) is a particular instance of the general phenomenon that hearers try to establish coherence between independent utterances. If it is actually possible to establish such a coherence, they build a *rhetorical relation*. Using the terminology of rhetorical relations, one can classify the relation between the two utterances in (1.1) as a a *speech act level explanation* (Jasinskaja and Karagjosova, 2015; Sweetser, 1990). This explanation basically captures the intuition that the second speech act provides an explanation for the performance of the first speech act.

In general, the terminology of rhetorical relations aims at delivering a method for describing and classifying coherence within a discourse, such as a dialogue or a text. However, the notion of rhetorical relations comes with some problems: The notion of rhetorical relations can be used to analyze the structure of coherent discourse. However, they do not provide any explanation for discourse coherence in the first place.

Secondly, there are some discourse phenomena (such as certain occurrences of non-local pronoun resolution, cp. page 6) which can be described in terms of rhetorical relations, but which cannot be explained by them.

This book, therefore, tries to address these two problems by establishing a notion of discourse coherence which incorporates rhetorical relations as an epiphenomenon and secondly, by providing a better understanding of some discourse phenomena wich can be described in terms of rhetorical relations. The general idea behind these attempts is to regard language actions and, hence, discourse, as a special case of general human action and, therefore, discourse coherence as a special case of general coherent behavior which is percieved as coherent because it is possible to recognize a plan which integrates individual actions. This means that a discourse is percieved as coherent if it is possible to identify a plan which connects the individual utterances. By integrating discourse structure into such an action theoretical framework, it will be possible to operate with notions such as actions, goals and plans, to apply these notions to speech acts, and gain new insights about the topic by taking a broader perspective.

The rest of this introductory chapter aims at giving an overview of the phenomena that are dealt with in this book, its general idea and its general structure: Section 1.1 provides a short introduction into rhetorical relations and discusses some discourse phenomena wich can be described in terms of rhetorical relations. Section 1.2 presents the main goals of this book, its main ideas and its most important results. Section 1.3 presents the most important theories and areas of research on which the ideas of this book are based. Finally, section 1.4 provides an overview of the content of the remaining chapters of this book.

1.1 The Topic: Rhetorical Relations

1.1.1 An utterance in its context

Any utterance has to be interpreted relative to its context and this context may equip the utterance with additional meaning. This is illustrated in the following examples:

(1.2) *A*: I'm going to water the plants.
B: It's raining. [*So you don't have to water them.*]

(1.3) *A*: Do you think I need sunglasses?
B: It's raining. [*So you don't need sunglasses.*]

(1.4) *A*: Let's go for a walk!
B: It's raining. [*So I don't want to go for a walk.*]

In all these examples, *B* produces exactly the same utterance but the intended effect of this utterance varies with respect to the given context, even though the semantic content of these utterances stays exactly the same.

This observation is not particularly new, of course. The best studied cases of the meaning of a sentence being enriched with additional information in a given context are propably Gricean implicatures (Grice, 1975). Grice' idea was that the hearer of

an utterance can expect it to be relevant at the given stage of a conversation. This assumption enables the hearer to calculate the intended meaning of an utterance and in particular meaning and information that goes beyond its pure semantic content. An agent *A* who wants to water plants can relate the information that it is raining to his plans and might come to the conclusion that carrying out his plan is superfluous.

1.1.2 Rhetorical Relations

In all the examples above, the context of the second utterance was constituted by an utterance made by another speaker. Of course, this does not need to be the case. The meaning of an utterance can also depend on a non-linguistic context or it can be constituted by a former utterance made by the same speaker, as in (1.5) and in (1.6). This case will be of particular interest for this book.

(1.5) a. John insulted Max.
b. Max pushed him.

(1.6) a. John fell.
b. Max pushed him.

These examples are similar to the examples (1.2)–(1.4) in the sense that the respective second utterances are identical. And they are similar in the sense that the hearer (or reader) of the examples enrich these pair of utterances with additional meaning. The event which is described in (1.5.b) is interpreted as a *result* of the event described in (1.5.a). On the other hand, the same utterance in (1.6) is interpreted as an *explanation* for the event in (1.6.a). However, these example differ from the examples in (1.2)–(1.4), since the context sentence of the second utterance was an utterance made by the same speaker!

In order to determine the meaning and the intended function of an utterance it has to be related to the context in which it was made. When this context is constituted by a former utterance, as in (1.5) and (1.6), this relation is called a *rhetorical relation*. Rhetorical relations can basically be classified by the type of information which is inferred while relating the given utterances. For example (1.5) is an example of a *Result* relation. In a *Result* relation, the hearer infers that the state of affairs which is described in the first utterance causes the state of affairs that is described in the second utterance. Example (1.6) on the other hand is an *Explanation*. In an explanation, the event described in the first sentence was caused by the event in the second sentence and, hence, *explains* it.

There are many more possible relations between two utterances. The following list shows examples of some further rhetorical relations.

(1.7) a. To bake bread you need lots and lots of salt. *Elaboration*
b. Round about 20 grams for one pound of flour.

(1.8) a. Bill must have left. *Evidence*
b. His bike is gone.

(1.9) a. Max painted the wall. *Parallel*

b. Chris repaired the roof.

(1.10) a. Max fell. *Narration*
b. John helped him up.

These examples show again that the interpreter has to infer additional information in order to understand the actual meaning of the discourse. For instance, the temporal structure in all these examples is basically not overtly expressed but has to be inferred: In (1.10) it is quite clear that John helped Max up *after* Max fell, even though the temporal order of the events does not have to be reflected by the their order of the occurrence of the utterances, as was shown in (1.6). However, the ability to draw these additional inferences, and hence to construct rhetorical relations from the given material, seems to belong to the interpreters linguistic competences (Zeevat, 2011).

1.1.3 Relations between non-adjacent utterances

All examples above were cases where the relation holds between two adjacent utterances. However, an utterance in a discourse can also be related to an utterance in a discourse which was performed several utterances before, as shown in (1.11):

(1.11) a. John fell.
b. Lena pushed him.
c. Now his nose is broken.

In (1.11), the utterance in (1.11.b) is related to (1.11.a) by an *Explanation* but (1.11.c) is also related to (1.11.a) by a *Result* relation. In such cases, the interpreter has to look further back in the discourse to find the context sentence to which the new utterance is to be related to.

Cases of non-adjacent utterances which are related by a rhetorical relation make it somewhat complicated to talk about the utterances involved. Therefore, it will be convenient to have a name, at least for one of the utterances involved. In the following, the sentence S' to which a new sentence S is connected by a rhetorical relation will be called the *pivot* of S. In the case of (1.11), this means that the utterance (1.11.a) is a pivot for (1.11.b) and also a pivot for (1.11.c).

1.1.4 Coordination and Subordination

The set of all rhetorical relations can be subdivided into two subclasses: *coordinating* and *subordinating* relations. The distinction between these two types of relations is based on an the intuition that in a text or discourse some parts (utterances, sentences, paragraphs) are less important than other parts and that the parts of lesser importance depend on the more important parts (Asher and Vieu, 2005). In order to capture this intuition, one can call the less important parts of a discourse are *subordinated* to the more important parts. The following table shows how the relations which have been presented so far fit into the dichotomy of subordinating and coordinating relations:

Coordinating Relations:	Subordinating Relations:
• *Result*	• *Explanation*
• *Parallel*	• *Elaboration*
• *Narration*	• *Evidence*
• ...	• ...

As consequence of the fact that a discourse contains less important parts and more important parts, a discourse is ordered not only by the time of the occurrence of its utterances or sentences but also hierarchically ordered by the importance of the material presented. This additional hierarchical structure of a discourse has direct linguistic consequences and is relevant for the description of actual linguistic phenomena: It has been found that subordinating and coordinating structures behave differently with regard to the ability of the given discourse material to serve as pivot for new discourse material. This behaviour can be described as follows:

(1.12) Coordination, Subordination and Pivots: Let U_1 and U_2 be two temporally ordered adjacent utterances which are connected by a rhetorical relation R and let U be a new sentence. If R is a coordinating relation than only U_2 can serve as a pivot for U. If R is subordinating then both U_1 and U_2 can serve as a pivot for U.

In less technical terms this means that a coordinating discourse structure prevents new discourse material from relating to older discourse material whereas a subordinating discourse structure allows new material to relate to older utterances. Asher and Vieu, (2005) used the ability of older discourse material to serve as pivot for new discourse material as a test to decide whether a given rhetorical relation is subordinating or coordinating and even proposed to turn this property of subordinating rhetorical relations into a definition.

Another way to think about this behavior of subordinating relations is to say that discourse subordination works in some sense "against" the natural tendency of interlocutors to talk about the most recent material. This phenomenon is already illustrated in (1.11): (1.11.c) relates to (1.11.a) by a *Result* relation. According to (1.12), this is possible only because the preceding two utterances are related by an *Explanation*, which is a subordinating relation.

Since the notions of subordination and coordination describe a hierarchal structure of the set of utterances in a discourse, it is, moreover, now possible to represent this hierarchy in a tree-like structure as shown in figure 1.1. In this figure the horizontal arrows indicate coordination and the vertical arrows indicate subordination.

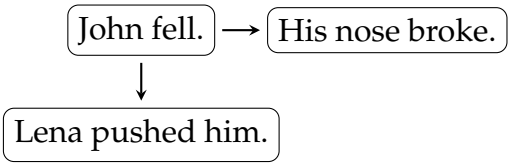
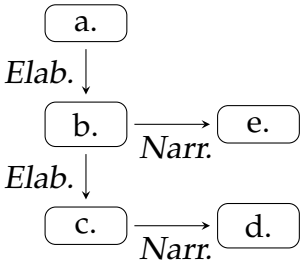


Figure 1.1: A discourse tree for (1.11).

Such tree-like structures also make it possible to represent larger segments of a discourse. The following example (from Asher and Lascarides, 2003) shows a larger discourse and the associated discourse tree.

(1.13)
a. Max had a great evening last night.
b. He had a great meal.
c. He ate salmon.
d. He devoured cheese.
e. He then won a dancing competition.

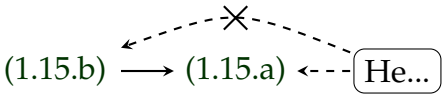


Such discourse trees provide a natural possibility to generalize the observation that different rhetorical relations have an impact on the discourse structure as formulated in (1.12): This generalization is called the *right frontier constraint* (RFC) (Polanyi, 1988; Jasinskaja and Karagjosova, 2015) and it will be one of the most essential target issues of this book:

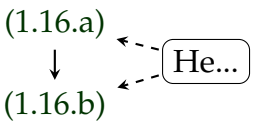
(1.14) Right Frontier Constraint: Only the nodes at the right frontier of the discourse graph are accessible for attachment of new discourse material.

If one again makes use of the additional assumption that non-local antecedents for third person pronouns can be only in the pivot of the utterance that contains the pronoun, one has another opportunity to actually see the impact of the RFC on discourse structure (Zeevat, 2011). As an explanatory example, consider (1.15) and (1.16). In (1.15), the first two utterances are coordinated by a *Parallel* relation which is coordinating. According to the RFC the pronoun in (1.15.c) can, therefore, only resolve to Chris, as can be seen by the associated discourse tree. In (1.16), on the other hand, the third person pronoun (1.16.c) can be resolved to both referents, Max and Chris, since both utterances are on the right frontier of the associated discourse tree.

(1.15)
a. Max$_m$ painted the wall.
b. Chris$_c$ repaired the roof.
c. He$_{*m/c}$ did a good job!



(1.16)
a. Max$_m$ painted the wall.
b. Chris$_c$ convinced him.
c. He$_{m/c}$ did a good job!



1.2 Goals of this Book

So far, we have seen that in a coherent discourse, utterances are related by rhetorical relations and that these relations can be subdivided into subordinating and coordinating rhetorical relations. Moreover, this distinction has a clear linguistic effect in terms of the ability of utterances in a discourse to serve as pivot for new utterances. However, up to now there are at least two open problems in this area: first, there is still no satisfactory intrinsic definition of subordinating rhetorical relations and, second, there is no satisfactory explanation for the right frontier constraint. The original goal of this work was to solve both problems, i.e. to develop a satisfactory definition of subordination and to show how the right frontier constraint can be derived on the basis of this definition.

Unfortunately, it was not possible to provide an intrinsic definition of subordination, that covers all known types of rhetorical subordination on the one hand, and provides an explanation for the RFC on the other hand. However, it was possible to develop a definition for a class of rhetorical relations which can be shown to be subordinating and for which it is possible to provide an explanation for the RFC. This class of rhetorical relations will be called *support* relations[1]. Support relations will be characterized by the property that they are means to achieve the goal of another utterance and, therefore, *support* that utterance. Support relations can be shown to have many properties of subordinating rhetorical relations and it is even possible to show why support relations are subject to the right frontier constraint. Moreover, this definition has the advantage that it is purely intrinsic, in the sense that one does not need to consider upcoming discourse to measure whether a given relation is a support relation.

Support relations are defined in terms of discourse participants goals and plans. This means that in order to provide a satisfactory definition of support relations, it might be a good idea to take a broader perspective, which includes not genuinely linguistic notions such as actions, goals of actions and plans that contain these actions. The idea behind the approach that is pursued in this book is the following: rhetorical relations are relations between utterances or speech acts. But speech acts are a particular type of actions. Actions, on the other hand, are normally performed in order to achieve some goal. Hence, in order to gain a better understanding of rhetorical relations, we first have to develop a better understanding of goals and actions, and, maybe even more importantly, a better understanding of how actions are perceived by an observer. In other words, in this book discourse is regarded as a special case of general human agency, which makes it possible to apply insight from general action theory to discourse and, hence, to obtain a better understanding of discourse structure. Therefore, the general idea for the work ahead can be summarized as follows: what happens if one integrates discourse structure and rhetorical relations into an action theoretical framework and what insights about these phenomena can be gained if one follows this approach!?

As can be seen in the course of this book, this approach turned out to be quite successful. It was not possible only to gain insights about rhetorical relations, but also to

[1] The term *Support Relations* is the result of joint work with Katja Jasinskaja.

provide explanations for other phenomena in communicative situations, as for example, the omnipresence of so called *acceptance moves* in oral conversation. Acceptance moves are rather short utterances, as for example *mm*, *aha* that are performed by the hearer of a speech act in order to signal that an utterance was accepted. Moreover, it is actually possible to show that support relations are not only a necessary tool for achieving a certain goal, but are actually the default case for rhetorical relations. In sum, on the basis of some action theoretical notions, it will be possible to achieve the following main results:

a. support relations as class of rhetorical relations will be defined,
b. it will be shown, that by default, one can assume that two utterances are related by a support relation,
c. an explanation for the necessity of acceptance moves in oral conversation will be provided,
d. using the notion of support relations, an explanation for the right frontier constraint can be provided.

1.3 Tools and Methods

In this book, ideas and tools from several scientific fields and areas will be used to solve a primarily linguistic problem. This section gives an overview of these different areas and emphasizes the connection between this book and other scientific traditions.

Discourse Representation Theory: Discourse representation theory (DRT) as it is presented by Kamp, Genabith, and Reyle, (2005), has proven to be extremely flexible as a modeling device for several linguistic phenomena. Moreover, it already provides a formal way to deal with propositional attitudes, the formal counterpart of emotions like desire or intentions, which will be used in order to define goals. Therefore, DRT is predestined to serve as the formal foundation of the theory in this book, and, hence, a great part of the formal work of this book will be formulated within discourse representation theory. DRT is not used as a device to model discourse content or discourse processing but as a way to represent discourse situations from the point of view of an omniscient observer of the given communicative situation who makes judgments about the speaker and the hearer of a discourse. In this regard, the theory which will be developed in this book is similar to an attempt by Poesio and Traum, (1997), who also represented actions, speech acts and belief states in DRT in order to obtain a more flexible notion of the of interlocutors common ground.

BDI Logic and Planning Theory: Since this book makes heavy use of propositional attitudes, it uses results and methods from *belief-desire-intention* logics (Meyer, Broersen, and Herzig, 2015). BDI logics deal with formalization and the interferences between propositional mental states and with the formalization of goals. Such logics have been applied in order to derive principles in communication by Cohen and J. Levesque, (1988), who showed that it is possible to derive the illocutionary force of a speech act by assumptions about the speaker's mental states. This book follows the tradition of these

authors, since they defined goals as configurations of certain mental states and derived communicative principles from the recognition of propositional attitudes. Grosz and Sidner, (1990) already formalized the assumption that any discourse segment is part of a plan of the speaker and that making sense of an utterance in its context is basically a plan recognition task for the hearer. Moreover, this book uses some results by Searle, (1983, 2005), who investigated the structure and the inferences of propositional attitudes and, in particular, of intentions from a more philosophical point of view.

Action Theory: Another basic notion which is used in this book is the notion of an *action*. Actions will be defined as particular events which are caused by certain belief-desire-intention combinations of the agent. The theory of actions which is developed and used in this book can therefore be classified as a *causal theory of actions* and, therefore, stands in the tradition of Davidson, (1980) and Bratman, (1987), who distinguished actions from mere motions of an agent based on particular propositional attitudes.

Speech Act Theory: Although this book does not adopt all claims and results of classical speech act theory in the sense of Austin, (1962), Searle, (1969), Searle and Vanderveken, (1985) and Grice, (1968), many of the core ideas of these theories will be integrated in this work, such as for example the idea that speech acts are *actions* and that some conditions have to hold in order for a speech act to be *sincere*. However, some (rather important) aspects of the work of Searle and Austin will not be used in this book. For example, hardly any use will be made of the notion of the illocutionary force of speech acts. The particular notion of the illocutionary force of a speech act will be replaced by the idea of Geurts, (2018) and Roberts, (2017), that the actual goal of any speech act is commitment sharing.

Dialogue Theory: The term *dialogue theory* refers to the work of Ginzburg, (2012), Clark and Brennan, (1991) and Clark, (1996), who investigated the mechanics of dialogue and in particular the mechanics of *grounding*, which basically is the question which conditions have to hold for interlocutors to agree upon propositions. Their work has been a great source of examples and inspiration. In particular Ginzburg represents whole communicative situations from the point of view of an omniscient observer. This book adopts this idea and extends it by also including interlocutors' propositional attitudes into the representation. Another important aspect which is exemplified by these authors is the fact that communication is not a one way road. Both agents, speaker and hearer, have to do their respective part in order to achieve a communicative goal. This book emphasizes this aspect by placing particular focus on the role of the hearer in a conversation.

Theories of Discourse Structure: *Discourse structure* refers to the structure of texts or dialogues beyond the sentence level which is constituted by rhetorical relations. Rhetorical relations have been investigated since the work of Hobbs, (1979) who also investigated the impact of such relations on coreference. *Rhetorical structure theory* (RST) by Mann and Thompson, (1988) was an early attempt to represent the structure of a whole text. Polanyi, (1988) finally established the notions of *subordinated* and *coordinated* discourse units, used these notions for establishing discourse trees, and formulated the right frontier constraint. Asher and Lascarides, (2003) finally formulated

an update semantics which takes discourse structure into account. This book builds upon the results of all these approaches to discourse structure. In particular the notion of *presentational relations* from RST will be of interest, since these relations can be regarded as a predecessor for the class of subordinating relations.

Prominence Theory: In recent years the notion of *prominence in language* has been proposed as a general organization principle of language (Himmelmann and Primus, 2013; Jasinskaja, Chiriacescu, et al., 2015; von Heusinger and Schumacher, 2019). Prominence is characterized as a) being relational, b) being dynamic and c) attracting linguistic operations. In this book, the notion of prominence will be applied to speech acts. A speech act is prominent at a given point of a conversation if its communicative goal is not yet achieved at that point. Using the notion of prominence of a speech act, one can derive the effect of rhetorical relations on discourse structure on the basis of the assumption that prominent speech acts are attractors for further linguistic operations.

1.4 Overview of the Individual Chapters

This book is organized as follows. The chapters 2–4 develop the conceptual apparatus needed to address the problem: mental states, actions in general, and speech acts in particular. Chapters 5 and 6 use this conceptual apparatus to address the central problem of this book, i.e. to explain the difference between coordinating and subordinating rhetorical relations and, moreover, to explain how this opposition affects prominence in discourse.

Chapter 2: Propositional attitudes: The first building block of the theory developed in this book are propositional attitudes (belief, desire, intention) which represent mental states of agents in communication. The chapter gives a brief introduction to the DRT formalism and the treatment of propositional attitudes in DRT. Propositional attitudes play a central role in defining the notion *goal* in chapter 3. Moreover, this chapter introduces the common sense principle of inertia, which is central for describing a human agent's assumptions about changes of states of affairs in the world and, therefore, also for describing a speaker's assumption on the mental states of a hearer. Inertia, therefore, plays a central role for predicting default interpretations for utterance pairs in chapter 5.

Chapter 3: Actions, goals and plans: Discourses consist of speech acts and speech acts are instances of human actions in a more general sense. Chapter 3 introduces a number of notions of a general theory of action, which will be instantiated in relation to speech acts in later chapters. These notions include the notion of goal, goal-directed act, the mental state of pursuing a goal, the relationship between goals and subgoals and how these are organized in plans, and the principle of commitment to a goal. All these concepts and principles constitute building blocks for reasoning about speech acts and larger discourses, as special cases of actions and plans in a more general sense.

Chapter 4: From acts to speech acts: The focus of chapter 4 lies on speech acts as a special type of acts. I argue that the goal of communication in general and of individual speech

acts in particular is to negotiate commitments to act. That is, to reach common ground means to reach consensus on a future course of the speaker's and the hearers' action. Speech acts of all types can be analyzed in this way, even assertions, the goal of which is to ensure that the participants in communication behave as if something was the case. An additional principle of sincerity ensures that this behavior is backed up by consistent beliefs.

However, common ground can be reached only if the hearer accepts the speaker's proposal to share a commitment. The chapter discusses how acceptance moves in dialogue lead to common ground updates.

Finally, presuppositions are treated along the same lines, namely as acts of presupposing. Presuppositions play an important role in building up connected discourse, especially one particular kind of rhetorical relations in discourse – coordinating relations, discussed in chapter 5.

Chapter 5: From speech acts to rhetorical relations: The goal of chapter 5 is to explain the difference between subordinating and coordinating rhetorical relations in terms of the speech-act theoretic notions developed in previous chapters (goals, plans, common ground updates).

First the problem and the existing approaches to it, in particular the principle known as the Right Frontier Constraint, are introduced in more detail. Then the notions of coordination and subordination in discourse are reanalyzed at the level of communicative goals of respective speech acts. On the basis of the notions of communicative goals and subgoals, a class of rhetorical relations – so called *support relations* – is defined which can be shown to be subordinating. The general action-theoretic principles developed in chapter 4 are used to derive a theorem that predicts that supporting relations (such as *Elaboration* and *Explanation*) will be established by default. It is then shown how discourse markers such as *also* that introduce presuppositions can override this default, giving rise to coordinating rhetorical relations.

Finally, the notion of discourse structural subordination is considered from the perspective of dialogue interaction. It is shown how lack of acceptance from the hearer and the resulting failure to update the common ground gives rise to subordination in discourse.

Chapter 6: Prominence and the Right Frontier: The goal of Chapter 6 is to integrate the the notions of goals of speech acts into the prominence framework. It is assumed that a speech act the goal of which is not achieved yet is open for attachment of new discourse material, either because it still has to be supported or because it has to be signalled that the goal of the speech act is achieved. A referent that occurs in a speech act the goal of which is not achieved yet is therefore more likely to be picked up in the upcoming discourse and is, therefore, more prominent. Hence, being a referent in a speech act the goal of which is not achieved yet is a prominence lending factor.

The chapter finally presents a rule for constructing discourse trees from a given set of speech acts, defines the right frontier of discourse trees and shows that the elements on the right frontier of a discourse tree correspond to the speech acts of the discourse the goal of which are not achieved yet. Therefore, this chapter shows that being an element of the right frontier of a discourse is a prominence lending factor and, hence,

draws a connection between the notion of discourse prominence to the notions of discourse trees and the right frontier.

Chapter 2

Propositional Attitudes

The basic building blocks of the linguistic theory which is going to be developed in this book are speech acts. Speech acts, however, are performed by speakers, who have (at least in most cases) a particular reason to perform a speech act, i.e. the speaker wants her speech act to have an impact on the course of things in some way: she may want to persuade her hearer, make him do something, make him regret something, make him believe something and so on. In many cases, the speaker wants to change the world by affecting a particular hearer, but there are also cases in which she even wants to change the world in a way that is, in some sense, independent of her hearers. When someone baptizes a ship, it is her goal to make the ship carry its given name, independent of the current audience. However, all these cases have in common that the speech acts are performed with a certain intention and to achieve some *goal*. In this book, it will therefore be assumed that any speech act is performed because the speaker wants to achieve some goal by means of the performance of the speech act. This assumption will be used in order to connect speech acts and the linguistic material they contain to the mental states of the speaker, and, hence, to embed individual speech acts of the speaker into a communicative plan, on the other hand, makes it possible to connect individual speech acts and to turn a set of speech acts into a coherent discourse. This means, that for constructing a coherent discourse from individual speech acts, the language perceiver has to make assumptions about the goals and mental states of the speaker.

In order to be able to put a speech act into its context and, hence, in order to construct a coherent discourse from a sequence of incoming speech acts, the hearer has to understand *why* the speaker produces this particular speech act at at this point and which goals the speaker tries is pursuing with it. But understanding why a speech act was performed means, that the hearer, who tries to construct a coherent discourse from the incoming utterances, has to make guesses about the speaker's *goals* and *plans*. A plan, however, can be regarded as a certain combination of propositional attitudes, such as desires, intentions and beliefs: the plan holder *desires* that some state of affairs hold, *believes* that some actions bring about that state and therefore, *intends* to perform these actions. The hearer, who wants to construct a coherent discourse, there-

fore, has to form beliefs about propositional attitudes, in particular about the speaker's *beliefs, desires, intentions* and, hence, her mental states.

By talking about things like mental states and propositional attitudes, this book enters the realm of epistemic logic. In epistemic logics, it is common practice to refer to the holder of a propositional attitude as the *agent*. This practice will be adopted in this book: subjects which are capable of having propositional attitudes and which are able to act are referred to as agents. There is a risk of confusing this term with the term *agent* as a semantic role, but since thematic relations will hardly play a role in the course of this book, this risk seems to be negligible.

As a consequence of the assumption that the construction of a coherent discourse is basically a plan recognition task, propositional attitudes play at least two different roles in this reconstruction process:

1. they are used to model the hearer's *assumptions* (i.e. beliefs) about the mental states of a speaker and
2. they are used to model the speaker's *goals* and *plans* (i.e. belief-desire-intention combinations)

Chapter 3 will investigate on plans and goals of agents and will define these notions in terms of the propositional attitudes. Since this book is interested not only in the development of a theory of discourse coherence on the basis of recognition of communicative plans but also in formalizing this theory and the arguments which are developed within this theory, this chapter is used for investigating propositional attitudes and for developing the formal framework in which this formalization will take place[1]. The formal part of this book will be implemented in Discourse Representation Theory (DRT) (Kamp, Genabith, and Reyle, 2005),. DRT was originally designed to model an agent's processing of linguistic input or to model the development of the common ground of a conversation, but has proved to be extremely flexible as a modeling device for all kinds of different linguistic phenomena. In this book, DRT will be used for modeling communicative situations from the point of view of an omniscient observer. This makes the theory of this book similar to Ginzburg's dialogue game boards, where communicative situations are represented from such a perspective (Ginzburg, 2012). In contrast to Ginzburg's framework, where utterances are the basic building blocks, this book uses assumptions about the mental states of agents as the basic blocks on which it develops its theory.

In order to be able to "calculate" with mental states, some rules or principles will be needed which enable the theory to model how agents draw conclusions about states of affairs in the world. As a toy example, consider an agent who sees a wet street and who used to know that this street was dry several hours ago. One can assume that the agent is able to draw some conclusions on the basis of this visual input. Maybe she will come to the conclusion that the street is wet because of rain. But how can this conclusion be modeled in terms of propositional attitudes? In this case, one needs a rule which models the basic assumption that "changes in the world" have "reasons":

[1]Since this book deals with formal properties of mental states, it follows the tradition of logics and beliefs (Hintikka, 1962), logics of intentions (Cohen and Levesque, 1990) and *belief-desire-intention* logics (Meyer, Broersen, and Herzig, 2015).

the agent, who believes that at some time point T, some state of affairs K was the case, and at some future time point T' believes that K is not the case any more, will come to the conclusion that there was something which caused K to change. In this book, such basic rules will be implemented and will enable the theory to model how agents draw conclusions from states of affairs in the world. Because these rules model phenomena which can be thought of as cognitive principles and because they will have an axiomatic character for the proposed theory, they will be referred to as *principles*.

These principles are supposed to model parts of a rational agent's cognition. However, a great deal of this work will be concerned with modeling an agent's assumptions about *another* agent's communicative plans and, hence, the agent's mental states. This means that there will be situations in which an agent's assumptions about another agent's mental states and his assumptions about states of affairs in the world have to be modeled. In order to do this, it will be assumed that the agent who draws assumptions about another agent's mental states believes that this agent's cognition is subject to the same cognitive principles. Theoretically speaking, this means that it will be assumed that 1. all agents under consideration are subject to the rules which will be proposed in this book and, moreover, that 2. all agents assume that respective other agents are subject to these rules. In non-theoretic terms, these (admittedly, rather strong) assumptions can be interpreted as follows: human interaction is based on the assumption that the cognition of all participants basically works in a similar way. If one cannot rely on the assumption that another agent, given similar information, will draw similar conclusions, the involved agents are not able to anticipate any further mental states of the other agents, and, moreover, they are not able to anticipate further actions. Under such pre-conditions social interaction would hardly be possible. Coming back to a theoretical level, on can say that these assumptions enable the theory to "control" an agent's mental state. For example, the observer and the involved agents can rely on the fact that mental states do not change randomly, which makes it possible for an agent to make guesses about another agent's mental states.

In sum, this chapter pursues three main goals: i) giving a short introduction to DRT, ii) discussing and formalizing some properties of propositional attitudes within DRT and iii) discussing and formalizing some cognitive principles which are going to be used in this book. This will be done in this chapter as follows: In section 2.1 a brief overview of the syntax of Discourse Representation Theory will be given. Moreover this section introduces temporal referents, since these will be of particular interest for a proper treatment of states (as for example mental states) and events (as for example speech acts). Section 2.2 introduces propositional attitudes. It will be discussed how they and their formal properties can be formalized in DRT. After having discussed how mental states can be formally presented, section 2.3 introduces some basic cognitive principles which are needed in order to perform calculations with mental states of agents. Section 2.4, finally, introduces two further basic common sense principles of cognition: *inertia* and *learning*. Both principles model the assumption that rational agents do not change their mind at will, but normally need a particular reason for such a change. In the section 2.5 we apply these principles to a situation of social interaction, in order to show how basic calculations can be performed using propositional attitudes and the principles which have been introduced so far.

2.1 Introduction to Discourse Representation Theory

This section is an introduction to the syntax of *discourse representation theory* (DRT). DRT will be used as a framework for most of the formalizations in this book. This section can only serve as a brief introduction to this theory. In particular, we will not deal with the semantics and the model theory but will just sketch the syntax. For a general introduction to DRT, its syntax and its semantics, see Kamp, Genabith, and Reyle, (2005) or Kamp and Reyle, (1993).

Originally, DRT was designed to model the processing of a new incoming utterance by a language perceiver and, in particular, to solve some problems of former formal approaches to natural language which came up in the context of non-local resolutions of referential expressions. But DRT also developed some very good theoretical tool to deal with other aspects of natural language, as for example temporal referents or sentences that contain claims about the mental states of agents. Especially these aspects of DRT will be of great importance for this book.

2.1.1 Syntax of Discourse Representation Theory

Discourse representation theory represents the content of a discourse in a so called *discourse representation system* (DRS). A DRS contains some variables which correspond to the referents of the discourse, i.e. the entities the discourse is about. These variables are also usually called *discourse referents*. Moreover, it contains some formulas which represent the information that has been given about the discourse referents. These formulas are called *conditions*. Formally speaking, a DRS is a pair (U, C), where U is the set of discourse referents , and where C is a set of DRS-*conditions*. DRS-conditions can consist of "simple" predications but also of logical operators, such as $\neg$, $\vee$, $\wedge$ and $\Rightarrow$. The predicates and the logical operators in a DRS represent the predications in a natural language sentence. As usual it is assumed that the natural language connectors *not, or, and* and *if..., then...* translate into the logical symbols $\neg$, $\vee$, $\wedge$ and $\Rightarrow$. The two sets U and C of a DRS are often represented in a horizontally divided box the upper part of which box contains the set of discourse referents and the lower part of which contains the set of conditions. The set of discourse referents is also called the *universe* of the DRS.

The rest of this section will be used to show how DRSs can be read and how a DRS can be constructed. This will be done on the basis of the following example:

(2.1) a. Pedro owns a donkey. He likes it.

b.

$x\ y\ z\ w$
$\mathsf{name}(\mathrm{pedro}, x)$
$\mathsf{donkey}(y)$
$\mathsf{own}(x, y)$
$w = x$
$y = z$
$\mathsf{like}(w, z)$

For every referential expression of a discourse, DRT introduces a new referent. Therefore, the upper part of the DRS in (2.1.b) (its universe) contains four different variables which correspond to the four referential expressions in the discourse in (2.1.a), i.e. they correspond to the expressions *Pedro*, *a donkey*, *he* and *it*. However, some of these referential expressions refer to the same entities: *he* refers to *Pedro* and *it* refers to *a donkey*. This equality is captured in the lower part of the DRS by the condition that the variables which are introduced for the referential expressions that refer to the same entities are equal. The DRS in (2.1.b), therefore, contains the conditions $w = x$ and the conditions $y = z$. The remaining conditions in this DRS are predicates that correspond to natural language predicates, i.e. in this case to *own* and to *like*.

But how can this DRS be constructed? (2.1.a) consists of two different sentences, and, as already explained, DRT was originally designed to model how new utterances can be integrated into the given discourse. So, how does DRT add the content of new utterances into an already existing DRS? Suppose, that there is already a DRS as in (2.2) which represents the content of the first utterance in (2.1.a).

(2.2)

$x\ y$
$\mathsf{name}(\text{pedro}, x)$
$\mathsf{donkey}(y)$
$\mathsf{own}(x, y)$

In some sense, DRT tries to imitate the cognitive processes of a hearer in order to add the the new information to the already existing DRS. For example, a human hearer uses information to calculate identity inferences and, hence, is able connect two independent sentences. In the case of the second sentence of (2.1.a), the referential expressions *he* and *it* convey some cues which can be used by the hearer to find the referent to which the speaker is referring: the pronoun *he* conveys the information that it refers to some human male referent, whearas *it* must refer to some non-human referent. The information that the referent who *likes* "it" is male and human is presupposed information (in the sense that it for example survives a negation, cp. Simons, 2013 for a discussion of presupposed information), and this presupposed information is essential for the hearer's successful resolution of the pronoun. DRT tries to model this resolution process as follows: A sentence which contains presupposed information is represented in DRT as a pair of DRSs, where one DRS contains the presupposed material while the other one contains the non-presupposed information. Such a DRS pair for the second utterance in (2.1.a) may look as follows:

(2.3)

$\Big\langle$ | $w\ z$ | / $\mathsf{human}(w)$ / $\mathsf{male}(w)$ / $\mathsf{non\text{-}human}(z)$ |, | $w\ z$ | / $\mathsf{like}(w, z)$ | $\Big\rangle$

The presuppositional part in this pair of DRSs now contains enough information to find the right referents for a proper resolution of the new variables: a donkey is the only non-human referent in the previous discourse and Pedro is the only male and human

referent in the previous discourse (this information of course has to be accomodated). Therefore, the presuppositional part of that pair of DRSs can now be resolved. Afterwards the actual information which is conveyed in the new sentence can be added to the previous DRS. These considerations can be summarized in the following calculation:

(2.4)
$$\begin{array}{|l|}\hline x\ y\ w\ z \\ \hline \mathsf{name}(\text{pedro}, x) \\ \mathsf{donkey}(y) \\ \mathsf{own}(x,y) \\ x = w, y = z \\ \hline \end{array} \uplus \begin{array}{|l|}\hline w\ z \\ \hline \mathsf{like}(w,z) \\ \hline \end{array} = \begin{array}{|l|}\hline x\ y\ z\ w \\ \hline \mathsf{name}(\text{pedro}, x) \\ \mathsf{donkey}(y) \\ \mathsf{own}(x,y) \\ w = x \\ y = z \\ \mathsf{like}(w,z) \\ \hline \end{array}$$

In the example above, all information had to be interpreted *globally*, i.e. as information about the actual world. However, natural language often introduces referents or propositions which only have a local meaning, i.e. a meaning within the scope of some operator. Standard examples for local interpretations in natural language are propositions that live in the scope of the antecendent of a conditional or a modal psych verb like *believe* or *imagine*. Take for example the proposition *Pedro has a donkey* in the following two examples:

(2.5) a. If Pedro has a donkey, he likes it.
b. Pedro believes that he has a donkey and that he likes it.

The important point in both examples is that the truth of the statement that Pedro has a donkey is neither presupposed nor is it necessary for either of these sentences to be true, to be able to have a truth value or even to be interpretable. In both cases, the clause which refers to the proposition *Pedro has a donkey* is introduced in the scope of a (syntactic) operator or of a modal expression which prevents it from being interpreted as an actual true statement about the world. In such cases one also says that the truth of such propositions does not *project* out of the scope of these operators (cp. Benz and Jasinskaja, (2017)). Operators which have the property of being able to cancel the truth of embedded propositions and the global existence of embedded referents are called *entailment cancelling operators*. In DRT such propositions are represented as a further DRS which also lies in the scope of a predicate which represents the operator that embeds the propositions. For example, a representation of the statement in (2.5.a) as a DRS may look as follows:

(2.6)
$$\begin{array}{|l|}\hline x \\ \hline \mathsf{name}(\text{pedro}, x) \\ \begin{array}{|l|}\hline y \\ \hline \mathsf{donkey}(y) \\ \mathsf{own}(x,y) \\ \hline \end{array} \Rightarrow \begin{array}{|l|}\hline w\ z \\ \hline \mathsf{like}(w,z) \\ w = x \\ y = z \\ \hline \end{array} \\ \hline \end{array}$$

As one can see in the box in (2.6), the expression *a donkey* introduces a referent in the universe of an embedded DRS which is the antecedent of a conditional. The proposition that Pedro likes that donkey is represented as a DRS in the consequent of that conditional. The only information which is contained in the outermost DRS is the information that there is somebody named Pedro.

After this informal overview of the basic properties of DRT, (2.7) presents the basic rules for the construction of a DRS in a more formal fashion.

(2.7) SYNTAX OF DISCOURSE REPRESENTATION THEORY: A *discourse representation system* (DRS) is a pair $K = (U_K, C_K)$ such that

1. U_K is a set (of discourse referents)
2. C_K is a set of DRS conditions. A DRS condition can have the following form: let K, K_1 and K_2 be DRSs, $x_1, ..., x_n$ be discourse referents, and P an n-place predicate, then $P(x_1, ..., x_n)$, $\neg K$ and $K_1 \bullet K_2$ are DRS-conditions, where $\bullet \in \{\vee, \Rightarrow\}$.
3. If $K_1 = (U_1, C_1)$ and $K_2 = (U_2, C_2)$ are DRSs, then $K_1 \uplus K_2 := (U_1 \cup U_2, C_1 \cup C_2)$ is a DRS.

In the proceedings of this book, the following typographical conventions will be used when it comes to formalizations in DRT: "normal" predicates, i.e. predicates that represent natural language predicates (such as for example $\mathsf{like}(x, y)$ for *he likes it*) will be written in a sans serif font. The discourse referents, on the other hand, will be written in *italics*. In the following chapters, some 'special' predicates, such as *events, speech acts* or *goals* will also be of interest. These will be written in CAPITALS.

2.1.2 Time, Events, and States

Natural language does not only provide information about physical referents (like a guy named *Pedro* or donkeys) but also provides information about temporal order. Every event must take place at some point in time and this information is also encoded in natural language, either implicitly or explicitly. For example, a sentence like *Pedro beats the donkey*, does not only talk about Pedro, a donkey and an abstract action of beating, but introduces a concrete beating event. The beating event is, in fact, so concrete that it can be picked up by a pronoun as in the sentence *It was painful*. Therefore, DRT actually provides a way to treat events in a similar fashion as any other physical referent.

In this book, this feature of DRT will mainly be used for the representation of events and states in communicative situations, such as for example mental states of agents or certain events such as speech acts. This section, however, introduces temporal referents such as events and states on the basis of natural language examples.

Things that happen in the world are somehow temporally ordered and natural languages have various ways to report temporal orders. Take for example (2.8):

(2.8) Max pushed John. Then John fell.

(2.8) does not only introduce two referents *Max* and *John* but also two events: a

push(Max,John) event and a fall(John) event. Moreover, the two sentences contain the information that the push event precedes the fall event. In order to be able to capture temporal information DRT introduces special temporal referents for time points, events and states. Using these referents, the temporal structure of (2.8) can be presented as follows:

(2.9)

$m\ i\ j\ E\ E'\ T\ T'$
$\mathsf{name}(\text{Max}, m)$
$\mathsf{name}(\text{John}, i)$
$E : \mathsf{push}(m, i)$
$E \subset T$
$\mathsf{name}(\text{John}, j)$
$E' : \mathsf{fell}(j)$
$i = j$
$E' \subset T'$
$T \prec T'$

The two sentences in (2.9) relate two events. However, there is a common assumption that languages do not only introduce events as temporal referents but also as *states*. The difference between states and events is based on the following intuition: Events happen in some way 'quickly', they are limited in time and have a clear beginning and a clear end. States, on the other hand, are in some sense extended in time, do not necessarily have a beginning or an end and are homogenous in the sense that any 'subpart' of a state can be described by the same predicates as the state itself. In that sense the difference between events and state can be compared to the difference between *things* and *stuff*, i.e. the objects count nouns refer to and mass nouns are referring to (Lambalgen and Hamm, 2005). Things can be individuated whereas stuff is, in some sense, homogeneous. Example (2.10) contains both, an event an a state:

(2.10) The sun was burning on the desert when John jumped on the back of his horse.

The description that the sun was burning on the desert in (2.10) is a description of a *state*. The sun burning on the desert exactly fits the intuition for states in the sense that it is homogenous and has no clear endpoints. John's jumping on the horse, on the other hand, is clearly an event which is located *in* this state. The box in (2.11) shows a possible representation of this situation:

(2.11)

$s\ d\ S\ T\ j\ h\ E$
$\mathsf{sun}(s), \mathsf{dessert}(d)$
$S : \mathsf{burn.on}(s, d)$
$T \subset S$
$\mathsf{name}(\text{John}, j)$
$\mathsf{horse}(h)$
$E : \mathsf{jump.on}(j, h)$
$E \subset T$

Even though, there is a clear intuition about what should count as events and what should count as states, language has means to "turn" events into states and vice versa. Take for example (2.12):

(2.12) Max pulled his gun while John was jumping on his horse.

In this example, John's jumping on his horse is told in the past progressive tense which gives it the character of an action which is still unfinished and in progression. In a proper presentation of this situation, this action should therefore be presented as a state. However, since this book basically does not use states and events in order to formalize problems of tense and aspect, but just to be able to locate special temporal referents (such as speech acts or mental states), a further discussion of such phenomena will be omitted here.

In this book, events and states will be treated as primitive objects. For a construction of a time structure on the basis events and events structures where events build the only primitive objects, consider Lambalgen and Hamm, (2005). However, there will be a need for some relations with which it is possible to relate temporal referents. The boxes in (2.9) and (2.11) already contain some relations for temporal referents: an "is contained in" relation $\subset$ and a "preceedes" relation $\prec$. The following shows all temporal relations that will be used in this book:

(2.13) RELATIONS FOR TEMPORAL REFERENTS: Let E be an event, S a state and let T and T' be time points. We define the following relations for temporal referents:

1. $E \subset T$ iff E takes place in T.
2. $T \prec T'$ iff T precedes T'.
3. $T \subset S$ iff the situation described by S holds at T.

Usually, events are located by the condition that they are contained in a time point or interval whereas states are located by the condition that they contain a time point or an intervall. However, it will turn out to be convenient to use notions for comparing states and events, for example when one wants to describe a situation where an event takes place within a state, when one wants to talk about the overlap of to states or want to describe a situation in which a state is completely located within another state. This can be done by using the function

$$|\cdot| : \text{STATEs} \to \mathcal{P}(\text{TIME}), \quad S \mapsto |S| := \bigcup\{T \mid T \subset S\}.$$

which maps from states to sets of time points by mapping a state to the set of time i intervals which are completely contained in that state. Given this function, we can define the following additional relations for comparing events and states:

(2.14) MORE RELATIONS FOR TEMPORAL REFERENTS: For T, T' time points, E, E' events and S and S' states, define the following notational conventions:

1. $S \subset S'$ iff $|S| \subset |S'|$
2. $S \bullet S'$ for $|S| \bullet |S'|$ and for $\bullet \in \{\cup, \cap\}$
3. $E \subset S$ iff $\exists T : (E \subset T) \wedge (T \subset S)$.
4. $E \prec E'$ iff $\exists T, T' :$ TIME such that $(E \subset T) \wedge (E' \subset T') \wedge (T \prec T')$

Note that neither $S \cup S'$ nor $S \cap S'$ in 2. of (2.14) is a state on its own but just a set of time intervals. However, in section 2.2 this notation will be used in order to model configurations of propositional attitudes of an attitude holder.

It was already stated that states have the property of being homogenous in the sense that any subpart of the state has the same properties as the state itself. This can be compared to *stuff* such as water or mud. Stuff, in this sense, also has the feature that whatever subpart one takes from it, it has the same properties as the whole, i.e. stuff (as for example water) is homogeneous. For theoretical reasons, states shall have the same properties, i.e. a state should have no "holes": if the sun is burning on the desert, there should be no time point where it isn't burning on the desert. This persistence property of states sounds almost tautological and has basically technical rather than philosophical advantages. However, it is introduced as a requirement for states and can be formulated as follows:

(2.15) STATES ARE PERSISTENT: Let T, T', T'' be time points and let S be a state. If $T \prec T' \prec T''$ and $T, T'' \subset S$ hold, then $T' \subset S$ holds, too.

Summary of this section: So far, some of the basic features of discourse representation theory have been introduced, in particular its syntax and some special referents for dealing with temporal phenomena. The next section introduces mental states of agents which will serve as basic building blocks for the formal work in later chapters.

2.2 Propositional Attitudes and their Properties

This section is concerned with a formalization of mental states in DRT. Many natural language sentences contain epistemic modal verbs such as *think* or *want* which also have to be represented in a DRS in some way. Epistemic modal verbs have some properties which make them special concerning their representation in a DRS. For example, they usually take a whole clause as a complement, i.e. something that denotes a whole proposition which has to be represented as a DRS on its own. In order to be able to represent them, DRT introduces so-called *attitude description sets* (ADSs) . An epistemic modal verb together with its complement forms a propositional attitude which can be represented in DRS. But since human agents usually have several propositional attitudes at once, these propositional attitudes can be collected in an attitude description set which finally may serve as a DRS condition (Kamp, 2015; Maier, 2017). This feature of DRT is of interest for this book since it allows a proper formalization of content of the interlocutors' "minds".

This section is structured as follows. 2.2.1 introduces attitude description sets as the

basic tool for treating epistemic modal verbs. 2.2.2 deals with *parasitism*. Parasitism is a feature of epistemic modal verbs which can be dealt with by ADSs. Parasitism describes the possibility that referents which are introduced in the scope of one epistemic modal verb occur also in the scope of another epistemic modal verb, event though epistemic modal verbs usually do not allow their complements to project. Finally, this section introduces some notational conventions which will turn out to be convenient in the proceedings of this book.

2.2.1 Attitude Description Sets

As already explained in the introduction of this section, attitude description sets were originally invented for modeling natural language sentences which contain epistemic modal verbs (such as for example *believe*, *think* or *want*) within DRT. In the following, it is assumed that such epistemic modal verbs denote *mode operators*, (in this case, a *belief* operator and a *desire* operator), and together with the denotation of their complement form so called *propositional attitudes*. These propositional attitudes can be collected in an attitude description set. Attitude description sets are designed to represent the mental states of an agent, i.e. an agent's beliefs, desires or intentions. Take for example the following sentence:

(2.16) John thinks that he will catch a fish.

The verb *think* in (2.16) refers to a mode operator *believe*, and its complement *that he will catch a fish* refes to the proposition $\mathsf{catch}(J, f)$ (where J stands for *John* and f for *a fish*). Together, they denote a *propositional attitude*, i.e. an attitude towards a certain proposition. Formally speaking, a propositional attitude as a part of a DRS is presented as a pair (MOD, K) where K is a DRS which represents the complement of the proposition and where MOD is a *mode indicator*, i.e. an indicator which represents the given attitude, such as INT for intention, BEL for believe, DES for desire etc. Using this notation, the propositional attitude which is presented in (2.16) can be represented as $(\text{BEL}, \mathsf{catch}(J, f))$. In the following, the proposition K in such a pair is called the *propositional content* of the attitude. An *attitude description set*, as a condition of a DRS, is simply a set which consists of such propositional attitudes.

(2.17) ATTITUDE DESCRIPTION SETS:
- a. A propositional attitude is a pair (MOD, K) where MOD is mode indicator and K a DRS which represents the content of the attitude.
- b. An attitude description set is a set of propositional attitudes.

The sentence given in (2.16) contains just one epistemic modal verb (*think*). An attitude description set which represents this sentence, therefore, contains only one element. (2.18) shows a possible ADS for describing the situation given in (2.16):

(2.18)
$$\left\{ \left(\text{BEL}, \begin{array}{|l} \hline E_1\ F \\ \hline T \prec E_1 \\ \mathsf{fish}(F) \\ E_1 : \mathsf{catch}(J,F) \end{array} \right) \right\}$$

The set in (2.18) represents what is described as John's mental state in (2.16). However, this presentation isn't a DRS yet (since it, for example, does not have a universe) and, moreover, lacks some basic information such as temporal information or, more importantly, information about the belief holder. In order to present all these information in a DRS, ADSs are usually used as an argument in a predicate *Att* which takes an agent, the ADS that describes the mental states of that agent holder and a *state* (which represents the fact that the agent is in the described mental state) as arguments. If A stands for some agent and $\mathbb{K}$ for an ADS, then a condition with *Att* as a part of a DRS looks as follows:

(2.19) $S : Att(A, \mathbb{K})$

With the help of this condition, it is now possible to give a complete presentation of (2.16), as shown in (2.20). This DRS now contains information about the holder of the attitude and the time in which the agent is in the described mental state (given the assumption that T is the time point of the utterance of (2.16)).

(2.20)
$$\begin{array}{|l} \hline S\ J \\ \hline T \subset S,\ \mathsf{name}(J, \text{John}), \\ S : Att \left(J, \left\{ \left(\text{BEL}, \begin{array}{|l} \hline E_1\ F \\ \hline T \prec E_1 \\ \mathsf{fish}(F) \\ E_1 : \mathsf{catch}(J,F) \end{array} \right) \right\} \right) \\ \hline \end{array}$$

Example (2.16) contains just one epistemic modal verb and, therefore, the attitude description set which describes John's mental state is a singleton set which contains just one element as well.

As already pointed out, an attitude condition is usually written as $S : Att(A, \mathbb{K})$, where S is a state, A an agent and $\mathbb{K}$ an ADS. However, in cases where the ADS is a singleton set (as in in (2.20)) we can use the following notational convention:

(2.21) NOTATIONAL CONVENTION: Let $\text{MOD} \in \{\text{BEL}, \text{DES}, \text{INT}\}$. Then
$$S : \text{MOD}(A, K) \ :\Leftrightarrow\ S : Att(A, (\text{MOD}, K))$$

This notation will turn out to be helpful when working with nested propositional attitudes, such as for example an agent's assumptions on another agent's beliefs. However, descriptions of mental states of agents, and of course the actual mental states of agents may contain several propositional attitudes. In these cases, the attitude description sets which describe them also contain more than one propositional attitude.

However, even in such cases it will be convenient to use the notation above. How this can be done in a well-defined way will be the subject of §2.2.2.

2.2.2 Parasitism in Attitude Description Sets

So far a natural language sentence that contained an epistemic modal verb was discussed and it was shown how this can be represented in a DRS. The example was simple in the sense that it contained just one verb. However, mental states of agents and descriptions of mental states of agents may contain several propositional attitudes:

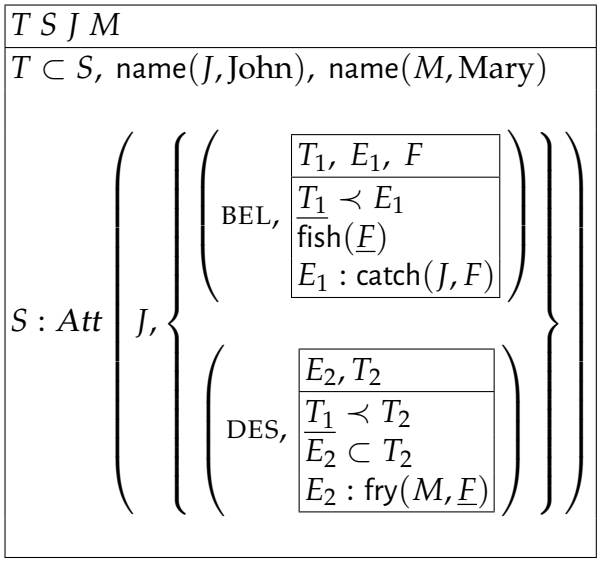
(2.22) a. John thinks he will catch a fish and he wants Mary to fry it. (Kamp, 2015)
b. Hans wants the ghost in his attic to be quiet tonight. (Maier, 2017)

Both examples in (2.22) describe mental states of an agent, which contain at least two propositional attitudes, namely a belief and a desire state. That the agent in (2.22.a) is in a certain belief state is overtly expressed by the verb *think*. The belief state of the agent in (2.22.b), on the other hand, is left implicit. Nevertheless, the fact that Hans is in a certain belief state is implied by (2.22.b), since Hans' desire that some particular ghost be quiet depends on his belief that this particular ghost exists. Both of these examples present not only a belief state but also a desire state. Furthermore, both examples have in common the referent which occurs in the scope of the desire operator depends on the referent which is introduced in the scope of the belief operator. On a syntactic level, this can be seen by observing that in both cases the referents in the scope of the desire operator are referred to by definite descriptions: *the ghost* for the ghost which must have been introduced in Hans' beliefs and *it* which refers to *a fish* which was actually overtly introduced in John's beliefs. The sentence (2.22.a) therefore describes the following two propositional attitudes:

(2.23) a. John believes that he will catch a fish_F.
b. John desires that Mary fry the fish F.

These examples show that the referents in the scope of the desire operator are actually introduced in the scope of the belief operator. Maier, (2017) called this behavior *parasitism of propositional attitudes*. For example the content of desire in (2.22.a) is parasitic to the content of the belief in the same example. In order to represent this parasitic behavior in a DRS, one allows that a referent in the scope of a mode operator referentially depends on a referent which lies in the scope of another mode operator. Example (2.24) shows how parasitism can be realized in an ADS condition on the basis of a representation of example (2.22.a). For easier recognition, the parasitic referents are underlined in this representation.

(2.24)



Parasitism of propositional attitudes provides great flexibility for the presentation of mental states of agents. However, this flexibility requires an elaborated formalism for its semantic interpretation. Normally, propositions can be interpreted as a set of possible worlds, i.e. if K is a proposition and W is the set of all possible worlds, K can be interpreted as denoting all possible worlds in which K is true. In the same way, an agent's desires could also be interpreted as a set of possible worlds $Dox \subset W$. Such a simple extensional semantics, however, is not able to deal with parasitism. Parasitic mental states contain DRSs which contain free variables (such as for example the variable F in in the scope of DES in (2.24)). Such a DRS however, does not denote a proper proposition (but rather a function from referents into propositions) and therefore does not possess a proper extension in terms of possible worlds. Kamp solved this problem by introducing an elaborated semantic interpretation of ADSs in terms of their *context change potential* (CCP) (cp. Heim, 1992). However, since this book does not deal with actual semantic interpretations, this complicated semantics can be skipped at this point. We just have to take care that semantic interpretations of the DRSs we are dealing with are still possible. For a detailed discussion of the semantic interpretation of ADSs consider again Kamp, Genabith, and Reyle, (2005).

Even though referential parasitism seems to be very flexible, not every referent which occurs in one propositional attitude can occur in another one. There is at least a temporal restriction: if a referent x which occurs in the scope of a mode indicator MOD_1 depends on a proposition in the scope of a mode indicator MOD_2 then the temporal extension of MOD_1 should be located *inside* the temporal extension of MOD_2.

For an ADS this requirement is alway fulfilled since an ADS is exactly *one* argument in the state *Att*. It is therefore guaranteed that all propositional attitudes in an ADS take place at the same time. However, if one uses the convention above (here repeated as (2.25)) this requirement is not autaomatically fullfilled.

(2.25) $\quad S : \text{MOD}(A, K) \;:\Leftrightarrow\; S : Att(A, \langle \text{MOD}, K \rangle).$

In order to use this convention in a well-defined way, one has to take care for this temporal condition on parasitism. In other words, any use of parasitism in this book

has to fulfill the following requirement:

(2.26) CONDITION ON PARASITISM: Let A be an agent, let S_1 and S_2 be mental states of A and let $\text{MOD}_1, \text{MOD}_2 \in \{\text{BEL}, \text{DES}, \text{INT}\}$ be two mode indicators. If S_2 is parasitic to S_1 then S_2 has to be temporally contained in S_1.

$$\boxed{\begin{array}{l} A \\ \hline \text{AGENT}(A) \\ \\ \boxed{\begin{array}{l} S_1\ S_2 \\ \hline S_1 : \text{MOD}_1 \left(A, \boxed{\begin{array}{c} x \\ \hline \mathsf{con}_1(x) \end{array}} \right) \\ S_2 : \text{MOD}_2 \left(A, \boxed{\mathsf{con}_2(x)} \right) \end{array}} \Rightarrow \boxed{S_2 \subset S_1} \end{array}}$$

Note, that in (2.26), the relations for temporal referents which we have defined in (2.13) and (2.14) were used. In order to show how the presentation of a mental state of an agent can be realized using the notational convention given in (2.25), consider again (2.22.a). If one wants to present this example in the new form, one has to take condition (2.26) into account, i.e. one has to make sure that John's desire state lies inside his belief state. Using the new conventions and conditions, the whole sentence can be presented as follows:

(2.27)
$$\boxed{\begin{array}{l} T\ J\ M\ S_2\ S_1 \\ \hline \mathsf{name}(J, \text{John}),\ \mathsf{name}(M, \text{Mary}),\ T \subset S_1,\ S_2 \subset S_1 \\ S_1 : \text{BEL} \left(A, \boxed{\begin{array}{l} E_1\ F \\ \hline T \prec E_1 \\ \mathsf{fish}(F) \\ E_1 : \mathsf{catch}(J, F) \end{array}} \right) \\ S_2 : \text{DES} \left(A, \boxed{\begin{array}{l} E_2, T_2 \\ \hline T_1 \prec T_2 \\ E_2 \subset T_2 \\ E_2 : \mathsf{fry}(M, F) \end{array}} \right) \end{array}}$$

So far, parasitism of propositional attitudes was investigated and a necessary condition for parasitism was introduced. These notational conventions make it possible to represent propositional attitudes as conditions without the need to represent them as an attitude description set. 2.2.3 will again be concerned with propositional attitudes; it will be shown how one can recover an ADS from several propositional attitudes given in the form as shown in (2.27).

2.2.3 Conjoint of Mental States

Some combinations of propositional attitudes will be of particular interest for this book. For example, in chapter 3 an agent's *goal* will be defined as a certain combination of beliefs, desires and intentions. One important condition on these combinations is that the involved propositional attitudes take place at the same time. It will therefore be interesting to be able to talk about the time intervals in which these mental states overlap. In these time intervals the involved propositional attitudes can be collected and put together in an attitude description set. This set and the time of the overlap of the individual propositional attitudes will be called the *conjoint* of the individual mental states.

The conjoint of two mental states is defined as the union of both attitude description sets on the temporal overlap of the two states. The word *conjoint* for this new mental state was chosen because it contains the *conjunction* of the propositional attitudes of both overlapping states. As an example for a conjoint of two mental states, consider an agent A who is in the state of desiring a cup of coffee and therefore goes into a bakery store. But while A is waiting to order his coffee A sees a nice piece of cake, begins to think about having cake and finally develops the desire to have a piece of cake. At the time point, when A may order his coffee, A has forgotten that he originally wanted to have a cup of coffee and just orders the cake. In this example A goes at least through three different mental states. First, he is in the state in which he just desires to have a cup of coffee, then, second, he is in a state where he desires to have a cup of coffee *and* desires to have a piece of cake and, finally, he is in a state where he forgot all about the coffee and just desires to have a piece of cake. Figure 3.1 illustrates this situation. In this figure, the states S_1 and S_2 denote the states where A wants to have a coffee or, respectively, a cake. The state $S_1 \oplus S_2$ denotes the overlap of these states where A wants to have both, a coffee and a cake.

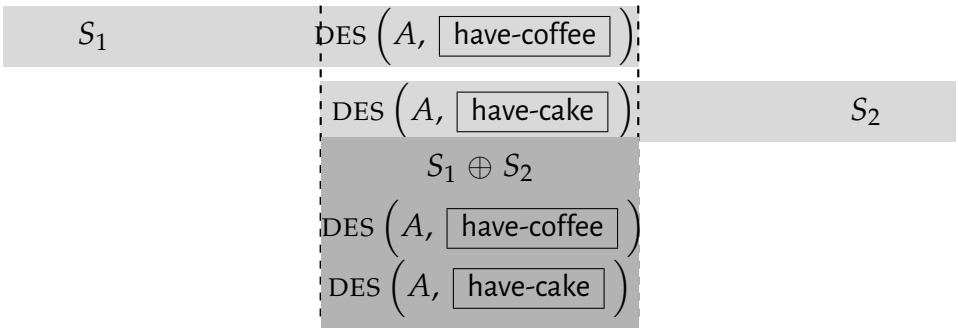


Figure 2.1: The conjoint $S_1 \oplus S_2$ of two mental states S_1 and S_2. The state $S_1 \oplus S_2$ consists of the union both ADSs on the temporal overlap of S_1 and S_2.

In the following, we will define the conjoint of two mental states S_1 and S_2 the union of both ADSs which hold in S_1 and S_2 on the intersection of these states:

(2.28) CONJOINT OF MENTAL STATES: Let A be an agent, let $\mathbb{K}_1$ and $\mathbb{K}_2$ be two attitude description sets and let $S_1 : Att(A, \mathbb{K}_1)$ and $S_2 : Att(A, \mathbb{K}_2)$ be two

mental states of A such that $S \cap S' \neq \emptyset$. Then we call the state

$$S \oplus S' : Att(A, \mathbb{K}_1 \cup \mathbb{K}_2),$$

where $|S \oplus S'| = S \cap S$, the *conjoint* of S and S'.

Recall that $|\cdot|$ is the function which assigns a state to the set of time points in which this state holds (see p. 21). Note that if one uses the notational convention (2.25) for the description of mental states, then for two states $S_1 : \text{MOD}_1(A, K_1)$ and $S_2 : \text{MOD}_2(A, K_2)$ the conjoint $S_1 \oplus S_2$ is defined as

(2.29) $$S_1 \oplus S_2 : Att(A, \{(\text{MOD}_1, K_1), (\text{MOD}_1, K_2)\}).$$

Note further, that parasitism is always well defined on the conjoint of mental states in the sense of condition (2.26), since by the definition of the conjoint of mental states, all propositional attitudes within the conjoint take place at the same time.

Summary of this section: So far we have introduced propositional attitudes and attitude description sets as conditions for DRSs. In particular, propositional attitudes have been defined as pairs of mode indicators and a proposition K. These attitudes can be collected in attitude description sets, where they can show a special parasitic behavior in the sense that a referent which occurs in the one propositional attitude referentially depends on a referent in another propositional attitude. Moreover, conditions for the possibility of parasitism have been discussed, Finally, we defined the conjoint of mental states s the union of two attitude description sets on the temporal overlap of two different mental states. These tools will prove very useful when dealing with nested propositional states.

The next section finally starts with the introduction of some basic rules for performing calculations with propositional attitudes. In particular, it discusses the interferences of propositional attitudes and negation. Moreover, it formalizes an agent's ability to draw conclusions from a given set of information within the framework of propositional attitudes.

2.3 Modeling Mental States

The last two sections were concerned with introducing DRT and its syntax and special referents, such as events and states and propositional attitudes. So far, DRT was used in a "classical" way, in so far as it was used in order to represent the content of natural language sentences. However, in this book DRT is rather used a modeling device for whole communicative situations and, in particular, for modeling an agent's assumptions about the mental states of another agent. This section, therefore, is not interested in the presentation of the content of natural language sentences any more but in presenting the content of communicative situations and mental processes.

In order to be able to model assumptions about mental states of agents, it is necessary to have some default rules for the behavior of mental states of agents. In other words, rules are needed on the basis of which one can model how an agent draws a

conclusion concerning another agent. Some of these rules will be introduced in this section. As it was already explained, these rules will be referred to as *principles*, since they are believed to model basic cognitive mechanisms. In this book, these principles have axiomatic character in so far as the calculus of mental states which is developed in this book will be built up on them.

The principles we are introducing in this section model basic cognitive processes, i.e. it is assumed that the cognition of every (human) agent behaves as described in these principles. Of course, the list of principles that is given in this book, is not meant as an exhaustive list, but just models a very small aspect of human cognition.

It will, moreover, be assumed that any agent believes that the cognition of any other agent behaves in the way described by the given cognitive principles. These assumptions can be summarized as follows:

1. By default, the mental states of any agent behave as described by the principles.
2. Any agent knows that 1.

The first assumption of this list ensures that we (as observers) can calculate with the mental states of agents, the second assumption ensures that we can model how an agent draws conclusions about another agent's mental state. Moreover, the second assumption models the fact that any agent has to believe that any other agent about which he wants to make guesses "thinks similar" to himself. This is an attempt to formally capture the following observation: there are many cases where we, as human agents, make guesses about the mental states of other agents. These guesses are based on the behavior of these agents. However, since we have never been able to actually *observe* the cognitive processes of another agent, these guesses have to be built up on the assumption that the cognitive processes in other agents are more or less comparable to our own. Without these assumption, guesses about another agent's mental states, would be indeed just pure guesses and would never be based on reasonable deduction.

To sum up, the principles which we are proposing in this book are means to model agents' cognitive processes and it is assumed that it is common knowledge of all agents that the cognition of all other agents is subject to these principles. This assumption makes it possible to apply these rules to other agents' minds in order to make guesses on their mental states. This section will be structured as follows: §2.3.1 will be concerned with the interference of propositional attitudes and negation. §2.3.2 is concerned with the question whether an agent's mental should be considered as being consistent and §2.3.3 introduces a version of modus ponens which depends on the particular belief states of an agent who draws a conclusion from a given set of information. Finally, a first example will be given which shows how these principles can be used to perform actual calculations.

2.3.1 Ignorance and Propositional Attitudes

In the following, negation of proposition attitudes will be investigated The problem which guides this part is the following: what is the relation between the negation of the propositional content of a propositional attitude and the negation of the whole

propositional attitude? In order to gain a better feeling for the difference between the two types of negations, consider the following advice by Searle, (2005), who claimed that once one is stuck with a problem of propositional attitudes (i.e. intentional states in Searle's parlance), one should look for a similar phenomenon in speech act theory. In speech act theory there are actually two different kinds of negations: One can negate the *propositional content* of a speech act and one can negate its *illocutionary force*. The best way to clarify the difference between these two types of negation is to consider performatives:

(2.30) a. I promise not to come.
b. I do not promise to come.

(2.31) a. I claim that O'Leary isn't an idiot.
b. I do not claim that O'Leary is an idiot.

In (2.30.a) the speaker indeed makes a promise: he promises *not to come*! In (2.30.b), on the other hand, the speaker makes it explicit that he does *not* promise to come. In this case, the speaker makes it explicit that he is not promising anything at all. The pair in (2.31) shows the same: in (2.31.a) the speaker indeed claims the propositional content of the speech act, whereas in (2.31.a) the speaker makes it clear that he does not claim anything at all (and in particular does not claim the propositional content of the speech act). Searle and Vanderveken, (1985) call the negation of the illocutionary force *illocutionary denegation* and characterize it as follows:

> We can say generally that an act of illocutionary denegation is one whose aim is to make it explicit that the speaker does not perform a certain illocutionary act. *(Searle and Vanderveken, 1985, p. 4)*

The fact that there is a difference between negation of the content of a speech act and illocutionary denegation can be formulated by saying that the illocutionary force operator and negation don't commute (i.e. are not interchangable) which can be formalized as follows:

(2.32) $\neg F(p) \neq F(\neg p)$.

Here, F denotes the illocutionary force operator and p the propositional content of a speech act.

One can either negate the propositional content of a speech act or one can negate the illocutionary force of a speech act. The analogous statement on the level of, let's say, beliefs is that there is a difference between believing that a certain proposition does not hold and *not* believing that a certain proposition holds. The latter condition can be reformulated by saying that the agent is *ignorant* about the proposition in question. He simply is not aware of the proposition in question at the given point. On a formal level, this can also be represented by saying that the content of a propositional attitude and the mode indicator show a similar behavior with respect to negation as speech acts, in the sense that they don't commute either. In this analogy, the mode operator of a propositional attitude corresponds to the illocutionary force of a speech act and the content of a propositional attitude corresponds to the propositional content of a speech act. It does make a difference whether one negates the propositional content of

a propositional attitude or whether one negates the mode operator of the propositional attitude. Take beliefs, for example: believing that O'Leary is *not* an idiot is something else than *not* believing that O'Leary is an idiot. In the first case, the attitude holder has an active belief concerning O'Leary, in the second case, the attitude holder can be considered to be ignorant about O'Leary.. Again this can be formulated by saying that mode indicators and negations do not commute:

(2.33) $\neg \text{MOD}(A, K) \neq \text{MOD}(A, \neg K)$.

Even though in both cases, i.e. in cases for speech acts and propositional attitudes, negation and illocutionary force or, respectively, mode operator do not commute, there is still an implication relation between the two types of negation. However, this implication is not based on the pure semantics of the two speech acts but rather on the *social effects* of speech acts. In recent years the opinion has become popular that while communicating speakers actually negotiate *social commitments*, rather then sharing information (Condoravdi and Lauer, 2012; Roberts, 2017; Geurts, 2018). This idea will also be adopted and formally implemented in chapter 4. However, at this point, the basic ideas of this theory will be sketched. The idea is that any speech act commits the interlocutors to a certain behavior: a promise commits the speaker to acting as promised, a question commits interlocutors to answering the question and so on. An assertive speech act, like a claim, on the other hand, commits the speaker to acting as if he believes the content of the claim. Given this assumption, one can see why there is at least one implication between the two different kinds of negation which we have discussed above. Consider again example (2.31.a). In this example, the speaker makes a claim about O'Leary and, therefore, *commits* to the content of this claim which means that he has to act as if O'Leary is not an idiot. But this implies that the speaker is also committed to *not claiming* that O'Leary is an idiot. On this basis one can explain why the following sentence is odd:

(2.34) O'Leary is no idiot but/and I claim that he is an idiot.

Note that this last example looks similar to the Moore paradox *p but I don't believe p*, which is a paradox since a priori the two statements *p* and *I don't believe p* are logically completely independent but, nevertheless, their conjunction sound odd. This can be explained on the basis of a commitment-based approach to communication since the two different statements give rise to inconsistent commitments. A similar approach can be used in our example: by uttering a speech act like (2.31.a) the speaker commits to act as if O'Leary is not an idiot. However, acting as if someone is not an idiot also implies to *not* explicitly claiming that this someone *is* an idiot, otherwise one would again be left with inconsistent commitments. In other words, if A claims that O'Leary isn't an idiot then A is committed to *not* claiming that O'Leary is an idiot. On a more abstract level, this means that claiming that $\neg K$ entails that the speaker doesn't claim that K.

(2.35) $F(\neg p) \Rightarrow \neg F(p)$

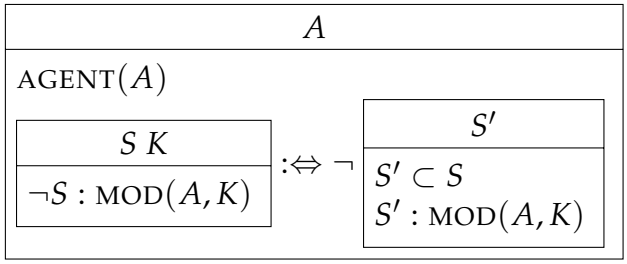
However, originally it was not the negation of speech acts that was of interest but for-

mal properties of negation of propositional attitudes. If one wants to use the analogy between illocutionary forces and mode operators in propositional attitudes, one can claim that a similar condition as in (2.35) holds for propositional attitudes. This can be expressed by saying that if an agent A believes $\neg K$, then A is *ignorant* about K (in the sense that A is not in an active belief state towards K). For example, if I believe that O'Leary is not an idiot then I do *not* believe at the same time that O'Leary is an idiot. This makes believing that some state of affairs does not hold a stronger condition than being ignorant about that states of affairs. However, before this assumption is actually formalized, a little bit more should be said about what exactly we mean by *ignorance*.

Ignorance is supposed to refer to a lack of knowledge. For example, being ignorant about the fact that it is raining simply means that one does not believe that it is actually raining. More generally, if an agent A is in the state of being ignorant about a state of affairs K, then A does not believe K. This formulation does not exclude the possibility that A is aware of K in the sense that A considers K to be possible. It simply states that ignorance towards K means not being in an active belief state towards K. Although so far only beliefs have been considered, this observation is not limited to beliefs. One can also be in a state in which one does not actively *desire* a state of affairs or *intend* a certain action. Ignorance can therefore easily be formulated for all these different types of mode indicators. In the following, ignorance will be defined as being a state in which the agent is not actively in a state of believing, desiring or intending something.

The definition given in (2.36) characterizes ignorance for all mode indicators of interest. In order to be able to differentiate between the different types of ignorance (i.e. not-believing, not-desiring, not-intending something) these different types of ignorance are called MOD-ignorance for $\text{MOD} \in \{\text{BEL}, \text{DES}, \text{INT}\}$, even though this term will rarely be used in this book.

(2.36) IGNORANCE: Let A be an agent and let $\text{MOD} \in \{\text{BEL}, \text{DES}, \text{INT}\}$. A is MOD-ignorant about a proposition K if A doesn't have the propositional attitude $\text{MOD}(A, K)$. Write $\neg S : \text{MOD}(A, K)$ for the state S in which A is MOD-ignorant towards K:



Given this characterization of ignorance, one can now formalize the assumption concerning negation of propositional attitudes, i.e. that believing that $\neg K$ implies being ignorant about K. Moreover, this assumption can be generalized to all mode indicators, which models the observation that agents can also be in the state of desiring or intending that $\neg K$ and therefore are in a state of not desiring or not intending K:

(2.37) NEGATION AND IGNORANCE:

Let A be an agent. If A has a propositional attitude $\text{MOD}(A, \neg K)$ for $\text{MOD} \in \{\text{BEL}, \text{DES}, \text{INT}\}$ towards a proposition $\neg K$, then A is MOD-ignorant about K:

$$\frac{\begin{array}{l}A \\ \hline \text{AGENT}(A) \\ \dfrac{S\ K}{S : \text{MOD}(A, \neg K)} \Rightarrow \neg\, S : \text{MOD}(A, K)\end{array}}{}$$

In the rest of this section, ignorance will will be used to model interferences between beliefs and desires.

In the following, the idea is formalized that an agent cannot be in a belief state and in a desire state towards the same proposition K *at the same time*. For example, if an agent believes that it's raining, the agent cannot at the same time desire that it rain. A desire is satisfied if the states of affairs of the desire hold true, or to be more precise, if the attitude holder *believes* that they hold true. This behavior of beliefs and desires can also be attributed to their *direction of fit*. Beliefs have the *mind-to-world* direction of fit and desires have the *world-to-mind* direction of fit (Searle, 2005). In other words, the content of beliefs has to fit the world (if true) whereas the world has to fit the content of a desire (if satisfied). This is the reason why beliefs can be *false*. If the content of a belief does not fit the state of affairs in the world, it has to be reconsidered as soon as the belief holder finds out. Desires on the other hand, have the *world-to-mind* direction of fit. If the world fits the content of the desire, then the desire is satisfied and, hence, one cannot desire it any more! If an agent believes that it's raining, then he cannot desire that it rain. Of course the agent can be pleased by the fact that it's raining and can desire that it rains for a longer period of time, but cannot desire that it rain anymore, because the world already fits the content of his desires. On the other hand, if one desires rain, one cannot believe that it rains at the same time. This means that if one desires a proposition K, then one has to be BEL-ignorant about K and vice versa. These considerations are summarized in the following principle:

(2.38) BELIEF-DESIRE INTERFERENCES:

Let A be an agent. If A has the propositional attitude $\text{BEL}(A, K)$ then A is DES-ignorant about K. On the other hand, if A has the propositional attitude $\text{DES}(A, K)$ then A is BEL-ignorant about K:

- $S : \text{BEL}(A, K) \Rightarrow \neg\, S : \text{DES}(A, K)$
- $S : \text{DES}(A, K) \Rightarrow \neg\, S : \text{BEL}(A, K)$

So far, this section was concerned with different types of interferences between negations and propositional attitudes. It was introduced and defined what it means to be ignorant about a proposition. On the basis of ignorance it was moreover possible to formulate a principle which represents the observation that being in a certain mental state (belief, desire, intention) towards a proposition $\neg K$ implies ignorance towards K. Moreover, it was possible to formalize an assumption about the interferences between beliefs and desires on the basis of the notion of ignorance.

2.3.2 Minimal Consistency

In the former parts of this section, it was claimed that a certain conjunction of certain sentences (*not p, but I claim that p*) are odd. The argument for these cases being odd was that such sentences lead to *inconsistent* commitments. To act in accordance with inconsistent commitments is basically not possible since it would require different actions which exclude each other. Take for example a (hypothetical) agent who is at the same committed to liking cheese and to not liking cheese. How would this agent possibly react if offered a plate of cheese? Inconsistent commitments give rise to problems in the realization of concrete actions. But what about mental states such as beliefs or desires? They can be inconsistent as well, but is that a problem? In general, one would assume that an agent's mind is consistent in the sense that the agent does not desire two opposing states of affairs or believes some p and at the same time believes $\neg p$. This assumption, as natural as it might seem, has some problems which we will discuss later on. However, it seems to be so natural to assume mental consistency (at least in a minimal sense) that it is included as a cognitive rule in the theory. This rule will be called *minimal consistency*. In the following, the question of how to model the assumption that an agent's mental states are consistent (in whatever sense) will be approached. In principle the rule which will be proposed can be regarded as a consequence of (2.37), but because of its importance it has a name on its own.

Even though, the assumption that mental states of rational agents are consistent at least in a minimal sense has some problems, it is a cornerstone for modeling assumptions about the mental states of rational agents. Without such an assumption, it would again be impossible to even make guesses about the mental states of agents. Suppose, for example, that we found out that some agent believes that it is not rain. Without assuming that this agent's mind is (minimally) consistent one would not be able to conclude that this agent does not believe opposing propositions at the same time and, even worse, that the agent draws additional conclusions from these beliefs. But even though a consistency requirement seems to be neccessary for dealing with other agents' minds, there is a long ongoing debate in philosophy of mind whether human convictions and cognition should be considered consistent at all, and if yes, in what sense. The basic problem is the following. It is very well possible for an agent A to believe a proposition K and at the same time a set of propositions φ which has $\neg K$ as a consequence, whatever notion of consequence one wants to use here (causation, implication, both ...). This is due to the fact that an agent with limited cognitive resources simply does not have access to the set of *all* the consequences of his beliefs and, therefore, the set of consequences may be inconsistent in the following sense: If P is an

agent's set of beliefs and $\bar{\varphi}$ is the set of its consequences (causation, implication, etc.), then it is possible that there is a $K \in \varphi$ such that also $\neg K \in \bar{\varphi}$. Such inconsistencies may occur, nevertheless, it seems to be a reasonable assumption that belief are consistent at least in a minimal form, i.e. in the sense that an agent who believes K cannot believe $\neg K$ *at the same time*! This assumptions does not exclude the possibility that the agent believes a set of propositions which contains K and has $\neg K$ as consequence. Such situations may occur, if the agent does not have the necessary knowledge about all the consequences of his beliefs or if the agent does not have enough time to consider all the consequences. However, it is reasonable to assume that an agent who finds out that his beliefs have contradictory consequences wants to change at least one of his beliefs.

The assumption that the mental states of an agent are consistent, at least in the minimal sense which was discussed above, will be called *minimal consistency*. Minimal consistency will be a cognitive principle which applies to all propositional attitudes which are of interest in this book, i.e. it applies also to *desire* and *intentions*. This means, for example, that according to minimal consistency it cannot be the case that an agent A desires rain and at the same time desires that it does *not* rain. A formalization of this principle can be given as follows:

(2.39) MINIMAL CONSISTENCY:
Let A be an agent. If A has a propositional attitude $S : \text{MOD}(A, K)$ for $\text{MOD} \in \{\text{BEL}, \text{DES}, \text{INT}\}$ with respect to a proposition K, then A cannot have the propositional attitude $\text{MOD}(A, \neg K)$ at the same time:

$$\boxed{\begin{array}{l} A \\ \hline \text{AGENT}(A) \\ \quad \dfrac{S\ K}{S : \text{MOD}(A, K)} \Rightarrow \neg S : \text{MOD}(A, \neg K) \end{array}}$$

Principle (2.39) can also be derived from (2.37) by replacing K in (2.37) with $\neg K$. However, since it is of particular importance it gets a name on its own.

It was already said that this principle may be problematic in several ways, even though at the first glance, it seems to be quite uncontroversial. At least, it seems to model assumptions of rational agents about the mental states of other rational agents. Nevertheless, at least one problem of this principle shall be addressed at this point. For an illustration of the problem, have a look at the following scenario:

(2.40) Suppose that there is an agent A and that the following conditions hold for A:

a. A loves pizza and therefore has the desire to eat lots of pizza.
b. A has the desire to get (or stay) slim
c. A believes that eating lots of pizza causes him to gain weight.

Under the condition that b. and c. hold, it would be quite natural for A to draw the consequence that he should not eat pizza and hence to not desire to eat pizza. But in combination with a. this would result in a situation in which A has the desire to eat pizza and the desire not to eat pizza at the same time. To be in similar scenarios is a common experience for most agents. However, this is exactly the kind of scenario which is excluded by the minimal consistency condition. Nevertheless, in principle it seems to be reasonable to include minimal consistency in the set of cognitive rules. However, this leads to the following problem: there is the every day experience that agents seem to be in inconsistent mental states and at the same time there is minimal consistency as a principle for rational cognition which seems to be necessary in order to be able to model assumptions about cognition. In the following, it is, therefore, tried to solve this tension and, in particular, to defend the minimal consistency assumption against objections such as provided by problematic situations such as given in example (2.40).

The following disussion will basically consist of two parts: The first part is rather technical in nature and simply states that a derivation such as the one in (2.40) is simply not admissible. The second argument is more conceptual and states that the inconsistent propositional attitudes in (2.40) are not the right description of this situation. These arguments do not strictly build up on each other. However, the second argument can be regarded as an extension of the first one.

First: Written down in its logical form, the deduction for the incoherent propositional attitudes of the mental state in (2.40) is basically the following:

$$\begin{array}{ll} \textsc{des}(A, \neg q) & \text{(assumption b.)} \\ \textsc{bel}(A, \neg q \rightarrow \neg p) & \text{(assumption c.)} \\ \hline \textsc{des}(A, \neg p) & \text{(conclusion),} \end{array} \tag{2.41}$$

where $p =$ [Eat Pizza] and $q =$[Gain weight]. The rule which was used to derive the conclusion in (2.41) is some kind of a "mixed" *modus ponens* (cp. section 2.3.3). In general, modus ponens refers to the inference rule which is described on the left side of (2.42) (the right side just shows a classical application of this rule).

$$\begin{array}{ll} P \rightarrow Q & \text{If it rains, then the street is wet.} \\ P & \text{It's raining.} \\ \hline Q. & \text{The street is wet.} \end{array} \tag{2.42}$$

Modus ponens as a deduction rule models the fact that if one knows that some p is the case and, moreover, knows that p has some consequence q then one also can conclude that q is the case. However, in (2.41) this rule was not applied to individual propositions but to the content in the scope of DES and BEL. But of course, this is not a case of modus ponens since modus ponens simply does not work for desire/belief/intention combinations (Searle, 2005). In order to illustrate a situation in which an application of modus ponens for belief/desire combinations produces counterintuitive results, Searle gives the following example (Searle, 2005, p. 53): An agent A may want a subway to be

less crowded and believes that if he kills all the other passengers it will actually be less crowded. However, in such a case it is absurd to assume that A is actually committed to forming the desire to kill all the passengers in the subway: Agent A simply does not have to form the desire from an intention/desire combination. The situation in the pizza example is similar in the sense that modus ponens was applied to belief/desire combinations.

However, one might object that situations as in (2.40) may occur even though an application of modus ponens to belief/desire combinations is in general not admissible. The next argument states that even in the situation of (2.40) it is actually *not the case* that the agent desires to eat pizza and at the same time desires to not eat pizza but that this situation has to be described in another way.

Second: Another possible reply to the problem posed by (2.40) is less technical but maybe more conceptual. The idea for this reply is that the agent does not necessarily have a desire on the content level, i.e. a desire concerning a description of the world, but has a desire which is concerned with his own mental states. But what does this mean? Assume again that all conditions in (2.40) are true. So far, it was assumed that as a consequence of b. and c. the given agent desires to not eat pizza and therefore is in the state of desiring to eat pizza and desiring to not eat pizza at the same time. Another possible consequence of this situation, however, is that the agent still desires to eat pizza and at the same time desires to *not desire to eat pizza*. This response to the objection given above provided by (2.40) has some advantages: first of all, it does not lead to a problem concerning minimal consistency. The statements $\text{DES}(A, \neg\text{DES}(A, p))$ (i.e. desiring to not desire to eat pizza) and $\text{DES}(A, p)$ (i.e. desiring to eat pizza) are completely independent and may occur at the same time without any problems. Furthermore, such a configuration of propositional attitudes is propably an even better description for a situation such as (2.40) than simply assuming inconsistent mental states. For example, an addicted smoker might have the desire to *not* want a cigarette while at the same time wanting a cigarette[2]. A PhD candidate may desire to desire to write the PhD thesis at the same time as actually desiring to play soccer. For situations in which an agent seems to be in an inconsistent mental state, it is therefore reasonable to assume that the agent is actually just in a situation in which she has *attitudes towards attitudes*, i.e. desires concerning desires or beliefs concerning desires.

Remark. Attitudes towards attitudes sometimes solve tensions which would give rise to consistency problems. However, this does not work with any intentional state. For beliefs, for example, it is reasonable to assume that they fullfill *introspectivity*, i.e. $\text{BEL}(\text{BEL}(A, p)) \rightarrow \text{BEL}(A, p)$ (if I believe that I believe p then I believe p). This makes it hard to argue that believing that one believes that p and believing $\neg p$ may occur at the same time. However, in the case of desires this assumption does not seem to be problematic.

None of these arguments above imply the neccessity of minimal consistency. However, it still seems to be necessary to assume such a principle in order to be able to

[2]The problem we have just sketched is one manifestation of the acrasia problem. The acrasia problem goes back to Aristotle and is a classical problem in practical philosophy which is concerned with the question how weakness of will (i.e. eating pizza even though one gains weight by eating pizza) is possible.

model the mental states of agents and, more importantly, to model assumptions of an observer about the mental state of an agent.

So far, consistency of an agent's mental states was discussed. It was discussed why it is reasonable to assume a minimal version of consistency and why such an assumption is, nevertheless, problematic. Moreover, a principle was introduced which models at least a minimal consistency assumption. Finally, an objection to the general assumption that mental states are consistent (even in a minimal sense) was discussed.

2.3.3 Modus Ponens for Beliefs

In 2.3.2 it was made very explicit that one cannot assume that modus ponens applies to propositional attitudes with mixed mode indicators. However, modus ponens describes the basic ability to derive new knowledge if one already has some knowledge and additional knowledge about consequences. Modus ponens, therefore, seems to be important for modeling something like an agent's ability to draw conclusions from given information. (2.43) shows again the general structure of modus ponens as a deduction rule.

$$\frac{\begin{array}{l} P \rightarrow Q \\ P \end{array}}{Q.} \tag{2.43}$$

In general, modus ponens describes an agent's ability to draw conclusions in cases where she has enough knowledge about a state of affairs and its consequences. However, given that human agents have limited resources to calculate logical implications from a given set of propositions, it is not quite clear how one should treat modus ponens in epistemic logics. It is surely not plausible to assume that an agent who knows or believes a certain proposition is also aware of all the consequences of that proposition simply because human agents do have access to a only a limited amount of cognitive resources. But even if an agent has knowlegde about all consequences of a proposition, the limitations of his cognitive resources still make it impossible to calculate all conclusions at once. On the other hand, it is still reasonable to assume that under certain circumstances an agent has knowledge, at least about some of the consequences of a proposition K. Hence, the problem is the following: given that an agent A knows a proposition K and given that A cannot know all consequences of K what consequences or conclusions can he draw from K and under which circumstances he can draw these conclusion? This section aims at giving a minimal solution for this problem and at formalizing this solution. This solution is minimal in the sense that it makes rather strong assumptions about the mental states of the agent we are dealing with.

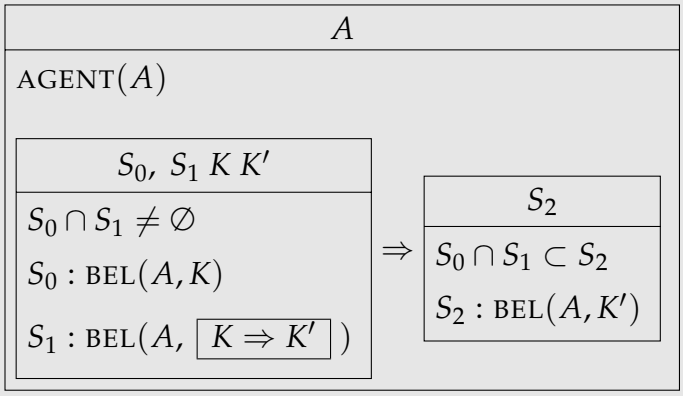
The rule for drawing conclusions which is developed here is basically a reformulation of the principle of *non-ignorance* (Rasmussen, 2015). Non-ignorance states that under certain circumstances, an agent simply cannot *ignore* the consequences of a given state of affairs:

(2.44) NON-IGNORANCE: If a resource-bounded agent knows the premises of a valid inference and knows the relevant inference rule, then, given sufficient resources, the agent can infer the conclusion.

A crucial point in the formulation of (2.44) is the requirement that the agent should be given *sufficient resources* in order to be able to infer a conclusion. This requirement will be translated in the calculus of propositional attitudes by the condition that the agent has to be in an *active* belief state towards both, the premises *and* the inference rule in question. This requirement corresponds to the condition of having sufficient resources in (2.44), in so far as it prevents the agent from being able to draw any possible conclusion at once, simply by forcing him to be in a certain mental state[3].

In the following, non-ignorance will be reformulated in the language which was developed so far (principle (2.45)). The additional assumption that the agent must be given sufficient resources will be captured by the requirement that the agent is in an *active* belief state towards the propositions involved. (2.45) is basically a version of modus ponens which takes an agent's beliefs into account and is, therefore, referred to as *modus ponens for beliefs*.

(2.45) MODUS PONENS FOR BELIEFS:
If an agent A believes a proposition K and at the same time believes that K implies a proposition K' then he believes the proposition K':



Modus ponens for beliefs is a weak version of modus ponens since it requires the agent to believe the antecedent and the relation between antecendent and consequent at the same time. However, it is a basic advantage of (2.45), that it allows the agent to make mistakes: if for some reason an agent believes believes a proposition p and some inference rule $p \rightarrow q$ at the same time and one of these beliefs is actually wrong, then modus ponens for beliefs still allows the agent to draw a conclusion from these rules, even though the conclusion is probably wrong. Principle (2.45) is important for performing calculations with propositional attitudes since it is the only principle so far which is able to model an agent coming to a new belief.

[3]Here, the tacit assumption that an agent cannot be in an infinite number of belief states at the same time is used.

2.3.4 A First Example

The rest of this section is devoted to the presentation of a first example in which some of the principles above are used. The set of tools which have been developed so far is still rather limited, which makes the presented calculation rather short and basic.

For carrying out this example, assume in the following that there is an agent A who utters (2.46) towards an agent B.

(2.46) A: I want to have cheese for dinner.

Assume further, that B was able to conclude from this linguistic input that A desires to have cheese for dinner. This means that B is in a belief state concerning A's mental state! Translated into the language which has been developed so far, this situation can be represented as follows:

(2.47)

$$\boxed{\begin{array}{l} A\ B\ S^1 \\ \hline \textsc{agent}(A),\ \textsc{agent}(B) \\ S_1 : \textsc{bel}\left(B, \boxed{\begin{array}{c} S_1^A \\ \hline S_A^1 \subset S \\ S_A^1 : \textsc{des}\left(\boxed{\mathsf{eat.cheese}(A)} \right) \end{array}} \right) \end{array}}$$

A's utterance (2.46) caused a mental state of B in which he has a belief about the mental state of A, i.e. in this case about a *desire* of A. Moreover B's belief state can temporally be located inside A's desire state[4]. It can be assumed that B assumes that A is in a consistent mental state. Because of the assumption that all agents are aware of all cognitive principles, B knows the structure of minimal consistency and, hence, is aware of what it means to be in a consistent mental state (cp. (2.39)). Therefore, B knows that if A desires to eat cheese, then at the same time A does *not desire* to *not* eat cheese. This situation can be presented as follows:

[4] This can be assumed because auf A's use of the present tense. If A had used a past tense, then B would have to locate B's desire state somewhere in the past as well.

(2.48)

$$
\begin{array}{|c|}
\hline
A\ B\ S_1\ S_2 \\
\hline
\textsc{agent}(A),\ \textsc{agent}(B) \\
S_1 : \textsc{bel}\left(B, \begin{array}{|c|}\hline S_1^A \\ \hline S_1^A \subset S \\ S_1^A : \textsc{des}\left(A, \boxed{\mathsf{eat.cheese}(A)}\right) \end{array}\right) \\
S_2 : \textsc{bel}\left(B, \begin{array}{|c|}\hline S_1^A : \textsc{des}\left(A, \boxed{\mathsf{eat.cheese}(A)}\right) \\ \Downarrow \\ \neg S_1^A : \textsc{des}\left(A, \neg \boxed{\mathsf{eat.cheese}(A)}\right) \end{array}\right) \\
\hline
\end{array}
$$

The state S_1 in (2.48) contains B's actual beliefs concerning B's desire to eat cheese. The state S_2 on the other hand, contains A's belief that minimal consistency holds. Translated into ordinary language, (2.48) describes the following scenario: agent B is aware of the fact that A desires to eat cheese and at the same time B is aware of the fact that if A desires to eat cheese A does not desire to *not* eat cheese. This, however, is exactly the precondition for applying modus ponens for beliefs (principle (2.45)), which means that B is now able to conclude that A *actually is* in a state in which he does *not desire to not eat cheese*. This situation is presented in (2.49). In order to save space, the presentation of (2.49) abbreviates the necessary mental state for applying modus ponens, with the conjoint (cp. principle (2.28)) of B's involved belief states. Hence, after performing some calculations, one can conclude that after A's utterance which was concerned with A's desire to eat cheese, B was able to conclude that at the time A produces her utterance she is not in the state of *not* desiring to eat cheese.

(2.49)

$$
\begin{array}{|c|}
\hline
A\ B\ S_1\ S_2\ S_3 \\
\hline
\textsc{agent}(A),\ \textsc{agent}(B) \\
S_1 \oplus S_2 \subset S_3 \\
S_1 \oplus S_2 \Rightarrow S_3 : \textsc{bel}\left(B, \boxed{\neg S_1^A : \textsc{des}\left(A, \neg \boxed{\mathsf{eat.cheese}(A)}\right)}\right) \\
\hline
\end{array}
$$

So far, this result is rather unspectacular. As already said, this is due to the fact that the set of tools for modeling assumptions about mental states is still yet very limited. However, the calculations which will be performed in this book will become more interesting once further tools are available. Section 2.4, for example, will introduce two further cognitive principles which make more interesting calculations possible.

Summary of this section: This section was concerned with the introduction of some basic principles which can be used for performing calculations with propositional attitudes and, hence, to model cognitive processes in an agent's mind and changes in his mental states. In particular we have been concerned with negation of propositional attitudes and the question whether an agent's mind can be assumed to be consistent

and how this can be modeled.

2.4 Inertia and Learning

In their popular scientific book "The evolution of Physics", Einstein and Infeld, (1960) defined the physical law of *inertia* as follows:

> Every body perseveres in its state of rest, or of uniform motion in a straight line, unless it is compelled to change that state by forces impressed thereon.

Originally, this law was formulated by Isaac Newton and is now known as the first axiom of classical mechanics, which introduces inertial frames into mechanics. Inertia, as an axiom in physics, (heuristically) states that a physical object does not move if it does not have a reason to move. This section introduces a similar principle for cognitive states: the idea of inertia as a cognitive principle is to apply it not to physical objects but to mental states of agents. This means that inertia - as a cognitive principle - states that mental states of agents do not change if there is no reason for them to change.

Inertia in mechanics guarantees stability of physical systems. It prevents physical objects from moving around randomly. Inertia as a cognitive principle, on the other hand, guarantees that minds of rational agents do not change randomly. What does this mean? Take for example, an agent who believes that there is a bakery store on a corner near his flat. Under normal circumstances, the agent will stick to this belief as long as he is not aware of any event which might have changed the fact that there is a bakery store on the corner near his flat. In other words, an agent sticks to his beliefs as long as he does not *learn* about an event which might have an effect on the content of his beliefs. Without having a reason for changing his mind, for example his beliefs concerning a bakery store, the agent will not change his mind.

Remark. In some sense, inertia as a cognitive principle can be regarded as an epiphenomenon of the everyday experience that physical objects are subject to inertia. A stone won't move without having a reason and, therefore, our belief concerning the whereabouts of the stone will not change either. Human agents are accustomed to the fact that physical objects are subject to inertia and therefore we, as human agents have established inertia as a general principle of thinking. This thought, however, is highly speculative.

In this book, inertia models the assumption that minds are stable in the sense that they do not change randomly and that any change in the mind of an agent needs a reason. In particular, inertia allows one to connect mental states of agents at different time points. Without inertia, an agent who wants to make reasonable guesses about another agent's mental state on the basis of his knowledge of the other agent's *former* mental states, needs the inertia assumption to be able to do so. For example, assume again that there is some agent A who knows that there is a bakery store near his flat and assume that there is an agent B who knows about A's knowledge at some time point T. It is now reasonable for B to assume at some time point T' after T that A still

posseses the knowledge about the bakery store. But why is that so? Again, B may assume that A's mind is subject to inertia (and, hence, stable) and therefore has certain assumptions about states of affairs in the world.

Even though the minds of human agents are stable in the sense above, sometimes beliefs do change. But this means that the agents actually *had* a reason for changing their beliefs. In this case, one can say that the agent *learned* something, namely that there was an actual change of states of affairs in the world. After having introduced inertia, the following section will, therefore, be concerned with *learning* as a form of inversion of inertia.

Inertia seems to be a basic principle for understanding cognition, and, therefore, 2.4.1 will be devoted to a further discussion of inertia and its formalization. After having discussed inertia, 2.4.2 will be concerned with a discussion of *learning*, wich can be considered the inversion of inertia: inertia models stability, learning on the other hand models that an agent has to change his mind if there is a reason for him to change his mind.

2.4.1 Inertia

As already mentioned, *inertia* can be characterized as the following principle: if there is no reason for a state of affairs to change, then it remains the same. This formulation can be applied to both physical objects and mental states. Inertia in physics is a very well established rule. Inertia as a cognitive principle was formulated more recently in the discussion of artificial intelligence (AI). In AI it was established in order to solve the following problem: given a description of the world in logical axioms, how it is possible to derive from these axioms which states of affairs in the world stay the same if suddenly something changes in this description. The classical example is the following: suppose there is an artificial agent who has knowledge about some cubes, their position and their color and that this knowledge is given as a set of axioms that constitutes the agent's world. Now, suppose that the position of one of the cubes changes. In that case, the artificial agent has to adjust its axiomatic system, but of course, the question is *how*? How does a change in the position of one cube have an effect on the position of the other cubes and does it affect their color or their size as well? Of course their color will not change, but maybe the position of the other cubes (for example, if the cubes form a stack). In any case, the agent has to adjust its axiomatics, but in a "minimal way" (for example, it does not have to reconsider the color of the cubes). Now the problem is how this "minimal" change of axiomatics can be found and how it can be performed. In AI this problem is referred to as the *frame problem*. Inertia allows us to model the fact that normally "most states of affairs in the world" do not change if one changes something. For more details on the frame problem see Shanahan, (1997).

Inertia as a cognitive principle has already found its way into models of human reasoning and human interaction. For example, McCready, (2014, p. 47) already implemented inertia into a game theoretic approach to reliability. However, inertia hat not yet been used for explaining phenomena in discourse structure, as it will be in this book. The idea for applying inertia to discourse structure will roughly be the following: since interlocutors are subject to inertia any interlocutor assumes that the

respective other interlocutor does not change her mind about a given theme or topic at will. These considerations result in the claim that, by default, discourse sticks more or less to the same topic and referents. This idea will be carried out in detail in 5.3.2.

The rest of this section discusses an example of an application of inertia to an everyday scenario and, afterwards, it abstracts over this scenario in order to obtain a general formulation of inertia. The example consists again of an agent who has knowledge about a bakery store near his flat:

(2.50) Assume that at time T an agent A needs some bread for breakfast and that A knows that the day before there was a bakery store near his flat. Moreover assume that A does not know about an event which might have caused the bakery store to be closed or destroyed (or similar). Under these conditions, A still believes at T that the bakery store is in the same spot as it was the day before T and therefore A can make a plan to buy bread at that bakery store.

A's plan to go to the bakery store in scenario (2.50) is based on the assumption that the store still exists and is located at the same point as it was the day before T *even though* A doesn't have actual evidence for this assumption. A's assumption, that the bakery store is still at the place where it used to be is based on A's *ignorance* about events which might have caused the bakery store to close or to move (or similar).

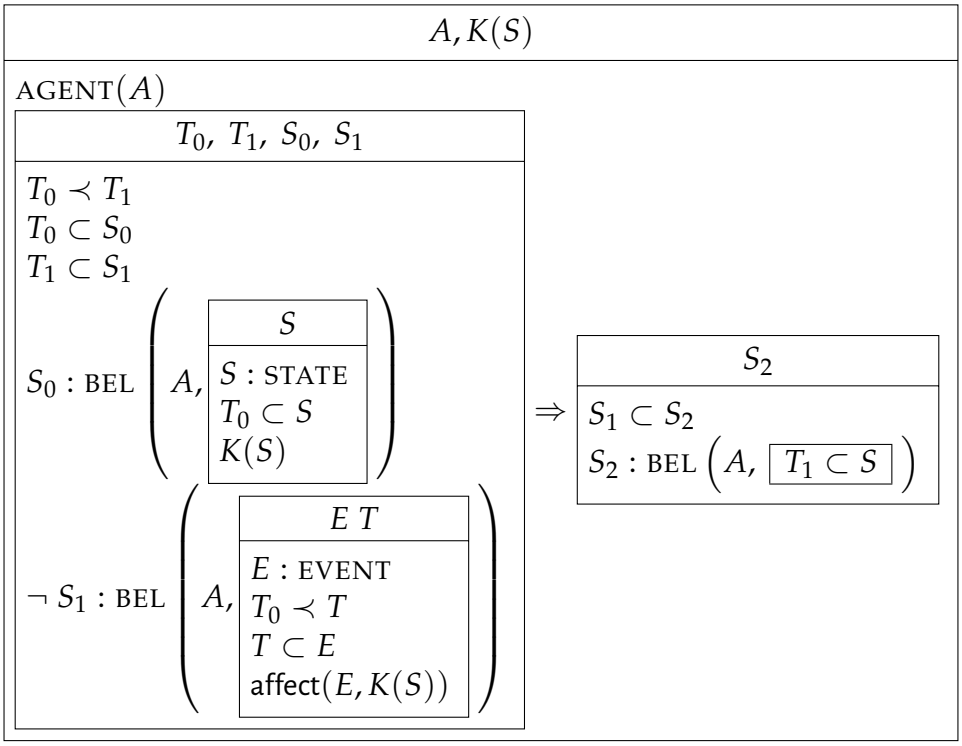
Even though there is basically an uncountable number of events and actions which happen every second of every day, agents have to be able to make plans, and this implies being ignorant about most of these events. In other words, inertia can also be regarded as a form of *organized ignorance*: Human agents have to live on the assumption that 'normally, given any event and any state of affairs, the event does not change it' (Shanahan, 1997, p. 57). This assumption can be paraphrased in terms of agents' beliefs by saying that as long as an agent is ignorant about an event which might have affected a particular state, the agent doesn't change his beliefs towards that state. (2.51) shows an informal formulation of *inertia* in terms of beliefs:

(2.51) If an agent A has a belief and if A has no reason to change this belief he will stick to it. One possible reason to change a belief about a state is knowledge about an event, which might have changed this state. Since normally, random events do not affect random states of affairs, an agent does not 'normally' have to change his beliefs about state of affairs.

Before this principle is formalized, consider again the bakery example: Assume that A wants to buy bread at some time point T_1. At T_1, A believes that around the corner from his flat there is a bakery store. A has this belief because there was a time point T_0 before T_1 where A also believed that the bakery store was around the corner and at T_1 A is ignorant about any event which might have had an effect on the bakery store. Even if there was a reason which caused the bakery store to be closed at T_1 (a flour dust explosion, lets say), it will not change A's beliefs concerning the bakery store since A simply does not know anything about this event. Hence, at T_1 A still has the belief that at T_1 there is a bakery store around the corner because 1. A believed at T_0 that there was a bakery store around the corner *and* because 2. A is ignorant about any event

which might have changed this. This whole scenario can be presented as in the box in (2.52) (where proposition $[S : [\mathsf{bakery.store}(x) \wedge \mathsf{around.corner}(x)]]$ is abbreviated by $K(S)$):

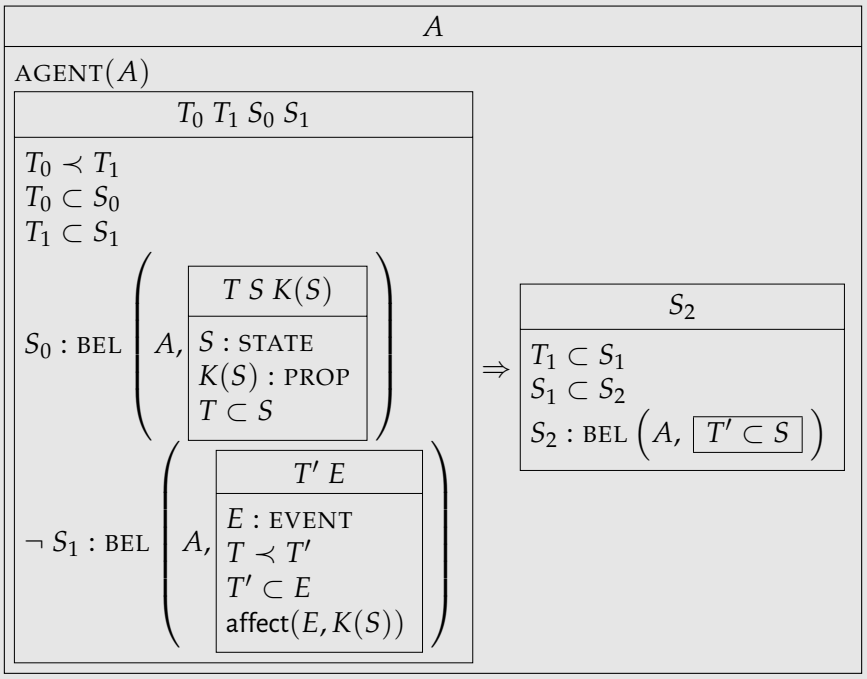
(2.52)



The presentation in (2.52) shows exactly the effect of inertia: A believes that at some time T_0 before T_1 a state S held true and is ignorant about any event E which might have changed this state. Therefore, A believes that S still holds true at T_1. The general formalization of inertia will basically be the same box as above but with some additional variables. These additional variables guarantee that it is possible to also apply inertia to states which do not lie in the agent's presence. For example, A may believe that there was a bakery store around the corner of his flat the day before yesterday and A does not have any evidence that this has changed in the meantime. Then A may also believe that there was a bakery store around the corner yesterday. In other words, the additional referents guarantee that the proposition in the scope of the belief operator can shift in time.

(2.53) INERTIA:

If an agent A is ignorant about an event which changes a state S, then A believes that S is still the case:



So far, this section was concerned with the motivation and formulation of inertia as a cognitive principle for human agents. The formulation of inertia which is proposed here is supposed to capture the intuition that as long as an agent is ignorant about an event which might have changed a state, the agent may assume that the state has not changed. However, in some cases states of affairs actually *do* change. For such cases, an inversed version of inertia is needed. This inversion will be called *learning*. This will be the topic of the next §2.4.2.

2.4.2 Learning

Consider again agent A who believed that there was a bakery store near his flat and who therefore planned to buy bread at that very place. But when A arrived at the bakery store, it is gone and replaced by a real estate agency. In this case, A must reconsider his beliefs and in particular, has to consider the fact that something has caused the bakery store to be gone. In other words, A *learned* that the bakery store is gone and that something must have affected the bakery store. In the following, this learning process is formalized.

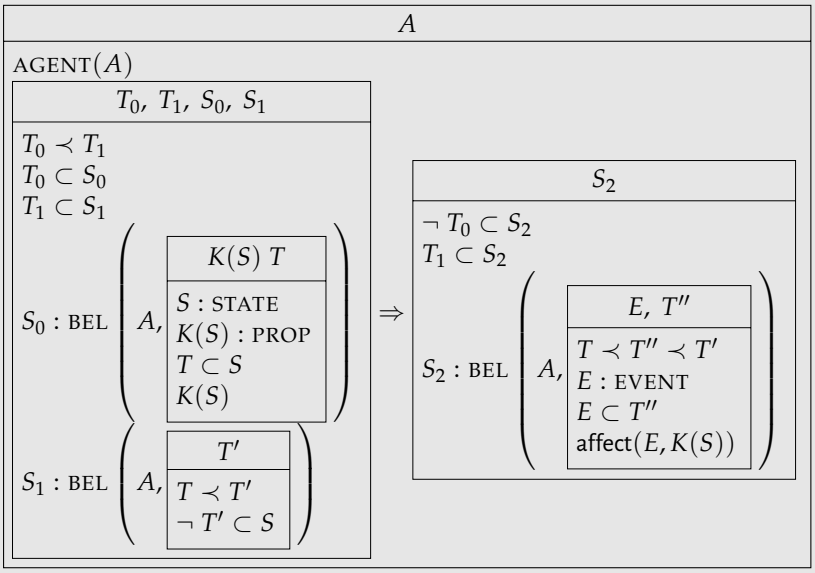
Learning is in some sense an inversion of inertia, as it models situations in which an agent finds out that the consequent of inertia turned out to be wrong, i.e. one of the conditions in the antecedent of inertia principle must have been wrong. In a more abstract version, learning models the following scenario: assume that an agent A is in a belief state concerning a state S and then learns that S is not the case anymore. But

since states do not end without a reason, A can assume that some event must have affected S. Take the bakery example again: at time T_1 agent A still believes that there is a bakery store around the corner of his flat. This is because A believed so at T_0 and since he has no counter-evidence that this state has changed. Hence at T_1, A decides to walk there to get some bread for breakfast. At the time point where A arrives at the bakery store, he sees that it is out of business. Therefore A has to reconsider his beliefs and can conclude that there must have been some event which caused the bakery store to be closed. In other words, A has to stop being ignorant about events which might have had an effect on the bakery store.

This can be formalized as follows:

(2.54) LEARNING:

If an agent learns that a state has changed, then the agent forms the belief that there was an event which must have affected this state:



Summary of this section: This section was concerned with the motivation and the formulation of two basic principles of rational agents' cognition, namely inertia and learning. Inertia models the assumption that an agent does not change his beliefs towards a particular state of affairs if there is no reason for the agent to change his beliefs. This principle states, in a certain sense, the "stability" of minds of human agents. Learning, on the other hand, models the assumption, that an agent who learns that some state has changed can conclude that there was a reason for this state to change.

2.5 Examples

This chapter closes by discussing some examples of human interaction and by applying the principles which have been developed so far for modeling the mental states

of the agents involved. The behind these examples is that human interaction is based on assumptions about the behavior of other agents. Agents and co-agents in social interaction make assumptions about the mental states of the respective other agents on the basis of the respective co-agents' behavior in order to make assumption on these other agent's future behavior. A crucial assumption is that the behavior of a rational agent is roughly in accordance with the mental states of this very agent. This last assumption will be explained in detail in 4.1.4. This section aims at showing how such assumptions of an agent about the mental states of a co-agent can be modeled in the framework which has been developed so far and, moreover, presents how calculations on the basis of the principles which already have been introduced can be performed. The calculative system which emerges from the principles above will be referred to as the *calculus of mental states*.

The rest of this section is concerned with the presentation of three examples. In 2.5.1 it will be shown how assertive speech acts restricts a hearer's assumptions about the speaker's behavior, and how this can be modeled using the cognitive principles above. §2.5.2 is concerned with assumptions about mental states of co-agents and shows that these are necessary in order to perform a joint activity. In particular, it shows that it is necessary for co-agency to assume that co-agents are subject to inertia. §2.5.3 finally, considers communication as a particular form of joint action in the sense of section 2.5.2.

2.5.1 First Application - Public Beliefs and Public Actions

In the following, it will be discussed how an agent's actions restrict another agent's assumptions about the first agent's possible behavior. In order to keep things simple, the action will be an assertive *speech act*. Chapter 3 will be concerned with general actions. At this point, it will be assumed that by asserting something the speaker adds the content of the assertion to the set of her public beliefs and that by default speakers act in accordance with their public beliefs. On the basis of this assumption, it is possible to perform calculations which have the mental states of the speaker as their subjects.

It is a common assumption in pragmatics that by asserting something, a speaker A adds the propositional content of the assertion to the set of propositions which can be called the set of her *public beliefs*. In other words, by asserting a proposition K, the speaker A is committed to the fact that she believes K. This effect of an assertion can be called the *assertive effect*. In chapter 4, this assumption will be replaced by the assumption that while communicating, interlocutors do not build up a set of shared beliefs but of mutual commitments. However, for simplicity reasons, this section is only concerned with the assertive effect. The following considerations are based on the formulation of the assertive effect by Condoravdi and Lauer, (2011):

(2.55) ASSERTIVE EFFECT: An assertion of a declarative φ in context C adds $[\varphi]_C$ to the public beliefs of the speaker, publicly committing the speaker to act as though he believes $[\varphi]_C$.

In order to make sense of (2.55), take the following situation as a non-example:

(2.56) Agent Ann utters a. on Monday towards Bob:

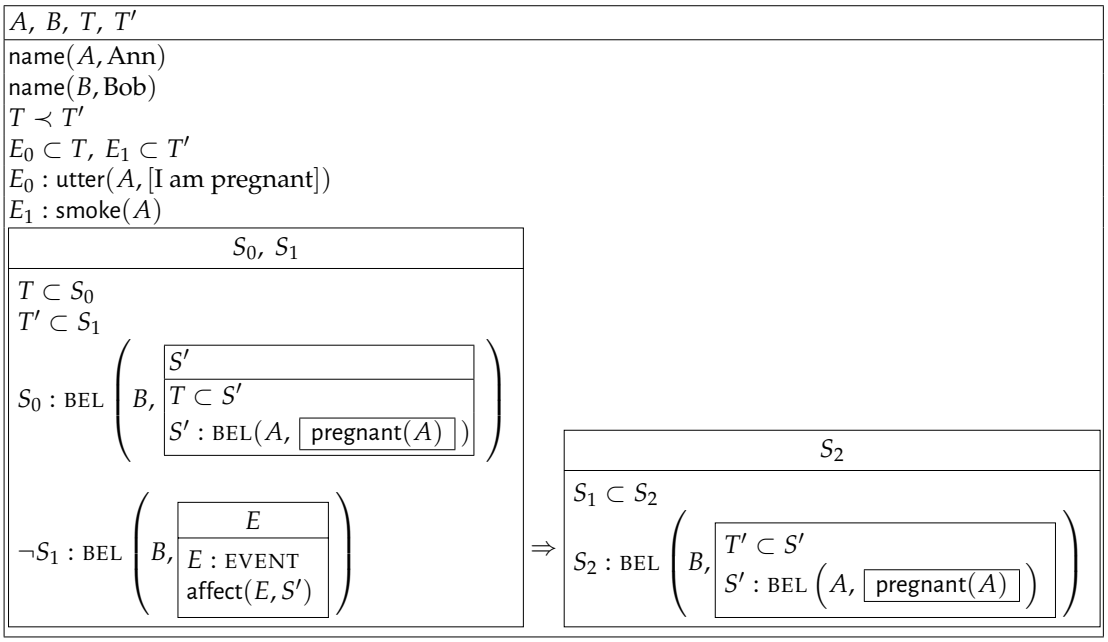
a. Ann: I am pregnant.

On Tuesday Bob sees Ann sitting in a pub, where she is smoking and drinking beer.

Probably Bob is surprised when he sees Ann in a pub where she is smoking and drinking. The rest of this section is devoted to explaining Bob's surprise on the basis of the cognitive principles which have been introduced so far. First, an informal discussion of the reasons for Bob's surprise is given. Afterwards, this discussion will be formalized in the calculus of mental states. According to (2.55) Ann commits herself to the proposition $K = [\text{BEL}(\text{Ann}, \mathsf{pregnant}(\text{Ann}))]$ by uttering something like (2.56.a), which in turn means that she has to act as though she believes K. Now Bob is surprised when he sees that Ann is drinking and smoking on Tuesday. In order to understand Bob's surprise, there are at least two stipulations which have to be made: 1. believing that one is pregnant nowadays implies that one has to try to live healthy and 2. drinking alcohol and smoking cigarettes is not considered healthy. Despite the fact that the meaning of *being surprised* cannot be implemented in the calculus so far, it can be explained precise why inertia and learning are important for understanding Bob's surprise in the situation of (2.56): First, assume that by uttering (2.56.a) at some time point T_0, Ann is committed to act on the proposition that at T_0 she is in a *belief state* that she is pregnant. Moreover assume that Ann is a person of which Bob believes that she cares for the health of her unborn. Under this assumption, Bob can now conclude that she is not acting on the proposition that while she is drinking she believes that she is pregnant. Therefore, Bob can conclude that Ann's belief that she is pregnant has ended and hence that she believes that something must have affected her pregnancy. Note that in this example we use learning twice: Right before Bob saw Ann drinking alcohol, Bob believed that Ann believed that Ann was pregnant, since up to this point Bob had had no reason to change his attitude towards Ann. Now, Bob can apply learning and can conclude that something in Ann's belief states must have changed and therefore, (again by learning, but now applied to Ann) Bob can conclude that Ann now believes, that something must have affected her pregnancy.

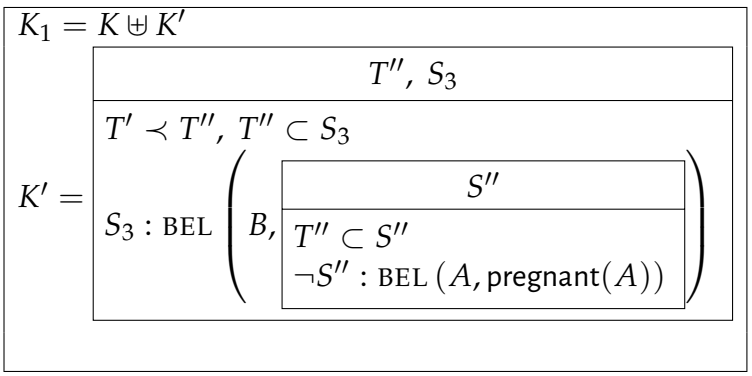
In the following this situation will be presented in the framework which has been developed so far. The DRS in (2.57) serves as an approximation for the scenario in (2.56) right before Bob was able to think about Ann's behavior. (just assume that Bob is not the fastest thinker and that he needed a while to process Ann's behavior):

(2.57)



So far, this presentation is just an application of inertia: Up to the point, when Bob was able to process Ann's behavior he had no reason to change his beliefs concerning Ann and, therefore, still believed that Ann believed she was pregnant. Now assume that at some time point T'' Bob inferred by Ann's behavior that at T' Ann does not believe that she is pregnant anymore. This means that at $T' \prec T''$ Bob is in a belief state S_3 in which he believes that Ann does not believe that she is pregnant anymore. In the following, let K denote the DRS in (2.57). The new situation can be described by the following presentation:

(2.58)



The situation in (2.58) is a situation in which *learning* can be applied: Bob believes that a certain state has ended and, therefore, he can conclude that there must have been an event, which has affected this state. After applying learning to $K_1 \cup K_2$, the following situation can be derived:

(2.59)
$$K_2 \Rightarrow \left[\begin{array}{l|l} & S_5 \\ \hline & T'' \subset S_5 \\ S_5 : \text{BEL} & \left(B, \left[\begin{array}{l} E,\ T'' \\ \hline T' \prec T''' \prec T'' \\ E : \text{EVENT} \\ E' \subset T''' \\ \mathsf{affect}(E', S' : \text{BEL}(A, \mathsf{pregnant}(A)) \end{array}\right) \right) \end{array}\right]$$

Hence, these calculations end in the result that after Bob learned that Ann was smoking and drinking, Bob is in a state of believing that something must have affected Ann's belief of being pregnant. Note that on the basis of this derivation, Bob cannot actually derive, that Ann is not pregnant, nor which exact event has affected her beliefs about her pregnancy. Bob can derive only, that there has been some event which has affected Ann's beliefs.

So far, a first application of the principles which were developed in the sections 2.3 and 2.4 was presented. It was shown, how an observer can derive assumptions about another agent's mental state on the basis of his actions. The next section is concerned with the question of co-agency, i.e. of cases of two agents performing a joint action (such as playing a flute and piano duet). There it will be shown, that for the performance of joint actions, inertia again plays an important role.

2.5.2 Second Application - Joint Action

In the following it will be shown, how inertia can be applied to an example of a joint action. In order to do this, consider the following example from Clark, (1996, p. 60). Take again Ann and Bob as agents and assume that they want to play a flute and piano duet together: Ann plays the flute while Bob plays the piano. In order to perform the joint action of playing a flute and piano duet, it is not enough for Ann to play the flute and for Bob to play the piano. In order for Ann to be able to play a flute and piano duet with Bob, she has to play the flute *on the assumption* that Bob will play the piano *in order* to play a flute and piano duet together with Ann, and vice versa. I.e. for the performance of a flute and piano duet, both participants have to assume the respective other participant will perform his or her part of the joint action. Clark, (1996, p. 61) formulated this condition on joint actions as follows:

(2.60) JOINT ACTION: Ensemble A-and-B is doing joint action k if and only if:

0. the action k includes 1 and 2;
1. A intends to be doing A's part of k and believes that 0;
2. B intends to be doing B's part of k and believes that 0.

The structure of this condition is circular and therefore it has similar problems as for example the classical definitions of common ground in update semantics: in order to be able to perform joint actions, both participants have to make an infinite number of assumptions on the respective other participant's mental states. A has to believe that

B has to believe that *A* has to believe ... that *B* is doing his part of *k*. The circularity in Clark's characterization of joint action may be another piece of advice for the assumption that *joint action* is generically different then single actions which happen to be performed simultaneously (Lauer, 2014; Rödl, 2018). On the other hand, Clark's assumptions in this characterization of joint action are reasonable from a practical point of view. Assume that Ann intends to do her part of the joint action of playing-a-piano-and-flute-duet, but that she, for example, believes also that Bob does not intend to do his part of the joint activity. In this case the event of both participants performing a piano-and-flute-duet will probably not come into existence. Ann's intention to perform a joint action will vanish if she does not believe that Bob does not intend to perform his part of the joint action. If Ann cannot rely about Bob's intention to perform his part of the duet, then she probably will not start playing her part of the duet either. Ann's assumption on Bob's mental state therefore is an actual obstruction for the performance of joint action and vice versa. Such obstructions have to be removed in order to successfully perform a joint activity.

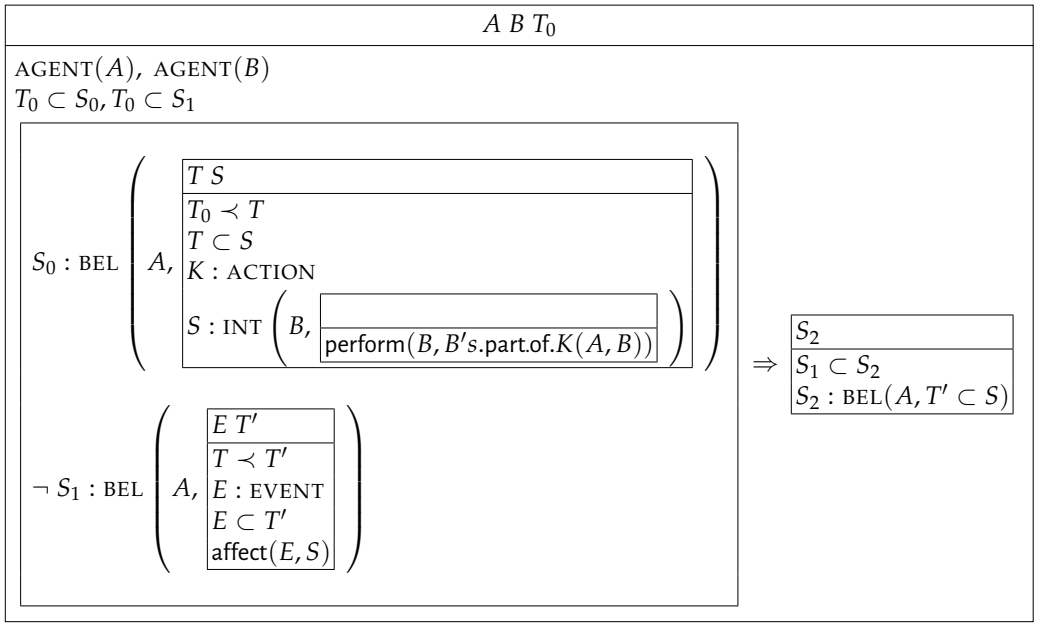
An aspect of joint action which is not explicitly mentioned in the definition above, but which also has to be taken care of, is the dynamic and temporal aspect of joint action. If a joint action is expanded in time, like for example a flute-and-piano-duet, the agents of the joint action have to believe that their respective partner intends to perform his or her part of the joint action as long as the joint action lasts. This implies that agents have to make assumptions about the future mental states of their co-agents. If agent *A* assumes that partner *B* will lose his intention to perform his part of the joint action in the next moment, partner *A* will probably lose his intention to perform his part of the joint action as well. Let's try to illustrate this with an example: assume that it is 7.55pm and Ann and Bob want to play a flute and piano duet together. Originally they planned to perform a whole piece with three movements in which each movement is about 5 minutes long. However, Ann knows that Bob usually leaves at 8pm and moreover that Bob has no idea what time it is (because Bob is sometimes a little bit confused and couldn't see the clock for quite a while). Moreover, assume that Ann knows that at 8pm the church bell from the church nearby will ringing and that Bob usually recognizes this bell. Therefore, Ann is in a situation in which she believes that Bob actually wants to perform his part of the joint action, but in which she also believes that there might be an event in the future (the church bell) that might have an effect on Bob's intentions. Therefore Ann can conclude that in the future there will be an event which might change a necessary condition for performing the joint action of playing flute and piano duet together: Bob could lose his intention to perform his part of the joint action when he hears the church bell! As a result, the ensemble have to communicate about their plans to perform a joint action. For example, they can agree to play only one of the three movements of the duet or they can agree to not start to play the duet at all, or Bob can decide to stay a little longer then usual.

After these considerations, let's come back to inertia. Inertia plays a central role in joint action. Before agents agree on the performance of a joint action, such as playing a piano-and-flute-duet (or playing soccer etc.) they have to make assumptions about the stability of the mental states of the other agents: they have to assume that there is not much fluctuation in the other agents' intention and belief states. Otherwise they

would not agree to performing a joint action. To make such an assumption is basically an application of inertia in a forward looking version. The agent has to assume that nothing will change the mind of his co-agent in the future, i.e. he has to assume that his co-agent's intention to perform his part of the joint action will remain stable. In the following, these condition for joint action will be discussed. For reasons of simplicity, the focus is on just one of the co-agents, lets say A. The conditions for joint action for the other co-agent are of course symmetrical. Note, moreover, that the following formulation of the conditions for a joint action does not take the circularity of (2.60) into account.

In order to perform a joint action at some point T, agent A has to believe that B intends to do his part of the joint action of A and B at T. Moreover, A has to assume that there is no future event which will change B's intention to perform his part of the joint action. In other words, in order to perform a joint action with B, A has to be ignorant about any events which might change B's mind and hence he can assume that B intends to do his part of the joint action, as long as the joint actions lasts. In order to formalize this, assume that there is a part-of-joint-action predicate part.of. The following representation describes the scenario at a time point T, right before the agent begins with the performance of the joint action.

(2.61)



In order to be able to perform a joint action, it is necessary for each of the agents to be able to rely on his co-agent, i.e. it is necessary for agents to be able to assume that their co-agents do not change their minds randomly. This in turn means that an agent has to assume, that his co-agent is subject to inertia.

Note that the box in (2.61) contains only first order recursions of mental state, i.e. it only contains an agents belief's about another agent's intentions. As already mentioned, Clark's definition of joint actions includes a cascade of recursive mental states. Given that agents have only a limited amount of cognitive resources, it is a reasonable question to ask whether such a cascade of recursive mental states as a condition for joint action matches reality. The question of how many recursions of mental states are

necessary for modeling human communication or human interaction is the source for a difficult debate in which I do not want to get involved at this point. An overview of this debate for the case of the recognition of speaker intentions can be found in Harras, (1983, Chap. II).

2.5.3 Third Application - Communication as Joint Action:

The rest of this section discusses informally the question what communication has in common with and how it differs from other prototypical examples of joint action. At the first glance, communication is a paradigmatic case of joint action. In pragmatics, it is even sometimes assumed that *any* communication serves the greater goal of finding an agreement about the states of affairs in the world (Roberts, 1996). In this sense, communication is a joint action in which the interlocutors have the common goal to learn about the state of affairs in the world. However, this is of course not always the goal of communication. In many cases, communication is not about general states of affairs of the world, but about coordination of future actions or just about arguing. At least cases of arguments between interlocutors do not seem to be examples of joint actions. Another aspect of communication which makes it a special case of joint action is the fact, that any interlocutor can pursue his own communicative strategy. The concrete goals and strategies which speakers use in order to achieve their communicative goals can change very fast. This fluctuation in the interlocutors' concrete goals an strategies turns communication into a very dynamic case of joint action. The exact notion of goals and communicative goals will be defined precisely in the next chapters (cp. 3.1 for the definition of *goals* and plans and 4.1 for the definition of a *communicative goal*). However, for the purpose of this discussion it will be enough to assume that the goal of an utterance is to persuade the hearer about the content of the utterance. The high fluctuation of goals and strategies in a conversation, in turn, means that conversational participants have to continuously be able to change their assumptions about the conversational goals of the other dialogue participants. In this sense, communication differs from other types of joint action, such as playing a flute-and-piano duet, where the goal of the joint action is 'more robust': the goal of the joint action of performing a piece of music stays the same during the joint activity, whereas in a communicative situation, goals can be achieved quickly fast, speakers can dismiss their goals, can construct new communicative strategies and so on.

In order to illustrate this, let us have a look at the following example of a conversational situation taken from Ginzburg, (2012, p. 82). It is assumed that *A* and *B* are preparing a guest list for a party:

(2.62)
*A*1: Who will agree to come?
*B*2: Helen.
*A*3: I doubt Helen will want to come after last time.
*B*4: Nah, I think she's forgiven and forgotten.
*A*5: No, she definitely has not.
*B*6: OK! What about Sue...

During this conversation, the interlocutors are dealing with different opinions and

hence different goals. This is a result of the fact that the interlocutors have completely different beliefs with respect to Helen: in $A3$ speaker A makes it public that he believes that Helen does not want to come to the party while B seems to have the opposite belief. This in turn means that at this point A and B have two different goals: A wants to make it a common belief, that Helen doesn't want to come, whereas B wants to make it a common belief that Helen wants to come. Hence, this is a situation where both participants in conversation seem to have completely different goals, and in particular, different intentions. However, in the definition of joint action, both agents have more or less the same intention, namely to perform their respective part of the joint action. Now the question arises, how conversation can be a joint action if there are competing goals in a conversation!? The tension between the intuition that communication is a case of joint action and the fact that in communication, agents pursue different goals by different utterances vanishes if one differentiates between *the goal of a conversation* and the *goal of an utterance*. All communication has a purpose, and this purpose coordinates the agents' actions in the communicative situation. Therefore, any utterance in the conversation can be regarded as an attempt to serve the purpose of the conversation. In the case of (2.62), it is the purpose of A's and B's actions to plan a party together and this purpose coordinates their actions. A and B have to manage the task of coming-up-with-a-guest-list and therefore it is A's and B's respective part in this joint action to say anything they know that can be relevant in this situation. Hence, after A and B have agreed upon the fact that they want to perform the joint action of making a guest list, it is their task to come up with names of people to put on this list. After Helen's name popped up they now have to decide whether to put Helen's name on the list. After accepting this as a subtask it is again A's and B's part of the joint action to say anything they know about Helen that might be relevant in this situation. After $A3$ the discourse participants have different opinions concerning Helen, and therefore different conversational goals. But despite the fact that the discourse participants' goals differ, the conversation does not come to an end because negotiating about Helen's case is a subtask of preparing a guest list. The interlocutors are in a joint action and therefore they can assume that the respective other interlocutor's intention will not suddenly come to an end, again because of the same argument as in 2.5.2: when performing a joint action, the co-agents have to rely on the fact that they will do their respective part in the action. This guarantees that the interlocutors will stick to their plan of making a guest list, even though their conversational goals for performing the subtask of this joint action may differ. The assumption that conversation is goal-directed in the sense that interlocutors (at least implicitly) agree upon which kind of conversation they are having, will be explained precisely in 6.2.1.

2.6 Summary

In this chapter, the basic rules for a calculus system of mental states were proposed. This calculus system is formulated within the framework of discourse representation theory, which was introduced in the beginning of the chapter. Moreover, the propositional attitudes and attitude description sets were introducedas a formal device to

model the mental states of agents within DRT. Afterwards, some cognitive principles which describe properties of propositional attitudes and which serve as axioms for the formal system in which calculations with propositional attitudes are performed were introduced. Finally, an example of a calculation within the developed calculus was given and the role of inertia for social interaction and communication was discussed.

The insights and rules which have been developed in this chapter will be the basis for the whole rest of this book, but in particular will be used in the next chapter where phenomena such as goals, plans or actions will be considered. In particular, goals will be defined as certain configurations of propositional attitudes. This convention will make goals mental states themselves. Therefore, it will be possible to perform calculations with these phenomena in a similar fashion as in 2.5.

Chapter 3

Actions, Goals and Plans

It is a major interest of this book to gain a better understanding of discourse structure and a better understanding of the recognition of rhetorical relations. The strategy which is pursued in this book is based on the assumption that recognizing a rhetorical relation is basically a plan recognition task. In order to recognize a rhetorical relation, the hearer has to form assumptions about the speaker's communicative goals and these assumptions enable the hearer to connect individual speech acts. The idea for this approach is the following: By observing an agent's actions, the observer makes guesses about the agent's goals and plans. If the observer has made an assumption on the goals of the agent she observed, she is able to bring the agent's various actions into a meaningful relation. This approach is of course not limited to the recognition of rhetorical relations. In fact, the recognition of rhetorical relations is just a special case of the plan recognition in general. Consider for example, the following non-linguistic situation: take an agent A who observes an agent B who is walking in the direction of a train station. Moreover, assume that B has a look at his watch and starts running in that very direction. In order to make sense of B's behavior, A has to relate B's different actions somehow. Given a little bit of world knowledge, it is not that complicated for A to do this: A may assume that B wants to arrive at the train station to catch a train which makes it possible to connect B's looking.at.watch and B's start.running. A just has to assume that B wants to catch a train and this assumption enables A to guess that the look at his watch told B that he is already late and that this knowledge caused B to start running. A's assumptions about B's goals make it possible for A to bring B's actions into a meaningful relation and enable A to construct a coherent behavior from these actions. Of course, A's assumptions about B's goals can in general be wrong. Maybe B does not want to get a train but wants to pick up a friend at the train station whose train will arrive in a few minutes or B is actually doing some interval training. Nevertheless, A's reasoning about B's goal still enables A to put B's action into a meaningful relation (even though A is wrong).

But what about discourse structure? The idea pursued in this work is that the process of recognizing rhetorical relations from distinct speech acts is similar to the process

of constructing a coherent behavior from distinct actions of agents. Take for example the following pair of speech acts:

(3.1) a. Could you close the window, please!?
b. It's cold in here.

In the case of (3.1) it is reasonable to assume that it is the purpose of (3.1.b) to explain why the speaker wants the window to be closed. But how is the hearer or the addressee of (3.1) able to draw this conclusion? Similar to the observer in the previous example who sees agent *B* looking at the watch and starting to run, the hearer of (3.1) has to bring the two distinct speech acts into a meaningful relation somehow. The problem is again, given a little bit of world knowledge, how is the hearer of (3.1) is able to construct a coherent discourse from the linguistic input? First of all, one can tell that the speaker has the goal of making the addressee close the window. In order to persuade his addressee of the fact that it is a good idea to close the window, the speaker gives additional reasons. Given the knowledge that a closed or open window influences the temperature in a room, the hearer can draw the conclusion that the speaker utters (3.1.b) *because* he wants his addressee to close the window. In other words, the two utterances in (3.1) are connected by the speaker's goal of making his addressee close the window. Similar to the non-linguistic example, these two utterances are connected by a common goal, which turn them from isolated speech acts into a coherent discourse. Therefore, it is the hearer's task to recognize the speaker's goals and plans in order to be able to construct a coherent discourse from isolated speech acts. In order to provide a better understanding of the recognition of a speaker's communicative goals and plans and, hence, for the reconstruction of rhetorical relations, this chapter seeks to provide a better understanding of the reconstruction of general coherent behavior of agents and to include notions, such as goals and plans in the theory.

Plans, however, are a very complicated matter. In the very first lines of his essay "Time, Rationality, and Self-Governance", Bratman, (2018). writes:

> Much adult human activity is temporally extended. It involves distinctive forms of cross-temporal guidance and organization. And the agent's plan-like commitment to these cross-temporal forms of organization is a central element in the explanation and understanding of her practical thinking and action. *Bratman, 2018*

In this quote, Bratman emphasizes the assumption that an agent's commitment to a plan is central for explaining the agent's actions. From the point of view of an observer, this means that plans are central for turning an agent's actions into a coherent behavior. Bratman continuous by giving the example of an agent who wants to build a house. Building a house is a complex action which involves lots of forms of organization: One has to make plans and making these plans again depends on knowledge about former events and achievements. An agent who wants to build a house has to form a set of goals which are somehow causally related to the goal of building a house. All goals which subordinated to the project of building a house can be referred to as *subgoals* of building a house. Some of these subgoals are independent of each other, some are related to each other in the sense that these goals determine a temporal order in which they can be achieved: the roof cannot be constructed before the load-bearing walls are

completed, the walls cannot be painted before they are constructed, etc. On the other hand, normally it makes no difference whether one paints the walls of the living room first or the walls of the toilet. Nevertheless, all these goals are related to each other by the larger plan of building a house.

Since building a house is a much to complicated example to examine in detail, consider first (3.2):

(3.2) Suppose that an agent A wants to have a cup of tea. Therefore, he develops the following plan:

1. fill the kettle with water
2. turn on the stove
3. put the kettle on the stove
4. put a tea bag in the cup
5. pour the boiling water into the cup

Some of the actions in (3.2) are independent from each other but some depend on former actions. For example, to make a cup of tea one needs hot water, which means that one has to put a kettle with water on the stove *before* one can use this water for the tea and pour it into the cup. On the other hand, putting a tea bag in a cup is independent of filling a kettle with water. But this means that the order in which these actions are performed doesn't matter at all. Hence, we see that some of the action in (3.2) depend on each other and some do not. These dependencies can be represented in a "dependency tree" as shown in figure 3.1.

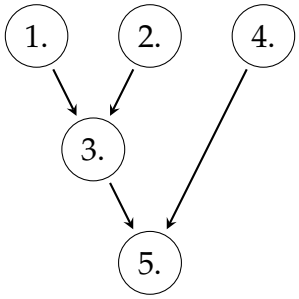


Figure 3.1: The dependencies among goals of A's plan in (3.2). The arrows indicate a direct dependency, i.e. the action given in 3. depends on action 1. and on action 2. Action 5. depends on action 3. and on action 4., but action 3. is independent from action 4.

The structure in figure 3.1 is of course already highly simplified. Each of the actions above can be subdivided into "smaller" action units which again constitute larger actions. For example 1. can be subdivided in the following actions:

1a. hold the kettle under the water tap
1b. turn on the water tap.

Moreover, 1a. can again be subdivided into a grabbing.the.kettle event and a carrying.the.kettle event and so on. The question whether there are atomic elements of an action and how one can identify these elements is another difficult problem which will

not be approached in this book[1]. For reasons of simplicity, it will be assumed that there are certain events which classify as actions. A collection of actions of which an agent believes that the performance of these actions in a particular order make the agent achieve a particular goal will be called a *plan* for that goal. Despite the fact that it is unclear when and whether the subdivision process for the actions of an agent stops, the collection of actions in (3.2) seems to be a reasonable plan for achieving the goal of having a cup of tea. The fact that distinct actions of the agent in (3.2) belong to the same plan turns these distinct actions into a coherent behavior of this agent.

This chapter aims at introducing the notions of actions, goals and plans. In section 3.1, the notions of a *goal* of an agent and of *pursuing a goal* will be introduced. Pursuing a goal will be defined as a particular combination of propositional attitudes towards the agent's goal. Moreover, the notion of a *subgoal* as a goal which helps to achieve the original goal of the agent will be introduced. Since an agent's goals and subgoals depend heavily on the agent's assumptions about causalities, this section is preceded by a short discussion of an agent's reasoning about causalities. After having introduced the notion of goals, section 3.2 is concerned with introducing the notions of an agent's plans. Moreover, so called goal-directed act will be introduced. These are certain actions of which the agent believes that they help to achieve his goals will be discussed. Finally, section 3.3 places a particular focus on an observer observing the agent an agent in order to understand how the observer is able to make guesses about the agent's goals and plans and how the observer turns distinct actions of the agent into a coherent behavior. Furthermore, this section introduces some additional common sense principles which model assumptions about an agent's reasons to act. The results of this part will be of particular interest in chapter 5 which will be concerned with the hearer's construction of a coherent discourse.

3.1 Goals and Subgoals as Mental Attitudes

As already pointed out, this chapter seeks to provide a better understanding of the mechanisms behind the recognition of another agent's goals and plans. In order to to do this, it will be assumed that agents in general act in order to achieve their goals. This book does not provide any theory of how exactly the agent's goals come into existence. But it will be assumed that, once an agent pursues a goal, he will perform actions in order to achieve that goal. Take again the example of an agent A who wants to have a cup of tea. At first, of course, this describes merely the agent's *desire* but under certain conditions (which will be clarified later) this desire may become one of A's goals and having this goal may force A to *act*, which could mean, for example, that A starts to fill a kettle with water, puts it on the stove and so on. The assumption that goals are the reasons for an agent to act makes it more or less neccessary to have an operationable notion of goals. This section, therefore, aims at providing such a notion.

Following Pollack, (1990), we will assume that goals are complex mental states. Formally, this can be represented by defining goals as configurations of propositional

[1]For an early discussion of the question, when actions can be considered as atomic actions, see Anscombe, (1957).

attitudes. (3.3) already shows an informal version of the notion goals which we want to develop in this section.

(3.3) GOALS: A state of affairs G is a *goal* of an agent A if the following conditions hold:

1. A believes that it is possible that G will hold at a time point in the future.
2. A believes that there are actions which can be performed by A and which will cause G to hold.
3. A desires G to hold in the future.
4. A intends to perform some of the actions of 2.

As already pointed out, it is assumed that goals are the reasons for agents to act: an agent performs a certain action because she believes that this action will cause a particular state to hold. This assumption is captured in (3.3) by the second condition. This assumption expresses a tight connection between the notion of goals and the notion of *causalities*. Having a goal *causes* an agent to act in a way of which she believes will *cause* a particular state of affairs. Hence, causation plays at least two roles for goals and acts: having a goal causes actions of which the agent believes that they are causally connected to the goal.

The fact that assumptions about the possible realization of goals play a central role in the definition of goals make it possible to distinguish goals from mere desires. For example, I may have a desire for sunshine but because of my inability to change the weather, this desire will not become a goal of mine and will not make me act in a certain way. In other words, a desire only has a chance to become an agent's goal if the agent has assumptions about actions of his own which cause the realization of that desire. It therefore seems to be imperative to have at least some basic ideas about reasoning about causalities. This section, therefore, begins with a discussion on general reasoning about causalities.

The general structure of this section is the following: In 3.1.1 a short discussion on reasoning about causalities is given. §3.1.2 defines *desired* goals as the class of goals which make agents act. §3.1.3 defines *necessary* goals as goals which are not primarily desired but which the agent believes will help him to achieve a desired goal. This notion is used again in 3.1.4, where the *subgoal relation* and *goal stacks* are introduced. The subgoal relation describes the fact that necessary goals are believed to be means to achieve a desired goal. Goal stacks, on the other hand, are defined as being finite sets of goals which are partially ordered by the subgoal relation. One can consider goal stacks as preliminary versions of *plans* which we will define in section 3.2. Finally, 3.1.5 places its focus on the mental state of an agent who *is pursuing* a goal which will be defined as a mental state an agent who has a goal is actually in. Having this additional notion makes it possible to differentiate between the agent's goal and his mental state which is connected with this goal.

3.1.1 Reasoning about Causalities

Agents act in order to achieve some goal – at least this is the basic assumption on which the theory of action which will be developed in this chapter is based. This will be made more precise in 3.2.1 where the notion of goal-directed acts will be introduced. Moreover, it is assumed that the agent believes that her actions help to *achieve* her goal. Under this assumption, one therefore has to deal with at least two different types of causalities: causalities which hold between the agent's mental states and her actions and the agent's assumptions about causalities in the outside world. As already pointed out, pursuing a goal implies that the agent believes that there are actions which she is able to perform and which she believes help to achieve her goals. Since we assume that certain mental states cause agents to perform certain actions the theory of actions which will be developed in this chapter can be classified as a *causal theory of action* (McCullagh, 1975).

For including causalities in the formal framework which is developed so far, the following notational convention will be used:

(3.4) For events E_1 and E_2 write $E_1 \twoheadrightarrow E_2$ iff an event E_1 causes an event E_2.

For a semantic interpretation of the symbol $\twoheadrightarrow$ in terms of possible worlds see Lewis, (1973). The rest of this part, however, will be concerned with the question of how agents come to assumptions about causalities.

As an example of a situation where an agent has to deal with causalities, take the following scenario: Suppose that an agent A wants to illuminate a room and that there is a light switch and bulb in this room. For A it is reasonable to assume that flipping the light switch will cause the bulb to go on and, hence, will cause the room to be illuminated. This situation can be described as follows:

(3.5) $$\text{BEL}\left(A, \boxed{\begin{array}{c}\text{flip light switch}\\ \downarrow\!\!\downarrow \\ \text{light goes on}\end{array}}\right)$$

However, it is of course well known that flipping the light switch is just one event in an infinite network of other states of affairs and events which stand in a causal relationship to the event of illuminating the room (Hobbs, 2005): the bulb has to be in working condition, the switch has to be connected to the lamp, the lamp has to be connected to a power plant ... Of course, in a situation in which A wants to illuminate a room and sees a light switch, A cannot consider all these different eventualities.

Following Hobbs, the set of all different eventualities which stand in a relation to another event will be called a *causal network*. A priori, causal networks have an infinite number of elements and therefore, almost all eventualities (i.e. all but a finite subset) in a causal network have to be ignored by an agent in order to be able to perform actions. A's action of flipping the light switch is based on the assumption that this action will cause the light to go on and in this case, A ignores other necessary eventualities in the causal network.

In his talk *Das Risiko der Kausalität* (*The Risk of Causality*) Luhmann, (2017) summarized the problem that we have to deal with infinite causal structures as follows:

> An sich weiß man, dass mit Kausalität eine nach zwei Richtungen offene Unendlichkeit gemeint ist - eine Unendlichkeit von voraussetzenden Ursachen und eine Unendlichkeit von weiteren Wirkungen. Dabei geht es nicht nur um lineare Strukturen, sondern um kaskadenförmige Vermehrungen von Mitursachen und Nebenwirkungen;
>
> *In principle it is known that causality describes an infinity with two open ends - one of which is an infinity of causes and the other one an infinity of further effects. This involves not only linear structures, but cascade shaped multiplications of contributory causes and side effects*[2];

Luhmann describes knowledge about causal relations as a form of *schematic ignorance* (schematisiertes Nichtwissen): in order to be able to make assumptions about causalities, one has to exclude almost all elements of a causal network. In this sense, the structure of reasoning about causality can be compared to the structure of the common sense principle of *inertia* (cp. section 2.4.1), which was also described as a form of organized (or schematic) ignorance: Inertia states that the agent is ignorant about almost all events in the world which can affect a certain state of affairs; reasoning about causality implies that the agent is ignorant about almost all eventualities in a causal network.

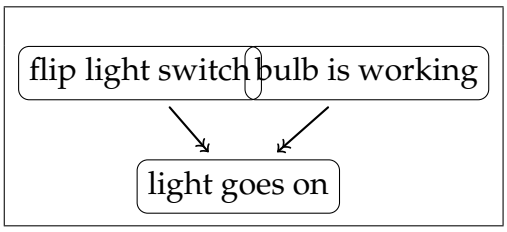
The fact that an agent has to ignore almost all eventualities in a causal network is again due to the agent's limited cognitive capacities. However, one still has to act on assumptions about causalities in order to be able to act at all. But given the complexity of causal networks, how can this be done at all? One way to deal with limited cognitive capacities and the lack of knowledge about causal relationships in the world is to move from a strict notion of causality to a notion of causality which involves *uncertainty*. A possible way to include uncertainty in a model of causality is to include probabilities in the calculation of causal relationships. This approach was proposed, for example, by Pearl, (2009). If one says that an event E causes another event E' *with some probability*, then it is the job of this probability to cover an infinite number of possible events and states which are "somehow" causally connected to E' but which have to be excluded from the causal relationship in question: in this case it is the job of the involved probabilities to cover the missing knowledge about other elements in the causal network. For rational agents' reasoning concerning causalities, such an approach can be modeled with probabilistic epistemic operators (Demey and Sack, 2015), i.e. with belief operators which assign a probability to propositions. For example, if P_A denotes a probabilistic belief operator for A, then the formula $P_A(\varphi \twoheadrightarrow \psi) = 0.95$ says that agent A is 95% sure that φ causes ψ. For example, if φ =[flip light switch] and ψ =[illuminate room], the expression above means that A is by 95% sure, that flipping the light switch will cause illumination of the room. The 5% which are left cover the missing knowledge about the condition of the wires, the light bulb, the power plant, the lack of knowledge about the physical properties of electricity and so on.

This book, however, pursues another approach to reasoning about causalities. In the following, it will be assumed that agents in general try to construct a "minimal"

[2] My translation (FZ).

causal network and that they only add new elements to this network in cases where the former causal network is not sufficient for understanding the given phenomena. In other words, agents in general try to keep things simple. This approach can be considered a causality theoretical version of *Occam's razor*: keep your explanations for the states of affairs in the world as simple as possible!

As an illustration of this principle, take again agent A who wants to illuminate a room by flipping the light switch. In cases where flipping the light switch results in illumination of the room, A can be completely satisfied with a version of a causal network where flipping the light switch results directly in illumination of the room. However, in cases where flipping the light switch does *not* result in illumination of the room, A has to reconsider her assumptions about the causal network in question. In this case, A has to take other eventualities in the causal network into account, but again in a reasonable *and* simple manner. For example, in a case where flipping the light switch does not result in an illuminated room, A should not question the condition of the power plant but rather the condition of the light bulb. This means that in cases where a room remains dark even though a light switch was flipped, the agent will adjust her mental model of the causal network in a way that keeps it as small as possible but includes the condition of the light bulb.

(3.6) BEL $\Big(A,$ 

$\Big)$

This type of reasoning about causalities turn it into a non-monotonic type of reasoning in the sense that agents have to reconsider their former assumptions about causal relations in the world (Shohan, 1990). Moreover this reconsideration process is a case of a counterfactual reasoning (cp. Lewis, 1973): the agent has to think about the states of affairs in which the light bulb *would have worked*. This makes reasoning about causal networks a modal operation since it involves reasoning about possible worlds. In a world in which the light bulb worked and everything else stays the same, flipping the light switch would have caused the light to go on. But since the light does not go on, one has to look for the closest possible world in which it actually would have gone on, and has to consider the differences between the two given worlds. Of course, the actual process of adjusting assumptions about a causal network depends heavily on the agent, his knowledge and his personal experience. But again, in order to keep things simple, this book deals with an "idealized" agent with basic assumptions about things in the world.

After having discussed assumptions about causal networks, it is possible to reformulate the notion of causalities in (3.4) into a version that fits the purposes of this book:

(3.7) An agent A is said to believe that an event E' causes an event E if A believes that E and E are elements of the same causal network and if A believes that E'

temporally precedes E. If this is the case, write $\text{BEL}(A, E' \twoheadrightarrow E)$.

In cases where causal networks actually have to be used, the agent's assumptions about causal networks will simply be stipulated. Reconsidering causal networks will play a role again in section 3.2, where reconsidering a *plan* will be discussed. Note that (3.7) did not actually make any claims or stipulations concerning real causalities in the world but that it deals only with an agent's *assumptions* about causalities. This means that no questions concerning the ontological nature of causalities have to be considered nor whether it can be experienced or be proven. Even the agent's philosophical position concerning causalities does not really play a role in this approach. The only causalities which are used in this book are actually an agent's assumptions about causalities in the world.

3.1.2 Goals which make Agents Act

After the discussion of the role of causalities for actions and goals in 3.1.1 this section seeks to provide the definition of the class of goals which are believed to be reasons for agents to act. Such goals are characterized by the property that they describe a state of affairs in the world which is actually *desired* by the agent in question. This stands in contrast to goals which have to be achieved by an agent in order to achieve a desired goal. For example, an agent may desire a cup of tea which can become a goal for that agent. But in order to have a cup of tea, the agent needs to have hot water. This makes the state of "having hot water" a necessary goal which has to be achieved in order to achieve the actual goal of having a cup of tea. However, having hot water cannot be considered an actual desire of that agent and, hence, it is not a goal which makes him act.

Therefore, two different types of goals will be discussed: goals wich are actually desired and make agents act and goals which are means to achive these goals. The first type of goals will be referred to as *desired goals*, the second type of goals, on the other hand, will be referred to as *necessary goals*. Neccessary goals will be the subject of 3.1.3.

By intuition, a goal is a *desired goal* of an agent if it is a state of affairs she would like to achieve by means of her actions. Take again the agent A who wants to illuminate a room. We may assume that A wishes the room to be illuminated which means that she is in a state of *desiring* illumination of the room. Moreover, assume that A actually believes that the illumination of the room is possible and can be achieved by flipping a light switch which will cause the light in the room to go on. This means that A is in a certain *belief* state: she believes that the desired state can be achieved and that she is able to perform an action which will cause the desired state of affairs to hold. Finally, A forms the intention actually flip the light switch, i.e. A forms a particular intention towards some of the actions which will help her achieving the desired state of affairs. In this example, A's mental state consists of a certain configuration of attitudes, i.e. a certain belief state, a desire state and an intention and, moreover, these states interact, in the sense that the content of one attitude depends on the content of each of the other attitudes. In this book, goals will , therefore, be assumed to be mental states

which consist of the configuration which was discussed for the case of the agent A who wants to illuminate the room. This can be summarized in a proper definition as in (3.8), where desired goals are also referred to as *d-goals*.

(3.8) D-GOALS: A state of affairs G is a *d-goal* of an agent A at time T if the following conditions hold:

1. A believes that it is possible for G to hold at a time point in the future.
2. A believes that there are actions which can be performed by A and which will cause G to hold.
3. A desires G to hold in the future.
4. A intends to perform some of the actions of 2:

$$
\begin{array}{|l|}
\hline
A\ T \\
\hline
\textsc{agent}(A) \\
T : \textsc{time} \\
\textsc{DGoal}_A^T(G) :\Leftrightarrow \\
\quad
\begin{array}{|l|}
\hline
S_{\textsc{DES}}\ S_{\textsc{BEL}}\ S_{\textsc{INT}} \\
\hline
T \subset S_G \\
S_G = S_{\textsc{DES}} \oplus S_{\textsc{BEL}} \oplus S_{\textsc{INT}} \\
S_{\textsc{BEL}} : \textsc{BEL} \left(A, \begin{array}{|l|}\hline G\ T'\ k \\ \hline T \prec T' \\ \Diamond G@T' \\ k : \textsc{ACTION} \\ \mathsf{perform}(A,k) \twoheadrightarrow G \\ \neg(T \subset G) \\ \hline \end{array} \right) \\
S_{\textsc{DES}} : \textsc{DES} \left(A, \boxed{G@T'} \right) \\
S_{\textsc{INT}} : \textsc{INT} \left(A, \begin{array}{|l|}\hline T''\ E' \\ \hline T \prec T'' \prec T' \\ E'' \subset T'' \\ E'' : \mathsf{perform}(A,K) \\ \hline \end{array} \right) \\
\hline
\end{array} \\
\hline
\end{array}
$$

Note that this formulation makes use of parasitism of propositional attitudes (cp. 2.2.2): some of the referents which are introduced in the scope of one propositional attitude occur also in the scope of another. Moreover, the notion of goals emphasizes the assumption that an agent's goal depends only upon the agent's mental states, i.e. his assumptions about states of affairs in the world, assumptions about causalities, his desires and intentions. As a consequence of this definition of goals, there may be states of affairs which are goals of an agent but which are actually not achievable. This captures the fact that there are cases in which agents *try* to achieve certain goals by means of certain actions but fail because the goal is actually not achievable. Note further that we have defined goals as a certain mental states directed at a particular state of affairs in the world. In the following, this state of affairs will be referred to as the *goal state* or the *goal* of the agent; the mental state which is described in (3.8), on the other hand, will be referred to as the state of *pursuing* a goal.

Of course, one may wonder whether all the conditions in (3.8) are actually needed in order to make sense of the notion of a desired goal. The rest of this part is used to defend the individual conditions one by one against this objection. However, all these defenses are based on intuition and can of course be the target of further objections themselves. In the following, desired goals are simply refereed to as goals:

1. *A goal is a state which the agent believes does not hold yet and that it is possible that it holds in the future*: This should be rather uncontroversial. If an agent already has a car, then he cannot say that it is his goal to have a car. (although he can, of course, pursue the goal of having *another* car). Only states which do not hold yet should be considered goals. This can also be regarded as a consequence of the assumption that beliefs and desires towards the same proposition cannot occur at the same time (cp. principle (2.38)). The assumption that the agent has to consider a goal to be at least possible should also be quite uncontroversial, and is also implied by 2.

2. *A goal is a state which the agent believes can be achieved by performing certain actions:* It is unintuitive to say that a state of affairs is an agent's goal if the agent isn't able to achieve the goal by any of her actions. For example, an agent may wish to have sunshine at her birthday party. This is certainly a plausible desire but is impossible for the agent to achieve by her actions. Therefore, having sunshine at the birthday party does not qualify as an agent's possible goal.

3. *A goal is a desired state*: This assumption is uncontroversial for this particular type of goal.

4. *The agent intends to perform some of the actions of 2.*: Assume that the agent desires a state and knows about certain actions he can perform in order to achieve this state but does *not* intend to perform these actions (maybe because he is lazy or because of a lack of time, etc.). In this case the agent is in a state of a mere belief-desire combination towards this state, but this does not have to lead to an *intention* (cp. Searle, 2005). Assume, for example, that a boy is in love with a girl and wants her to be his girlfriend. Moreover, assume that he considers this state to actually be possible. However, the boy can actually be very shy, and in fact so shy that he does not even dare to talk to the girl. In this case the boy is in a desire-belief combination towards a particular state, but does not develop the intention to perform actions which might result in the desired state. In this case all the conditions 1.,2. and 3. hold, however, having the girl as a girlfriend cannot be considered a goal but stays a mere desire.

This section was concerned with *desired* goals as a special class of goals. Desired goals are assumed to be the class of goals which actually make agents *act*, by which it is meant that agents perform actions in order to achieve these goals. But of course, there are states of affairs which have to be achieved by an agent but which are not directly *desired* by that agent. These states are just means to an end. However they can also be considered goals of the agent since, nevertheless, they have to be achieved by the agent. In other words, these "goals" are necessary intermediate steps for achieving a particular desire. These goals will be referred to as *necessary* goals. Necessary goals and their properties will be the subject of 3.1.3.

3.1.3 Necessary Goals

Achieving a desired goal may also require the achievement of several other goals – goals which are not desired in the first place but which need to be achieved in order to achieve the desired state of affairs. Consider again the agent A who has the *desired* goal of having a cup of tea. This being a goal implies that the agent knows actions which he is able to perform and which he believes will help him achieve the state of having a cup of tea. However, some of these actions do not directly result in having a cup of tea. For example, in order to have a cup of tea, A has to put a tea bag into a cup and has to prepare hot water. Although hot water is not directly desired by A, A still has to make sure to get hot water since hot water is a necessary ingredient for making tea. In other words, the agent has to perform certain actions which result in necessary intermediate steps in order to finally have a cup of tea. In the following, these intermediate steps will be called *necessary goals*. Hence, necessary goals are goals but without the condition of actually being *desired* by the agent. The agent can have the desired goal of having a cup of tea and since he knows that hot water is a necessary ingredient for making tea, he will have the additional goal of having hot water. However, the agent does *not* actually desire to have hot water. Necessary goals are, therefore, only means to achieve a desired goal. (3.9) shows a definition of necessary goals in the formal framework we have developed so far. As one can see, it looks similar to the definition of desired goals given in (3.8) but without the desire component in the agent's mental state.

(3.9) N-GOALS: A state of affairs G is am *n-goal* of an agent A at time T iff:

1. there is a *d-goal* G' such that $\text{BEL}(A, G \twoheadrightarrow G')$
2. A believes that there are actions which he can perform which will cause G to hold.
3. A intends to perform some of the actions of 2.

$$
\begin{array}{|l|}
\hline
A\ T \\
\hline
\textsc{agent}(A) \\
T : \textsc{time} \\
\qquad\qquad
\begin{array}{|l|}
\hline
S_{\textsc{bel}}\ S_{\textsc{int}}\ G', T' \\
\hline
T \subset S,\ T \subset T' \\
\textsc{dGoal}_A^{T'}(G') \\
S_G = S_{\textsc{bel}} \oplus S_{\textsc{int}} \\
\textsc{pGoal}_A^{T}(G) :\Leftrightarrow \quad S_{\textsc{bel}} : \textsc{bel} \left(A, \begin{array}{|l|}\hline G\ T'\ k \\ \hline T \prec T' \\ \Diamond G@T' \\ k : \textsc{action} \\ \mathsf{perform}(A,k) \twoheadrightarrow G \\ G \twoheadrightarrow G' \\ \neg(T \subset G) \\ \hline \end{array} \right) \\
\qquad\qquad S_{\textsc{int}} : \textsc{int} \left(A, \begin{array}{|l|}\hline T''\ E' \\ \hline T \prec T'' \prec T' \\ E'' \subset T'' \\ E'' : \mathsf{perform}(A,K) \\ \hline \end{array} \right) \\
\hline
\end{array} \\
\hline
\end{array}
$$

Neccessary goals will also be referred to as *n-goals*. So far, two different types of goals have been introduced: desired goals and necessary goals. Both types of goals were defined as certain configurations of mental states directed towards a particular state of affairs in the world - the *goal state*. In the next section, goals will be used to define *goal stacks*. Goals stacks can be defined as ordered sets of goals for achieving a particular desired goal. These goal stacks can be regarded as a preliminary version of the notion of a *plan*.

3.1.4 Subgoals and Goal Stacks

The sections 3.1.2 and 3.1.3 introduced two different types of goals: *desired goals* and *necessary goals*. Desired goals represent states of affairs which are desired by an agent and which the agent believes can be achieved. Moreover, it was assumed that desired goals are goals which make an agent act. Necessary goals, on the other hand, are intermediate steps towards achieving a desired goal and do not necessarily have to represent *desired* states of affairs. This distinction describes different phenomena which can be distinguished. However, in the following, it will again be useful to have a notion of goals which refers to both types distinguished so far since goal stacks are sets of goals that contain both types of goals. Therefore, it will again be useful to be able to talk about goals that make agents act (desired goals) and goals which are part of a plan to achieve a desired goal (necessary goals) at the same time. Therefore, the term *goal* will be used to refer to both types of goals again:

(3.10) GOALS: A state of affairs G is called a *goal* of an agent A at time T if it is either a *d-goal* or *n-goal* of A at time T.
If G is a goal of A at time T, write $\text{GOAL}_A^T(G)$.

Goals in general (i.e. d-goals and n-goals) can be complex in the sense that their achievement requires the achievement of several other goals. As an example, take again the agent A who has the goal of having a cup of tea. In order to have a cup of tea, the agent needs hot water and in order to have hot water, the agent has to have a kettle which is filled with water and a hot stove and so on. Therefore, in order to have a cup of tea, the agent has to achieve several *subgoals*. Some of these subgoals depend one another, some do not. These dependencies were already discussed in the introduction to this chapter. Figure 3.2 shows a set of A's goals and subgoals and dependencies of these goals.

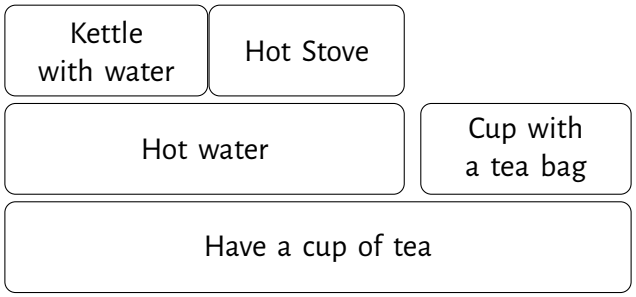


Figure 3.2: The dependencies A's goals for the main goal of having a cup of tea. The goals form a *goal stack*.

Every box in figure 3.2 represents one of A's goals. This pictorial representation of goal stacks will use the convention that a box which is drawn on top of another box represents a subgoal which has to be achieved for in order to achieve the goal which is represented by the lower box. This graphical way of representing the dependencies of goals shows that a set of goals builds a stack-like object. Therefore, the set of goals with a dependency structure as in 3.2 is called a *goal stack*. Of course, like goals themselves, goal stacks depend on the agent's assumptions about causalities and on her own abilities and, hence, the particular form of the goal stack depends on the agent who has constructed the goal stack.

It is possible to say that a goal stack for a goal G' lies on top of another goal stack for the goal G. In this case G' will be called a *subgoal* of G. For example, having a kettle with water in 3.2 is a subgoal of having hot water but also a subgoal of the overarching goal of having a cup of tea. In the following, a relation for goals will be introduced which makes this notion of subgoals precise. (3.11) shows a possible defintion of a subgoal relation. This relation again depends on the agent's assumptions about causaltities. The idea of this relation is that an agent A's goal G' is a subgoal of a goal G if A believes that achieving G' helps him to achieve G:

(3.11) SUBGOAL: Let G and G' be goals of an agent A at time T. G' is a *subgoal* of

G if A believes that G' causes G:

$A\ G\ G'\ T$	
T : TIME $\text{GOAL}_A^T(G)$ $\text{GOAL}_A^T(G')$	
$\text{SUBGOAL}_A^T(G, G') :\Leftrightarrow$	S_{BEL} $T \subset S_{\text{BEL}}$ $S_{\text{BEL}} : \text{BEL}(A, G \twoheadrightarrow G')$

The notion of causation in (3.11) is again to be understood in a rather broad and individualistic sense: $G' \twoheadrightarrow G$ refers to the fact that A believes that G' is an element of a causal network for G (cp. page 64). Based on the notion of subgoals, it is possible to formalize the notion of a *goal stack*. As was already said, a goal stack is supposed to be a (partially) ordered set of goals the order of which is given by the subgoal relations. Given the notion of the subgoal relation, this idea can now easily be made precise. A goal stack can be considered as a preliminary version of a *plan* (cp. 3.2) in the sense that a goal stack already contains an organized structure of intermediate goals but the execution of the associated actions are not temporally ordered yet. Given a concrete goal stack, the maximal element with respect to the subgoal relation will be called the *maximal goal* of the goal stack. For example, the maximal goal of the goal stack in 3.2 is the goal of having a cup of tea.

(3.12) GOAL STACKS: A *goal stack* $\mathbb{S}_A^T$ of an agent A at time T is a set of goals of A such that $\mathbb{S}_A^T$ is partially ordered by SUBGOAL_A^T and such that there is a unique element $S_{Max} \in \mathbb{S}_A^T$ with $\text{SUBGOAL}_A^T(S_{Max}, S_G)$ for all $S_G \in \mathbb{S}_A^T$. We write $max(\mathbb{S}_A^T)$ for the maximal element of a goal stack and call it the *maximal goal* of $\mathbb{S}_A^T$.

For example, the maximal goal of the goal stack in 3.2 is the goal of having a cup of tea.

This section has been concerned with a formalization of a particular relation within the set of possible goals. It defined the subgoal relation as a relation which models an agent assuming that achieving one goal helps him to achieve another goal. Moreover, goal stacks as a sets of goals which are partially ordered by the subgoal relation have been introduced. In §3.1.5, the focus will again be placed on a single goal and on goals as mental states.

3.1.5 Pursuing a Goal

Sections 3.1.2 and 3.1.3 introduced the notions of desired goals and of necessary goals and section 3.1.4 introduced the notion of a goal as a cover term for both of these cases. In both cases of goals, a goal is defined as being a mental representation of a state of affairs with some additional conditions on the particular mental state of the agent who has the goal (see (3.8) and (3.9)). In the following, the complex mental state in which

an agent is while she has a goal will be referred to as the state of *pursuing the goal* G. This distinction makes it possible to keep separate an agent's goals – the states of affairs in the world which the agent wants to achieve – and the mental state an agent is while pursuing a goal. These two different states (the goal state and the associated mental state) heavily depend on each other and can be consider two sides of the same coin:

1. *Pursuing a goal* is a complex mental state which takes the goal of the agent as an argument.
2. A *goal* is a state of affairs which, by definition, occurs as argument of *pursuing a goal.*

The assumption that an agent who has a goal is in the appropriate mental state towards this goal is summarized in (3.13).

(3.13) PURSUING A GOAL: Let A be an agent and let G be a goal of A. We call the complex mental state S_G in the definitions of (3.8) and (3.9) the state of *pursuing a goal*. If A is in the state of pursuing the goal G, we write $S_G : \text{PURSUE}(A, G)$.

At the end of this section, a further constraint on the temporal structure of goals should be added. In the definitions of goals, a goal G depends on a time point T which is temporally located inside S_G, i.e. inside the agent's mental state of pursuing G. However, it is natural to assume that G is a goal of agent A, as long as A pursues this goal. Therefore, we add the following condition to the description of goals:

(3.14) If A is in the state S_G of pursuing a goal G then G is a goal of A at any time $T \subset S_G$:

$$S_G : \text{PURSUE}(A, G) \Rightarrow \forall T(T \subset S_G \Rightarrow \text{GOAL}_A^T(G))$$

With the notion of pursuing a goal, it is more easy to distinguish the mental states an agent is in while having a goal and the goal state as a state of affairs which the agent desires to achieve.

Summary of this section: This section introduced goals as mental representations of states of affairs in the world with some additional properties. It furthermore introduced the notions of desired goals and necessary goals which are both covered by the notion of goals which will be used in this book. Moreover, the subgoal relation as a relation between goals and, using this relation, goal stacks as partially ordered sets of goals were defined. In chapter 4, this notion of goals will be used to define goals of *speech acts*. But first, a more general notion of actions as means to achieve goals will be established. This will be the subject of section 3.2. Another goal of section 3.2 will be to introduce a notion of *plans* as temporally ordered sets of actions, in order to achieve the goals of a goal stack.

3.2 Actions and Plans

So far, it has been assumed that goals are the reasons for agents to act, regardless of the fact that the question of how goals come into existence has not been discussed. In order to achieve a goal, agents have to perform actions which they believe will finally result in the achievement of the goal. But often, goals are complex, in the sense that they cannot be achieved by a single action. This was already demonstrated by the tea example where several subgoals need to be achieved in order to achieve the main goal (in this case, the goal of having a cup of tea). In order to have a cup of tea an agent has to achieve several subgoals and hence, several actions are needed which help to achieve these subgoals. Therefore, in the case of a complex goal the agent has to come up with a whole sequence of goals and actions, which finally result in the achievement of the main goal. This can be paraphrased by saying that the agent has to construct a *plan* for carrying out particular actions in a specific order in order to achieve his goal. The actions which occur in a plan are special in so far as they are directed towards a specific goal.

This section is devoted to the definition of plans and goal-directed acts as special events which are caused by the agent and of which the agent believes will bring about a particular goal. These goal-directed acts can be regarded as means to carry out a plan. In section 3.2.1 the notion of goal-directed acts will be introduced. Based on this notion, section 3.2.2 will define the notion of plans. Plans can be regarded as temporally extended goal stacks in which each of the goals of the stack is associated with a goal-directed act which brings about that very goal. Section 3.2.3 is concerned with the actual execution of a plan and defines the notion of a *successful* plan. This notion will capture the idea that a plan is successful if any action belonging to the plan decreases the size of the plan. Section 3.2.4 finally is concerned with unsuccessful plans, and with the question of how an agent can "repair" an unsuccessful plan.

3.2.1 Goal-Directed Acts

This book adopts the view that actions are caused by goals (or, to be more precise, by the mental state of pursuing a goal): in order to achieve their goals, agents have to *act*. For example, an agent A who has the goal of greeting somebody may raise his arm in order to achieve that goal. The intuition that agents act in order to achieve goals will be captured by the notion of a *goal-directed act*. The definitions of goals which were proposed so far include the condition that the agent believes that there are actions which he is able to perform that result in the given goal. Goal-directed acts are realizations of these actions. A goal-directed act is, therefore, an event caused by the agent which satisfies the following two conditions:

a. The mental state of pursuing a goal caused the action of the agent.
b. The agent believes that performing the action causes the goal.

As an example, take again an agent A who wants to illuminate a room and assume that A believes that flipping the light switch will cause a light bulb in the room to go on which will finally cause the room to be illuminated. Therefore, the agent flips the light

switch. In this case, the event E_G of flipping the light switch is a goal-directed act which pursues the goal of illuminating the room. This event fits the informal conditions above since the action was caused by the fact that the agent was in the state of pursuing a particular goal and since the action was believed to be an action which would cause the light to go on. Moreover, note that also actions which did not achieve desired effect can be classified as goal-directed acts if they fulfill the two conditions above. This would be the case if the agent had made wrong assumptions about the possible outcome of an action. For example, the agent's action of flipping the light switch is still a goal-directed act towards the goal of illuminating the room, even if it does not cause the light to go on. In particular, this means that the definition of goal-directed acts which is proposed in this section allow to model *mistakes* of an agent. The two conditions above can be formalized as in (3.15):

(3.15) GOAL-DIRECTED ACT: A *goal-directed act* E_G of an agent A with goal G at time T is an action of A such that

1. A believes that the realization of E_G will cause G
2. being in the state of pursuing goal G causes A to perform E_G:

$A\ T\ S_G\ E_G\ G$	
$\text{AGENT}(A)$ $S_G : \text{PURSUE}(A, G)$ $T \subset S_G,\ E_G \subset T$	
$E_G : \text{GOAL-DIRECTED ACT}(A, G) :\Leftrightarrow$	k : ACTION $E : \mathsf{perform}(A, k)$ $S_G \twoheadrightarrow E_G$ $\text{BEL}\left(A, \boxed{E \twoheadrightarrow G}\right)$

We call G the *goal of the act* E_G.

The notion of goal-directed acts will make it possible to define plans which will be the subject of 3.2.2. The idea of plans which is pursued in this book is that plans are goal stacks in which each subgoal of the goal stack is already "connected" to a certain goal-directed act. In other words, the agent who has a plan knows not only which subgoal he has to achieve in order to achieve the goal of the plan but knows also by which actions the agent achieve these goals.

3.2.2 Plans

Section 3.1.4 was concerned with the definition of goal stacks (definition (3.12)) and the last section 3.2.1 was concerned with the definition of goal-directed acts. A goal stack can be interpreted as a preliminary version of a something which can be called a *plan*. A goal stack already contains the agent's assumptions about causalities. However, it

does not include actual actions which are necessary to achieve the goals in a goal stack. This gap can now be closed on the basis of the notion of goal-directed acts.

In the following, a plan will be defined as a temporally ordered sequence of goal directed acts meant to bring about the subgoals of a goal stack. A plan includes not only goals or goal-stacks but also specific actions which bring about the subgoals of the goal of the plan, as well as knowledge about a meaningful order in which to perform these actions. These aspects of plans are captured in (3.16):

(3.16) PLAN: A *plan* $\mathbb{P}$ of an agent A at time T, is a finite sequence $(E_i, G_i)_{i=1}^n$ of goal-directed acts E_i with goals G_i such that

- $T \prec E_1$ and $E_i \prec E_{i+1}$ for all $i \in \{1, \ldots, n-1\}$.
- the set $\mathbb{S} := \{G_i\}_{i=1}^n$ is a goal stack of A at T
- if $\text{SUBGOAL}_A^T(G_i, G_k)$, then $E_k \prec E_i$.

If a plan involves n actions, it is said to be *of size* n. The goal $max(\mathbb{S})$ is called the *goal of the plan* $\mathbb{P}$ and $\mathbb{P}$ is a *plan for* $max(\mathbb{S})$.

A plan can be considered a goal stack, together with some actions which the agent believes to bring about the individual goals of the goal stack. The notion of a plan given in (3.16) is rather strict one since it requires the agent to already have complete knowledge about all the subgoals which he has to achieve and about all goal-directed acts which he wants to perform in order to achieve the respective subgoals. In this sense, the given notion of a plan can also be referred to as a *complete plan*. Of course, in general agents do not have complete plans. Bratman, (1990, p. 19) claims that limited agents' plans are typically just *partial* plans in the sense that they do not have to be completely worked out. For example, I might have the plan to go to Paris by train and therefore go to the train station but without knowing which train I have to take. In this case I first have to find out which train I have to take and to which platform I have to go. Having the plan to go to Paris without knowing which train I have to take is just a partial plan since I do not know about all the actions I might have to perform in order to achieve my goal. In this example, one could even say that making a plan is a part of my original plan, i.e. go to the train station and decide at the train station which train I to take. The assumption that agents in general don't have complete plans but just partial plans is in many cases probably a better description of states of affairs in the world then assuming that agents have complete plans. However, this book will only be concerned with plans of rather small size which can be assumed to be complete. Therefore, incomplete plans will not be considered at all.

In later sections of this book it will be convenient to not only be able to talk about plans but also to be able to talk about *subplans*. A subplan of a plan will be defined as a plan for achieving one of the subgoals of the plan. For example, take again the agent who wants to have a cup of tea and who has already come up with a plan that contains the subgoal of having hot water and, therefore, contained some actions which result in the state of having hot water. In other words, this plan contains a subplan for making hot water. A subplan can be defined as a subsequence of the sequence of goal-directed acts in a plan which is again a plan for the highest ranked elements with respect to the

subgoal relation. In (3.17), these assumptions about subplans are made precise:

(3.17) SUBPLAN: A subsequence $\mathbb{P}' = (E_{i_k}, G_{i_k})_{k=1}^{j}$ of a plan $\mathbb{P} = (E_i, G_i)_{i=1}^{n}$ is a *subplan* of $\mathbb{P}$ if

- $\mathbb{S}' := \{G_{i_k}\}_{k=1}^{j}$ is a goal stack and
- $\mathbb{P}'$ is a plan for $max(\mathbb{S}')$.

The definition in (3.17) also implies that any plan $\mathbb{P}$ has at least two trivial subplans: the empty plan and the plan $\mathbb{P}$ itself are subplans of $\mathbb{P}$. The notion of subplans will turn out to be of interest for a theory of actions and plans in cases in which a plan does achieve its goals. In this case, the agent has to reconsider his plan and has to replace the non-functioning subplan by another subplan. Note that for any goal G of a plan $\mathbb{P}$ there must be a subplan which has G as its main goal:

(3.18) Let $\mathbb{P}$ be a plan and let G be a goal in $\mathbb{P}$. Then there is a subplan $\mathbb{P}'$ of $\mathbb{P}$ which has G as its main goal. We write $\mathbb{P}_G$ for this subplan of $\mathbb{P}$.

This section was concerned with plans and subplans. §3.2.3 will be concerned with the question of how a plan is executed. For the execution of a plan, there are basically two possibilities: Either the plan is successful, i.e. the performance of all planned actions is successful and has the desired effect or the plan is not successful. In this case, the plan has to be reconsidered somehow. This will be the point of interest in the next sections.

3.2.3 Executing a Plan

After the notions of plans and subplans have been introduced, this section is finally concerned with the execution of a plan. If an agent executes a plan, the agent executes actions which are part of the plan and if these actions are successful (i.e. the goals of the actions are achieved) the actions can be removed from the plan. This means that the agent does not have to take care for the achievement of the respective goal anymore. Hence, during the execution of a plan one achieves subgoals of the plan and these subgoals can therefore, be excluded from the plan. This means that the size of the plan decreases while executing a plan.

Consider again the example of agent A who wants to make a cup of tea. Making a cup of tea already involves a plan. By executing the plan of preparing a cup of tea, the agent will achieve some of the subgoals of the plan – for example having a tea bag in a cup. If a subgoal is achieved, it is no longer necessary to keep this goal as an element of the plan: the agent no longer has to take care getting a cup with a tea bag in it and, therefore, this subgoal can be removed from the plan. In order to illustrate this development of a plan, assume that at time T, agent A came up with the following goal stack for his goal of having a cup of tea:

G_0: Have a cup of tea.

- G_1: Have a cup with a tea bag in it.
- G_2: Have hot water.

G_{21}: Have a kettle filled with water.
G_{22}: Have a hot stove

Recall that goal stacks can be visualized as boxes, each box representing a goal in the plan and where one box is place on top of another box if a subgoal relation holds between the respective goals. Figure 3.2 already showed a visualization of the goal stack above. So the agent already has a goal stack for achieving his goal. But a goal stack is not a plan yet. The agent has to come up with a series of goal-directed acts in order to achieve the individual subgoals of the plan. Therefore, assume that A has associated the actions given in the list in 3.2.3 with the corresponding goals of the goal stack, which means that A plans to perform the following actions E_i in the given order.

1. E_{21}: hold the kettle under the tap and fill it with water (to achieve G_{21})
2. E_{22}: turn on the stove (to achieve G_{22})
3. E_2: put the kettle on the stove (to achieve G_2)
4. E_1: put a tea bag in the cup (to achieve G_1)
5. E_0: pour hot water into the cup (to achieve G_0)

The list in 3.2.3 is a particular sequence of actions of which A believes that their execution in the given order will bring about the state of having a cup of tea. This turns the sequence (E_k, G_k) into a *plan* in the sense of definition (3.16). Now assume that A actually performs the actions E_k in the given order and that A performs these actions successfully, i.e. assume that the performance of any of these actions brings about the corresponding subgoal. This means that with the performance of an action, the corresponding subgoal can be removed from the goal stack and that the plan decreases in size (which again corresponds to the fact that the agent comes closer to achieving his final goal). For example, after the performance of the action E_{21} agent A has achieved the subgoal G_{21}. This subgoal therefore can be removed from the given goal stack. After A's performance of E_{22} the goal G_{22} can be removed from the goal stack and so on. This means that the successful performance of the actions of a plan reduces the number of elements of the plan's goal stack. The development at of a successful plan can, therefore, be illustrated as follows.

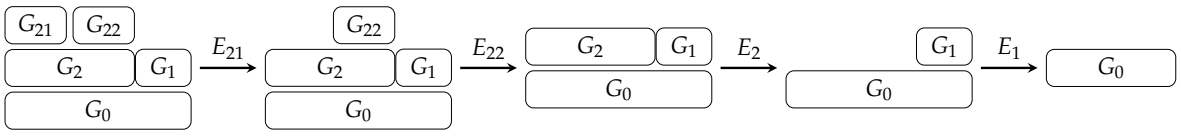


Figure 3.3: The development of a successful plan. The performance of each goal-directed act of the plan removes one of the subgoals of the goal stack.

In the case of a plan an action of which is successful, the performance of any particular action removes one particular subgoal from the goal stack of the plan: performing E_{21} removes G_{21}, E_{22} removes G_{22} and so on. In other words, in the case of a successful plan, the performance of each action decreases the size of the given goal stack and, hence, brings the agent closer to his final goal. These observations should match the

following intuition: a plan is considered to be successful if it works out, i.e. if any action belonging to the plan brings the agent closer to the achievement of the goal of the plan. (3.19) turns this observation into a definition by defining a plan as being successful performing the action involved removes the corresponding subgoal from the top of the goal stack.

Unfortunately, this heuristics has a formal problem: the definition of goal stacks and, therefore, the definition of plans is static, i.e. it is not able on its own to reflect the fact that one of its elements gets removed by a certain action. Therefore, in order to capture this intuition, definition (3.19) has to use not one plan, but a sequence of plans, with one plan emerging from another by removing the performed goal-directed act and the associated subgoal from the former plan:

(3.19) SUCCESSFUL PLAN: A plan $\mathbb{P} = (E_i, G_i)_{i=1}^{n}$ of agent A of size n is *successful*, if A performs E_i at time T_i and if there is a sequence of plans $(\mathbb{P}_i)_{i=1}^{n+1}$ such that $\mathbb{P}_{n+1} = \mathbb{P}$ and such that $\mathbb{P}_i = \mathbb{P}_{i+1} \setminus \{(E_i, G_i)\}$.

In the case of a successful plan, the size of the goal stack decreases with the performance of every action. With this notion in mind, it is now possible to make sense of figure 3.3. Each action removes one goal from the corresponding goal stack. Therefore it shows how the development of a successful plan can be visualized.

This section was concerned with the execution of a plan and, in particular, with the execution of a successful plan. But as a matter of fact, not all plans are successful. In fact, in probably most cases, something goes wrong in a plan and the plan has to be repaired. Unsuccessful plans will be of particular interest for this book since certain repair moves in discourse can be interpreted as the realization of reconsidered communicative plans. This assumption will be formulated in section 5.4. The next section will, therefore, be concerned with unsuccessful plans and how agents can react if they learn that a plan is unsuccessful.

3.2.4 Executing an Unsuccessful Plan

This section is concerned with the execution of unsuccessful plans. We have defined successful plans as plans in which the performance of each goal-directed act decreases the size of the plan. An unsuccessful plan can, therefore, be defined as a plan which contains at least one action which does not decrease the size of the plan. This section places a particular focus on the question of an agent's reaction when he finds out that his plan isn't successful.

In general, the success of a plan is not guaranteed. There are several reasons why plans fail to achieve their goals. While pursuing a plan, the agent may learn that some of the actions which he wanted to perform do not have the desired effect or that some of the causalities he assumed don't actually hold or that he is simply not able to perform one of the goal-directed acts which are part of the plan. If an agent learns that his original plan will turn out to be unsuccessful, it is in principle not necessary to give up the whole plan or even to give up the goal of the plan. In many cases, the agent has still the opportunity to *reconsider* his plan. Given the assumption that the agent does

not want to give up the goal of his plan, he may set new subgoals on top of the original goal stack of the plan or he even may replace subplans of the original plan by other subplans which he believes are easier to achieve.

The following situation illustrates a case of a plan failing and the agent reacting by placing a new subgoal on top of the original goal stack of the plan. Take again the agent A who wants to make a cup of tea. Originally, A had the plan which was already discussed extensively in the former sections: filling a kettle with water, taking a cup from the cupboard and so on. However, when A wanted to take a cup from the cupboard A notices that there are no clean cups left (they are all dirty and still in the dishwasher) and that, moreover, he is not willing to prepare his tea in a dirty cup. In this case, A has to include the additional subgoal of having a clean cup into his plan. In the terminology of goal stacks, this means that the goal of having a clean cup has to be placed on top of the goal of having a tea bag in a cup. This means that in this case, the subgoal of having a tea bag in a cup suddenly requires a non-trivial plan. In the following, the situation in which a goal enters a plan will be illustrated in a similar way as in figure 3.3. The symbol R_A will be used for the process of A's reconsideration of the plan. Figure 3.4 shows how such a reconsideration of a plan can be illustrated. This figure uses the following abbreviations:

- Having a tea bag in a cup. G_1
- Having a clean cup. G_{11}
- Washing a cup. E_{11}

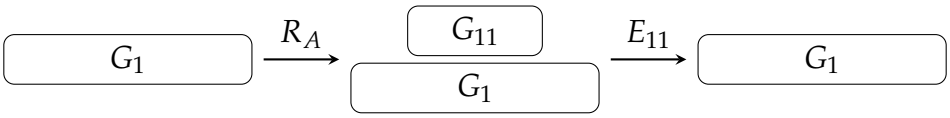


Figure 3.4: A *reconsideration of a plan*. In this case a new subgoal comes on top of a goal, which turned out not to be achievable the way it was originally planed.

The process, denoted by R_A in the figure above, will be referred to a *reconsideration of a plan*. Like the process of constructing a plan, the exact process of reconsidering a plan cannot be part of this book, since a huge amount of knowledge is needed for such a reconsideration process. However, it is still possible to define a formal version of plan reconsideration. This is done in (3.20). This definition aims to capture the intuition that whenever an agent makes a plan for something but learns that some subplan has failed, the agent can still replace that subplan and does not have to abandon the whole plan.

(3.20) RECONSIDERING A PLAN: Let $\mathbb{P}$ be a plan of agent A at time T with goal G. A plan $\mathbb{P}'$ of A at time T' with $T \prec T'$ is a reconsidered version of $\mathbb{P}$ if $\mathbb{P}'$ also has goal G and if there is a subplan $\mathbb{P}_S \subset \mathbb{P}$ and a subplan $\mathbb{P}'_S \subset \mathbb{P}'$ such that $\mathbb{P}' = (\mathbb{P} \setminus \mathbb{P}_S) \cup \mathbb{P}'_S$. If a plan $\mathbb{P}'$ is a reconsidered version of a plan $\mathbb{P}$, we write $\mathbb{P} \Rightarrow \mathbb{P}'$.

Definition (3.20) does not exclude the possibility that plan $\mathbb{P}_S$ is actually the empty subplan. In this case, the agent just puts new subgoals on top of the original goal stack. This was the case in the example above where A first has to achieve the subgoal having a clean cup and therefore put the goal of having a clean cup on top of his plan.

But under which conditions should an agent reconsider his plans at all? The tea example above showed a case in which the agent did not want to perform a certain action because a condition for the performance of this action was not fulfilled (in the case of the tea example, the cup wasn't clean which was a condition for putting a tea bag in the cup – at least for this agent). But of course, there are other cases which might result in the necessity of reconsidering a plan. Some of these are collected in the following short list:

(3.21) CONDITIONS FOR RECONSIDERING A PLAN: An agent A with plan $\mathbb{P}$ has to reconsider $\mathbb{P}$ if there is an action-goal pair (E_k, G_k) of $\mathbb{P}$ such that

1. the performance of E_k did not show the expected outcome *or*
2. A was not able to perform E_k *or*
3. A lost the intention to perform E_k *or*
4. A's assumption about causalities concerning the state S_k have changed.

Of course, this list is not exhaustive but note that these conditions are similar to the conditions for a goal-directed act (cp. definition (3.15)). Of course, this is expectable, since by definition goal-directed acts are means to carry out a plan and, hence, if the performance of a goal-directed act fails, the plan in which it occurs has to be reconsidered.

Summary of this section: This section was concerned with plans and actions as means to carry out a plan. The idea which was proposed in this section was that agents act because they are in a state of pursuing a goal. Plans were defined as sequences of goal stacks and goal-directed acts such that each act removes a goal from the given goal stack. Moreover, the notion of a successful plan was introduced. A plan can be called successful if its size decreases with the performance of any goal-directed act belonging to the plan. The final part of this section was concerned with the execution of unsuccessful plans and with the reconsideration of plans. Reconsidering a plan means that an agent replaces subplans which he believes will fail with new subplans which will hopefully be successful. The insights of this section will play a role again in chapter 4, when *communicative plans* are introduced and in section 5.4, when certain discourse moves (such as self-repair moves) are explained as reconsiderations of plans.

The next section shifts its focus from the agent himself to the observer of an agent's actions who wants to make assumptions about the mental states of an agent on the basis of his actions. The *observer's perspective* will be of particular interest for the linguistic application of the theory developed in this chapter. This is due to the assumption that the observer of actions plays a similar role to the hearer of a discourse in the sense that both have to turn distinct actions of an agent into a coherent behavior.

3.3 The Observer's Perspective.

This section places its focus on the observer who observes an agent's actions and who wants to understand why an agent acts the way he does in a given situation. It will turn out that understanding how an observer is able to draw conclusions about another agent's goals and actions is of linguistic relevance since understanding the communicative goals of a speaker seems to be a necessary condition for understanding a discourse. In situations in which a hearer has to reconstruct a rhetorical relation, the hearer has to make assumptions as to *why* a speaker performs a certain speech act in a given situation. This means that the hearer, when trying to interpret speech acts, has a similar task to fulfill as an observer of an agent's actions. This section, therefore, is concerned with the question of how an observer can draw conclusions about another agent's goals on the basis of his actions. In the following, this process is referred to as a *goal recognition task*.

In order to make sense of the mechanisms behind a goal recognition task, take again the following example:

(3.22) Assume that an observer sees a scene in a silent movie in which somebody goes into a kitchen, opens a cupboard, takes out a cup and puts a tea bag in the cup. Afterwards A takes a kettle and fills it with water.

By observing the scene above, an observer can conclude that the agent has a goal and is in the process of executing a plan for achieving this goal. First of all, the agent has the immediate goal of having a cup with a tea bag in it. However, is it reasonable to assume that a tea bag in a cup is of no value on its own, which means that it was not the agent's maximal goal to have a tea bag in the cup. But this means that a tea bag in a cup is just an intermediate step for achieving another goal. Given a little bit of general knowledge about how one makes tea, the observer can conclude that the agent is actually in the process of preparing a cup of tea. Now, the agent fills a kettle with water. Since the observer already knows that the agent is preparing a cup of tea, he is able to connect this action of the agent to the agent's preceding action. Given the knowledge that the agent is preparing a cup of tea, the observer can conclude that filling the kettle with water is a subgoal of having a cup of tea. Under these conditions, the observer is able to connect two completely separate actions and can conclude that they are part of a common plan.

The rest of this section is basically concerned with the question of how these inferences were possible. In order to answer this question, further common sense principles of human reasoning are introduced. Section 3.3.1 introduces the general assumption that any action is performed because it is connected to some goal. Section 3.3.2 is concerned with the observer's ability to draw conclusions about the agent's general goals. In terms of example (3.22), section 3.3.2 is concerned with the question of how the observer was able to guess that the agent was actually in the process of preparing a cup of tea. Section 3.3.3, finally, deals with the ability of an observer to conclude that an agent's distinct actions belong to the same plan.

3.3.1 The Act-Goal Connection

This section is concerned with the question how an observer can make guesses about an agent's goals on the basis of the agent's actions. The problem in this task is the following: an action basically consists of body movements (or of a speech act, in the next chapter), a goal on the other hand, is something in the agent's mind. So the observer has the task of "looking" in the agent's mind, based on his body movements. This is possible on the assumption that there is some kind of connection between an agent's action and his goals. In the following, this connection will be called the *act-goal connection*. The act-goal connection basically states that an agent who performs a certain action normally pursues a goal by performing this action.

At the first glance, this principle seems to be almost tautological. However, it is not as tautological as it seems. The fact, that this section deals with the perspective of an observer turns this principle into a useful assumption about connections between actions and goals. Of course, an agent who is action knows why he performs a particular action, i.e. he knows the goal which he is pursuing and believes that the action he is performing will help him to achieve this goal. An observer, on the other hand, is left with elaborated guesses concerning the reasons for an action. Of course, these guesses are still based on the actions observed. If the observer has an idea *why* the agent is performing a particular action, he can, nevertheless, make assumptions about the agent's mental states. For example, if the observer sees an agent filling a kettle with water, the observer can assume that the agent is filling the kettle with water because the agent is in the state of pursuing a particular goal with that action. In the following, this intuition is formalized in principle (3.23):

(3.23) ACT-GOAL CONNECTION:

If an Agent performs a goal-directed act, then he is in the state of pursuing a goal:

$A\ E\ G$	
$\text{AGENT}(A)$ $E \subset T$ $E : \text{GOAL-DIRECTED ACT}(A, G) \Rightarrow$	S_G $T \subset S_G$ $S_G : \text{PURSUE}(A, G)$

The act-goal connection enables the observer to draw a connection between an agent's actions and the agent's mental states. It can, in particular, be used in cases in which the observer has already made assumptions about the goal of a particular action and where he wants to derive the plan underlying this action. For example, an observer who sees an agent picking apples from an apple tree can conclude that the agent wants to have apples. However, there are cases in which an agent can make guesses about the goals of an agent by observing his actions but in which it is not as clear how the action is connected to the goal as it is in the case of an agent picking apples. Take again, for

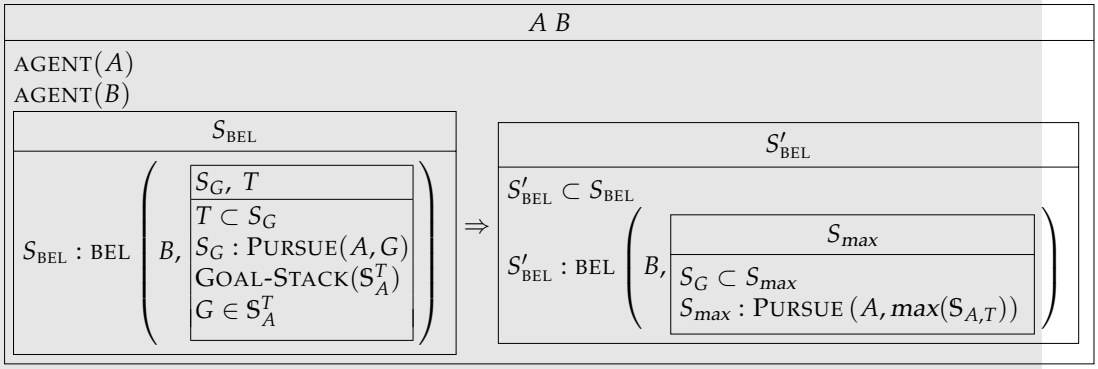
example, the agent, who wants to have a cup of tea and therefore puts a tea bag in a cup. An observer who sees this action can normally guess that the agent is in the process of preparing a cup of tea. Preparing a cup of tea requires many different actions which depend on each other, picking apples, on the other hand, consists of just one action which is repeated several times. However, the observer is able to generalize over the action of putting a tea bag in a cup and is able to make assumptions about a superordinated goal of the agent's. 3.3.2, therefore, investigates the rules which underly the observer's additional assumptions.

3.3.2 Generalized Goals and Scripts

Rule (3.23) provides a way to model an observer's ability to make assumptions on another agent's mental states, given that he has already observed an action of the agent and made guesses about the goals of that action. However, in many cases, the observer is also able to make guesses about a superordinated goal of that agent. Take the tea example again: as already mentioned, in normal circumstances, a tea bag in a cup is of no value on its own. Therefore, an observer can assume that the tea bag in the cup is just a subgoal for another goal. In this case, the observer perceives one of the agent's actions and concludes from this action that the agent is actually in the state of pursuing a goal which is superordinated to the goal of that action: From observing the agent putting a tea bag in the cup, the observer can conclude that the agent wants to have a cup of tea. Other examples in which observers are able to make guesses about agent's goals based on the agents actions are the following: An agent who carries a ladder to an apple tree probably wants to pick apples from the tree. An agent who enters a restaurant probably wants to eat something. An agent who is running in the direction of a train station probably wants to get a train and so on. In principle these inferences can be wrong. An agent can also carry a ladder to an apple tree because he wants to cut some branches, an agent may enter a restaurant because he wants to use the bathroom and so on. However, observers still make such assumptions based the agent's behavior. These assumptions, of course, depend heavily on the observers knowledge of the world and his general assumptions about "normal" behavior of agents, i.e. the observer assumes that agents *normally* enter restaurants because they want to eat there.

The observer generalizes about the particular behavior of agents in order to be able to make assumptions about an individual agent's behavior. The structure of this reasoning can be summarized as follows: the observer sees an agent who is performing certain actions and, moreover, assumes that the goals of these actions are part of a plan or part of the goal stack of a plan. Therefore, the observer is able to derive that the agent is actually in the state of pursuing the maximal goal of this plan. The structure of this reasoning can be formulated in terms of goal stacks as follows:

(3.24) GENERALIZED GOALS:
If an observer B believes that an agent A has a goal stack $\mathbb{S}_A^T$ and that a goal G of A is an element of $\mathbb{S}_A^T$ then B can deduce that A is pursuing the goal $max(\mathbb{S}_A^T)$:



The exact process of how an observer can construct another agent's goal stack is again left as a black box. In cases in which (3.24) is actually applied to an agents goal, the goal stack has to be stipulated. Nevertheless, the rest of this section is devoted to a short discussion of the question how an observer can come to his assumptions about an agent's goal stacks? Following Schank and Abelson, (1977), one can answer this question by stating that human agency is organized in *scripts*. A script is a set of beliefs about the 'normal' order of events in a particular situation which one has encountered many times before. According to Schank and Abelson, scripts can be characterized as follows:

> A script is a structure that describes appropriate sequences of events in a particular context. A script is made up of slots and requirements about what can fill those slots. [...] Scripts are not subject to much change, nor do they provide the apparatus for handling totally novel situations. Thus a script is a predetermined, stereotyped sequence of actions that defines a well-known situation. *(Schank and Abelson, 1977, p. 41)*

The standard example of a script which is used by Schank and Abelson is the *restaurant-script*. People have quite a robust intuition about events that are involved when agents visit restaurants. Agents who go to a restaurant usually take a seat, wait for the waiter to bring them the menu, the waiter takes orders for drinks and so on. A situation which makes reference to a frequently occurring script, such as for example visiting a restaurant or making a cup of tea, can easily be dealt with by an observer. In situations in which agents' actions activate an available script, the observer can make guesses about the actions and their underlying goals, since the observer has already encountered such situations many times before and is, therefore, able to make guesses about the agent's plans.

From the ideas of Schank and Abelson, we can adopt the idea that most human action is organized in scripts, i.e. in sequences of events which apply in appropriate contexts and that, moreover, these scripts enable observers to make assumptions about plans underlying the agent's actions. For more details on scripts, consider Schank and Abelson, (1977).

This section was concerned with an observer's ability to make assumptions about an agent's goals based on his actions without being able to derive them directly from the goal of the actions. The problem which was approached in this section is basically a bottom-up problem: given knowledge of an action, how it is possible to derive the goal of the plan to which this action belongs? The next section deals with a problem which can be classified as a top-down problem: given knowledge of an agent's goal, how can an agent make sense of further actions performed by the agent, with respect to that goal? The next section answers this question in terms of a principle which describes the assumption that if an agent has a goal than the agent is *committed* to achieving this goal.

3.3.3 Commitment to a Goal

This section finally deals with the following problem: how can an observer connect two distinct actions of the same agent's given that the observer already has some knowledge about the agent's goals? The idea for an answer to this question is the following: an agent who is in an active state of pursuing a goal will work to achieve this gaol as long as it is achieved. To be a little bit more concrete, assume that an observer knows that an agent A is pursuing a particular goal G. Moreover, assume that the agent performs an action E. Given the assumption that A is in the state of pursuing goal G, the observer can conclude that E belongs to a plan for achieving G. In other words, as long as an agent A is in the state of pursuing a goal G, the observer can assume that A is *committed* to G in the sense that all his actions are directed to G.

In order to give a better motivation for this idea, take again the agent A who wants to make a cup of tea and assume that the observer already has knowledge about this goal of A. Now assume that the observer sees A filling a kettle with water. A priori, this can be done for many different reasons. Maybe A wants to water his plants or A wants to prepare hot water for cooking spaghetti. Nevertheless, for the observer it is reasonable to assume that filling the kettle is part of the process of making a cup of tea *because* at this point the observer may assume that A is still in the state of pursuing the goal of having a cup of tea. The observer's assumption about the agent's goals caused him to think that the agent is performing an action in order to achieve that goal.

The general reasoning for the conclusion that A is filling the kettle with water because he wants to have a cup of tea is summarized in (3.25):

(3.25) COMMITMENT TO A GOAL:
As long as an agent is in the state of pursuing a goal G, then the goal of any goal-directed act is a subgoal of G:

$$\boxed{\begin{array}{l} A \\ \hline \text{AGENT}(A) \\ \boxed{\begin{array}{l} E\ S \\ \hline E \subset S, \\ S : \text{PURSUE}(A, G) \\ E : \text{GOAL-DIRECTED ACT}(A, G') \end{array}} \Rightarrow \boxed{\begin{array}{l} T \\ \hline E \subset T \\ \text{SUBGOAL}_A^T(G, G') \end{array}} \end{array}}$$

Principle (3.25) can be considered a principle which connects distinct actions of an agent on the basis of a common goal. Given the knowledge about the agent's goal and the assumption that the agent is committed to a goal in the sense of (3.25), the observer is able to interpret these actions as elements of a plan of the agent's and, therefore, is able to construct a coherent behavior from the individual actions of the agent. Note again that in principle the observer's assumption can be wrong. As already mentioned, the agent is not actually committed to filling the kettle with water in order to make a cup of tea. The fact that the agent pursues a particular goal is just a good reason for the observer to assume that the agent's actions are means to achieve that goal.

Summary of this section: This section was concerned with the question of how an observer can come to assumptions about an agent's mental states and about the connection between different action of his. This section introduced three principles which model an observer's ability to draw conclusions about an observed agent's plans and goals from the actions of that agent. These principles are referred to as the *act-goal connection* (principle (3.23)), which models the assumption that any action is performed in order to achieve some goal, the *generalized goal principle* (principle (3.24)), which models the assumption that if G' can be assumed to be a subgoal of G, then the agent is pursuing G as well, and finally *commitment to a goal* (principle (3.25)) which models the assumption that as long as an agent is in the state of pursuing a goal, all of his actions are directed to this goal. In the next section, principles will be applied in order to see how they can be used to represent an observer's reasoning about an agent's mental state. These principles will be used again in chapter 5, where they will be central for deriving a claim about default assumptions about rhetorical relations.

3.4 A Cup of Tea - Formalized

This section finally aims to apply the tools which have been developed so far in order to show how they can be used to represent an observer's inferences about an agent's goals and plans and, hence, about his mental states. The example which will be discussed for this presentation is again the tea example (3.22). In the following calculations, it will be assumed that there is an observer B who observes an agent who performs the

actions described in (3.22) and that this observer just saw that agent A putting a tea bag in a cup. Under these conditions, B can assume that this action actually has the goal of having a tea bag in a cup. This whole situation can be presented as in (3.26):

(3.26)

$$
\begin{array}{|c|}
\hline
A\ B\ T_0\ E_0\ E_1\ S_1\ x\ y \\
\hline
\textsc{agent}(A) \\
\textsc{agent}(B) \\
T_0 \subset S_1 \\
E_0, E_1 \subset T_0 \\
\textsf{tea-bag}(x) \\
\textsf{cup}(y) \\
K = \boxed{E_0 : \textsf{put.in}(A, x, y)} \\
G_1 = \begin{array}{|l|}\hline S_{G_1} \\ \hline E_0 \prec S_{G_1} \\ S_{G_1} : \textsf{in}(x, y) \\ \hline \end{array} \\
E_1 : \textsf{see}(B, K) \\
S_1 : \textsc{bel}\left(B, \boxed{E_0 : \textsc{goal-directed act}\,(A, G_1)}\right) \\
\hline
\end{array}
$$

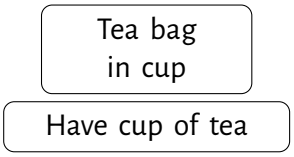
Since B has assumed that event E_0 was a goal-directed act of A, B can apply act-goal connection (principle (3.23)) in order to derive that A was actually in the state of pursuing the goal of having a tea bag in the cup:

(3.27)

$$
\begin{array}{|l|}
\hline
(3.26) \Rightarrow K_2 \\
K_2 = \begin{array}{|c|}\hline S_2 \\ \hline S_2 : \textsc{bel}\left(B, \begin{array}{|c|}\hline S_A \\ \hline E_0 \subset S_A \\ S_A : \textsc{Pursue}\,(A, G_1) \\ \hline \end{array}\right) \\ \hline \end{array} \\
\hline
\end{array}
$$

In the situation shown in (3.27), B can use generalized-goals (principle (3.24)) in order to make assumptions about A's primary goals. However, to use this goal it is necessary to make some stipulations concerning the observer's assumptions about A's goal stacks since having assumptions about an agent's goal stacks is necessary for making assumptions about the maximal goal in this stack. The actual process of guessing a goal stack requires a huge amount of knowledge about the world and some concrete knowledge about the agent. Since these factors cannot be included in the calculation at this point, it will be assumed that B constructed the goal stack in (3.28) for A:

(3.28) $\mathbb{S}_A^{T_0} =$



Now, B is in a situation where he has certain beliefs about agent A's goal stack. It is therefore now possible to apply generalized goals (principle (3.24)) to this situation. This means that B was able to infer that A actually had the goal of having a cup of tea. This can now be presented as follows:

(3.29)

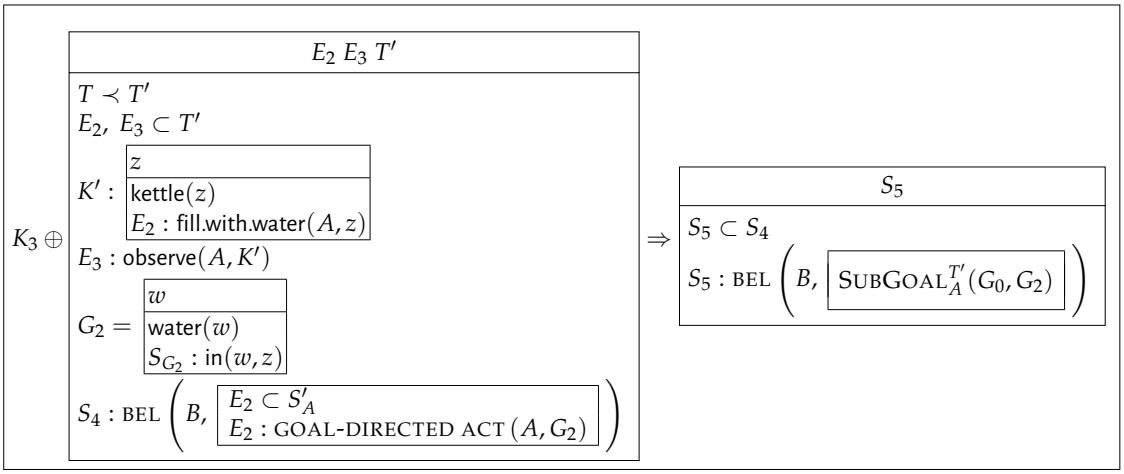
$$K_2 \oplus \boxed{\begin{array}{c} S_2' \\ \hline E_0 \subset S_2' \\ S_2' : \textsc{bel}(B, \mathbb{S}_A^{T_0}) \end{array}} \Rightarrow K_3$$

$$K_3 = \boxed{\begin{array}{c} S_3 \\ \hline E_0 \subset S_3 \\ S_3 : \textsc{bel}\left(B, \boxed{\begin{array}{c} S_A' \\ \hline E_0 \subset S_A' \\ G_0 = \boxed{S_{G0} : \text{have.cup.of.tea}} \\ S_A' : \textsc{pursue}\,(A, G_0) \end{array}}\right) \end{array}}$$

The box in (3.29) represents a situation in which B has assumptions about A's main goals. In this case, B assumes that A is in the state of pursuing the goal of having a cup of tea.

Now assume that B observes the second event of (3.22), i.e. assume that B sees A filling a kettle with water. As an interpreter of the situation, B has to fulfill the task of bringing this action in a reasonable relation to A's former actions or A's former goals. Therefore, B has to make assumptions about A's current goals. In order to come up with assumptions about A's former goals, B can assume that A is subject to inertia (principle (2.53)): According to inertia, an agent's mental states do not change if there is no reason for them to change. At the time of A's second action, B has no evidence that anything has affected A's mental states and, in particular, B has no evidence that anything has affected A's state of pursuing the goal of having a cup of tea. Therefore, B can assume that while A fills the kettle with water, A is still in the state of pursuing the goal of having a cup of tea. This assumption, however, is exactly the precondition for applying commitment to a goal. Hence, B can apply commitment to a goal in order to conclude that the goal of A's action of filling the kettle with water is a subgoal of having a cup of tea. This last step is presented in (3.30):

(3.30)



The box in (3.30) shows B's final inference, i.e. B's assumption that A is filling a kettle with water in order to make a cup of tea. Note again that this conclusion could be wrong. Maybe A was filling the kettle with water in order to water his flowers. Nevertheless, the general structure of B's reasoning and its outcome still seems to be a reasonable model for B's ability to draw conclusions about A's mental states. The individual steps of B's reasoning are summarized again in the following short list:

1. B saw A performing a goal-directed at time T, and hence by act-goal connection B may assume that A is indeed in the state of pursuing a goal at T (i.e. having a tea bag in a cup).
2. Using additional assumptions about A's goal stack and using the generalized goals principle, B was able to conclude that it was actually A's main goal to have a cup of tea.
3. By inertia B was able to derive that at time T' (when A filled the kettle with water) A was still in the state of pursuing the goal of having a cup of tea.
4. Finally, because of the commitment to a goal principle, B was able to conclude that filling the kettle with water still serves the goal of having a cup of tea, i.e. having a kettle filled with water is a subgoal of having a cup of tea.

Summary of this section: This section was concerned with the presentation of an application of the principles which were developed in this chapter. The principles were applied to an example and it was shown how they can be used in order to calculate an observer's assumptions about another agent's goals and plans.

At this point, it has to be remarked that putting a tea bag in a cup is, of course, an action which gives a very strong hint as to the agent's goal. This is of course the very reason why this particular action was used as a motivation and as a starting point for most considerations of this chapter. Even though in this example, it is very clear what the agent's maximal goal is, it can be assumed that additional knowledge about the context of any action makes it possible to make assumptions about the agent's goal stack and, hence, to make assumptions about his maximal goals.

3.5 Summary

This chapter placed its focus on the theoretical basics which will be used for analyzing sequences of speech acts and, hence, rhetorical relations in the next chapters. The notions of goals and of pursuing a goal were introduced. On the basis of the agent's assumptions about causalities, the notions of subgoals and goal stacks were introduced. The subgoal relation was defined as a relation which holds between two goals if the agent believes that one goal helps to achieve another goal. Goal stacks were defined as sets of goals which are partially ordered by the subgoal relation and which contain a maximal element with respect to the subgoal relation.

Afterwards, goal-directed acts and plans were introduced. Goal-directed acts were defined as events which are performed by the agent and which are believed to cause a particular goal. Plans can be regarded as temporally extended versions of goal stacks which are realized by goal-directed acts. Moreover, the question of when an an agent's plan can be considered successful was specified and the question of how an agent can react to an unsuccessful plan was approached. Finally, this chapter placed a particular focus on the observer of an agent's actions and was concerned with the question of how the observer is able to draw conclusions on the agent's plans and goals. In order to be able to model these assumptions, further common sense principles were introduced which made it possible to model an observer's reasoning about the observed agent's actions. The final section of this chapter was concerned with the application of these principles to an example, showing how they can be used to derive an observer's assumptions about an agent's reasons for performing certain actions. In particular, the final action showed how an observer is able to connect distinct actions performed by an agent and how they can be regarded as the individual steps in the agent's plan. This means that the observer was able to construct a "coherent behavior" from distinct actions performed by the agent. This feature will be of particular interest for developing a theory of rhetorical relations.

The definitions of goals and plans will be used again in chapter 4, where communicative plans and goals of speech acts will be defined. The part of this chapter which was concerned with the observer's perspective will be used again in chapter 5 where, the results of this chapter will be applied in order to understand the reconstruction of rhetorical relations by the hearer.

Chapter 4

From Acts to Speech Acts

This chapter shifts its focus from actions in general to linguistic actions, i.e. to speech acts. It is one of the key concepts and insights from speech act theory that using language is a form fo performing actions and, moreover, that the use of language is not limited to the description of states of affairs in the world but also has direct impact on the world. This observation was made very explicit by Austin, (1962). The most impressive examples of the power of language to change the world are *performative* speech acts such as (4.1):

(4.1) I hereby declare you man and wife!

The speaker of an utterance like (4.1) does not describe a state of affairs in the world but, if performed under the right circumstances, he *creates* a state of affairs! These words complete a process which makes the man and the woman in front of the speaker to spouses. The utterance is the actual performance of a wedding.

The observation that speakers actually change things in the world while they are using language is not limited to performatives. By using language, speakers in general want to produce a certain effect in the world, or at least in a part of the world. This part of the world is most of the time the *audience* of the speaker. However, there are exceptions to this assumption: the performance of non-volitional exclamatives like shouting *Ouch*! is not necessarily meant to produce an effect in the speaker's audience (even though, if audience is present, it may have a strong effect on it). Other interesting examples for the use of language which is *not* directed at an audience, but which is meant to have an impact in the world are ceremonial speech acts as they occur in the celebration of a catholic mass for example. A priest may celebrate a private mass and, therefore, speaks the words of institution, which changes the substance of bread and wine, at least according to the dogmatics of the catholic church. The words of institution, are produced because they are meant to have a direct impact on states of affairs in the world but they do not depend on an audience.

Although, there are utterances that are not meant to have an effect on some audience, it can be assumed that such utterances occur sporadically. Therefore, this book

places its focus on utterances which are performed in order cause an effect in the audience. From here on, speech acts are, therefore, regarded as utterances which are performed in order to cause a certain effect in the speaker's audience. This assumption turns speech acts into a special case of goal-directed acts. It implies that some other cases of language use, like senile chatter or talking in the sleep can be excluded from the following considerations.

This chapter formalizes the concept of speech acts and integrates it in the theory of goals and acts which was developed in the preceding chapters. In order to be able to distinguish between speech acts and trivial utterances, speech acts are defined as *goal-directed utterances*. This assumption is formalized in definition (4.2). In this definition, utterances are treated as primitive event type predicates, which take a speaker and some phonological material as arguments.

(4.2) SPEECH ACTS: A *speech act* of agent A with respect to agent B with goal G is an utterance and a *goal-directed act* with goal G:

$A\ B\ E\ U$	
$\text{AGENT}(A)$ $\text{AGENT}(B)$ $\text{PHON}(U)$ $E : \text{EVENT}$	
$E : \text{SPEECH ACT}(A, B, U, G) :\Leftrightarrow$	$E : \text{UTTER}(A, U)$ $E : \text{GOAL-DIRECTED ACT}(A, G)$

This chapter investigates the consequences of this assumption. It is organized in three major parts: Section 4.1 is concerned with the following problem: given the assumption that a speech act is a goal-directed act, then any speech act has a goal. But what types of goals can be pursued by a speech act? Section 4.1 proposes that it is a speech act's general goal to make all interlocutors *committed* to a particular proposition. This section defines the notion of commitments and formulates a version of the common ground of interlocutors in terms of mutual commitments. Section 4.2 is concerned with the role of the hearer for a successful commitment uptake. It discusses different states of grounding and explains why acceptance moves, as certain speech acts signaling the acceptance of a former speech act are necessary for a successful common ground update. The final section 4.3 is concerned with *presuppositions*. It proposes that presupposing is a distinct communicative event with which the speaker makes her assumptions about some content of the common ground public.

4.1 The Goal of a Speech Act: Sharing Commitments

So far, the only characterization of speech acts has been given in (4.2) where it was proposed that a speech act is a goal-directed utterance. This characterization implies that any speech act is performed in order to achieve a particular goal. Given this assumption, a natural question that can be asked at this point is the following: what is

the goal of a speech act? This section aims at giving an answer to this question. It will be proposed that it is the goal of any speech act to share commitments among the interlocutors.

The idea that communication is not a process of sharing information but a process of *sharing commitments* has become more and more popular in recent years. (Roberts, 1996; Condoravdi and Lauer, 2012; Krifka, 2015; Roberts, 2017; Geurts, 2018). This approach to communication has some interesting aspects. First of all, it emphasizes the social nature of communication, since commitments are always commitments with respect to an *audience*. Another very interesting aspect of this view on speech acts, which will be of particular interest in this book, is that it allows a unified treatment of speech acts with different illocutionary forces. If one regards communication as a process of negotiating commitments, it is not important, for example, whether a speech act is uttered with an *assertive* as in (4.3.a) or a *commissive* illocution as in (4.3.b), since both cases can be regarded as means to negotiate commitments.

(4.3) a. I don't like cheese.
b. I'll pick you up at the airport.

After the speaker uttered (4.3.a), he is committed to the fact that he doesn't like cheese; and after (4.3.b), the speaker is committed to pick up the hearer at the airport. Hence, the adaptation of the view that speech acts negotiate commitments makes it possible to give a unified description of the goal of speech acts for *all* speech acts, which can be formulated as follows:

(4.4) GOAL OF A SPEECH ACT: It is the goal of a speech act to negotiate commitments for the interlocutors.

This section pursues the idea formulated in (4.4), i.e. it gives further motivation for this assumption, formalizes this idea, and shows advantages (and disadvantages) of this approach to goals of speech acts. In 4.1.1, the notion of a commitment will be made precise. The idea from Condoravdi and Lauer, (2011) that commitments are always commitments to act in a certain way will be elaborated. In 4.1.2, it will be discussed how interlocutors can reconstruct the content of the commitment negotiated by a speech act from the concrete realization of the speech act. This section has to end with the unsatisfactory conclusion that this problem cannot be addressed in this book. Following the ideas of Geurts, (2018), 4.1.3 redefines the common ground of interlocutors as a set of mutually shared commitments and formulates the assumption that it is the goal of a speech act to perform a common ground update with the content of that speech act. 4.1.4 is concerned with the connection of commitments and mental states of agents. It addresses the problem that, in principle, one cannot make assumptions on an agent's mental state on the basis of her commitments. Section 4.1.5 is a remark on greetings. It shows how greetings, as a very special type of speech acts, can be captured by an approach that assumes that communication is a form of commitment sharing. Finally, 4.1.6 is a remark on the consequence of the assumption that communication is a form of commitment sharing for the direction of fit of speech acts. The two subsections 4.1.5 and 4.1.6 can be regarded as additional remarks on the main topic of

this section and are actually not needed in the proceeding of this book anymore.

4.1.1 Commitments as Commitments to Act

So far, it was proposed that any speech act serves as a device to negotiate commitments. However, it was not yet clarified what the term "commitment" actually means. Clarifying this notion will be the subject of this section. It will be proposed that commitments are always commitments *to act*.

There are many speech acts that rather obviously aim aim at the performance of some action. For example, *commissive* speech acts, such as in (4.5), are clearly cases in which the performance of the speech act makes one of the interlocutors (in this case, the speaker) committed to the performance of some action.

(4.5) I'll pick you up at the airport tomorrow.

In the case of (4.5), it is quite intuitive to say that the performance of this speech commits A to pick up B at the airport at the day after the utterance was made: B is committed to *act upon this promise*. Another example where it seems to be quite clear that it is the job of the involved speech acts to share commitments (but with the roles of hearer and speaker interchanged compared to commissives) are *requests* as in (4.6):

(4.6) a. A: Please pick me up at the airport tomorrow.
b. B: Sure!

In example (4.6), A is first requesting a certain action from hearer B and B agrees on the performance of this action. Agreeing on a request can be regarded as a similar act as promising and, therefore, it is again rather intuitive to assume that after his speech act, B is committed with respect to A to act upon this promise. In both cases, in (4.5) and in (4.6), one of the interlocutors is committed to the performance of a certain action. This observation is in accordance with the work of Condoravdi and Lauer, (2012). They claim that commitments are *always* commitments to act in some way. The speaker who promises to pick up the hearer at the airport is committed to act upon this promise, and, hence, to actually pick up the hearer at the airport. The hearer, on the other hand, can now exclude all possibilities that the speaker does not at least try to pick up the hearer at the airport from his considerations. As soon as the speaker is committed to act upon his promise, all possibilities that the agent does not act in accordance with his commitments are excluded. This means that commitments *restrict future states of the world*.

(4.7) PUBLIC COMMITMENTS: An agent A is publicly committed to a proposition K with respect to an audience B iff, by encountering this commitment, A excludes all possible future states for the audience B, in which A acts as if he does not believe that K.
If A is publicly committed to a proposition K with respect to an audience B, write $PC_{A,B}K$. The proposition K is called the *content of the commitment*.

In the following, the focus is placed on the question of what it means for the speaker to

become committed to a proposition after the performance of a speech act. The hearer's role in this process and what *negotiation* in this context means will be the subject of section 4.2.

So far, only cases of speech acts where it is more or less clear that these speech acts are directed at certain actions of one of the interlocutors have been considered. The examples (4.5) and (4.6) show situations where it seems to be obvious that the conversation is directed towards a certain action of one of the interlocutors and, therefore, it is not surprising that at least one of the interlocutors is committed to that action. However, there are of course many speech acts where it is not as clear that they negotiate commitments as it is in the case of requests or promises. The rest of this section tries to show that actually *all* speech acts can be interpreted as means to negotiate commitments. Take the following examples as starting point for this endeavor:

(4.8) Speaker A to hearer B:

a.	I pick you up at the airport tomorrow.	*promise*
b.	Is it raining?	*question*
c.	I hate cheese.	*assertion*
d.	It is raining.	*assertion*
e.	73 is the 21st prime number.	*assertion*

The list obove is ordered from cases where it is rather obvious that they negotiate commitments *to act* to cases where this is less obvious. It was claimed that all utterances negotiate commitments and, according to (4.7), this means commitments to *act*. So, where are the commitments to act in the examples above? It was already discussed how (4.8.a) determines future actions of the speaker. For a promise, it is rather clear that the speaker is committed to act in a certain way after it was uttered and if it was accepted. Example (4.8.b) is a question. Questions serve as devices to signal that the speaker wants to be given a certain information while the question is uttered. A question, therefore, restricts future states of the world in the sense that the speaker cannot act as if he did not want to know this information and, in that sense, determine his future actions. In the case of (4.8.b), this makes A at least committed to the proposition K =[At the time of the utterance, A wants to know whether it is raining at the time of the utterance]. This implies for example that A cannot react as in (4.9) without expecting some social consequences:

(4.9) A: Is it raining?
B: Yes, it is raining.
A: (?) Why the hell do you tell me that?/I don't care!

Therefore, uttering a question excludes future states of the world in which the speaker acts as if he doesn't care for the answer. Of course, there are different types of questions: there are questions which directly address a hearer and where the speaker wants to have an answer exclusively from the hearer. And there are questions where the speaker just generally asks for information. (4.10.a) shows an example where the speaker directly addresses a certain hearer. (4.10.b) on the other hand shows an example where the speaker does not seem to directly care which hearer provides the

answer to the question.

(4.10) a. Say, is it raining?
b. I wonder: is it raining?

The difference between these types of questions can be expressed in the particular form of the proposition the speaker is committed to. In the case of (4.10.a) the speaker becomes committed to the proposition that he wants to know whether it is raining and that he wants to have this information *from the hearer*. In (4.10.b), on the other hand, the speaker just wants to generally know whether it is raining. Other approaches to questions, such as e.g. Farkas and Bruce, (2009) and Roberts, (2017) propose that questions are just attempts to increase the set of the commonly believed set of propositions (i.e. the context set). However, such approaches ignore, that the fact of the speaker asking something can itself become a subject of conversation. For example, let's assume that A asked B how to make spaghetti carbonara, and that B gives a very extensive answer to that question; and in fact so extensive that, after some minutes, A gets annoyed. In this case the following little dialogue may happen.

(4.11) a. A: Ok...I got it.
b. B: I'm not done yet!
c. A: Yea...but let's talk about something else.
d. B: But YOU ASKED ME for the recepy.

In this case, A looses the interest in the conversation which he started. However, B's reaction is completely legitimate, since it was actually A who asked a question. Given the assumption, that A committed himself to the fact that he wants to know how to make spaghetti carbonara, B's reaction is understandable and, at least to some extend, predictable.

In (4.8.c), the speaker claims a proposition that contains an information about himself. In fact, this makes this speech act, in fact, easy to analyze as a proposal for a public commitment. If A claims that he hates cheese, then he is committed to act upon the proposition $K =$[A hates cheese.] For example, this implies that A excludes future states of the world where he publicly devours cheese. This means that hearer B can count upon the fact that he will never see how A joyfully eats cheese.

Examples (4.8.d) and (4.8.e) are more abstract then (4.8.c) in the sense that the speaker is not part of the content of the proposition of the speech act. But even such abstract cases can be interpreted as attempts of the interlocutor to negotiate commitments. The fact that in (4.8.d) the speaker claims that it is raining commits the speaker to act as if he believes that it is raining. This implies, for example, that the speaker avoids to leave the house (at least if he is known to hate rain) or that he doesn't water the plants. Even in the most abstract case of (4.8), i.e. in (4.8.e), which is a statement about prime numbers, the speaker becomes committed to something and, hence, excludes future states of affairs in the world: if B hears that A is claiming that 73 is the 21st prime number, B can expect that A will not try to perform an integer division of 73 without a reminder. In other words, even speech acts with a rather abstract content, as for example a statement about primes restricts the future actions of the speaker and,

therefore, future states of the world.

On the basis of the discussion above, it will be assumed that all speech acts have the goal to commit the speaker to a certain proposition. However, it is not clear yet to what exact proposition the speaker becomes committed once a speech act was performed. In the case of a question as for example in (4.8.b), it was claimed that the speaker commits to the proposition that he wants to know the answer to that question. In the case of an assertion, on the other hand, A is committed to the propositional content of the assertion. Hence, different types of speech acts negotiate different types of commitments. 4.1.2 places its focus on the link between the content of the commitments and the propositional content of speech acts.

4.1.2 The Content of a Commitment

In the following, the connection between a speech act and the form of the commitment which is negotiated by the speech act is investigated. The problem is that the content of the commitment can differ from the propositional content of the speech act in which the commitment is negotiated. The propositional content (in the sense of Searle, (1969)) of the question in (4.8.b), for example, is something like [It is raining (at the utterance time).]. The content of the commitment on the other hand is K =[At the time of the utterance, the speaker wants to know whether it is raining.]. Hence, there is a gap between the propositional content of a speech act in the sense of Searle and the content of the commitment which is negotiated in the speech act in the sense of this book. Nevertheless, it would be helpful to be able to tell *who* is committed to *what* after a successful speech act. But at least at this point, it seems to be the cost of the assumption that it is the goal of all speech acts to negotiate commitments. Without further knowledge on the context in which a speech act was uttered, it is in many cases very difficult to construct the content of the commitment which is negotiated from the given speech act. This observation was already made by Wittgenstein, (1953). In the beginning of the *Philosophical Investigations* (Philosophische Untersuchungen), and in particular in the paragraphs §§19-21, Wittgenstein raised the question of the actual meaning of certain expressions and how it is possible to learn a certain language game. Take the following (mild) variation of Wittgenstein's own example:

(4.12) Consider a craftsman (let's say a carpenter) who gives orders to his assistant by just naming the tools which he needs. The carpenter just has to say "Hammer" and the assistant brings a hammer; "Screw driver" and the assistant brings the carpenter a screw driver; and so on. Moreover, assume that if the carpenter points at something and says the name of the object he is pointing on, the assistant knows that this object is called that name. So if the carpenter points on something and says "Saw", the assistant knows that this object is a saw and so on.

What exactly does this example tell us? In the context of (4.12), all interlocutors are completely aware of what a particular speech act means, even though in all cases it just consists of one word. However, the carpenter could also have uttered one of the following speech acts:

(4.13) a. Bring me a hammer!
b. I order you to bring me a hammer.
c. Can you pass me the hammer!?

In all of these cases, the content of the commitment is the same as in the one-word speech act "Hammer": the carpenter wants his assistant to give him a hammer. In terms of commitments, this means that it is the purpose of the speech act to make the assistant committed to the proposition [Assistant brings carpenter a hammer.]. Thus a one-word speech act can be used to negotiate the same commitment as a speech act which consists of a longer sentence.

But it gets even worse: one and the same one word utterance (even uttered with the same prosodic contour) can be used to negotiate *different* commitments in different contexts. Take again the situation of (4.13) but assume that in cases where the assistant does not know the requested tool he also may ask what the carpenter is referring to by just recalling the name.

(4.14) *A*: Hammer!
B: Hammer?
A: Hammer. *pointing at a hammer.*

In this situation, one and the same utterance has a completely different intended effect. The first use of *Hammer* is an order, the second use is a question and the third is an explanation. Given this observation, one can come to the unsatisfactory conclusion that it is at least really complicated to derive the content of the intended commitment from the concrete realization of the speech act. Moreover, tt is basically not even possible to derive the content of the negotiated commitment from the propositional content of the speech act. Actually, a huge amount of scholars working on the fields of syntax, morphology, semantics and pragmatics, tries to answer the question, how the actual "meaning" of a sentence can be derived from its form. This might serve as further evidence for the claim that it is quite complicated to derive the form of the commitment which is negotiated by a speech act from the form of the speech act.

However, at this point this prize can be paid in order to gain the possibility of a unified treatment for all speech acts, regardless of their concrete realization and their illocutionary force. The question of how one can derive a goal of a speech act from its concrete realization will not be approached anymore. In most cases interlocutors have quite a robust intuition concerning the actual reason for a speech act's performance, i.e. hearers are often able to guess the content of a given speech act. In order to be able to develop the theory further, the connection of the commitment of a speech act and its concrete realization will be stipulated and will be used as a black box.

4.1.3 Commitment and Common Ground

So far, commitments have been regarded as *speaker* commitments. In (4.7) commitments in general have been defined as restrictions on possible future states of the world: If an agent A is committed to a proposition K with respect to a hearer B, the hearer can rely on the fact that the speaker acts on this proposition. But negotiating

commitments is not a one-way road. The hearer of a promise for example can, and in fact *has* to act upon the promise in the same way as the speaker has to act upon the promise. The agent A, who gave the promise to pick up agent B at the airport, has to expect social consequences if A is not at the agreed place at the agreed time. But the same holds for B: if B is not at the agreed place at the agreed time, he has to expect similar social consequences: A invested time and maybe money in order to pick up B and, hence, in a case where B is *not* at the agreed location, it is legitimate for A to be mad at B. In other words, not just one but *both* agents are committed to the proposition that A picks up B at the airport and both agents have to *act* upon this commitment.

Therefore, a speech act commits both interlocutors to the proposition of the speech act, i.e. the interlocutors are *mutually* committed to that proposition. Geurts, (2018) elaborated on this idea and called the set of mutually shared commitments their *common ground*. This section follows his example and formalizes the idea that the common ground of agents is the set of mutual commitments.

The notion of the common ground of interlocutors was originally established in order to model mutually shared *beliefs* (Stalnaker, 1978). In the following the common ground will be abbreviated as cg. In its classical version, the common ground, is defined as a set of mutual shared beliefs:

$$\varphi \in cg :\Leftrightarrow \forall a \in \mathcal{D}\ (\textsc{bel}(a, \varphi) \wedge \textsc{bel}(a, \varphi \in cg)). \tag{4.15}$$

This notion of the common ground will be referred to as *epistemic common ground.* The notion of epistemic common ground can be used to model the effect of assertions: a felicitous performance of an assertion updates the common ground with the propositional content of the assertion. As already mentioned, this section proposes a notion of common ground which does not model mutual beliefs but *mutual commitments*. Regarding the common ground as a set of commitments has some advantages over the epistemic common ground. For example, it makes it possible to model a common ground update for *all* speech acts in a discourse and not just a common ground update with the content of assertions.

The rest of this section is concerned with a formal definition of the common ground in terms of public commitments. Afterwards, this notion of common ground can be used in order to give a general description of the goal of a speech act.

Recall that the predicate $PC_{A,B}(K)$ means that A is publicly committed to K with respect to an audience B. For a set of discourse participants $\mathcal{D}$, the predicate $PC_{\mathcal{D}}(K)$ will be used to describe the fact that K is a public commitment among a group of agents $\mathcal{D}$, i.e. $PC_{\mathcal{D}}(K)$ holds if any member of $\mathcal{D}$ is publicly committed to K with respect to any other member in $\mathcal{D}$. Based on the predicate $PC_{A,B}(K)$, this can be formalized as in (4.16):

$$PC_{\mathcal{D}}(K) := \bigwedge_{A \in \mathcal{D}} \bigwedge_{B \in \mathcal{D} \setminus \{A\}} PC_{A,B}(K). \tag{4.16}$$

Epistemic common grounds operate with the notion of mutual beliefs. A proposition K is assumed to be mutually believed among a group of agents $\mathcal{D}$ if every member of $\mathcal{D}$ believes K, and if every member of $\mathcal{D}$ believes that every member of $\mathcal{D}$ believes K, and so on. In order to be able to define the common ground in terms of commitments,

the notion of mutual beliefs has to be replaced by the notion of *mutual commitments*: K is a mutual commitment if it is a public commitment and if it is a public commitment that it is a public commitment and so on. This can be made precise by saying that a mutual commitment is an infinitely often iterated version of $PC_{\mathcal{D}}$:

$$MC_{\mathcal{D}}(K) := \bigwedge_{n=1}^{\infty} PC_{\mathcal{D}}^{n}(K), \tag{4.17}$$

where $PC_{\mathcal{D}}^{n}(K) := \underbrace{PC_{\mathcal{D}}(PC_{\mathcal{D}}(\dots(PC_{\mathcal{D}}}_{n \text{ times}}(K))\dots)$. This can (maybe more elegantly) be rewritten in the following iterative version:

$$MC_{\mathcal{D}}(K) :\Leftrightarrow PC_{\mathcal{D}}(K) \wedge PC_{\mathcal{D}}(PC_{\mathcal{D}}(K)). \tag{4.18}$$

The common ground of a group $\mathcal{D}$ can now be defined as the set of propositions to which the group $\mathcal{D}$ is mutually committed to:

(4.19) COMMON GROUND: Let $\mathcal{D}$ be a group of agents. A proposition K is an element of the common ground of $\mathcal{D}$ iff $\mathcal{D}$ is mutually committed to K. The common ground of a group $\mathcal{D}$ will be denoted $\text{CG}_{\mathcal{D}}$:

$$K \in \text{CG}_{\mathcal{D}} \Leftrightarrow MC_{\mathcal{D}}(K).$$

At first glance, this definition of common ground with its iterative structure seems to be quite similar to its epistemic counterpart. Therefore, one could expect that the notion of mutual commitments is subject to similar theoretical problems as the notion of mutual beliefs. For example, in both cases there is an infinite number of conditions which have to hold for a proposition in order to be an element of the common ground. Nevertheless, there are crucial differences between commitments and beliefs which make a commitment based notion of common ground more manageable than its epistemic counterpart. The difference of the two notions of common ground was already made precise by Geurts, (2018). Geurts argued as follows: *First:* Beliefs are *private* by their very nature. It is simply impossible for one agent to inspect the mental states of another agent. Public commitments, on the other hand, are by their very nature *public*: By observing the actions of another agent B, an agent A can check whether B acts in accordance with A's and B's commitments.
Second: If an agent A is publicly committed to K with respect to an agent B, then B can exclude all future states where A acts as though he does not believe K. Hence, it is again rational *for* B to act on A's commitment. In order to illustrate this, consider again the following example:

(4.20) A to B: It is raining

With the utterance of (4.20), A becomes committed with respect to B to the proposition [It is raining] and as a result excludes all worlds where he acts as if he doesn't believe that it is raining (for example A excludes a future where he goes out wearing sunglasses). Therefore, it is now rational for B to act on the proposition $PC_{A,B}$([It is raining.]). In other words, if A is publicly committed to [It is raining], B can act as though he believes that A acts in accordance with his public commitments, which in turn makes

B committed to the proposition $PC_{A,B}(\ $[It is raining.]$\ )$. This last observation can be generalized as follows:

(4.21) If A is committed to K with respect to B, than B is committed to A's commitment to K with respect to A.

This observation will be called *transmission of commitments* and will be formalized in (4.22).
Third: If A is publicly committed to K with respect to B, A has to act as if A believes K. However, this implies that A has to act as if A believes that A is committed to K. In other words, one can assume that if A is committed to K, than A is committed to this commitment. This assumption has an epistemic counterpart which is called *positive introspectivity*[1].

These observations and assumptions are summarized in the following rule (Geurts, 2018):

(4.22) TRANSMISSION OF COMMITMENTS: Let A and B be agents and K a proposition. If A is committed to K with respect to A then

1. B is committed to A's commitment to K and
2. A is committed to A's commitment to K:

$$PC_{A,B}(K) \Rightarrow PC_{B,A}(PC_{A,B}(K)) \wedge PC_{A,B}(PC_{A,B}(K)).$$

The assumptions in (4.22) imply that "half" of the conditions for a proposition to be an element of the common ground given in (4.19) already hold if one of the interlocutors is committed to that proposition with respect to the other interlocutor. Therefore, if A is committed to K with respect to B, the only thing which is left to do in order to make K an element of the common ground is to commit B to K with respect to A.

The fact that commitments are the basic notion of this version of common ground, and the fact that it was assumed that it is the goal of any speech act to negotiate commitments, makes it possible to reformulate the classical assumption that agents communicate in order to update their common ground. This assumption is summarized in (4.23):

(4.23) THE GOAL OF A SPEECH ACT: For any speech act E, there is a proposition K_E such that it is the speech act's goal to update the common ground with K_E. The proposition K_E will be referred to as the *content of the speech act* E:

$A\ B\ E$	
$\text{AGENT}(A)$ $\text{AGENT}(B)$	
$E : \text{SPEECH ACT}(A, B, U, G_E) \Rightarrow$	K_E $K_E : \text{PROP}$ $G_E = \boxed{K_E \in \text{CG}_{A,B}}$

[1] See Fagin, (2003) for an introduction to the axiomatics of epistemic logics.

Remark Since the this terminology has a potential for confusion, a remark on the terminology seems to be adequate. Searle uses the notion of the *propositional content of a speech act* in order to refer to the predicated content. It only depends on the predications in a given speech act. For example, the speech acts in (4.24) all have the propositional content [smoke(Sam)].

(4.24) a. Does Sam smoke?
b. Sam smokes.
c. Sam, smoke!
d. Would that Sam smoked.

The notion of the *content of a speech act* as it is used in this book, on the other hand, refers to propositions with which the speaker wants to update the common ground. Hence the content of the speech act (4.24.a) is [Speaker wants to know, whether Sam smokes].

This section defined the goal of a speech act as a common ground update with the content of the speech act. To be an element of the common ground in the sense of this book means that both, speaker and hearer, are mutually committed to the propositions in the common ground. According the characterization of commitments given in (4.7), a committed agent has to act *as if* he believes the content of the commitment. This formulation makes it possible to measure the success of a speech act on the basis of an agent's behavior. Moreover, it makes it also possible to deal with a whole class of *insincere* speech acts, such as lies[2]. A lie can only work if the liar acts in the presence of his audience as if he believes his lies. Otherwise, there would be no change for the lies to be successful in the sense that the target audience acts upon the lie's content. However, lies will do not play a role in this book anymore.

So far, it was claimed that speakers communicate in order to negotiate commitments and that commitments are commitments to *act*. However, in principle, actions do not have to be in accordance with the mental states of agents. But in general, one would suppose that there is at least some kind of connection between the behavior of an agent and the agent's mental states and, in particular, that the mental states of an agent guide his actions, at least to some extent. The problem of the connection between the mental states of an agent and his commitments will be discussed in 4.1.4

4.1.4 Sincerity and Commitments

It was assumed that it is the goal of a speech act to establish a mutual commitment to the content of the speech act among the interlocutors. This approach to speech acts has some advantages. For example, it allows unified treatment of all kinds of speech acts

[2]There are other cases of speech acts which might provide a problem for the account of speech acts as commitment sharing. The most challenging examples for a commitment based notion of a common ground are probably fictional discourses and speech acts which occur as parts of theater plays. Such examples are of course excluded from our considerations. However, I believe that even these cases can be captured by the commitment based approach, but this will not be part of this book.

and one does not have to differentiate between speech acts with different illocutionary forces. It emphasizes the social nature of communication since commitments are, by definition, public. However, there are some problems: A commitment based approach to communication makes it difficult to recover the content of the commitment from a given speech act. This was already discussed in 4.1.3. Moreover, modeling communication only in terms of public commitments allows for some unintuitive phenomena concerning an agent's actions and his mental states. These aspects will be targeted in this section.

The main problem which will be addressed is the following: in principle, there is no connection between a commitment and the mental states of an agent. If an agent claims that it is raining, he is committed to the proposition that it is raining an, therefore he has to act *as if* he actually believes that it is raining. Nevertheless, he does not actually have to believe that it is raining while being committed to act as if it is raining. This mere observation is no problem in principle, since this is what makes lying possible. An agent can claim against his own beliefs that it is raining and thereby becomes committed to that proposition. Moreover, such a speech act can actually be successful in the sense that it commits all interlocutors to that proposition. Such a speech act can be even successful a priori in the sense that it establishes a commitment if no interlocutor actually believes this proposition. In such a situation, all interlocutors have to act as if they believe that it is raining, but at the same time, no one actually believes that it is raining. In other words, everybody would just *pretend* that it is raining.

Such a situation was described in the short tale *The Emperor's New Clothes* by Hans Christian Andersen. In this tale, two weavers promise new clothes for an emperor but they claim that these clothes are invisible for everybody who is stupid or incompetent. But the truth is that they made no clothes at all. The emperor, however, does not want to seem incompetent and, therefore, he pretends to act as if he actually is able to see his new clothes. This in turn made all his subjects pretend to act as if they all see the emperor's new clothes. In the terminology of this book this can be restated by saying that everybody near the emperor and the emperor himself is committed to the proposition [one can see the emperor's new clothes] but at the same time no one actually *believes* this proposition. In Andersen's short tale, the whole situation breaks down in the middle of a procession when a child shouts that the emperor is actually wearing no clothes at all. Even though the emperor realizes that the child is right, he continues his procession in order to keep his face.

A priori, it is possible to act like the emperor and his subjects in Andersen's tale, i.e. one can act against one's actual beliefs. Nevertheless, it seems to be natural to assume that keeping up such a charade for a longer period of time is a difficult task; it appears to be extremely fragile and unnatural. In Andersen's tale this fragility is expressed by the fact, that a single child can tear down the whole situation.

Recent psychological studies support the claim that acting against own beliefs is difficult consumes more cognitive resources than acting in accordance with own believes. Lying, which in these studies is defined as making statements that do not match with actual beliefs, can be considered as a prototypical phenomenon for *acting* against own beliefs. Despite the fact that one can actually *learn* to lie in certain situations (van Bockstaele et al., 2012), it can be considered as the default case that successful decep-

tion requires a greater amount of cognitive resources then telling the truth (Vrij et al., 2006). There even is empirical evidence that agents tend to be more honest when put under high cognitive load then under low cognitive load (Veer, Stel, and van Beest, 2013), which again provides evidence for the claim that acting against own beliefs requires more cognitive resources then acting in accordance with own believes. Given these findings and the limitation of cognitive resources for human agents, it is, therefore, reasonable to assume that acting in accordance with own beliefs is the default case for human agents.

In the following, this assumption is incorporated in the theory developed in this book. This new principle models the assumption that if agents are publicly committed to some proposition, then by default they are in an "appropriate mental state" towards this proposition. This principle is called *sincerity* since it corresponds roughly to the sincerity condition in speech act theory from Searle, (1969). In order to make sense out of "appropriate mental states", consider again the conditions for public commitments (4.7), here repeated as (4.25):

(4.25) An agent A is publicly committed to a proposition K iff A excludes all possible future states, where A acts *as if he does not believe* that K is true.

In other words, an agent has to act as if he believes K. On the basis of the considerations above, it is reasonable to assume that acting in accordance with own beliefs is the default case for human agents. This implies that if an agent is publicly committed to a proposition K, then by default the agent actually *believes* K. This assumption is summarized in principle (4.26):

(4.26) SINCERITY:
If an agent A is publicly committed to a proposition K with respect to B, then A also believes K:

$$\frac{\begin{array}{l} A\ B\ K \\ \hline \text{AGENT}(A),\ \text{AGENT}(B),\ K : \text{PROP} \\ PC_{A,B}K \Rightarrow \dfrac{S}{S : \text{BEL}(A,K)} \end{array}}{}$$

In contrast to public commitments, sincerity is not a public condition in the sense that interlocutors can't tell whether other interlocutors are actually sincere since mental states are still private. The problem that beliefs are private and therefore not verifiable by other agents was actually overcome by making public commitments the central notion of communication. Adding sincerity to the given framework has the flavor of this advantage being lost. However, this is actually not the case: public commitments restrict *actions* of agents and these actions are observable by other agents, and it can therefore be still be controlled whether an agent acts in accordance with his public commitments. The observer's ability to make judgments about the agent's mental states can be seen as an epiphenomenon of public commitments. Adding sincerity

to the given framework enriches it with further realistic constraints for commitment uptakes, as for example some propositional mismatches in the mental states of the interlocutors. Take for example a situation in which an agent believes that another agent does *not* believe a certain proposition. If this is the case, this should prevent agents from a common ground update with that proposition: If A believes that B does not believe that K and if A believes that B is a sincere person then A cannot expect that B will act in accordance with K. Of course, this might happen for complicated social reasons, as for example in *The emperor's new clothes*, but in general not believing a given proposition is a constraint for a commitment uptake.

So far, only the speaker's part of communication was considered. The speaker has the goal of negotiating a mutual commitment and, therefore, performs a speech act, of which he believes that it helps to bring about that commitment. However, establishing a mutual commitment, i.e. performing a common ground update, is not only the speaker's task. A common ground update does not come for free: next to the conditions in the definition of the common ground there are several other conditions that have to hold for a successful commitment uptake. These conditions will be investigated in section 4.2. However, this section ends with a remark on commitments in the case of *greetings* and with a philosophical remark on the consequences for the *direction of fit* of speech acts. These remarks do not have any consequences for the proceedings of this book. The remark on greetings is more of a case study for a speech act which is in some sense very special as will be shown in that section, and the remark on direction of fit is just of philosophical interest.

4.1.5 Greetings and Commitments

This section is a remark on *greetings* and how greetings fit into a framework of speech acts as means for commitment sharing. It was stated that any speech act has the goal to make the interlocutors committed to a certain proposition, the content of the speech act, and that this proposition has to be recovered from the realization of the speech act somehow. Moreover, in section 4.1.2 it was claimed that there is no general procedure to construct the content of the speech act from its realization. The argument for this rather pessimistic claim was based on one-word speech acts which showed that the performance of one and the same utterance can be used to negotiate completely different commitments. Another interesting and, in some sense, *extreme* case of speech acts where one does not see the intended commitment on the nose are greetings. Greetings are interesting for a commitment based approach to communication since they represent an extreme case of speech acts. Their position of a conversation literally extremal in the sense, that they normally constitute the first speech acts of a conversation, they do not seem to contain any actual information and they are usually very short and can contain even just one syllable. Hence, it is reasonable to ask whether greetings can be captured in terms of public commitments. This section investigates greetings and shows how they can be treated on the basis of public commitments.

According to Searle, (1969, p. 67), the propositional content of greetings is vacuous. In Searle's terminology, this assumption makes completely sense, at least for a simple greeting as *Hello*, since such a speech act does not involve any actual pred-

ication. In terms of public commitments, however, this assumption seems to be too weak. This section pursues the idea that speakers actually undergo very strong commitments when greeting someone. It will be proposed that by greeting someone the speaker actually encounters a commitment to the proposition that in principle he is ready to communicate:

(4.27) It is the purpose of greeting to signal the speaker's they readiness to communicate.

The assumption in (4.27) captures the intuition that a greeting commits the speaker, at least to some extent, to tolerate communicational attempts of the greeted. If agent A greets agent B and B then starts a conversation, it would be considered to be very rude if A stopped the communication without any further excuses. Take for example the following conversation(s):

(4.28) A: Hi, George!
B: Hi! How is your son?
A: ??I don't want to talk.
?I don't have time.
Sorry, but I have to catch the bus. We can talk later...

A's first answer (indicated with ??) would probably be considered very rude, A's second answer (indicated with ?) is slightly better since it at least gives a rough explanation for A's behavior, but would still be considered to be rude. In both of these cases, A acts against own public commitments. A's third answer is the best. In this answer, A signals that he is in principle ready for conversation and he therefore has to talk at least as much as not to be considered rude. Furthermore, the observation above is symmetrical: if the hearer of a greeting doesn't ignore it, but accepts the greeting in some way, then the hearer also has to tolerate at least a minimal amount of conversation; i.e. a greeting makes indeed *both* agents committed to the fact that they are ready for communication. Hence, greetings can be given the following semantic interpretation:

(4.29) GREETINGS: A greeting is a speech act which pursuesthe goal to make interlocutors publicly committed to the proposition that they may engage in communication:

$A\ B\ E\ U$	
AGENT(A) AGENT(B) U : PHON	
E : GREETING$(A, B, U) \Rightarrow$	K
	$K =$ ready.to.communicate$\{A, B\}$
	E : SPEECH ACT (A, B, U, K)

The condition in (4.29) is of course not sufficient to describe all aspects of greetings. Greeting is a complicated act which involves many different social aspects. One aspect of greeting which is missing in (4.29) is the fact that a greeting should normally be the first speech act in a conversation. However, this condition can easily be added, if a

mechanism is included which keeps track of discourse moves as it was done e.g. by Ginzburg, (2012, p. 75).

So far, greetings have been concerned and the question of how one can integrate greetings in the framework of mutual commitments was approached. It was possible to show that even extreme cases of speech acts, such as greetings, can be captured with this framework. The fact, that a commitment-based approach to speech acts is actually able to deal very special types of speech acts such as greetings provides further evidence for the flexibility of this approach.

4.1.6 A Remark on Directions of Fit

In this book, it is assumed that it is any speech act's goal to achieve commitments among all interlocutors; Furthermore, it is assumed that such commitments restricts future states of the world. This section, finally, wants to comment on the consequences of this assumption for the *direction of fit* of speech acts.

Classical speech act theory in the sense of Searle and Vanderveken, (1985) assumes that different speech acts have different directions of fit. An assertive speech act, for example, has the *word-to-world* direction of fit, i.e. the words should fit the world, in the sense that the states of affairs which are described in an assertive act should actually hold in the world. A commissive speech act like a promise, on the other hand, has the *world-to-word* direction of fit: The promise works out, if the world, fits the content of the promise. However, it was claimed that it is the goal of a speech act to negotiate commitments, and that commitments restrict future states of the world. But saying that a speech act restricts future states of the world means that the future states of the world have to fit the content of the speech act. In the terminology of classical speech act theory, this means that *any* speech act has a world-to-word direction of fit. In cases of directives and commissives this is in accordance with the classical assumption from speech act theory but not in the case of assertive speech acts which are assumed to have the word-to-world direction of fit.

Nevertheless, these observations should cause no problems, neither for a theory of speech acts, which operates with illocutionary forces, nor for a theory which operates with public commitments as goals of speech acts. Direction of fit is a main feature of different illocutionary forces. But illocutionary forces are private in the sense that the hearer does not have to know about the illocutionary force of a speech act for the existence of that force. The following example, taken from Condoravdi and Lauer, (2011), illustrates this behavior of illocutionary forces:

(4.30) He ordered Kornilov to desist but either the message failed to reach the general or he ignored it.

Even though Kornilov's speech act failed, it was still an order and, hence, had the illocutionary force of an order.

Commitments, on the other hand, are public and therefore verifiable. The assumption that speech acts negotiate commitments, makes the effect of the speech act on its audience crucial for the speech act. However, the illocutionary force of speech act exists even if the speech act doesn't have an impact on the world. This means that these

two approaches to speech acts, an approach which focuses on the illocutionary force of a speech act the approach that focuses on its social effect in terms of public commitments, are not exclusive, but just emphasize different aspects of speech acts.

Summary of this section: This section was concerned with goals of a speech act. It was proposed that it is the general goal of a speech act to negotiate mutual commitments. Commitments have been regarded as commitments to act: the committed agent has to act as if he believes the content of the commitment. A notion of the common ground was established according to which the common ground of a group of agents is a set of propositions to which all agents are mutually committed to. Given this notion of common ground it was able to formulate that it is the goal of a speech act to update the common ground with the content of the speech act. Moreover, the problem was addressed that it is in principle it is not possible to make assumptions on an agent's mental states just on the basis of the commitments he is engaged in. Nevertheless, it is reasonable to assume that (most of the time) agents act in accordance with their own beliefs. In order to model this intuition, a principle was introduced which states that by default an agent actually acts in accordance with his beliefs. This principle provides an honest constraint for a commitment uptake since it forces agents to only be committed to propositions which the believe. Finally, this section was concerned with greetings. It was shown how the effect of a greeting can be modeled in terms of commitments. The last section of this section discussed some consequences of the assumption that communication is a process of commitment sharing in terms of the direction of fit of speech acts. It was emphasized that a commitment based approach to speech acts and illocutionary forces of speech acts are not exclusive but just stress different aspects of speech acts.

4.2 Commitment Uptakes

Section (4.23) formulated the assumption that it is in general a speech act's goal to perform a common ground update. However, communication and hence a common ground update is not a one-way road and cannot be performed by the speaker alone. A common ground update is a coordinated process which involves all interlocutors: speaker *and* his audience have to do their respective part for the success of a speech act. This section aims at discussing the hearer's part of communication and, in particular, the hearer's role in a common ground update.

The discussion of this section proceeds in the spirit of Clark and Brennan, (1991). Clark and Brennan emphasized the fact that communication is a coordination process which aims at establishing a common ground. Note that Clark and Brennan thought about the common ground as an epistemic common ground. This coordination process was called *grounding*.

According to Clark and Brennan, grounding in discourse consists of at least two phases:

(4.31) PHASES OF COMMUNICATION: Assume that A is the speaker and B is the hearer of an utterance U. The grounding process can be divided into the

following phases

1. *presentational phase*: A presents utterance U for B to consider. He does so on the assumption, that if B gives some evidence e, A can believe that B understood, what A means by U.
2. *acceptance phase*: B accepts utterance U by giving evidence e that B believes that he understands, what A means by U. B does so on the assumption, that once A registers e, A will also believe that B understands U.

(Clark and Brennan, 1991, p. 130)

In order to emphasize the dynamic nature of grounding as a coordination process, speech acts in a discourse will also be called discourse *moves*, as is common practice in the work of Clark and Brennan. The assumption that grounding is divided in these two phases is based on empirical evidence coming from natural dialogues in which moves occur frequently that have the only purpose to signal the hearer's acceptance. Take the example (4.32) from Ginzburg, (2012, p. 101):

(4.32) a. A: Several people showed up. Bill did.
b. B: Aha.
c. A: Max did.
d. B: I see.

This example shows both phases of grounding: the moves in (4.32.a) and (4.32.c) are part of the presentational phase, whereas the moves in (4.32.b) and (4.32.d) are part of the acceptance phase. Moves which have the only purpose to constitute the acceptance phase such as (4.32.b) and (4.32.d) will be called *acceptance moves*.

In addition to the two phases of communication presented in (4.31), Clark and Brennan present a list of necessary conditions of an utterance before it can be accepted. These grounding conditions (or states of grounding as they are originally called) lie, so to speak "in between" the two phases of grounding. An utterance U can only satisfy this conditions *after* it was uttered, but these conditions have to be satisfied *before* the hearer can perform an acceptance move for U. However, they constitute necessary conditions for a successful grounding of an utterance. The list in (4.33) presents a slightly modified version of these conditions.

(4.33) CONDITIONS FOR GROUNDING: For a speaker A, a hearer B and an utterance U, the following conditions are necessary for a successful grounding of the utterance U

State 1 : B noticed that A uttered U
State 2 : B correctly heard U
State 3 : B understood what A meant by U

(Clark and Brennan, 1991, p. 130)

Section 4.1 was basically concerned with the first phase of grounding, i.e. with the goals of the presentational phase. This section, on the other hand, investigates more on the conditions of the list in (4.33) and on the *acceptance phase*. It will be structured

as follows: 4.2.1 will discuss what it actually means to accept a speech act in terms of mental states of the hearer. 4.2.2 tries to give an answer to the question of why evidence for acceptance in the form of certain acceptance moves is needed and why acceptance is not the default case. Since acceptance does not always have do be signaled by an acceptance move, 4.2.3 consists of a brief discussion on other possible forms of acceptance. Finally, 4.2.4 will be concerned with the question of how a common ground update after the performance of an acceptance move can be modeled and presents a possible (yet still problematic) solution to this problem.

4.2.1 Acceptance of Speech Acts

Clark and Brennan, (1991) pointed out that a common ground update does *not* come for free. A speech act has to satisfy at least the list of conditions in (4.33) in order to have a chance to update the common ground. One basic requirement for a successful common ground update is of course that the intended hearer of the speech act notices that he was meant as the addressee of the speech act. If this is not the case, the speech act is stuck before it even satisfied state 1 of the list in (4.33). In this case, the speech act is basically lost. It cannot achieve anything since the hearer is simply ignorant about it. Hence, for the mechanics of the grounding this case is not of interest and can be ignored in the following considerations. Hence, it will be assumed that E already achieved state 1, or higher. This means that hearer B noticed the utterance of a E and, hence, can conclude that speaker A pursues some goal with that utterance. However, even if this is the case, it is still not neccessarily the case that B already knows which exact goal A pursued with the utterance of E, i.e. with which proposition A wants to update the common ground. Even worse, B does not even have to have correctly heard the phonetic realization of the given speech act. However, in the formal framework which was developed so far, it is still possible to represent the first state of the list given in (4.33). Such a representation may look as follows:

(4.34) ACHIEVING STATE 1: Let A and B be agents and let E be a speech act of A with addressee B. The speech act is in state 1 of the list in (4.33) if the following holds:

$$\begin{array}{|c|}\hline A\ B\ E\ U\ K\ S \\ \hline \textsc{agent}(A), \textsc{agent}(B) \\ E \subset S \\ E : \textsc{Speech Act}(A, B, U, K) \\ S : \textsc{bel}\left(B, \begin{array}{|c|}\hline U',\ K' \\ \hline E : \textsc{Speech Act}(A, B, U', K') \\ \hline \end{array}\right) \\ \hline \end{array}$$

In this case, B can already conclude that A pursued a goal with the performance of E.

Note that the formulation of state 1 given in (4.34) contains two pairs of variables U and K and U' and K'. U represents the actual phonetic material of the speech act E and K

the actual content of that speech act, i.e. the proposition with which the speaker wants to update the common ground. U' and K' on the other hand, is the material which is perceived by the hearer and these can differ from U and K. It is simply possible that the hearer misheard[3] some of the phonetic input. This can be represented by the condition that $U \neq U'$. Another possibility is that the hearer misinterprets the intendend meaning of the given speech act, which can be represented by the condition that $K \neq K'$. If $U = U'$, then the speech act achieved state 2 of (4.34); if in addition to that, $K = K'$, the speech act achieved state 3 of that list. This means in particular that the presentation in (4.34) can be used to easily formalize the three different states of grounding:

1. $U \neq U'$ and $K \neq K'$: The hearer correctly heard that she is the actual addressee of the speech act but doesn't correctly understand the utterance and therefore is not able to construct the goal of the utterance. The speech act is in state 1. but not higher.
2. $U = U'$ and $K \neq K'$: The hearer understands the utterance but is not able to reconstruct the goal of the speaker. The speech act is in state 2. but not higher.
3. $U = U'$ and $K = K'$: The hearer correctly heard the utterance of the speech act and is able to reconstruct the speaker's communicative intentions from the utterance. The speech act is in state 3.

In addition to (4.34), which formalized the most basic step of grounding, these different phases can be formalized as in (4.35). For reasons of simplicity, it is assumed in (4.35) that these formulas are embedded in a DRS as in (4.34), which contains the universe and predications over the speech act event, phonological content, propositions, etc.:

(4.35) CONDITIONS FOR GROUNDING: Let E : SPEECH ACT(A, B, U, K) be a speech act of A towards B with content K and phonological realization U. Let $\psi_1, \ldots, \psi_3$ denote the following propositions:

State 1: $\psi_1(E) =$

$$\boxed{\begin{array}{l} S_B \\ \hline S_B : \text{BEL}\left(B, \boxed{\begin{array}{l} U'\ K' \\ \hline E : \text{SPEECH ACT}(A, B, U', K') \end{array}}\right) \end{array}}$$

State 2: $\psi_2(E) =$

$$\boxed{\begin{array}{l} S_B \\ \hline S_B : \text{BEL}\left(B, \boxed{\begin{array}{l} U'\ K' \\ \hline E : \text{SPEECH ACT}(A, B, U', K') \end{array}}\right) \\ U = U' \end{array}}$$

[3]The possible situation that the realization of the speech act differs from the speaker's intended realization, i.e. the case of wrong pronunciation of the speech act, is not considered in this book, although it is in principle possible to represent this case in a similar form as in (4.34).

$$\text{State 3: } \psi_3(E) = \left[\begin{array}{l} S_B \\ \hline S_B : \text{BEL}\left(B, \left[\begin{array}{l} U'\ K' \\ \hline E : \text{SPEECH ACT}(A, B, U', K') \end{array}\right]\right) \\ U = U',\ K = K' \end{array}\right]$$

If a speech act satisfies the conditions in state 3, it is possible that this speech act achieves its goal. However, there are further necessary conditions for a successful common ground update. One of these conditions is that a speech act in a conversation has to be *accepted* in some way. Clark and Brennan, (1991) emphasized in their definition of the *acceptance phase* that a hearer accepts a speech act by giving evidence for his willingness to accept a speech act. Section 4.2.2 investigates the question of why giving evidence for acceptance is actually necessary for a successful common ground update. But before question will be addressed, the notion of *acceptance* in general will be clarified. This will be the topic of the rest of this section..

It was already mentioned that acceptance moves occur quite frequently and, hence, seem to be important. Acceptance moves provide evidence for the willingness of the hearer to accept a speech act. In the previous section, it was claimed that speech acts are performed in order to achieve a certain goal, namely to make a certain proposition an element of the common ground. But how do acceptance moves fit in this framework?

If it is a speech act's goal to make a proposition an element of the common ground, the hearer has to pursue the same goal at a given point, in order for the speech act to be successful. This means that accepting a speech act can be identified with the process of adopting the speaker's goal to make the content of the speech act an element of the common ground. This idea is pursued, from here on, i.e. it will be assumed that accepting a speech act means to pursue the goal the speech act which is to update the common ground with its content:

(4.36) ACCEPTANCE: Let A an B be agents and let E : SPEECH ACT(A, B, U, K) a speech act of A. B *accepts* E if $\psi_3(E)$ of the list given in (4.35) holds and if B also pursued A's goal to make K an element of their common ground. In the following, $\psi_4(E)$ will denote the state where speech act E is accepted:

$$\left[\begin{array}{l} A, B, \\ \hline \psi_4(E) = \left[\begin{array}{l} S_B,\ S'_B \\ \hline S_B : \text{BEL}\left(B, \left[\begin{array}{l} U'\ K' \\ \hline E : \text{SPEECH ACT}(A, B, U', K') \end{array}\right]\right) \\ U = U',\ K = K' \\ S'_B : \text{PURSUE}(B, K \in \text{CG}) \end{array}\right] \end{array}\right]$$

Definition (4.36) states that an agent *accepts* a speech act if he adopts the goal to make the content of the speech act an element of the common ground. This definition has a similar taste as the *cooperativitiy* [4] axiom by Asher and Lascarides, (2003, p. 391).

[4] This axiom is called *cooperativity* since it is actually supposed to model the Gricean axiom with

This axiom states that, by default, both agents adopt the goals of the respective other agent. If this is not the case, one of the agents has to make the respective other agent believe that she does *not* adopt the goals of the other agents (i.e. this agent has to mark disagreement somehow). However, the fact that Clark and Brennan declare the existence of an acceptance phase for as a necessary part of grounding provides a problem for this axiom. According to cooperativity in the sense of Asher and Lascarides, the hearer would only have to signal *disagreement*. The version of SDRT which was developed in 2003 wrongly "stipulates that lack of disagreement implies agreement" (Asher and Lascarides, 2009). The empirical work of Clark, however, actually shows that *both* cases, i.e. acceptance and disagreement, have to be marked. Asher and Lascarides, (2009) noticed this error of SDRT in its original conception and implemented a notion of acceptance and grounding in SDRT.

So far, it was clarified how acceptance can be modeled in terms of goals of agents. However, acceptance in this sense is still a private issue. What is still missing is the acceptance phase and its realization. This will be the subject of 4.2.2. There it will also be shown why a lack of disagreement does not imply agreement. In particular, it shows why acceptance moves in the sense of Clark and Brennan are actually necessary for a successful grounding.

4.2.2 Acceptance Moves

Clark and Brennan, (1991) divided communication into two phases: the *presentational phase* and the *acceptance phase*. In the presentational phase, the speaker presents his speech act, and in the acceptance phase, the hearer has to give some evidence for the acceptance of the speech act. But accepting a speech act implies that the hearer is actually willing to accept the given speech act, which means that the hearer adopted the goal of the speech act, at least in the terminology of this book. In other words, accepting a speech act and signaling acceptance are again two different things. This means that the acceptance phase can again be subdivided into the hearer's *acceptance* in the sense of (4.36) and the performance of an *acceptance move*:

(4.37) TWO STEPS OF THE ACCEPTANCE PHASE:

1. The hearer *accepts* a speech act if he adopted the goal to update the common ground with the content of the speech act. This was formalized in (4.36).
2. An *acceptance move* is a speech act with which the hearer gives the speaker evidence for the fact that he has accepted a speech act.

This makes an acceptance move a special type of speech act with the goal to update the common ground with the content of one (or more) preceding speech acts. This section is devoted to the question of why such acceptance moves exists, why they seem to be necessary for communication, and how a common ground update can be modeled after such an acceptance move was uttered. The approach to acceptance moves developed in this section is heavily inspired by the work of Ginzburg, (2012), who already

the same name.

included acceptance moves in his theory. Therefore, it seems to be a good idea to have a short look on Ginzburg's approach to acceptance moves.

In general, acceptance moves, as evidence for acceptance, can have different realizations. The hearer can utter something as *OK* or *aha* or he can just nod his head. The following dialogue shows different realizations of acceptance moves in (4.38.b), (4.38.d) and (4.38.f).

(4.38)
a. *A*: Helen is coming tomorrow.
b. *B*: OK.
c. *A*: She wants to go to the movies that night.
d. *B*: mhm.
e. *A*: I think, we have to join her.
f. *B*: OK.

In (4.38.a), *A* made a proposal to update $\text{CG}_{A,B}$ with the proposition *K*=[Helen will come the day after the utterance]. By uttering *OK* in (4.38.b), *B* signals that he accepted *K*. The dialogue continues in a similar way. As already pointed out, such moves are going to be interpreted as an attempt to signal that it is also *B*'s goal to make the content of *A*'s speech act an element of the common ground.

Ginzburg, (2012, p. 85) already formalized acceptance and acceptance moves in his theory of conversation. He proposes the following semantics for an acceptance move.

(4.39)
$$\texttt{Accept move} := \left[\begin{array}{ll} \texttt{pre} : & \left[\begin{array}{lll} \texttt{spkr} & : & \texttt{Ind} \\ \texttt{addr} & : & \texttt{Ind} \\ \texttt{LatestMove} = \texttt{Assert}(\texttt{spkr},\texttt{addr},p) & : & \text{IllocProp} \\ \texttt{qud} = \langle p?, \texttt{pre.qud}\rangle & : & \text{poset(Question)} \end{array}\right] \\ \texttt{effects} : & \left[\begin{array}{lll} \texttt{spkr} = \texttt{pre.addr} & : & \texttt{Ind} \\ \texttt{addr} = \texttt{pre.spkr} & : & \texttt{Ind} \\ \texttt{LatestMove} = \texttt{Accept}(\texttt{spkr},\texttt{addr},p) & : & \text{IllocProp} \end{array}\right] \end{array}\right]$$

Translated in ordinary language, this schema roughly means the following:

a. An acceptance move has to satisfy certain *preconditions* (`pre`) in order to be felicitously performed:
   (i) there is some material p? on the `qud`-stack
   (ii) the assertion of p was the last move in the dialogue
b. The *effect* of an acceptance move is:
   (i) the former hearer accepts p
   (ii) speaker and hearer switch places.

As usual, `qud`s are questions which are raised in the conversation and where all interlocutors are committed to answer that question. Hence, an accept move has the effect that `Accept` was the last move of the conversation. This, on the other hand, is a precondition for applying `FactsUpdate`:

(4.40) $$\texttt{FactsUpdate} := \begin{bmatrix} \texttt{pre} & : & \begin{bmatrix} p & : & \text{Prop} \\ \texttt{LatestMove} = \texttt{Accept}(\texttt{spkr}, \texttt{addr}, p) & : & \text{IllocProp} \\ \texttt{qud} = \langle p?, \texttt{pre.qud} \rangle & : & \text{poset(Question)} \end{bmatrix} \\ \texttt{effects} & : & \begin{bmatrix} \texttt{facts} = \texttt{pre.facts} \cup \{p\} & : & \text{Set(Prop)} \\ \texttt{qud} = \texttt{NonResolve}(\texttt{pre.qud}, \texttt{facts}) & : & \text{poset(Question)} \end{bmatrix} \end{bmatrix}$$

FactsUpdate removes p? from the qud-stack, updates Facts (which is basically Ginzburg's version of a common ground), with the material of the last discourse move:

a. In order to perform FactsUpdate with p, the following preconditions have to be satisfied:
   (i) p? on the qud-stack
   (ii) the acceptance of p was the last move in the dialogue

b. The *effect* of an acceptance move is:
   (i) Facts gets updated with p
   (ii) all questions, which can be resolved by the new facts are removed from the qud-stack

The effect of an acceptance move is the concatenation of these two rules to the former communicative situation. The result of an acceptance move in Ginzburg's framework is, therefore, an update of facts with p. Ginzburg only considers the overt material of a dialogue, i.e. only material which is actually realized. In particular, the mental states of the interlocutors are not really mentioned in his framework. The only aspect of his framework which at least makes reference to beliefs is the set Facts.

Ginzburg's formalization of acceptance moves captures the simple intuition that the performance of an acceptance move is evidence is evidence for accepting the content of some former speech act which in turn means that this content can become an element of the common ground. However, the question is: why is acceptance not the default case? In principle, it is possible to assume that only disagreement has to be marked, as it actually was predicted by the early versions of SDRT. But it seems to be the case that giving evidence for acceptance is still needed for a successful grounding of the speech act. The theory developed so far is actually able to explain the fact that the speaker needs some kind of acceptance move by the hearer. Acceptance moves can be explained as an epiphenomenon of the common sense principle of inertia (principle (2.53)). The argument for this explanation is roughly the following: the speaker doesn't have direct access to the hearer's mental state. Because of inertia, the speaker has to assume that nothing in the hearer's mental states changes until the speaker has some evidence for assuming the opposite. The hearer is also aware of the fact that both interlocutors are subject to inertia. Therefore, the hearer can conclude that the speaker believes that nothing in the hearer's mental state has changed until evidence is provided for such a change. This means that the hearer has to provide such evidence and, hence, an acceptance move.

This argument depends on the assumption that the hearer knows that the speaker cannot directly "look in the hearer's mind". However, a change in the hearer's mind is necessary for a common ground update. Therefore, the hearer has to give evidence

for a change in his mind, i.e. has to perform an acceptance move. A similar argument will again be used in section 5.3 where the default interpretation of the relation of two adjacent speech acts will be discussed, i.e. the default relations in the *absence* of an accept move.

The rest of this section is used to formalize the argument above. In order to do this, assume that A performed some speech act E towards hearer B with content K. Moreover, assume that B was already able to decode proposition K from U, i.e. assume that B is in state 3. Since B is aware of the fact that A performed a speech act, B can conclude that A actually pursues the goal to update the common ground with the content of the speech act. This situation can be described as follows:

(4.41)

$A\ B\ E\ T\ S_A\ S_B\ S'_B$
$\textsc{agent}(A),\ \textsc{agent}(B)$
$E \subset T$
$E \subset S_A$
$E : \textsc{Speech Act}(A, B, U, K)$
$S_A : \textsc{Pursue}(A, K \in \text{CG}_{A,B})$
$E \subset S_B,\ S'_B \subset S_B$
$S_B : \textsc{bel}(A, E : \textsc{Speech Act}(A, B, U, K)) \Rightarrow S'_B : \textsc{bel}\big(A, S_A : \textsc{Pursue}(B, K \in \text{CG}_{A,B})\big)$

Now assume, that B actually accepted A's speech, i.e. that the condition $\psi_4(E)$ of (4.36) holds. This means that A and B want to update CG with K. However, B did not provide any evidence for the fact that he wants to update CG yet. Therefore, on the basis of inertia, B may assume that A is still ignorant about the fact that B wants to update the common ground with K. This situation can be presented as in (4.42):

(4.42)

$S_A\ S'_B\ S''_B\ T\ T'\ E$
$E \subset T$
$T \prec T'$
$E : \textsc{Speech Act}(A, B, U, K)$
$S_A : \textsc{Pursue}(A, K \in \text{CG}_{A,B})$
$S'_B : \textsc{bel}\big(B, S_A : \textsc{Pursue}(A, K \in \text{CG}_{A,B})\big)$
$S''_B : \textsc{Pursue}(B, K \in \text{CG}_{A,B})$
$S'''_B : \textsc{bel}\left(B,\ \begin{array}{\|l\|}\hline S_{A1} \\ \hline T' \subset S_{A1} \\ \neg S_{A1} : \textsc{bel}\left(A,\ \begin{array}{\|l\|}\hline S' \\ \hline T' \subset S' \\ S' : \textsc{Pursue}(B, K \in \text{CG}_{A,B}) \\ \hline \end{array}\right) \\ \hline \end{array}\right)$
$T' \subset S_A \cap S'_B \cap S''_B$

The representation in (4.42) shows a mismatch between assumptions about the mental states, (i.e. in S'''_B B believes that A is ignorant about B's intention to make K an element of CG) and in the actual mental states of the interlocutors (since B is actually willing to make K an element of CG). Such a situation is an obstruction for a common

ground update: in order to perform a common ground update, both interlocutors have to agree on this update. In B's assumptions on A's mental states this is not the case and this mismatch has to be repaired. Therefore, B has to signal that he actually wants to perform a common ground update. To do that, B can now signal acceptance in order to make clear that B is willing to perform a common ground update with the content of the former speech act. This means that an acceptance move depends on the former speech act and its content. This assumption can be implemented in a definition, as shown in (4.43):

(4.43) ACCEPTANCE MOVES: An acceptance move of a speech act E is a speech act E' with which the hearer signals his intention to make the content of E an element of the common ground:

$A\ B\ E\ E'\ U\ K$	
$\text{AGENT}(A), \text{AGENT}(B)$ $E \prec E'$ $E : \text{SPEECH ACT}(A, B, U, K)$	
$E' : \text{ACCEPTMOVE}(B, E) :\Leftrightarrow$	U' $E' : \text{SPEECH ACT}(B, A, U', K)$

This definition can now be applied to the situation above in order to see how an acceptance move actually results in a common ground update. In order to do that, take again the situation described in (4.42) and assume that B performed an acceptance move and that A is actually aware of the fact that B performed an acceptance move. This situation can be described as follows:

(4.44)

$A\ B\ E\ E'\ S_A\ S'_A S_B$
$\text{AGENT}(A), \text{AGENT}(B)$ $E \prec E',\ E \subset S_A, E' \subset S_B, E' \subset S'_A$ $E : \text{SPEECH ACT}(A, B, U, K)$ $E' : \text{ACCEPTMOVE}(B, E)$ $S_A : \text{PURSUE}(A, K \in \text{CG})$ $S_B : \text{PURSUE}(B, K \in \text{CG})$ $S'_A : \text{BEL}(A, S_B : \text{PURSUE}(B, K \in \text{CG}))S'_B : \text{BEL}(A, S_A : \text{PURSUE}(A, K \in \text{CG}))$

Given such a situation, the question is of course, whether K is now actually an element of the common ground!? In a situation like (4.44), one may actually assume that K has good chances to become such an element. However, this situation does not imply that B is aware of the fact that A noticed B's acceptance move. In this case, a similar argument as above would have to be applied to the given situation, which would mean that A actually would have to signal that he noticed B's acceptance move. In other words, acceptance moves would have to be accepted. If this is true, one has to answer the question, *how* acceptance moves can be accepted. At least, acceptance moves do not have to be accepted by another acceptance move, since this would end up in an

infinite loop of acceptance moves, where each move accepts its predecessor: B would have to accept A's speech act by uttering m which would have be accepted by A with the utterance of something like m again, and so on. This problem was already seen by Clark and Schaefer, (1989, p. 268). As a solution for this problem, they proposed the following principle:

(4.45) STRENGTH OF EVIDENCE: The participants expect that, if evidence e_0 is needed for accepting presentation u_0, and e_1 for accepting the presentation of e_0, then e_1 will be weaker than e_0.

This principle implies that an acceptance move does not have to be accepted by another acceptance move but that continued attention or the initiation of the next turn suffices as further acceptance. This observation is going to be included into the above version of acceptance moves by adding the following condition: if an acceptance move is not rejected or disagreed upon, then it is accepted[5]. A lack of disagreement can for example be expressed simply by *keep talking*, as it was done after the first two acceptance moves in example (4.38). The continuation of the discourse suffices for the acceptance of an acceptance move. The exact time point when the content of the accepted speech act actually becomes an element of the common ground cannot be made explicit with the tools developed so far. A possible solution to this problem is discussed in section 4.2.4.

So far, the necessity of acceptance moves was discussed. Giving evidence for acceptance is necessary since all interlocutors are subject to inertia which makes all interlocutors thinking of their respective counterparts not changing their mental states the default case. Therefore, a change in the mental state has to be signaled. Acceptance moves have been defined as speech acts with the function to signal that a mental state has changed. In the terminology of this book, an acceptance move signals that the hearer of a speech act with some content K adopted the goal to make K an element of the common ground.

4.2.3 Other Forms of Acceptance

4.2.2 might have suggested that acceptance moves are the only way for agents to signal acceptance. Of course, acceptance moves are a frequently used tool in dialogues. However, they are not the only possibility to signal acceptance. For example, if the acceptance of a speech act implies the immediate performance of a particular action of one of the interlocutors, acceptance can also simply be expressed by performing the required action. This is for example the case in question-answer pairs.

(4.46) a. A: Did Bo leave?
b. B: Yes. He left at 7:00 am already.

In such cases, acceptance is simply signaled by giving the answer to the question. However, even in cases where the speech act aims at making the hearer do something,

[5] This assumption is actually stronger as Clark and Schaefer's strength of evidence principle. It is formulated this strong since the framework of this book is to coarse to include a principle as (4.45).

acceptance moves *can*, and often *do* follow up that speech act. In such cases, they even may accompany the performance of the required action. For example, situations like the following occur quite often in natural communication:

(4.47) a. *A*: Give me the salt, please.
b. *B*: OK. [*uttered as B passes the salt*]

In this situation, two acceptance moves are performed at once: there is a linguistic acceptance move which consists of the answer *OK* and a non-linguistic acceptance of *A*'s speech act which consists of the performance of the required action. Despite the redundancy of these 'double acceptance' moves, such conversational situations may occur. The following example shows a similar phenomenon. This example is of particular interest, since the accepted speech act (4.48.a) is an example of an *indirect speech act*:

(4.48) Asher and Lascarides, 2003, p. 308
a. *A*: Can you please pass the salt?
b. *B*: Sure. [*uttered as B passes the salt*]

Asher and Lascarides, (2003) take such phenomena as evidence for the claim that indirect speech acts have multiple illocutionary forces. (4.48.a) does not serve only as a question but also as a request, and *B*'s simultaneous performance of two actions serves to satisfy both intended effects of the speech act[6]. Such "double acceptance moves" can be explained with similar arguments to the general argument for acceptance moves: in a small time gap between the request and the execution of the required action, the speaker might not be sure about the hearer's reaction. In order to make the speaker aware of the fact that the hearer actually did understand him and is willing to perform the requested action, the speaker performs an acceptance move.

4.2.4 Common Ground Update

The box in (4.44) already showed situation in which both interlocutors believe of the respective other interlocutor that s/he wants to make the same proposition *K* an element of their common ground. In this case, *K* should have good chances to actually become such an element. However, at this point it is not clear yet whether a common ground update actually takes place. This section finally proposes one possibility to model a common ground update. The idea which is pursued is the following: assume that a speaker *A* performs a speech act $E : \textsc{Speech Act}(A, B, U, K)$ towards a hearer *B* and assume that *B* is in state 3 with respect to this speech act (i.e. *B* heard *U* correctly and decoded *K* from *U*). It is the purpose of a speech act to propose a common ground update with the content of the speech act. But for a common ground update with a proposition *K*, it is a necessary condition that the speaker who performed the

[6]For Asher and Lascarides, the illocutionary force of a speech act is a semantic entity. In order to give such mixed speech acts a semantic interpretation, they introduce •-type expression to model such copredication. In this formalism, (4.48.a), has the type *question* • *request*. For a very interesting semantic interpretation of such dot-types in terms of Cartesian closed categories, see Asher, (2014).

speech act with content K is committed to that content. Therefore, proposing a common ground update with some proposition K implies that the speaker commits to that proposition with respect to B. This effect of a speech act can be formalized as follows:

(4.49) SPEECH ACTS AND COMMITMENTS: If A utters a speech act E towards B and B is in state 3. of the list given in (4.33), then A is committed towards B to the content of E:

$A\ B\ E$	
$\text{AGENT}(A), \text{AGENT}(B)$ $E : \text{SPEECH ACT}(A, B, U, K)$ $S : \text{BEL}\,(B, E : \text{SPEECH ACT}(A, B, U, K))$	$\Rightarrow PC_{A,B}K$

As an example, take again the situation in which a speaker A claims towards another agent B that he likes cheese. Regardless of the question of whether B accepted this claim, A has to act as if he does not like cheese in order for that claim to even have a chance to become an element of the common ground.

Now assume that B performed an acceptance move after A's speech act and that A actually noticed B's acceptance move. Given the assumption about the effect of a speech act in (4.49) and given the semantics of acceptance moves in (4.43), the situation after a speech act was accepted by an acceptance move can be described as follows:

(4.50)

$A\ B\ E\ E'\ U\ K$
$E \prec E'$ $E : \text{SPEECH ACT}(A, B, U, K) \Rightarrow PC_{A,B}K$ $E' : \text{ACCEPTMOVE}(B, E) \Rightarrow PC_{A,B}K$

Consider again the example above. B has accepted A's claim that A does not like cheese. This means that B wants to make that proposition an element of the common ground. But this means that B also has to act as if A does not like cheese. In other words, after the performance of an acceptance move, B is committed to the same proposition as A is.

The presentation in (4.50) shows a situation in which A is publicly committed to K with respect to B and where B is publicly committed to K with respect to A. However, according to the definition of transmission of commitments (4.22), this exactly means that K becomes an element of the common ground.

So far it was claimed that in order to perform a successful common ground update, the speaker of a speech act E with content K has to act as if he believes that content. This makes him publicly committed to K with respect of his audience B, regardless of the fact whether B has heard or even accepted B. If B wants to make K an element of the common ground as well, B can perform an acceptance move, which makes B also committed to K. This means that both agents are committed to K with respect to the respective other agent. But because of transmission of commitments, K becomes an element of the common ground. In other words, they have succesfully performed a common ground update with K.

This argument was based on assumption (4.49) which says that a speaker is committed to the content of a speech act if the hearer heard the speech act correctly and was able to decode its content correctly. This assumption is quite strong. It is again possible to argue that the speaker must have some evidence for the fact that the hearer understood him in order to be actually committed to the content of his speech act. And vice versa, the hearer has to be somehow aware of the fact that his acceptance move was noticed and so on. However, this problem will not be addressed further in this book.

Summary of this section: This section was concerned with the question of how a commitment uptake is actually possible and, hence, it was concerned with grounding of a speech act. Some necessary steps for an utterance to become an element of the common ground were discussed and formalized. Moreover, this section targeted the problem of what it actually means to accept a speech act and how acceptance can be expressed. The common-sense principle of inertia was used in order to show why overtly expressed acceptance in discourse in the form of acceptance moves seems to be necessary for making propositions an element of the common ground. After this discussion, a semantic interpretation of acceptance moves in the framework of goals of speech acts was given. Acceptance moves have been defined as speech acts with the same goal as the speech act they are accepting. Finally, a new but yet problematic way to model common ground updates on the basis of acceptance moves was proposed.

4.3 Presuppositions

This section aims at incorporating presuppositions into the theory of speech acts that was developed in this chapter so far. The debate on the role, function, and the meaning of presuppositions is as old as modern semantics and logics itself. Historical milestones in this debate were Frege, (1892) and his influential essay *Sinn und Bedeutung* and Russell and his essay *On Denoting*. Although these authors have different opinions on the question of what a presupposition is supposed to be, both treat presuppositions from the point of view of their logical implications. For an overview on the historical development of the debate on presuppositions, see Simons, (2013). The view on presuppositions that will be adopted in this book is the assumption that presuppositions are *speaker presuppositions* in the sense of Stalnaker, (1973). According to this view, presuppositions express beliefs on the content of the common ground: a speaker who presupposes content makes it public that he believes that the presupposed content is an element of the common ground. Take (4.51) as a classical example for presupposed content. In (4.51) the speaker presupposes the existence of the king of France and states that this king is bald. According to Stalnaker, the speaker communicates with this presupposition that he believes that there is a unique king of France is already an element of the common ground. This assumption can be formalized as in (4.51.c).

(4.51) a. A: The King of France is bald.
b. (4.51.a) $>>$ $\exists!x$ King.of.France(x)

c. (4.51.a) $\Rightarrow$ BEL$(A, [\exists!x\ \mathsf{King.of.France}(x)] \in \mathrm{CG})$

There are some linguistic constructions which are commonly believed to be designed as triggers for presuppositions. The following (not exhaustive) list (here cited from Beaver, 1997) contains some of these presupposition triggers:

1. Definite *NP*s
2. Quantificational *NP*s
3. Factive verbs
4. Iterative adverbs
5. Cleft sentences
6. Etc.

The example (4.51.a) shows a case where the existence of a unique King of France is triggered by a definite *NP*, i.e. *the king*. The sentence in (4.52.a) shows an example that contains the factive verb *to know*, which presupposes the truth of its complement and, hence, the truth of the proposition given in (4.52.b).

(4.52) a. George knows that the King of France is bald.
b. (4.52.a) $>>$ $\exists!x\ (\mathsf{King.of.France}(x)\ \wedge\ \mathsf{bald}(x))$

This section wants to elaborate on the assumption that presupposing, which is believed to be realized by one of the devices in the list above, can be regarded as a particular type of speech act. Presupposing content means that the speaker performs an action with which he makes his beliefs on the content of the common ground public. As already pointed out, this view on presuppositions stands in the tradition of Stalnaker's approach, who claimed that presuppositions are beliefs about the content of the common ground. In addition to the fact that presuppositions are important and interesting phenomena on their own, they will again play an important role in chapter 5 of this book, where they are used in order to explain some phenomena concerning rhetorical relations.

The rest of this section is organized as follows: 4.3.1 presents some classical ways for treating presuppositions in update semantics. 4.3.2 incorporates presuppositions in the theory of speech acts developed so far by formalizing the assumption that presupposition triggers are communicative events with which the speaker makes some of his beliefs on the content of the common ground public. Moreover, these formalizations will be applied to actual communicative situations.

4.3.1 Presuppositions in Update Semantics

In the following, a traditional way to deal with presuppositions in update semantics will be presented. The focus will lie on the presentation of the idea of Heim, (1992) who treats presuppositions as restrictions of the context in which a sentence with presuppositions can be uttered. Moreover, it will be shown how this assumption was formalized in DRT.

Heim, (1992) established the notion of a *context change potential* (CCP) of a sentence. The CCP of a sentence is its ability to update a context with the information encoded by the sentence. Heim formalizes the context as a set of possible worlds established by the former discourse. Given the notion of context, one can formalize the CCP of a sentence as a partial function from sets of possible worlds to sets of possible worlds: a CCP of an utterance takes a context and returns the context with the additional information of the given sentence. For a context c and a CCP φ one usually writes $c + \varphi$ for the value of c after one has applied φ. The CCP of a sentence that doesn't contain any presuppositions is defined for any context and its effect is to add the information of the sentence to the given context. This assumption reflects the intuition that presupposing content is only felicitous if the content of the presupposition is already was already mentioned in the discourse (i.e. it is an element of CG in the terminology of this book). A sentence which does not contain any presupposition triggers is, therefore, felicitous in any context. As an illustration, consider an example and take a sentence like *It is raining*. This sentence excludes all worlds from the context in which it isn't raining. This effect can be formalized as in (4.53.b):

(4.53) a. $\varphi =$ [It is raining.]
b. $c + \varphi = \{w \in c \mid \text{it is raining in } w\}$

A sentence which contains presuppositions, on the other hand, is *not* defined on all possible contexts anymore. For example, a sentence like *John's cat is hungry* presupposes that John has a unique cat and is only interpretable in worlds where this presupposition is satisfied. In update semantics, this is intuition reflected by the requirement that the domain of the CCP of sentences containing presupposition is restricted to sets of worlds where these presuppositions are true. This assumption is exemplified in (4.54.b). This condition states that the CCP of the sentence *John's cat is hungry* is only defined on contexts where all worlds are worlds in which John has a unique cat:

(4.54) a. $\varphi =$ [John's cat is hungry.]
b. $\text{Dom}(\varphi) = \{c \mid w \in c \Rightarrow \text{John has a unique cat in } w\}$
c. $c + \varphi = \{w \in c \mid \text{John's cat is hungry in } w\}$

In update semantics, presuppositions basically restrict the domain of the CCP of sentences, i.e. the CCP of a sentence with presuppositions is only defined on contexts which satisfy these presuppositions. On the level of the natural language, this means that presuppositions can only felicitously be used in cases where they are already common ground.

In DRT, this assumption is formally realized by means of a two-stage construction for an update of sentences with presuppositions. How this works was already sketched in section 2.1.1. The DRT-algorithm for applying a sentence which contains presuppositions to its context is summarized in the following two steps in an informal way:

1. Build a preliminary sentence representation using a *pair* of DRSs such that one of the DRSs contains the presupposed content of the given sentence and the other one the non-presupposed part (*asserted part*) of the sentence.

2. Verify the presuppositional part in the given context, i.e. try to resolve it, and apply the non-presuppositional part to the new context.

The requirement that the CCP of a sentence including presuppositions can only be applied to contexts satisfying these presuppositions is realized in DRT by the condition that the presupposed content has to be resolved to the context.

Below, it is exemplified how presuppositions are treated in DRT on the basis of the following example:

(4.55) a. A boy owns a donkey.
b. He beats it.

Since (4.55.a) doesn't contain presupposed content, it can be applied to the trivial context that contains no information (i.e. all possible worlds). The DRS in (4.56) shows a possible presentation for (4.55.a).

(4.56)

$b\ d$
$\mathsf{boy}(b)$
$\mathsf{donkey}(d)$
$\mathsf{own}(P, d)$

The sentence in (4.55.b) contains two presupposition which are triggered by the two pronouns *he* and *it*. These pronouns presuppose the existence of a uniquely identifiable male referent and the existence of a uniquely identifiable non-human referent. The asserted part of (4.55.b) consists of the information that the male referent beats the non-human referent. According to the algorithm given above, in order to update the context given in (4.56) with the CCP of (4.55.b), a pair of DRSs has to be constructed. In this pair, one DRS contains the presupposed content and the other one contains the asserted content. The pair of DRSs in (4.57) shows a possible representation of (4.55.b). The first DRS in this pair represents the presupposed part and the DRS in the second entry represents the asserted part of (4.55.b).

(4.57) $\left\langle \begin{array}{|l} \hline x\ y \\ \hline \mathsf{male}(x) \\ \mathsf{non\text{-}human}(y) \\ \hline \end{array} , \begin{array}{|l} \hline x\ y \\ \hline \mathsf{beat}(x, y) \\ \hline \end{array} \right\rangle$

Before the context can be updated with (4.57), the presuppositional part has to be verified, i.e. the content of the presuppositional part has to be resolved in the context. In the case of (4.57), pronoun resolution is fairly easy since possible antecedents for the two pronouns *he* and *it* are fully specified by their grammatical gender. The presupposed content of (4.55.b) can therefore easily be resolved to the context given in (4.56). Afterwards, the context updated with the asserted part of (4.57) can be performed:

(4.58)

$b\ d$
$\mathsf{boy}(b)$
$\mathsf{donkey}(d)$
$\mathsf{own}(b,d)$
$\mathsf{male}(b)$
$\mathsf{non\text{-}human}(d)$

$\uplus$

$x\ y$
$\mathsf{beat}(x,y)$
$x = b$
$y = d$

$=$

$P\ d\ x\ y$
$\mathsf{boy}, b)$
$\mathsf{donkey}(d)$
$\mathsf{own}(b,d)$
$\mathsf{male}(b)$
$\mathsf{non\text{-}human}(d)$
$\mathsf{beat}(x,y)$
$x = b$
$y = d$

The box on the right side of the equation in (4.58) is the result of applying the asserted content of (4.55.b) to the context with the resolved presuppositions.

DRT captures the idea that presuppositions are only interpretable under certain conditions by the requirement that the presupposed content has to be resolved to material in the context. In update semantics, this is modeled by the assumption that the CCP of sentences with presuppositions is only defined on contexts which satisfy the presuppositions (Heim, 1992).

So far it was presented how presuppositions are treated by DRT and how DRT formalizes the idea that presuppositions restrict the domain for the context change potential of sentences. 4.3.2 will take this idea as a starting point to integrate presuppositions in the theory of speech acts as goal-directed acts. The assumption that presupposing is a speech act with which the speaker makes own assumptions about the content of the common ground public will be discussed and formalized.

4.3.2 Presuppositions and Communicative Goals

In the following, presuppositions will be included in the theory of goals of speech acts. It will be proposed that presupposing is a communicative action of the speaker with which she makes some of her beliefs on the content of the common ground public. This assumption will again be used in section 5.3 where the impact of presuppositions on discourse structure will be discussed.

In this book, it was assumed that speakers pursue goals by performing speech acts. When encountering presuppositions, this assumption triggers the following question: what does a speaker want to communicate by presupposing content and how does that fit into a framework of goals of speech acts. §4.3.1 showed a classical semantic point of view according to which one can only felicitously presuppose content if this content is already contained in the context in which the current utterance was formulated. This is formally captured by the assumption that presuppositions restrict the domain for the CCP of their hosting sentence. The appropriate counterpart of the context of a sentence in the terminology of this work is the *common ground*. A proper translation of this assumption in the terminology of this book, therefore, is the statement that speakers can presuppose content if it is already an element of the common ground. Although this requirement on presupposed content actually captures the condition for a *felicitous* use of presuppositions, it seems that it is too strong, at least for the goals of this work. In communication, it may happen that speakers presuppose content because they just

believe that this content is already an element of the common ground. In this regard, it seems to be more natural to say that, by presupposing content, the speaker communicates that he takes it for granted that the presupposed content is an element of the common ground. This assumption makes presuppositions *speaker presuppositions* in the sense of Stalnaker, (1973): a speaker who presupposes content communicates that he *believes* that this content is already an element of the common ground. This assumption is captured in the following characterization of presuppositions.

(4.59) PRESUPPOSING: Let A and B be agents. If A presupposes a proposition K to hearer B, then A believes that K is an element of the common ground of A and B.

$$\boxed{\begin{array}{l|l} A\ B\ E & \\ \hline \textsc{agent}(A) & \\ \textsc{agent}(B) & \\ & \\ E : \textsc{Presuppose}(A, B, K) \Rightarrow & \boxed{\begin{array}{l|l} S & \\ \hline E \subset S & \\ S : \textsc{bel}(A, K \in \textsc{cg}_{A,B}) & \end{array}} \end{array}}$$

In (4.59), the action of presupposing is regarded as a communicative event. However, acts of presupposing are embedded in speech acts. Therefore, they have to be regarded as "sub events" of their hosting speech act. The rest of this section will be used to go through an example that contains presupposed content. It will be shown how presuppositions can be modeled on the basis of the characterization given in (4.59). Consider the sentence given in (4.60)

(4.60) A to B: The boy has a donkey.

The definite *NP* "the boy" in example (4.60) presupposes the existence of a uniquely identifiable boy, at least in the context of the communicative situation. According to the characterization of presuppositions in (4.59), this means that speaker A believes that the fact that there is a uniquely identifiable boy is already an element of the common ground, regardless whether it is an element of the common ground or not. The situation in which this speech act was performed can be represented as follows:

(4.61)
$$\boxed{\begin{array}{l|l} A\ B\ E_1\ E_{11}\ U & \\ \hline \textsc{agent}(A), \textsc{agent}(B) & \\ E_{11} \subset E_1 & \\ E_{11} : \textsc{Presuppose}\left(A, B, \boxed{\begin{array}{l|l} b & \\ \hline \mathsf{boy}(b) & \end{array}}\right) & \\ E_1 : \textsc{Speech Act}\left(A, B, U, \boxed{\begin{array}{l|l} d & \\ \hline \mathsf{donkey}(d) & \\ \mathsf{own}(b, d) & \end{array}}\right) & \end{array}}$$

Assume now, that the speech act E is in state 3 of the grounding conditions described in the list in (4.33). Recall that this means that hearer B noticed that A talked to him and that he was able to decode the content from the perceived utterance. This means in particular that B is able to tell that A wants to perform a common ground update and that B is aware of the fact that A believes that the presupposed content (i.e. *the boy*) is already an element of the common ground. Hence, given the assumption that E is already in state 3 of the list given in (4.33), B's mental states after the speech act in (4.60) can be presented as follows:

(4.62)

$$\boxed{\begin{array}{l} A\ B\ E\ ST\ T' \\ \hline \begin{array}{l} \textsc{agent}(A), \textsc{agent}(B) \\ E_{11} \subset E_1,\ E_1 \subset T,\ T \prec T',\ T' \subset S \\ E_{11} : \textsc{Presuppose}\left(A, B, \boxed{\begin{array}{l} b \\ \hline \mathsf{boy}(b) \end{array}}\right) \\ E_1 : \textsc{Speech Act}\left(A, B, U, \boxed{\begin{array}{l} d \\ \hline \mathsf{donkey}(d) \\ \mathsf{own}(b,d) \end{array}}\right) \\ S : \textsc{bel}\left(B, \boxed{\begin{array}{l} S_1^A\ S_2^A \\ \hline E_1 \subset S_1^A,\ E_1 \subset S_2^A \\ S_1^A : \textsc{bel}\left(A, \boxed{\begin{array}{l} b \\ \hline \mathsf{boy}(b) \end{array}} \in \mathrm{CG}_{A,B}\right) \\ S_2^A : \textsc{Pursue}\left(A, \boxed{\begin{array}{l} d \\ \hline \mathsf{donkey}(d) \\ \mathsf{own}(b,d) \end{array}} \in \mathrm{CG}_{A,B}\right) \end{array}}\right) \end{array} \end{array}}$$

The box in (4.62) describes a situation where B knows that A believes that *the boy* is already an element of the common ground. Moreover, it describes that A pursues the goal to make the information [*the boy owns a donkey*] an element of the common ground. The fact that A believes that *the boy* is already an element of the common ground has an interesting implication: Given that A believes that *the boy* is common ground, B can conclude that A does not pursue the goal to make *the boy* an element of the common ground. In section 5.3, this observation will be used to explain how hearers can infer coordinating rhetorical relations. The argument which will be developed in more details in section 5.3 roughly works as follows: if a speaker uses a discourse connective that presupposes the content of a former speech act, such as for example *then* or *also*, the hearer can conclude that the speaker believes that his former goal is achieved and, therefore, the speaker does not have to talk about the former events anymore. Hence, the speaker can precede in the discourse without supporting the former speech act, i.e. without having the need to ensure that the former speech act was successful. But this means that the speaker can continue in the discourse without having the need to add further information to old speech act. In this case, the speaker can continue with the narration or can set contrasts and so on. This argument will be laid

out in detail in section 5.3.

At this point, it has to be remarked that treating speech acts as devices for negotiating commitments has the further advantage that presuppositions occurring in non-assertive utterances can be treated exactly the same way as presuppositions occurring in non-assertive speech acts. Take for example (4.63).

(4.63) A to B: Kill your donkey.

The communicative situation of this speech act can be represented in exactly the same way as the situation of (4.60):

(4.64)
$$\begin{array}{|l|} \hline \qquad A\ B\ E_1\ E_{11}\ U \\ \hline \mathrm{AGENT}(A), \mathrm{AGENT}(B) \\ E_{11} \subset E_1 \\ E_{11} : \mathrm{PRESUPPOSE}\left(A, B, \begin{array}{|l|}\hline d \\ \hline \text{donkey}(d) \\ \text{own}(B,d) \\ \hline\end{array}\right) \\ E_1 : \mathrm{SPEECH\ ACT}\left(A, B, U, \begin{array}{|l|}\hline E' \\ \hline E \prec E' \\ \text{kill}(B,d) \\ \hline\end{array}\right) \\ \hline \end{array}$$

Truth conditional semantics, such as the context change semantics of Heim or DRT, are only able to treat presuppositions in sentences that allow a truth conditional interpretation (such as for example an interpretation as sets of possible worlds). However, content can be presupposed in all kinds of speech acts, as for example in orders such as (4.63). The given notion of speaker presuppositions, as proposed in (4.59), just contains assumptions on the common ground and, hence, is able to treat presuppositions in all kind of speech acts. This observation will also be of interest when the given theory of presuppositions is used in order to analyse rhetorical relations.

This section was concerned with incorporating presuppositions into the theory of speech acts that was developed in this chapter. We discussed the fact that presuppositions can be regarded as domain restrictors for the context change potential of their hosting sentences and showed how this assumption is formalized in DRT. Afterwards presuppositions were characterized as a special type of speech act event with which the speaker makes it public that he believes that the content of the presupposition is already an element of the common ground. Finally, we discussed an example which showed how presupposed content can actually be represented in the formal framework of this book.

4.4 Summary

It was the main purpose of this chapter to include speech acts in the theory of goal-directed acts. It was proposed to regard speech acts as goal-directed utterances and

that it is the goal of a speech act to perform a common ground update. The common ground of interlocutors was defined as a set of mutual commitments. The assumption, that it is the goal of a speech act to update the common ground with further commitments makes it possible to state that it is the goal of *any* speech act to update the common ground with the content of the speech act, independently of the illocutionary force of the speech act. However, for this advantage one has to pay the prize that it becomes very difficult to derive the content of a speech act from its concrete realization, at least in a unified way. Furthermore, the claim that any speech act negotiates commitments was applied to the rather special speech acts of *greetings*. It was shown that the performance of a greeting signals the speaker's willingness for communication.

Since the common ground and hence a common ground update depends on both agents, i.e. speaker and hearer, the role of the hearer for a successful common ground update was examined. Therefore, acceptance moves have been investigated and it was shown why such acceptance moves occur quite frequently. Moreover, it was shown how the performance of such an acceptance move can actually end up in a common ground update. Finally, this chapter examined presuppositions and their function in communication. Presuppositions have been regarded as distinct communicative acts within a hosting speech act with which the speaker communicates that he believes that the presupposed content is already an element of the common ground.

Chapter 5

From Speech Acts to Rhetorical Relations

Chapter 4 investigated isolated speech acts and their goals. In contrast to the last chapter, this chapter is concerned with larger discourse units and, in particular, with the question of how the meaning of speech acts varies in the context of other speech acts. Take for example (5.1):

(5.1) a. Max fell. John pushed him.
b. Max insulted John. John pushed him.

The respective second sentences in (5.1) are exactly identical, but they play completely different roles with respect to their context sentence. In (5.1.a), the second sentence *explains* the content of the first one, whereas in (5.1.b), the second sentence describes a *result* of the event which is described in the first one. This means that the contribution of a sentence to the discourse varies depending on the the context in which it is uttered. The interpretation of a sentence differs depending on how a sentence relates to its context. The relation of a sentence to its context sentence is referred to as a *rhetorical relation*, *discourse relation* or *coherence relation*. The term *rhetorical relation* describes how sentences are interpreted with respect to their preceding sentences and refers to the type of additional information of a sentence if one relates it to its context. For example, the relation in (5.1.a) is referred to as an *Explanation*, since the second sentence explained the first one. The relation in (5.1.b), on the other hand, is a *Result*, since the event of first sentence results in the event of the second sentence. The term rhetorical relations, therefore, makes reference to additional semantic interpretation of sentences.

In this book, rhetorical relations are approached from the perspective of goals, acts and plans. In particular, it is assumed that the recognition of a rhetorical relation is is the recognition of a certain communicative plan. This shifts the subject of the investigation from the semantic contribution of rhetorical relations (as it is pursued for example by SDRT) to the communicative *function* of a speech act in the speaker's communicative plan. Therefore, this chapter is interested in speech acts as parts of communicative plans and in the integration of rhetorical relations in the theory of goals, plans and actions.

This chapter will be structured as follows. Section 5.1 gives an overview of rhetorical relations and how one can describe the structure of a whole discourse on the basis of rhetorical relations. Moreover, this section introduces the idea of the *right frontier constraint* (RFC). The RFC is a claim about the effect of rhetorical relations on further discourse structure in terms of the ability of new discourse material to relate to older discourse material. This effect can, for example, be measured in terms of pronoun resolution. Moreover, the notions of coordinating and subordinating rhetorical relations are discussed. Section 5.2 makes first attempts to integrate rhetorical relations in a theory of goals and actions. This section defines the notions of *communicative plans*, *support relations*, and a notion of *communicative coordination*. Section 5.3 discusses default relations in communication. The main question of this chapter is the following: given an unmarked pair of utterances, what kind of relation between these utterances can be assumed? This question will be answered with the claim that, by default, interlocutors assume that two unmarked utterances are related by a support relation. Finally, section 5.4 is concerned with reasons to produce such support relations. It elaborates on the idea that support relations are very often realizations of self repair moves, i.e. of discourse moves with the purpose to remove some obstructions for the grounding of some speech act.

5.1 Discourse Structure - Coordination and Subordination

A text or dialogue is more than the sum of its parts. Depending on the context in which a sentence is produced, a sentence receives additional meaning that goes beyond the meaning which is encoded in its pure linguistic material. As an illustration for this phenomenon, consider the following discourse example, taken from Asher and Lascarides, (2003):

(5.2) a. Max fell.
b. John pushed him.

If one interprets the whole discourse given in (5.2), it is reasonable to assume that Max fell *because* John pushed him, even though no explicit causal relation is marked in that discourse. In other words, the speech act (5.2.b) gains additional meaning in the context of the speech act (5.2.a). The discourse was enriched with the additional information of a causal relation between the two described events. The causal relation between the two described events implies also that the event described in (5.2.b) took place *before* the event described in (5.2.a). This, in particular means that the events described in the two statement above must have occurred in the opposite order from the occurrence of the speech acts. The content of the given discourse can be represented in DRT as in (5.3):

(5.3)

$M\ J\ T_1\ T_2\ E_1\ E_2$
$E_1 \subset T_1,\ E_2 \subset T_2$
$\mathsf{name}(M, \mathrm{Max})$
$\mathsf{name}(J, \mathrm{John})$
$E_1 : \mathsf{fell}(M)$
$E_2 : \mathsf{push}(J, M)$
$T_2 \prec T_1$
$E_2 \twoheadrightarrow E_1$

This DRS already contains the additional information about the causal relationship between the presented events. But as a matter of fact, DRT itself does not provide any tool with which such inferences can be modeled. This was one of the reasons for the development of SDRT as an extension of DRT. SDRT contains predicates that connect speech acts and enriches the discourse with implicit information, as for example the causal relation in (5.2).

Example (5.2) shows that utterances gain additional meaning when they are related to their context. If the context is given by another speech act, this relation is called a *rhetorical relation*. Rhetorical relations describe dependencies of sentences or speech acts in discourse. The whole discourse is structured and segmented by rhetorical relations in a hierarchical order. This structure is mostly referred to as *discourse structure*.

This section is concerned with introducing the basic notions of discourse structure. In 5.1.1, rhetorical relations and some examples are discussed in greater detail. 5.1.2 introduces the classical dichotomy of rhetorical relations into two major classes: *subordinating* and *coordinating* relations. In 5.1.3, discourse graphs are introduced. Discourse graphs are means to represent the hierarchical structure of a discourse in a pictorial way. On the basis of these graphs, section 5.1.4 introduces a constraint for the ability of discourse material to serve as an anchor point for the attachment of new discourse material. This constraint is called the *right frontier constraint* and states that new discourse material can only relate to discourse material on the right frontier of a discourse graph. This claim can also be formulated in terms of the resolution of third-person pronouns. This shows that discourse structure also has an impact on the resolution behavior of anaphoric devices.

5.1.1 Rhetorical Relations

In different contexts, speech acts fulfill different functions. Depending on their function in discourse, pairs of sentences are related by different rhetorical relations. These relations connect speech acts and turn a set of isolated speech acts into a coherent discourse. Different theories of discourse coherence assume different sets of relations. However, most approaches to discourse relations agree on the following list of rhetorical relations, which is taken from Jasinskaja and Karagjosova, (2015):

(5.4) a. For baking bread you need lots and lots of salt. *Elaboration*
b. Round about 20 grams for one pound of flour.

(5.5) a. Bill must have left. *Evidence*

b. His bike is gone.

(5.6) a. Max painted the wall. *Parallel*
b. Chris repaired the roof.

(5.7) a. Max fell. *Narration*
b. John helped him up.

(5.8) a. John likes classical music. *Contrast*
b. But he hates Mozart.

(5.9) a. John pushed Max. *Result*
b. He fell.

The names of these relations refer to their intuitive function in discourse and to the additional semantic information which is inferred by the interpreter of the relations. For example, the second utterance in (5.4) basically states that one needs 20 grams of salt for just one pound of flour which can be considered to be a great amount of salt. Therefore, this sentence *elaborates* on the statement that bread needs lots of salt. In (5.6), the second utterance is interpreted as providing evidence for the content of the first sentence. The second utterance in (5.8), on the other hand, contrasts with its preceding utterance, since people who like classical music usually are expected to like Mozart's music as well.

This (non-exhaustive) list is just a set of rhetorical relations. But the set of all rhetorical relations has more structure. Rhetorical relations are usually divided into two subsets: coordinating and subordinating relations. The motivation of this subdivision is the topic of 5.1.2.

5.1.2 Coordination and Subordination

The set of all rhetorical relations can be bisected into *coordinating* relations and *subordinating* relations. This distinction roughly corresponds to the distinction of *multinuclear* and *nucleus-satellite* relations in rhetorical structure theory (RST). In the list (5.4) – (5.9) above, the relations *Explanation*, *Elaboration* and *Evidence* are typically regarded as subordinating relations, whereas *Parallel*, *Narration*, *Contrast* and *Result* are normally classified as coordinations:

COORDINATING RELATIONS:	SUBORDINATING RELATIONS:
• *Parallel*	• *Explanation*
• *Narration*	• *Elaboration*
• *Contrast*	• *Evidence*
• *Result*	• ...
• ...	

The distinction between coordinating and subordinating relations originates from the intuition, that in subordinating relations the speaker 'sticks more to the topic' then in

coordinating relations (the word *topic* is used here in non-technical way). Take the following examples as an illustration.

(5.10) a. Max painted the wall.
b. Chris repaired the roof.

(5.11) a. Max painted the wall.
b. It took him a while.

Example (5.10) shows a *Parallel* relation. The two utterances present two distinct events which are (more or less) independent from each other. Of course, both utterances can be unified under the common topic 'several people are constructing a house', but in the frame of this topic, the two presented events are independent from each other.

Example (5.11), on the other hand, is an *Elaboration*. The speaker elaborates on Max' action of painting the wall. Additional information is added, but the speaker 'sticks' to the theme of (5.11.a). In the case of *Elaboration*s, this intuition was captured by Hobbs, (1979, p. 73), who stated that in an *Elaboration*

> at a sufficiently deep level the two sentences say the same thing. In the typical case, new information is conveyed by the second sentence, since there must be some reason for saying it again.

According to this quote, two sentences in an *Elaboration* basically convey the same information and, hence, fulfill the intuitive condition that the speaker 'sticks' to the same topic. Therefore, *Elaboration*s are a paradigmatic case for subordinating rhetorical relations. RST-terminology reflects the intuition that the content of a subordinated utterance in subordinating relations does not move to far away from the content of the utterances it is subordinated to. In RST the subordinated utterance is called the *satellite* and the subordinating utterance is the *nucleus* of the relation[1]. This metaphorical terminology reflects the intuition that the 'core' (i.e. the nucleus) of the relation is the superordinated utterance and that the subordinated utterance (i.e. the satellite) does only make sense in the company of its nucleus.

So far, the distinction between subordinating and coordinating rhetorical relations was motivated on an intuitive level. However, no actual *definition* for one of these classes was provided. This is due to the fact that the distinction between subordination and coordination is basically driven by intuition. This problem was already seen by Polanyi, (1988, p. 611), who wrote:

> Ultimately, the decision of whether to subordinate or coordinate a given unit must be made using real world knowledge and inferential procedures.

One of the approaches from the classic literature, which aims at capturing the nature of subordination relations and which almost can serve as a definition, has its origin again in RST. RST defines *presentational relations* as a subclass of rhetorical relations. These are characterized as follows:

[1] RST itself actually does not have the terminology of "subordinations" and "coordinations". However, almost all nucleus-satellite relations fit to the description of subordinating rhetorical relations.

(5.12) PRESENTATIONAL RELATIONS: Presentational relations are those whose intended effect is to increase some inclination in the reader, such as the desire to act or the degree of positive regard for, belief in, or acceptance of the nucleus. (Mann and Thompson, 1988)

Although this characterization of presentational relations does not cover all nucleus-satellite relations in RST and, given the assumption that nucleus-satellite relations roughly correspond to subordinating relations (not all subordinating relations either) it already defies a fairly distinct class of relations which are subordinating.

At the moment, however, the distinction between subordinating and coordinating rhetorical relations is based on the intuition that subordinating rhetorical relations 'stick more to the topic' than coordinating relations. According to the characterization of presentational relations above, the related utterances in a presentational relation is in some sense *about* the nucleus of the relation. This suggests that the set of presentational relations is a subset of the set of subordinating relations. In addition to fact that there is an actual *definition* for presentational relations, there is another reason that makes them attractive for this work as a subject of consideration: In the characterization of presentational relations it is made explicit that it is the purpose of the satellite to support the nucleus. This can perfectly be translated in the framework of acts, goals and speech acts, which was developed in the last chapters. When one abstracts over the different possible "inclinations of the reader", as Mann and Thompson phrased it, it is possible to say that in the case of presentational relations it is the purpose of the satellite to help to achieve the goal of the nucleus. This can again be translated in the terminology from section 3.1 by saying that the goal of the satellite is a *subgoal* of the goal of the nucleus. Jasinskaja and Karagjosova, (2015) proceeded in this spirit and actually proposed a definition for subordinating relations. These authors generalized over the different possible purposes of an utterance (such as making the hearer to act, to believe, ...) and used the abstract notion of *communicative goals*:

(5.13) SUBORDINATION I: In a sequence of discourse units (U_1, U_2), U_2 is subordinate to U_1 whenever the communicative goal of U_1 cannot be reached before the communicative goal of U_2 is reached.

As in the characterization of presentational relations, this characterization of subordinating relations captures the intuition that it is the function of the subordinated part in a subordinating rhetorical relation to support the superordinated part. Moreover, (5.13) has the further advantage of capturing the intuition that the subordinated part cannot move thematically too far away from its context sentence: If it is the purpose of one utterance to support its preceding utterance (whatever this means), then the second utterance must have 'something to do' with the first utterance, i.e. it should be about the same issue in some sense. Section 5.3 defines a class of rhetorical relations which will be referred to as *support* relations. Support relations can be shown to be subordinating and they function in the spirit of Jasinskaja and Karagjosova's approach to subordination in the sense that in these relations the second speech act supports its preceding speech act (in this case, by definition).

So far, this section was concerned with rhetorical relations and a bisection of rhetorical relations into subordinating and coordinating relations. Although this distinction is still motivated by general intuitions on the function of the individual utterances, it has a measurable linguistic effect in terms of the ability of the involved utterances to serve as anchor point for the attachment of new discourse material. This will be laid out in detail in 5.1.4. However, in order to make this statement precise, the notion of discourse graphs is needed. These will be introduced in 5.1.3.

5.1.3 Discourse Graphs

So far, discourse relations have been regarded as local phenomena, i.e. as relations which hold between individual sentences or utterances. But discourse relations also have an impact on the global structure of a discourse. In the following, it will be shown how the structure of a whole discourse can be presented on the basis of discourse relations.

At least since RST (Mann and Thompson, 1988) entered the market, it is assumed that discourses or texts are hierarchically structured in the sense that some parts of a text are "more important" and some are "less important". This assumption is reflected in many theories which are concerned with rhetorical relations and/or discourse coherence in the sense that they represent the hierarchical structure of a discourse by means of *discourse graphs*. Such discourse graphs reflect both - the local and the global structure of a text. The ideas for these discourse graphs are basically the same in all of these theories. Some nodes of the graph correspond to the sentences or speech acts of the discourse and these are connected by edges that correspond to the respective rhetorical relations that hold between the speech acts corresponding to these nodes. However, the exact formats of discourse graphs vary across different theories. In the following, different formats of discourse graphs on the basis of the following example from Asher and Lascarides, (2013) will be presented.

(5.14)
a. Max had a great evening last night. (π_1)
b. He had a great meal. (π_2)
c. He ate salmon. (π_3)
d. He devoured cheese. (π_4)
e. He then won a dancing competition. (π_5)

The linguistic discourse model (LDM, Polanyi, 1988) distinguishes basically just between subordinating and coordinating relations. The figure in 5.2 shows an LDM discourse tree for the discourse above. The leaves of that tree correspond to the utterances of the discourse. The inner nodes indicate the type of relations which hold between the leaves or subtrees which they are connecting. The vocabulary for rhetorical relations in SDRT is more fine grained. In contrast to most other theories, SDRT-discourse graphs are not necessarily trees. Similar to LDM, SDRT uses additional nodes in its discourse graphs that do not correspond to single sentences or speech acts but which represent larger discourse segments related to other discourse segments. For example, the node π' in 5.1 represents a discourse segment that consists of the utterances (5.14.c) and (5.14.d). These units are locally related by *Narration* but both constitute an *Elabora-*

tion for (5.14.b). Even though the SDRT graph looks a little bit more elaborated than its LDM counterpart in figure 5.2, from a discourse structural point of view, both graphs carry the same information. With the exception of the labels of the arrows (which are missing in the LDM graph), these graphs can be easily translated into each other. RST also has a pictorial way of representing discourse structure as tree-like objects. Figure 5.3 shows the RST-version of a graphical representation of a discourse. The horizontal lines correspond to the discourse segments in SDRT, i.e. the line indexed with the numbers 2 - 5 corresponds to the discourse segment π in figure 5.1, the line which is only numbered with 4 corresponds to the node p_4 in 5.1, and so on. The arrows indicate whether a relation is a nucleus-satellite relation or a multi-nucleus relation. For simplicity reasons, the *Narration* relation is also called *Narration*, although in RST this relation is called *sequence*. Figure 5.4 finally shows the most basic and simple possibility for a discourse tree which is in use. In this graph, any node corresponds to a speech act and the edges are again labeled by the relation which holds between the corresponding speech acts.

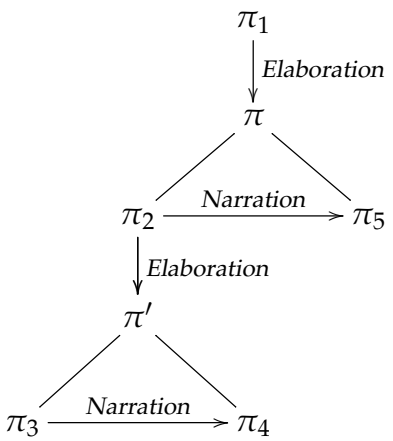


Figure 5.1: A SDRT discourse graph for (1.13). The horizontal arrows indicate a coordinating relation, whereas the vertical arrows indicate a subordinating relation. The edges with no arrow tips indicate that the connected nodes belong to the same discourse segment.

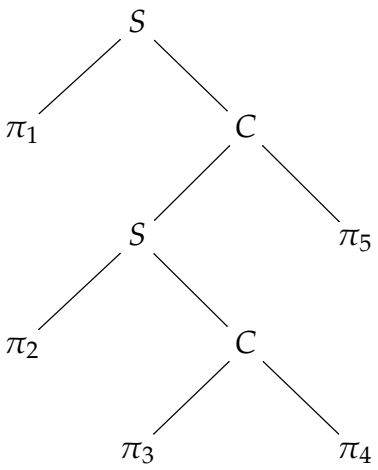


Figure 5.2: A discourse tree for (10) in LDM. The leaves of the tree correspond to utterances, the inner nodes specify the class of rhetorical relations that hold between their daughters.

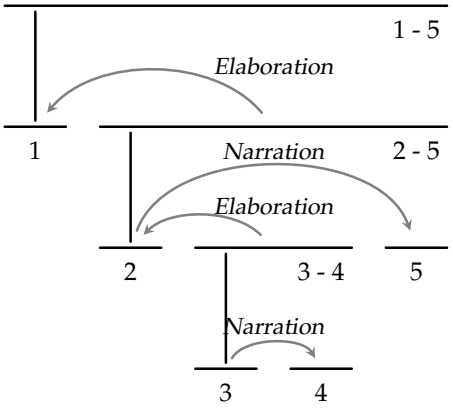


Figure 5.3: A RST graph for (1.13). The horizontal lines correspond to the segmentation of the discourse in SDRT. The numbers below the horizontal lines are the numbers of the sentences, which are represented by the lines. The direction of the arrows indicate the type of the relation.

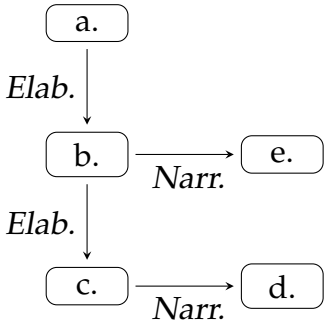


Figure 5.4: A simple discourse graph for (1.13). Subordination is indicated by vertical edges, coordination by horizontal edges.

So far, discourse graphs have been introduced for several theories as a device to represent the global structure of discourse. These discourse graphs will play a major role for understanding the impact of subordinating rhetorical relations on the discourse structure. On the basis of discourse graphs one can formulate the right frontier constraint, which makes predictions about the possibility of discourse material to serve as pivots (i.e. anchor points) for new discourse material. The exact formulation of the right frontier constraint will be the topic of 5.1.4.

5.1.4 The Right Frontier Constraint

It is one of the basic observation in the analysis of discourse structur that not all utterances in a discourse serve equally well as anchor points for the attachment of new discourse material. This observation was formulated in terms of the *right frontier constraint* (RFC). In the following, this constraint will be formulated andd it will be shown how its effect can be measured using pronoun resolution.

Jasinskaja and Karagjosova, (2015) formulate a version of the RFC as in (5.15).

(5.15) RIGHT FRONTIER CONSTRAINT: Only the nodes at the right frontier of the discourse graph are accessible for attachment of new discourse material.

A node in the discourse graph is 'accessible for attachment' if new discourse material can end up in a rhetorical relation with the utterance which is represented by this node. The utterance to which the new discourse material stands in a direct rhetorical relation is often called the *pivot* of the new discourse material. A possible method to measure whether an utterance can serve as a pivot for new discourse material is provided by pronoun resolution. It is assumed that a third person pronoun in an utterance can only

be resolved to an antecedent in the pivot of the utterance. Given this assumption, one can formulate the RFC as follows:

(5.16) RIGHT FRONTIER CONSTRAINT II: Non-local antecedents for third person pronouns in an utterance can only be in the pivot of the utterance. The pivot is always on the right frontier of the discourse tree. (Zeevat, 2011)

This formulation of the RFC is less general than the version given in (5.15). However, it has the advantage of providing a concrete method to measure whether an utterance in a discourse can serve as a pivot for new discourse material. In order to make sense of this claim, compare the following examples:

(5.17) a. Max_m repaired the roof.
b. Chris_c painted the wall.
c. $\text{He}_{*m/c}$ did a good job!

(5.18) a. Max_m repaired the roof.
b. Chris_c convinced him.
c. $\text{He}_{m/c}$ did a good job!

Example (5.17) shows a *Parallel* relation, (5.18) is an example for an *Explanation*. Since *Parallel* is coordinating and *Explanation* is subordinating, (5.16) predicts that the pronoun in (5.17.c) can be resolved only to Chris. This is the case because the utterance which contains Chris is on the right frontier of the discourse graph. On the other hand, it predicts that the pronoun in (5.18.c) can be resolved to both referents, Max and Chris, since both referents are contained in an utterance on the right frontier of the discourse graph:

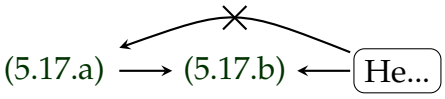


Figure 5.5: A discourse tree for (5.17). The attachment of the new discourse material to (5.17.a) is blocked by the coordinated structure, since (5.17.a) is not at the right frontier of the graph. In particular, the pronoun *he* can only be resolved to Chris.

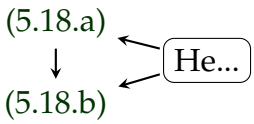


Figure 5.6: A discourse tree for (5.18). The subordinated structure of this discourse makes it possible that the new discourse material can attach to both previous utterances. In particular *he* can be resolved to either Chris or Max.

If one considers just two utterances, as in (5.17) or in (5.18), the effect of subordination and coordination can directly be measured on the basis of the RFC. Asher and Vieu, (2005) even used this effect in order to classify relations as coordinating and subordinating, depending on their behavior with respect to discourse anaphoricity. However, until today there is no satisfactory explanation of the fact that some relations allow more discourse units to serve as pivot for new discourse material then other relations. Section 6 provides an explanation to this problem for a distinct subclass of rhetorical relations, namely so-called *support* relations. Support relations will be the subject of the next section.

Remark: In general there is a strong recency effect in communication. Interlocutors

tend to talk about the newest information or about the newest utterance or speech act in a discourse. Given the right frontier constraint, one can interpret subordinating rhetorical relations as 'working against' the recency effect: the RFC basically says that subordinating rhetorical relations allow that new discourse material attaches to older discourse material.

Summary of this section: This section introduced rhetorical relations as relations between speech acts or sentences which may enrich the discourse with further information. Moreover, it discussed the distinction between *subordinating* and *coordinating* relations via some rather informal properties of subordinating and coordinating relations. Afterwards, this distinction was used for introducing discourse graphs and discourse trees. On the basis of these trees, the *right frontier constraint* was introduced as a restriction for the ability of utterances to serve as a pivot for new discourse material. However, even though there seems to be a quite robust intuition about which relations should be considered to be subordination, there is no satisfactory intrinsic definition for *subordination* so far. Nevertheless, the RFC makes predictions about the linguistic effect of subordination on the discourse structure. The RFC predicts that in subordinating rhetorical relations both discourse units can serve as a possible pivot for new discourse material. This behavior of subordinating rhetorical relations can at least serve as a criterion for subordination:

(5.19) SUBORDINATION CRITERION: In a sequence of discourse units (U_1, U_2), U_2 is subordinate to U_1 whenever both, U_1 and U_2, may serve as a pivot for new discourse material.

Turning the linguistic effect of coordinating and subordinating rhetorical relations into a definition was already proposed by Asher and Vieu, (2005). This criterion will be of interest for this book in section 6.2, where it will again be concerned with the right frontier of discourse graphs. The rest of this chapter is more interested in capturing the intuitions behind subordination in the terminology of communicative goals and speech acts developed in the last chapters.

5.2 Coordination and Subordination of Communicative Goals

Section 3.1 introduced the notion of *subgoals*, *goal stacks*, and *plans*. A subgoal was defined as a goal of which the agent believes that it helps him to achieve another goal. A goal stack was defined as a set of goals which is partially ordered by the subgoal relation and which contains a maximal element with respect to this relation. Finally, *plans* were defined as goal stacks together with certain actions, with which the agent wants to achieve the respective goals. Originally, these notions were introduced to be able to model an agent's plan and to connect individual actions of an agent to elements of a common plan and, therefore, to construct a coherent behavior from the agent's individual actions. In this section, these notions will be applied to communicative

goals, i.e. to goals of speech acts. In order to be able to talk about communicative subgoals, the following convention will be used:

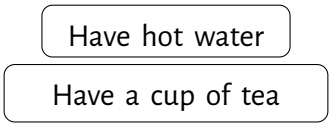
(5.20) COMMUNICATIVE SUBGOALS: Let A be an agent and let E_1 and E_2 be speech acts of A with goals G_{E_1} and G_{E_2}. The goal of E_2 is a communicative subgoal of the goal of E_2 if $\text{SUBGOAL}_A(G_{E_1}, G_{E_2})$. In this case, it will also be said that E_2 *supports* E_1 and that G_2 is communicatively subordinated to G_1.

This notion of communicative subordination must not be confused with the notions of subordination for rhetorical relations, although there are some similarities between these notions.

This section is devoted to the discussion of different types of subgoals and in particular to the question of how the notion of subgoals can be applied to communicative goals. In 5.2.1, the special case of 'mixed' goals will be discussed, i.e. cases where one goal is a communicative goal and where the other goal is a regular goal. After having established the notion of communicative subgoals, 5.2.2 places its focus on the following question: how can the notion of subordination of communicative goals be used to obtain a notion of *coordination* of communicative goals? Moreover, the question is addressed whether it is possible to apply such notions on problems of discourse coherence.

5.2.1 Different Types of Subgoals

In many cases, communicative goals are subgoals of non-communicative goals and vice versa. The following section is meant as a short discussion of such cases. In order to do this, recapitulate shortly what was said about subgoals in chapter 3. Subgoals are goals with the property of being means to achieve another goal. Consider again the example of an agent A who wants have a cup of tea and who believes that in order to make a cup of tea he needs hot water. In order to make a cup of tea, A has to prepare a certain amount of hot water. This makes the goal of having hot water a *subgoal* of the agent's goal of having a cup of tea. Therefore, it also belongs to a goal stack with the maximal goal of having a cup of tea. In chapter 3, goal stacks were represented as a set of boxes where each box represents a goal of the goal stack and where a goal G' is on top of a goal G, if G' is a subgoal of G. Given this convention for representing goal stacks, one can represent the goal stack of the tea example as follows:



Moreover, recall that speech acts have been defined as goal-directed utterances. This means in particular that the speaker pursues a certain goal, the *goal of the speech act*, with the performance of any speech act. In the following, such goals are also referred to as *communicative goals*. Since communicative goals are still goals, the notions of subgoals and goal stacks can easily be applied to communicative goals.

In principle, a subgoal relation can relate to both types of goals, i.e. communicative and non-communicative goals. That makes it possible that a non-communicative goal is a subgoal of a communicative goal and vice versa. Although this chapter is not primarily concerned with such 'mixed' cases of subgoals but with cases where both goals are communicative goals, such cases shall briefly be discussed. In the following, let G_1 and G_2 be two goals such that $\text{SUBGOAL}_A(G_1, G_2)$. Then there are four possibilities for G_1 and G_2 which can be considered:
First: G_1 is a non-communicative goal and G_2 is a communicative goal:
Such situations occur quite often. If A wants to have some bread from some bakery store, she has to *order* the bread. In this case, the communicative goal of ordering some bread is subordinated to A's goal of having bread. Or take the following situation: assume that A is in a coffee break at a conference and that A wants to get a cup of coffee but is involved in a boring discussion with another conference participant. Then (for reasons of social conventions) A is forced to say something like:

(5.21) Would you excuse me, I just want to get a cup of coffee.

The speech act in (5.21) also serves A's goal of having a cup of coffee and, therefore, the communicative goal of that speech act is a subgoal of the non-communicative goal of having a coffee.
Second: G_1 is a communicative goal and G_2 is a non-communicative goal:
Such situations occur for example in cases where communication involves technical devices. If A wants to write an e-mail, she has to first turn on the computer; if A wants to talk to somebody via mobile phone, she has take it out of the pocket, and so on. Another example for such cases include official talks or speeches. In such cases, the speaker simply has to first arrive at a certain location in order to be able to achieve his communicative goals. A priest, who wants to give a sermon has to climb the pulpit first, and so on.
Third: G_1 and G_2 are both non-communicative goals:
This is the case of pure actions where no speech acts are involved. This case was investigated in section 3.1.
G_1 and G_2 are both communicative goals: The case where both goals are communicative goals is the case where one speech act supports the other one. Presentational relations in RST are a particular class of relations which fits this condition. Remember that these relations are characterized as relations 'whose intended effect is to increase some inclination in the reader, such as the desire to act or the degree of positive regard for, belief in, or acceptance of the nucleus.' In other words, one of the involved speech acts is meant as a support for the other speech act. Speech acts which support other speech acts are going to be the main subject of this chapter.

This section was concerned with discussing different configurations of subgoals. It showed that, in many cases, communicative goals are subgoals of non communicative goals and vice versa. The next section will only be concerned with communicative goals and subgoals. It tries to establish a reasonable notion of *coordination* of communicative goals.

5.2.2 Coordination of Communicative Goals

This section aims at introducing a notion of coordination of communicative goals and a general idea for discourse coherence, that underlies the notion of coordination. The idea that will be carried out in this section is that communicative goals can be assumed to be coordinated if they are not subordinated and if they belong to the same *communicative plan*. The arguments provided in this section are based on the assumption that a notion of communicative coordination should also satisfy the condition that two speech acts that are communicatively coordinated are also perceived as coherent. It will, moreover, be assumed that a discourse is perceived as coherent if hearers are able to reconstruct a common plan connecting the individual speech acts. This assumption is similar to the assumption that individual actions of agents are perceived as coherent behavior if one is able to construct a plan which connects these actions.

In (5.20), the notion of a communicative subgoal was defined using the *subgoal*-relation. On the basis of this notion, one can assume that it is possible to define communicative coordination by exclusion: two communicative goals are coordinated if they are not communicative subordinated. However, this definition would not be very helpful for making claims or predictions about discourse coherence. The problem of this definition is that it coordinates *any* pairs of speech acts which would make any discourse coherent, at least under the assumption that the notion of communicative coordination implies coherence. For example, given this definition, the communicative goals of the following two utterances are coordinated, although this discourse should not be considered as coherent, at least without any further conditions on the context:

(5.22) ?John broke his leg. I like plums. (Knott and Dale, 1994)

On the basis of example (5.22), Knott and Dale argued that the set of discourse relations should somehow be restricted to relations 'which make sense'. The discourse in (5.22) could be considered to be coherent if one introduces a discourse relation of the form *inform-accident-and-mention-fruit*. However, there are good reasons why one should not introduce such a relation. For similar reasons, the notion of coordination of communicative goals without any further restrictions should not be used to make predictions about coherence.

In order to obtain a notion of communicative coordination that implies coherence, it will be imperative to have a look at larger discourse segments. As already mentioned, this section wants to establish the following idea: two speech acts should be considered to be communicatively coordinated if they are not communicatively subordinated and if they belong to the same communicative *plan*. Therefore, the notion of communicative coordination will be based on the notion of a communicative plan. But in order to get a better idea of communicative plans, larger discourse units should be considered.

Normally, speech acts are not isolated but part of a greater discourse, either in conversations or in monologues. This makes goals of speech acts a part of some greater conversational strategy or plan. For example, assume that someone tells the following (admittedly, not very interesting) short story:

(5.23) a. John filled a kettle with water and put it on the stove. E_1
b. Then, he put a tea bag in a cup. E_2

c. Afterwards, he poured the hot water in the cup. E_3
d. He waited for about five minutes and, finally, he had his cup of tea. E_4

Using the terminology of rhetorical relations, the individual sentences in (5.23) are basically connected by *Narration*. The speech act in (5.23.d) can be considered a *Result* of (5.23.c), but the general structure of this short discourse is narrative. Moreover, it will be assumed that all the communicative goals of the utterances in (5.23) are coordinated in the sense that they are no subgoals of further communicative goals. This last claim is derived from the fact that all utterances in (5.23) contain discourse connectives, such as *then* or *afterwards*. The exact reasoning for why these items signal coordination of communicative goals will be discussed in 5.3.3. Hence, no utterance in (5.23) supports a preceding utterance, but this discourse can nevertheless be perceived as coherent.

In the case of (5.23), this impression of coherence can be explained on the basis of the assumption that the individual sentences are tied together by the speaker's goal to *tell a story about John making a cup of tea*. Even though this goal is not overtly expressed in (5.23), one can assume that most hearers derive this conclusion from the individual speech acts.

The conclusion that it is the speaker's goal to tell a story about John and a cup of tea can again be explained as an application of the generalized goals principle (principle (3.24)). However, at this point the arguments how a hearer can come to that conclusion will not be laid out in detail.

Given the assumption that the speaker in (5.23) has the goal to tell a story about John, all utterances in (5.23) support the goal of telling a story about John, and therefore the hearer perceives this sequence of utterances as coherent. The following figure shows a possible goal stack which connects the sequence of utterances in (5.23).

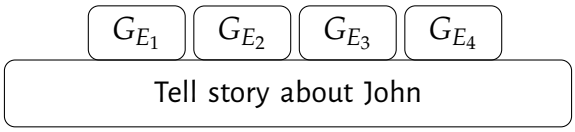


Figure 5.7: A stack of communicative goals for (5.23).

The individual utterances are tied together by a common goal, which means that they belong to a common communicative plan. Therefore, they are perceived as coherent.

In the following, the notion of a communicative plan will be used to define a more restricted and more useful notion of communicative coordination. To make this more precise, the following definition for a communicative plan will be used:

(5.24) COMMUNICATIVE PLAN: A communicative plan is a plan where all goals are communicative goals and all goal-directed acts are speech acts.

On the basis of communicative plans, it is possible to define a more restricted and a more useful notion of communicative coordination:

(5.25) COMMUNICATIVE COORDINATION: Two communicative goals G_1 and G_2 are coordinated if they belong to the same communicative plan and if they are not communicatively subordinated.

As a further example for an application of this notion of communicative coordination, consider again (5.6) (here repeated as (5.26)):

(5.26) a. Max painted the wall. F_1
b. Chris repaired the roof. F_2

In terms of rhetorical relations, the two utterances are related by a *Parallel*. Although the utterances above do not contain any connectives that signal coherence and since it is reasonable to assume that the second utterance is not meant as a support for the first one, it is perceived as a coherent discourse. Again, the coherence in this example can be explained by the fact that the hearer is able to reconstruct a superordinated communicative goal G from the given speech acts, which can roughly be described as G =[Inform about construction site]. Given this goal, one can conclude that both utterances belong to the same communicative plan, which establishes coherence between these utterances. The following figure shows a goal stack for that plan:

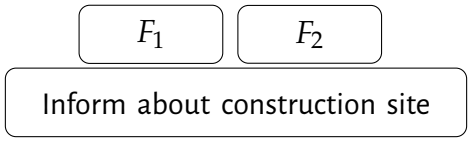


Figure 5.8: A stack of communicative goals for (5.26).

After the idea of communicative coordination was discussed, the rest of this section is devoted to a general discussion of discourse coherence.

So far, the considerations above were based on the idea that a discourse is perceived as coherent if the hearer of the discourse is able to relate the individual discourse units by a common plan. This assumption is based on the similar assumption that general actions of agents are perceived as coherent if they seem to belong to some plan of the agent.

The rest of this section discusses some other advantages of the idea that a discourse is coherent if it can be assigned to a plan. There are former theories which make claims about discourse coherence, such as *givenness theory* or *centering theory* (Averintseva-Klisch, 2013, Ch. 8). These theories will be discussed in further detail in section 6.1.1 where discourse prominence will be discussed. Both theories assume that coherence of a discourse depends on the choice of certain referential expressions. In particular, centering theory measures coherence of texts on the basis of referential chains. According to centering theory, an example like (5.23) is coherent because the referent *John* is picked up by the right referential expression *he* in all sentences of the text. However, just considering referential expressions, as it is done in centering theory, makes wrong predictions about the coherence of discourse. The modified example (5.27) from Knott and Dale should still be considered to be incoherent, even though it contains a referential chain.

(5.27) ?John broke his leg. He likes plums.

According to centering theory, (5.27) should be considered as coherent since it contains a perfect referential chain with an appropriate referential expression in the second utterance. On the other hand, centering theory predicts that some coherent discourses like (5.28) should be perceived as incoherent:

(5.28) Ann repaired the roof. Bob painted the wall. Chris lays the parquet...

However, on the basis of the idea that a discourse is perceived as coherent if the hearer is able to recognize the individual speech acts as parts of a common communicative plan, both examples can be explained. In example (5.27), there is no communicative plan that can be inferred without any further assumptions on John or the speaker from the given speech acts, which turn them into a coherent discourse. In (5.28), on the other hand, it is easy to infer the common goal [Inform about construction site] to which the individual speech acts are communicatively subordinated to and which turn them into a coherent discourse.

Finally, it has to be remarked that even a discourse like (5.27) could be considered coherent if the hearer is able to include the individual speech acts into a common plan. For example, in (5.29) the hearer is able to relate the two speech acts after he was able to construct the speaker's plan, which was to explain why the speaker goes to the hospital with plums.

(5.29) John broke his leg. He likes plums. I think I'll visit him at the hospital and bring him some plums.

Summary of this section: In this section, it was shown that it is easily possible to apply the notion of subgoals to the set of communicative goals and it was discussed that this application makes it possible to relate communicative goals to non communicative goals via the subgoal relation. A speech act which has a goal that is communicatively subordinated to the goal of another speech act can also be said to *support* that speech act. Moreover, this section was concerned with discussing communicative subordination and with the introduction of a reasonable notion of communicative coordination. This notion was established on the basis of communicative plans. The next section will be concerned with default assumptions in communication. This section claims (and proves) that, by default, the second utterance in a pair of utterances is interpreted as supporting the first one.

5.3 Defaults in Communication

This section is interested in the following question: given two speech acts, what is the default interpretation for the rhetorical relation relating them? The term *default interpretation* refers to the hearer's assumption on the function of speech acts in the case of the absence of any linguistic devices which signals that the speaker wants to produce a particular rhetorical relation. It is claimed in this section that, by default in a

pair of speech act, the second speech act is interpreted as supporting the first one (i.e. the goal of the second speech act is a subgoal of the goal of the first speech act). In the following, this claim is referred to as the *support default* or the *subgoal default*.

(5.30) SUBGOAL DEFAULT: Let (E_1, E_2) be an adjacent pair of speech acts. By default, the hearer interprets E_2 as a support for E_1.

Relations, in which the second speech act supports the first one will also be called *support relation*. Proving the subgoal default will be the main content of section 5.3.2. This prove requires almost all principles which have been developed in the last chapters. The subgoal default can be regarded as one of the basic claims of this book.

This section is structured as follows: Section 5.3.1 will introduce the *Elaboration*-default as a precursor of the support default. As one can already guess by the name, the *Elaboration* default claims, that *Elaboration* is the default relation for pairs of utterances. However, there are cases which are problematic for the *Elaboration* default but which can be captured by the support default. Section 5.3.2 is concerned with the prove of the subgoal default. However, not all relations are support relations. Section 5.3.3, will, therefore, target the question by which linguistic means this default can be overwritten.

5.3.1 Elaboration-Default

Given two adjacent and unmarked speech acts, which rhetorical relation holds between these speech acts? According to Jasinskaja, (2010), such a pair of speech acts is interpreted as an *Elaboration*. Other relations like *Narration*, *Parallel* etc. have to be marked somehow, either by discourse particles (such as *then, but...*) by a special syntactic structure, or by certain prosodic features. This section is devoted to a short sketch of the derivation of the *Elaboration*-default. Even though the *Elaboration*-default is in some sense the natural predecessor of the subgoal default, there are cases where it fails to predict the right relation. This problem will be solved by the subgoal default.

The derivation of the *Elaboration*-default basically uses two principles of discourse interpretation:

1. THE PRINCIPLE OF TOPIC CONTINUITY:
   By default, the discourse topic does not change.
2. THE PRINCIPLE OF EXHAUSTIVE INTERPRETATION:
   By default, an utterance is interpreted exhaustively.

The term *topic* used in the first principle refers to the discourse topic, i.e. to a constituent answering the current QUD[2]. Exhaustive interpretation, on the other hand, states that hearers try to extract as much information from a given sentence as possible (with respect to a given QUD). According to this principle, hearers interpret a sentence as if it gives an complete, i.e. exhaustive, answer to a given QUD. In particular, both principles depend on a given QUD.

[2]For a general discussion on the use of the term *topic*, see Roberts, (2011).

1. The discourse topic can be modeled as the *question predicate*, i.e. the part of the given QUD without the *wh*-element. This can be formalized by a λ-term where the *wh*-element is abstracted away. For example, the discourse topic for the QUD q =[*Who did Bob kill?*] can be formalized as $\lambda x\ [killed(\text{Bob}, x)]$.

2. According to *Exhaustive Interpretation*, the hearer interprets a sentence in a way that maximizes the information which he can obtain from a given utterance with respect to the discourse topic. For example, if the sentence A=[Bob killed Ann.] is supposed to be an answer to the QUD q from above, a hearer interprets this answer as [Bob killed Ann (and no one else).].

The *Elaboration* default can be derived from these two principles as follows. Given two utterances (U_1, U_2), the principle of topic continuity states that both utterances have the same topic. In particular, both utterances try to answer the same question under discussion. Moreover, both utterances are interpreted as if they give an exhaustive answer to the given question under discussion. But if both utterances are exhaustive with respect to the given QUD, both utterances give a complete answer to the current QUD, and, hence, both answers *coincide*. This, in particular, implies that both utterances refer to exactly the same eventualities, but this is exactly the definition of *Elaboration*, as it was for example characterized by Hobbs, (1979) (cp. page 137).

In order to see how these principles can be applied to an actual pair of natural language utterances, consider the following example:

(5.31) 1. Bob met a great actress at the party. Jasinskaja, 2010
2. He met Vanessa Redgrave.

According to the elaboration default and according to the intuition of most interpreters both sentences in (5.31) describe the same event: Bob met exactly one great actress at the party and this actress was Vanessa Redgrave. How can this interpretation be obtained on the basis of the two principles above? According to *topic continuity*, both utterances have the same topic which means that both utterances answer the same question $q = [\textit{Who did Bob meet?}] = \lambda x\ [meet(\text{Bob}, x)]$. Since both utterances give a complete answer to the question q, the event of Bob meeting a great actress and the event of Bob meeting Vanessa Redgrave have to coincide:

$$\left.\begin{array}{l} [(5.31.1)]^q_{exh} = \{\text{great actress}\} \\ [(5.31.2)]^q_{exh} = \{\text{Vanessa Redgrave}\} \end{array}\right\} \Rightarrow \{\text{great actress}\} = \{\text{Vanessa Redgrave}\}$$

Hence, the two sentences in (5.31) are interpreted as if they describe the same event and, therefore, they are interpreted as an *Elaboration*.

However, there are cases of utterance pairs where Jasinskaja's approach to defaults for rhetorical relations, makes wrong predictions. As examples of unmarked pairs of speech acts which can't be classified as *Elaboration*s, and hence which provide a

problem for the *Elaboration* default, take the following cases:

(5.32)
1. I think Bill is not at home.
2. There is no bike in front of the house. *Evidence*

(5.33) director A to assistant B
1. Would you please send these letters.
2. You will find some stamps on my desk. *Enablement*

(5.34)
1. We should go to the movies tonight.
2. The new Star Wars is on. *Motivation*

All these examples are clearly unmarked cases of pairs of speech acts but none of them can be classified as an *Elaboration*. However, *Evidence*, *Enablement* and *Motivation* (and *Elaboration* itself) are examples of *presentational relations* of RST(cp. definition (5.12)). These relations will be discussed in detail in section 5.4. Presentational relations, however, are typical examples for support relations. In other words, the examples above are all unmarked pairs of utterances which form a support relation. But this is exactly what is predicted by the support default, which will be the topic of 5.3.2.

5.3.2 The Support Default

This section wants to show how one can derive the support default from the principles developed so far. The support default claims that by default the goal of the second utterance of an utterance pair is interpreted as a subgoal of the first utterance of that pair. (5.35) formulated an informal version of this claim. Unlike the *Elaboration*-default, the support default does not make a precise statement about the particular rhetorical relation which relates an utterance pair, nor does it make predictions about its content. It rather emphasizes the *function* of the second speech act with respect to its context speech act. (5.35) shows an informal version of the subgoal default:

(5.35) SUBGOAL DEFAULT - INFORMAL VERSION:
Let (E_1, E_2) be an unmarked pair of adjacent speech-acts. By default, the goal of E_2 is interpreted as a subgoal of E_1.

Examples (5.32.1)-(5.34) already showed some applications of this rule. The rest of this section is now finally devoted to a proof of this claim. Since this proof is rather long, the following list presents the individual steps of this proof. These steps will be formalized and laid out in greater detail after the formal version of the support default will be given in (5.36). In the following, A refers to the speaker of the utterance pair (E_1, E_2) and B to the interpreter of that pair.

1. A utters the two speech acts E_1 and E_2.
2. B recognizes that A utters E_1 and E_2. Therefore B can conclude that A pursues goals G_{E_1} and G_{E_2} with the performance of these speech acts.
3. Moreover, by *act-goal connection* (principle (3.23)), B can conclude that A is in states S_{E_1} and S_{E_2} of pursuing the goals G_{E_1} and G_{E_2}.

4. By the assumption that E_1 and E_2 are unmarked, B has no evidence that A is not in the state S_{E_1} at the performance of E_2 anymore.
5. Therefore, by *inertia* (2.53), B can conclude that S_{E_1} still holds during the performance of E_2.
6. Therefore, B can conclude that $S_{E_2} \subset S_{E_1}$ and, by *commitment to a goal* (principle (3.25)), B can conclude that $\textsc{SubGoal}_A(S_{E_1}, S_{E_2})$.

The last point of the list above is exactly the statement of the support default. The use of *inertia* is again crucial in this argumentation and it is basically the only way to connect both states of pursuing goals. Both utterances are unmarked, which in particular means that A did not use any linguistic items with which he signals that he believes that the goal of E_1 is achieved at the time of E_2. Such items will be considered in section 5.3.3. The absence of such items makes it possible to apply inertia to A.

(5.36) shows a formal version of the subgoal default.

(5.36) SUBGOAL-DEFAULT:

Let A and B be agents. Assume that A performs an adjacent pair of speech acts (E_1, E_2). Then, by default, B assumes that $\textsc{SubGoal}_A(S_{E_1}, S_{E_2})$:

$$\boxed{\begin{array}{c|c} A\ B & \\ \hline \textsc{agent}(A),\ \textsc{agent}(B) & \\ \boxed{\begin{array}{l} E_1\ E_2 \\ \hline E_1 : \textsc{Speech Act}(A, B, U_1, G_1) \\ E_2 : \textsc{Speech Act}(A, B, U_2, G_2) \\ E_1 \prec E_2 \end{array}} & \Rightarrow \boxed{\begin{array}{l} S \\ \hline E_2 \subset S \\ S : \textsc{bel}\,(B, \textsc{SubGoal}_A(G_2, G_1)) \end{array}} \end{array}}$$

The rest of this section is concerned with the details of the proof of (5.36). In the following it will be assumed that some speaker A performs a pair of unmarked speech acts (E_1, E_2)towards hearer B. Given this situation, B can recognize that A performed some speech act, which means that B can conclude that A pursues some goals with the performance of these speech acts. This situation is represented by (5.37).

(5.37)
$$\boxed{\begin{array}{l} A\ B\ E_1\ E_2\ S_2\ S_2 \\ \hline \textsc{agent}(A)\ \textsc{agent}(B) \\ E_1 : \textsc{Speech Act}(A, B, U_1, G_1) \\ E_2 : \textsc{Speech Act}(A, B, U_2, G_2) \\ S_2 \subset S_1 \\ S_1 : \textsc{bel}\left(B, \boxed{\begin{array}{l} E_1,\ E_2 \\ \hline E_1 : \textsc{Speech Act}(A, B, U_1, G_1) \\ E_1 : \textsc{Speech Act}(A, B, U_2, G_2) \end{array}}\right) \Rightarrow S_2 : \textsc{bel}\left(B, \boxed{\begin{array}{l} S_{E_1}\ S_{E_2}\ K_1\ K_2 \\ \hline E_1 \subset S_{E_1}\ E_2 \subset S_{E_2} \\ S_{E_1} : \textsc{Pursue}\,(A, G_1) \\ S_{E_2} : \textsc{Pursue}\,(A, G_2) \end{array}}\right) \end{array}}$$

B is now in a situation where he believes that the speaker pursues two different goals at two points in time. However, at the time of the performance of E_2, B has no evidence

that some event may have affected the speaker's state of pursuing the goal G_1. Because of inertia, B can still believe that in the time of the event E_2, A is still in the state S_{E_1} of pursuing G_1. In this step, the following assumptions were used: 1. B did not show any reaction (cp, acceptance moves in section 4.2) to E_1 and 2. both speech acts do not contain any linguistic devices which signal that the speaker believes that one of his goal is achieved. (5.38) shows a formalization of this application of inertia.

(5.38)
$$S_2 : \text{BEL}\left(B, \begin{array}{|l} \hline S_{E_1}\ S_{E_2}\ G_1\ G_2 \\ \hline E_1 \subset S_{E_1}\ E_2 \subset S_{E_2} \\ S_{E_1} : \text{PURSUE}\,(A, G_1)) \\ S_{E_2} : \text{PURSUE}\,(A, G_2)) \end{array}\right) \Rightarrow \begin{array}{|l} \hline S_3 \\ \hline S_3 \subset S_2 \\ E_2 \subset S_3 \\ S_3 : \text{BEL}\left(B, \boxed{S_{E_2} \subset S_{E_1}}\right) \end{array}$$

The box in (5.38) describes a situation where B believes that A is in two states of pursuing a goal, namely S_{E_1} and S_{E_2} and that $S_{E_2} \subset S_{E_1}$. Moreover, (5.38) shows that $E_2 \subset S_{E_2}$ holds and, therefore, that $E_2 \subset S_{E_1}$ also holds. This, however, are exactly the preconditions for applying of *commitment to a goal* (cp. (3.25)). According to commitment to a goal, it is now possible to conclude that $\text{SUBGOAL}_A(G_1, G_2)$ holds:

(5.39)
$$\begin{array}{|c|} \hline S_4 \\ \hline S_4 \subset S_2 \cap S_3 \\ S_2 \oplus S_3 \Rightarrow S_4 : \text{BEL}\,(B, \text{SUBGOAL}_A(G_1, G_2)) \\ \hline \end{array}$$

The box in (5.39), shows a situation in which B is in a state of believing that $\text{SUBGOAL}_A(G_1, G_2)$, but this is exactly the statement of the support default.

Despite the fact that the *subgoal*-default seems to be a strong claim, it has to be stressed that (unlike the *Elaboration*-default) it makes no predictions about a concrete rhetorical relation. Moreover, it does not provide any tools to calculate any further inferences which are typical for certain relations (for example an identity inference for *Elaboration*s as in the case of example (5.31) or a causal relation as in the case of *Explanation*s). The support default rather makes claims about the *function* that a hearer has assigned to an utterance.

This section was concerned with a formulation and a proof of the support default. However, not all relations in a discourse are support relations. Hence, there must be some way to overwrite the support default. Section 5.3.3 will be concerned with some linguistic cues having the effect of overwriting the support default.

5.3.3 Overwriting the Default

In 5.3.2 it was claimed that, by default, an adjacent pair of utterances is interpreted as if the goal of the second utterance is a subgoal of the goal of the first utterance. However, not all pairs of utterances stand in such a relation. So, there must be some way to prevent this relation. This section analyses some means to overwrite the default. The idea proposed in this section is roughly the following: in the proof of the subgoal default, the assumption was used that the speaker did not signal that she believes that

a former communicative goal is achieved. But if a speaker does indeed signal that some former goals are achieved, the hearer can conclude that the speaker is not in the state of pursuing the goal of the first speech act anymore and, hence, commitment to a goal cannot be applied.

So the question is, how can speakers signal that they believe some former speech acts to be successful? In principle, there are many ways to do that. However, this section places a particular focus on certain presupposition triggers. The idea pursued in this section is the following: content can only be presupposed, if the speaker believes that this content is already an element of the common ground (cp. section 4.3). This, on the other hand means, that some former goal of one of the interlocutors is achieved. This implies that in cases where the presupposition has to be resolved to the content of the former utterance, the hearer can assume that the speaker believes that the goal of a former utterance is achieved. But this means that the speaker is not in the state of pursuing that goal anymore. The following short list summarizes this idea. Again A denotes the speaker and B the hearer of the utterances.

1. Assume that speaker A utters two speech acts (E_1, E_2) and that E_2 contains a presupposition which has to be resolved to some content of E_1
2. B knows that presupposing content is only felicitous, if the presupposed content is already an element of the common ground.
3. Therefore, B knows that A believes that the presupposed content is already an element of the common ground.
4. Therefore, B knows that A believes that communicative goal of E_1 is already achieved.
5. Therefore, B believes that E_2 cannot be a communicative subgoal of E_1.

This is the general structure of inferring that the speaker does not produce a support relation on the basis of presuppositions. In the rest of this section, this mechanism is exemplified on the basis of an example that contains *also* as presupposition trigger. Afterwards, the presupposition triggers *and* and *then* will be discussed but in lesser detail.

Since the argument above uses concrete presupposed content, it will be useful to represent whole propositions. As an example for the effect of presuppositions on the interpretation of a pair of utterances, take (5.40):

(5.40) Speaker A, Hearer B

a. John can open Chris' safe.
b. He can also open Diana's safe.

The utterance (5.40.b) contains *also* as a discourse connector. *Also* has the function to signal that there is a referent which is known to all interlocutors and which has similar or the same properties as one of the referent currently under discussion. In other words, *also* presupposes the existence of a referent which has the same properties of a referent in the current sentence. Moreover, one can assume that this referent is denoted by the focus constituent of the given sentence. The term *focus* refers again to the *wh*-element of the corresponding QUD (cp. page 150). This interpretation of *also* is

basically due to Heim, (1992) who proposed the following semantics for *also*[3]:

(5.41) $\textit{also}[\varphi(\alpha_F)] \gg \exists X \neq \alpha \ \wedge \ \varphi(X).$

In (5.41), φ denotes the current QUD and α_F the focus element (i.e. the answer to the current QUD) of the sentence which contains *also*. Hence, *also* presupposes the existence of a referent which is not identical to α but which has the same property φ. This analysis is of course oversimplified. The use of particles like *too* or *also* is in general more restricted than proposed by this analysis (Zeevat, 2004). For example, it is common knowledge that millions of people are having dinner in New York every day. However, without any further restrictions on the context, it is in general infelicitous to utter a sentence like (5.42):

(5.42) John had dinner in New York, too. Zeevat, 2004

Although the semantics of *also* given in (5.41) is not complete yet, it still models a necessary condition for a felicitous use. In particular, it includes a condition of *also* which is of interest for working out the current example.

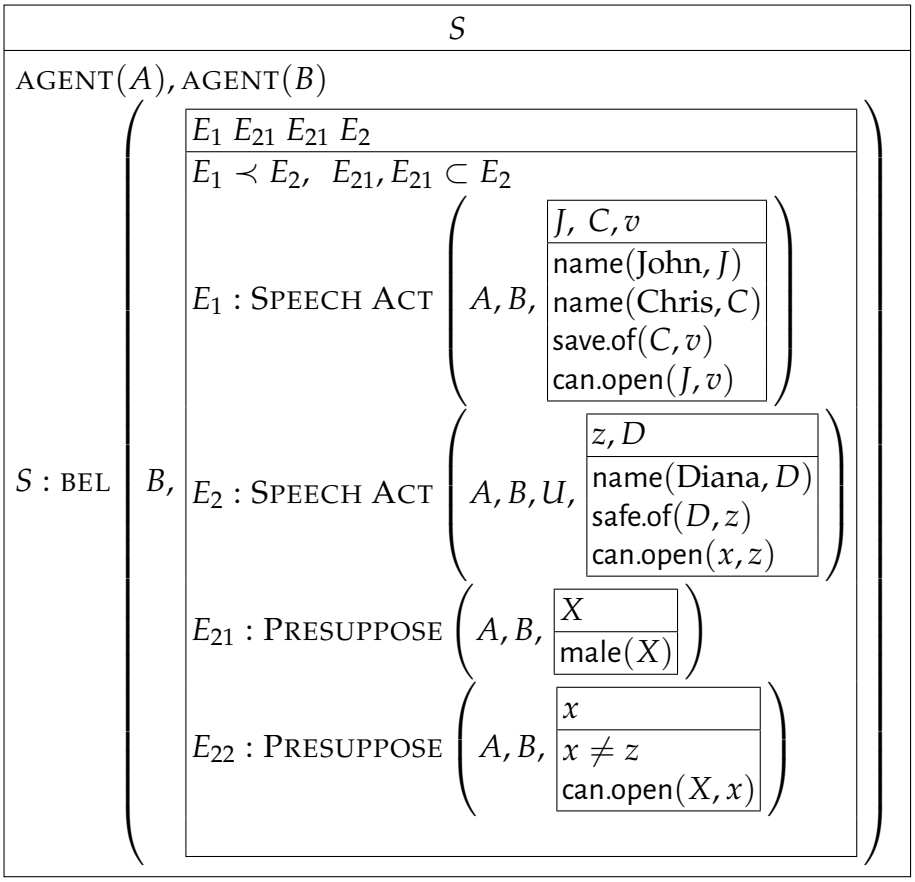
For the analysis of example (5.40), one may assume that the focus constituent of (5.40.b) is the whole *NP Diana's safe*, although this assumption is a mild oversimplification of the actual states of affairs. Moreover, proper names will be treated as simple predications and not as presupposition triggers as well. Given this assumption, the sentence in (5.40.b) makes at least the following two contributions to the discourse:

- There is some male individual Y who can open $[\textit{Diana's safe}]_F$ asserted content
- Y is already known and can open X and $X \neq F$. presupposed content

The sentence in (5.40.b), therefore, triggers at least two presuppositions: the existence of a male individual Y and that there is some thing X with the property that Y can open X. The situation after the utterance of (5.40.b) can be described as in (5.43):

[3]Originally, Heim proposed this analysis for *too*. However, at this point we assume that *too* and *also* are semantically equivalent.

(5.43)



One may assume that the resolution of the presuppositions of E_{21} and of E_{22} to the former context does not provide any problems for B. This means that B is able to conclude that $X = \text{John}$ and that $x = v$. According to the analysis of presuppositions in 4.3, a speaker who presupposes content already believes that this content is an element of the common ground. This assumption, and the assumption that it is the goal of a speech act to update the common ground with its content, makes it possible to derive the situation described in (5.44) from (5.43):

(5.44)
$$\begin{array}{|l|}\hline S' \\ \hline S' \subset S \\ S' : \text{BEL}\left(B, \begin{array}{|l|}\hline S_1, S_2 \\ \hline E_1, E_2 \subset S_1,\ E_{21}, E_{22}, E_2 \subset S_2 \\ S_1 : \text{BEL}\left(A, \begin{array}{|l|}\hline J\ C\ D\ X\ x\ v \\ \hline \text{name}(\text{John}, J) \\ \text{name}(\text{Chris}, C) \\ \text{safe.of}(C, v) \\ \text{can.open}(X, x) \\ X = J \\ x = v \\ \hline \end{array} \in \text{CG}_{A,B}\right) \\ S_2 : \text{PURSUE}\left(A, \begin{array}{|l|}\hline D\ z \\ \hline \text{name}(\text{Diana}, D) \\ \text{safe.of}(D, z) \\ \text{can.open}(J, z) \\ \hline \end{array} \in \text{CG}_{A,B}\right) \\ \hline \end{array}\right) \\ \hline \end{array}$$

In particular, this presentation states that B was able to conclude that A thinks that his former goal to make [*John can open Chris' safe*] an element of the common ground is already achieved. From the presentation above, this can be derived by the fact that the belief state S_1 contains the proposition given in (5.45).

(5.45)
$$\begin{array}{|l|}\hline J\ C\ X\ x\ v \\ \hline \text{name}(\text{John}, J) \\ \text{name}(\text{Chris}, C) \\ \text{safe.of}(X, v) \\ \text{can.open}(x, a) \\ X = C \\ x = J \\ \hline \end{array} \in \text{CG}_{A,B}.$$

After (5.40.b), B can conclude that A is not in the state of pursuing the goal of speech act (5.40.a) anymore and, hence, that the goal of the second speech act is not a subgoal of (5.40.a).

This argument does not only work for the connective*also*. There are many different linguistic devices with which one can overwrite the subgoal default. Other examples for morphological items whith this property are the connectives *then, therefore, but* and *and*. All of these connectives can have a similar function with respect to the status of communicative goals. All these particles presuppose content of the previous discourse, which signals that some former goal is believed to be achieved. The rest of this section is devoted to a discussion of the connectives *then* and *and*.

Then: In its non-conditional function, *then* is used in cases where a speaker wants to signal that the event which is described in a speech act follows an event which was described in some preceding speech act. This basic intuition is captured by (5.46):

(5.46) $$\textit{then}[\varphi(\alpha_e)] >> \exists e : \text{EVENT} \wedge e \prec \alpha.$$

In (5.46), the event which is described by sentence φ is called α. The connector *then* presupposes the existence of a preceding event, which has to be mentioned in some utterance in the preceding discourse. Therefore, by using *then*, the speaker assumes that the presupposed event, is already an element of the common ground and, hence, that the goal of the speech act in which the presupposed event is described is achieved. In order to illustrate this function of *then*, consider the following pair of examples:

(5.47) a. Max had a great evening last night.
b. He won a dancing competition.
b'. He *then* won a dancing competition.

The connector *then* in utterance (5.47.b') forces a narrative interpretation. In particular it forces the interpretation that Max' winning of the dancing competition is NOT necessarily part of his great evening anymore. Utterance (5.47.b), on the other hand, is interpreted as an elaboration of Max' great evening, or, in order to stay in the vocabulary of this work, supports the speech act given in (5.47.a), which is predicted by the support default. This forces the interpretation that the event of Max winning the dancing competition is interpreted as an event, which happened *within* this evening. However, the use of *then* in (5.47.b') blocks this interpretation.
And: *And* in sentence-initial position has indeed the effect to signal that the goal of the former sentence is believed to be achieved. In order to see the effect of the use of *and* in this position on the interpretation of the discourse, have a look at the following minimal pair:

(5.48) a. Bob met a great actress at the party.
b. He met Vanessa Redgrave.
b' And he met Vanessa Redgrave. Jasinskaja and Karagjosova, 2015

Example (5.48) shows a similar contrast as (5.47). In (5.47), the inference that winning a dancing competition is a sub event of Max' great evening is blocked by the use of *then* which presupposes that a former event is already an element of the common ground. This is very similar to the use of *and* in the former example: Jasinskaja and Karagjosova noted that the use of *and* in (5.48.b') blocks the inference that Vanessa Redgrave is identical with the great actress which was mentioned in (5.48.a. This blocking effect can be interpreted as an epiphenomenon of the general function of *and* as a discourse connective, which is to signal that the speaker believes that the goal of some former speech act is achieved. This implies that *and* operates on the level of *speech act*s, i.e. its contribution to the discourse is the communicative coordination of the connected speech acts (Txurruka, 2003). Hence, *and* as a discourse connective has essentially the function to signal that the speaker believes that some of his former goals are achieved. In contrast to *then*, which primarily makes a semantic contribution with an additional pragmatic effect, *and* directly operates on the level of speech acts. This claim is summarized by proposing the following semantics for *and*:

(5.49) SEMANTICS OF ***AND***: *And* in sentence initial position signals that the speaker believes that the goal of a former speech act is achieved.

$A\ B$	
$\text{AGENT}(A), \text{AGENT}(B)$	
$E : \text{SPEECH ACT}(A, B, \textit{and}[U], K) \Rightarrow$	$E'\ S'$ $E' \prec E, E \subset S$ $E' : \text{SPEECH ACT}(A, B, U', K')$ $S : \text{BEL}(A, K' \in \text{CG}_{A,B})$

Actually, it is not even necessary that the speaker of E coincides with the speaker of E'. A speaker may even use *and* in sentence-initial position in order to signal that the goal of some preceding speech act, of which he was *not* the speaker, is achieved and this also blocks identity inferences. This behavior of *and* explains the mechanics behind the following German joke:

(5.50) Wife and husband are having breakfast together. Husband suddenly spills some coffee over his best shirt.

a. husband: Damn. I look like shit.

b. wife: And your shirt is dirty.

The use of *and* in the second utterance blocks the causal inference that the husband looks "like shit" because he has coffee over his shirt. In fact, it forces the discourse to the interpretation, that the husband stained his shirt and *in addition* looks "like shit".

At this point it has to be remarked that the simple occurrence of a presupposition trigger does not suffice for triggering a coordinated interpretation. As an illustration, take again the following example:

(5.51) Max must have left. His bike is gone.

This is a perfect example for a support relation, even though the second utterance contains a third person pronoun as presupposition trigger. The important point of this example is that the main goal of the first utterance was not to introduce Max as an element of the common ground but to introduce the information that Max *left*. This, however, is not necessarily an element of the common ground after the performance of the first utterance and, hence, the goal of the speech act is not neccessarily achieved. One can compare this example with the following one:

(5.52) Max must have left. Alex has left, too.

In this example, the *too* presupposes an event with the same property as the event which is described in the sentence that contains *too*. This, however, means that the presupposed event (in this case, the element which is described in the former speech act) is an element of the common ground, which means that the goal of the former speech act is achieved.

Summary of this section: This section was concerned with the default interpretation of pairs of utterances. The *Elaboration* default as a former approach to this problem was discussed. Moreover, some examples have been introduced where the *Elaboration* default makes wrong predictions. Afterwards, the support default as a

default for the interpretation for unmarked pairs of utterances has been introduced and was derived from the principles that have been developed so far. The support default is able to cover more examples than the *Elaboration* default. However, it does not predict a specific rhetorical relation but rather makes claims about the specific function of speech acts in discourse. Finally, some linguistic means with which the support default can be overwritten have been discussed. Certain presuppositions can only be resolved to content contained in former speech acts which means that and using these presuppositions signals that the speaker believes that the content of these speech acts is already an element of the common ground. But becoming an element of the common ground was the goal of these speech acts which means that these speech acts don't need support anymore. As a result, the hearer can conclude that the second speech act was not meant to support its preceding speech act.

5.4 Repair Moves in Discourse

This section investigates some cases of so called *self-repair moves* and tries to explain such moves as realization of plan updates. In order to define such moves, consider the following examples.

(5.53) At that time, at Christmas, he, my brother, still lived in Cologne.

(5.54) It is there. On the refrigerator.

(5.55) Let us go to the movies tonight. The new Star Wars is on.

(5.56) Send these letters please. Stamps are on my desk.

In all these example there are units with which the speaker seems to *repair* a problem of the utterance. What does this mean? Consider (5.53). What makes a speaker produce such weird appositives as in (5.53)? A possible reason for such appositives may be the following: the speaker doubts that the hearer is able to resolve the presuppositions which are triggered by the phrases [*at that time*] or by *he*. In order to keep the discourse effective and avoid an upcoming clarification question, the speaker decides to resolve the problematic phrases himself. The utterance pair in (5.54) can be interpreted in a similar fashion: for some reason the speaker believes, that the hearer did not understand the content of the first sentence. As a result, he reformulates the part of the former speech act of which the speaker believes that it was responsible for the hearer's lack of understanding. In this sense, both examples can be regarded as cases of *self repair*: the speaker tries to solve possible problems *before* they are explicated by the hearer. In example (5.53), the speaker repairs a possible misinterpretation of some of the involved *NP*s, and in (5.54), the speaker repairs a possible lack of knowledge. These repairs concern the content of the speech act in the sense that they try to ensure that the hearer is able to identify the intended content of the speech act. For that reason Such repair moves will be called *content level* repairs.

The examples (5.53) and (5.54) are somewhat more abstract than the former two examples, but one can explain them in a similar fashion. A discourse like (5.55) can be

uttered if the speaker believes that the hearer is not willing to act as it was requested. Therefore, the speaker tries to increase the hearer's intention to perform the requested action. Example (5.56) shows a case where the speaker believes that the hearer is in principle ready to perform the requested action but is not *able* to perform it (in this case because the hearer lacks some knowledge). These examples show repair moves on a different level: the repair moves don't operate on the content level but try to ensure that the hearer *accepts* the speech act (in the sense of section 4.2). Such repair moves will therefore be referred to as repair moves on the *acceptance level*. A more precise definition of repair moves on the basis of goals of speech acts will be given in 5.4.1.

Self-repair moves have been, in particular, been investigated by Ginzburg, (2012). He emphasizes the analogy between self-repair and clarification requests (Ginzburg, 2012, p. 284):

(5.57) a. *A*: Did Bo leave? *B*: Bo? *A*: Your cousin.
b. *A*: [Did Bo + {I mean} your cousin] leave?

In both cases, a problem with the given material is detected and must therefore be repaired. The main difference between these examples is the following: in the first case it is made public that there is some problem with the given material. In this case *B* is not able to make sense out of the word *Bo*. In the second case, the hearer did not make it public that there might be a problem with the presented material. Nevertheless, it is repaired. The reason for this kind of repair can be found in a recursion of mental states: in (5.57.b), the speaker seems to have certain assumptions about the hearer's abilities to decode the propositional content from the given phonological material. According to Ginzburg, a clarification as in (5.57.b), can be interpreted as an answer to an implicit clarification request. But how can a speaker construct such a clarification request? Section 4.2 investigated commitment uptakes and grounding. In particular, the following list of necessary conditions for a successful grounding process from Clark and Brennan, (1991) was discussed:

(5.58) CONDITIONS FOR GROUNDING: For a speaker *A*, a hearer *B*, and an utterance *U*, the following conditions are necessary for a successful grounding of the utterance *U*:
State 1 : *B* noticed that *A* uttered *U*
State 2 : *B* correctly heard *U*
State 3 : *B* understood what *A* meant by *U*

This section pursues the following idea: a speaker has to produce a self-repair move if he believes that one of the conditions above does not hold yet. In other words, the speaker constructs the implicit clarification question on the basis of his assumptions on the status of the conditions in (5.58). This idea has the advantage of making it possible to classify repair moves on the basis of the list above. In §5.4.1 this list will be extended by an additional point that reflects acceptance of a speech act. Nevertheless, it is important to note that this section does not claim to give a full lists of all possible repair moves and of their recognition, but rather is meant as a presentation of a possible road map to provide a more fine-grained analysis and classification of rhetorical

relations. Moreover, it is rather speculative and makes some strong assumptions about the connection of certain mental configurations and predicted rhetorical relations.

The rest of this section is structured as follows: §5.4.1 defines the general notion of self-repairs as a reconsidered version of a communicative plan. 5.4.2 discusses conditions for such a plan update. These conditions will be an extended list of the list in (5.58). Section 5.4.3 is concerned with repair moves beneath the content level. These are cases where state 1 of the list above is not achieved yet (ignorance about a speech act) or where it is achieved but state 2 is not achieved yet. Section 5.4.4 investigates repair moves on the content level and shows that content level repairs can be realized as *Reformulation*s. Finally, section 5.4.5 investigates repairs on the acceptance level and shows that some presentational relations can be explained as realizations of self-repairs on the acceptance level.

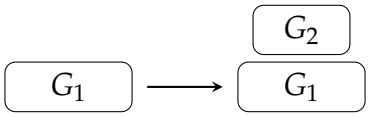
5.4.1 What is Self Repair

This section aims at giving a definition of repair moves in the terminology developed so far. Self-repair moves have been characterized as speech acts with the additional property of being performed in order to solve possible problems in the grounding process of a speech act but are realized *before* a possible grounding problem was made explicit. This makes-self repair moves similar to repair moves which give an answer to a clarification request as in (5.57.a). In both cases, a possible grounding problem is addressed, but with the difference that in the one case the grounding problem is overtly expressed and in the other case it is left implicit. Both cases have in common that the grounding problem is a problem for the speaker's communicative plan. A speaker performs a speech act in order propose a common ground update with the content of the speech act. The speaker is aware of the fact that the speech act can fail in achieving that goal for many reasons. The speech act can fail in fulfilling the conditions for grounding as in (5.58) and it can fail to be accepted. However, in both cases, the speech act did not show the intended effect, i.e. it did not result in a common ground update. In this case, the speaker has spent more work on achieving the original goal of the speech act than it was originally expected. But this means that the speaker has to reconsider his original communicative plan. In order to make sense out of this claim, consider the following example:

(5.59) a. He is going to come to the party.
b. I mean Bo.

Utterance in (5.59.b) can be regarded as a self-repair move. It clarifies to which entity the pronoun *he* is supposed to be referring and, hence, repairs a possible comprehension problem, i.e. a problem on step 3. of the list in (5.58). If the speaker would have guessed that referring to Bo with a pronoun may result in an understanding problem in the first place, he directly would have chosen to refer to Bo with the proper name *Bo*. However, the speaker did not do so and, hence, has to repair this problem *after* the utterance was performed. In other words, in (5.59) it was the speaker's original goal to inform the hearer that Bo is going to come to the party and, therefore, the speaker constructed the plan to utter (5.59.a) in order to achieve that goal. After the utterance

of (5.59.a), the speaker may have realized that this goal cannot be achieved by means of that utterance alone. The speaker's original plan will not work out as expected in the first place and this plan has to be reconsidered. In the case of (5.59), the speaker detects the problem in the hearer's ability to resolve the third person pronoun to the right referent. He has to give additional information for making the hearer understand the content of the first speech act. In section 3.2, a pictorial way to deal with plans and plan updates in form of goal stacks was developed. This can be used to illustrate the plan update in (5.59) as follows: let G_1 denote the goal of the first speech act and let G_2 denote the goal of the second speech act in (5.59). The following figure shows a development of the speaker's plan:



In a repair move, the goal of the new speech act comes on top of the goal of the other speech act in order to achieve the older goal.

As a result, self-repair moves can be defined as speech act which belong to an updated version of the plan to achieve the goal of the original speech act and which support its former speech act:

(5.60) SELF REPAIR MOVE: A speech act E' is a *self-repair move* for a speech act E if it is a new action in the reconsidered version of the speaker's former communicative plan to achieve goal G_E of E and if $\text{SUBGOAL}(G_E, G_{E'})$.

Definition (5.60) does not only cover self-repair moves on the content level as for example in (5.59). According to (5.60), speech acts that try to solve problems on the acceptance level can also be regarded as repair moves. Take for example the following *Motivation*:

(5.61) Eat your carrots. Carrots are good for you.

In (5.61), the speaker does not believe that the performance of the first speech act causes the intended result. Therefore, he gives additional reasons for the hearer to act as expected. According to the definition in (5.60) the second speech act can be classified as a self-repair move.

5.4.2 Precondition for Repairs

So far, self-repair moves have been regarded as realizations of plan updates but it was not clarified how the speaker comes up with the idea of updating the communicative plan in the first place. This section targets this problem, by elaborating on the following idea. Recall that section 4.2 already discussed necessary conditions for grounding: A speaker tries to achieve a certain goal with the performance of an utterance. However, the speaker knows that the success of an utterance depends on the hearer of the utterance. In particular, the speaker knows that the utterance has to fulfill certain grounding conditions, as repeated in (5.58), in order to be able to be successful. If the speaker now

believes that at least one of these conditions does not hold, he has to recapitulate his actions and adjust his plans.

The conditions given in (5.58) are all necessary conditions for an utterance to be accepted. Acceptance, however, is the final goal of a speech act and, therefore, the definition (5.60) of self-repair also covers repair moves on the acceptance level. In order to be able to talk about plan updates for solving problems on the acceptance level, the list in (5.58) will be extended by the following state 4.:

(5.62) CONDITIONS FOR GROUNDING - EXTENDED: Let A be a speaker of U and B the hearer.
State 4 : B is willing to accept U.

The idea for the preconditions that lead to repair moves is now quite simple. The speaker knows that it is necessary to achieve all states 1. – 4. for a successful common ground update. Hence, if the speaker does believes that there is one condition which doesn't hold after the performance of a speech act, the speech act is problematic and the speaker has to do something to make this condition hold! This means that the speaker has to perform a plan update for achieving his goal, which by definition (5.60) is a self-repair move.

Section 4.2 already formalized the conditions above. (5.63) repeats this formalizations:

(5.63) CONDITIONS FOR GROUNDING: Let $E : \text{SPEECH ACT}(A, B, U, K)$ be a speech act of A towards B with content K and phonological realization U. Let $\psi_0, \ldots, \psi_3$ denote the following propositions:

0. $$\psi_0(E) = \left[\begin{array}{l} S_B \\ \hline \neg S_B : \text{BEL}\left(B, \left[\begin{array}{l} U'\ K' \\ \hline E : \text{SPEECH ACT}(A, B, U', K') \end{array}\right]\right) \end{array}\right]$$

1. $$\psi_1(E) = \left[\begin{array}{l} S_B \\ \hline S_B : \text{BEL}\left(B, \left[\begin{array}{l} U'\ K' \\ \hline E : \text{SPEECH ACT}(A, B, U', K') \end{array}\right]\right) \end{array}\right]$$

2. $$\psi_2(E) = \left[\begin{array}{l} S_B \\ \hline S_B : \text{BEL}\left(B, \left[\begin{array}{l} U'\ K' \\ \hline E : \text{SPEECH ACT}(A, B, U', K') \end{array}\right]\right) \\ U = U' \end{array}\right]$$

3. $$\psi_3(E) = \left[\begin{array}{l} S_B \\ \hline S_B : \text{BEL}\left(B, \left[\begin{array}{l} U'\ K' \\ \hline E : \text{SPEECH ACT}(A, B, U', K') \end{array}\right]\right) \\ U = U',\ K = K' \end{array}\right]$$

$$4.\quad \psi_4(E) = \left[\begin{array}{l} S_B,\ S'_B \\ \hline S_B : \textsc{bel}\left(B, \left[\begin{array}{l} U'\ K' \\ \hline E : \textsc{Speech Act}(A, B, U', K') \end{array}\right]\right) \\ U = U',\ K = K' \\ S'_B : \textsc{Pursue}(B, K \in CG) \end{array}\right]$$

In the following, these abbreviations will be used in order to model the speaker's assumptions on the hearer's mental states. Note that the list above also contained a proposition ψ_0. This proposition basically states that the hearer is ignorant about the speech act. It plays a slightly different role than the other conditions of the list since it basically states that communication did not work at all. Nevertheless, it will turn out to be very handy to have ψ_0 for modeling problems in communication. This proposition models the case where the target of a speech act is not engaged in communication at all and this occurs quite frequently of course. However, it has to treated with a little bit more care then the other conditions. For example, it does not fit into a sequence of implications: for all $i = 2, 3, 4$, the implication $\psi_i \Rightarrow \psi_{i-1}$ holds. If one assumes that interlocutors know about this chain of implications, also the implications $\textsc{bel}(A, \psi_i) \Rightarrow \textsc{bel}(A, \psi_{i-1})$. But, ψ_1 does not imply ψ_0. This is of course just a theoretical problem which can be solved by introducing strict conditions.

In order to be able to model the situation in which an agent is in state i but not in state $i+1$, it is useful to have the following *strict conditions* on grounding.

(5.64) Strict Conditions for Grounding:

0. $\widehat{\psi_0} = \psi_0$
1. $\widehat{\psi_1} = \psi_1 \wedge (U \neq U') \wedge (K \neq K') \wedge \neg S'_B : \textsc{Pursue}(B, K \in CG)$
2. $\widehat{\psi_2} = \psi_2 \wedge (K \neq K') \wedge \neg S'_B : \textsc{Pursue}(B, K \in CG)$
3. $\widehat{\psi_3} = \psi_3 \wedge \neg S'_B : \textsc{Pursue}(B, K \in CG)$
4. $\widehat{\psi_4} = \psi_4$.

These conditions are strict in the sense that $\widehat{\psi_i}$ doesn't imply $\widehat{\psi}_{i-1}$. In particular, an agent satisfying $\widehat{\psi_i}$ is in state i of the list in (5.63) but not in state $i-1$ anymore.

As already pointed out, the conditions in 1.-4. are known to be necessary conditions for grounding. Therefore, if one of these conditions is believed to be not true, it has to be repaired. This is again assumed to be common knowledge among all language users and will be formulate as a principle:

(5.65) PRECONDITION FOR REPAIRS:
Let A be the speaker and B the hearer of a speech act E : SPEECH ACT(A, B, U, K). If there is an $i \in \{1, \ldots, 4\}$ such that A believes that $\neg\psi_i(E)$ then A has to repair E:

$A\ B\ T\ T'\ E\ S$	
AGENT(A), AGENT(B)	
$E \subset T, T \prec T'\ T' \subset S$	
E : SPEECH ACT(A, B, U, K)	
S : BEL $(A, \neg\psi_i(E)) \Rightarrow$	E'
	$E \prec E'$
	E' : SPEECH ACT(A, B, U, K)
	SUBGOAL$_A(G_E, G_{E'})$

The formulation of principle (5.65) uses the notational convention that the goal of a speech act E is denoted G_E.

So far, self-repair moves were introduced as plan updates. It was discussed which conditions might lead to the fact that the speaker comes to the conclusion that a self-repair move is necessary. Since it is assumed that self-repair moves are produces in order to solve a possible grounding problem, the grounding conditions from Clark and Brennan have been reformulated as propositions in the formal framework of this book. Note that in the formulation of the two phases of grounding in (4.31) of Clark and Brennan, it is assumed that the speaker produces an utterances under the assumption that the hearer produces evidence that the utterance was understood or even accepted. For an analysis of self-repair moves it will be assumed that the hearer does not produce any evidence for anything, neither for misunderstanding nor for rejecting the utterance. This means that, in the case of self-repair moves, the speaker has to make own assumptions concerning the grounding of his speech acts.

So far, the definition and the general set up for analyzing self-repair moves have been established. The rest of this section will be concerned with the question, of how a speaker concretely reacts if he believes that one of the conditions 1-4 does not hold. For this analysis, the following cases can be distinguished: *First – Repairs beneath the Content Level:* Such repair moves are performed in order to solve problems on stage 0, 1, or 2. In these cases the content of the speech act did not even had a chance to be recovered by the hearer. Therefore, repair moves for these stages are referred to as repair moves beneath the content level.
Second – Repairs on the Content Level: These are self-repairs for problems on stage 3, which concerns the content of a speech act.
Third – Repairs on the Acceptance Level: Finally, repair moves on the acceptance level have the purpose to increase the hearer's willingness to accept the speech act. If one is interested in rhetorical relations, these repair moves are the most interesting since many types of rhetorical relations are realizations of repair moves at the acceptance

level.
The rest of this section aims at analyzing these different types of repairs. In particular, the question will be discussed which relations can be predicted in cases where the speaker believes that there are problems in one of these states.

5.4.3 Repairs beneath the Content Level

Repair moves beneath the content level repair the most basic obstructions for successful communication. They occur in cases where the hearer is completely ignorant towards the performed speech act or did not perceive the right acoustic material. There are several possibilities for repairing a speech act with fails on this level. the speaker can repeat his speech act, reformulate it, or he can call for attention. The fact that an obstruction beneath the content level is rather basic in the sense that it is beneath the level of language recognition makes it hard to make any predictions on the particular form of a possible repair move. Nevertheless, some possible solutions for this problems can be identified.

The list of conditions given in (5.63) says that repair moves beneath the content level have the purpose to ensure that ψ_1 and ψ_2 hold. In both cases, it is a possible reaction of the speaker to just speak louder and/or more articulated, as in (5.66):

(5.66) a. Did Bo leave?
b. DID BO LEAVE?

In casese where the speaker simply believes that the hearer is *ignorant* about the given speech act (i.e. the speech act is not even in state 1 yet) it may also be the case that the hearer is actually not engaged in communication. In this case, the speaker can try to get the hearer's attention:

(5.67) a. Did Bo leave? [*short break*]
b. HEY!
b'. Do you listen?
b''. Hello?

In particular, a discourse move like *Hello* seems to occur quite often in conversations on the phone. In cases where the connection is assumed to be bad (or where it is in fact bad) one can ask whether the hearer is still engaged in communication, simply by saying "Hello?".

Due to a lack of better terminology, speech acts like in (5.66.b) will be called *rearticulation*. A speech act as in (5.67.b), will be referred to as a *call for attention*. A call for attention can only be a repair move for $\neg\widehat{\psi}_1$, a rearticulation can repair both obstructions, $\neg\widehat{\psi}_1$ and $\neg\widehat{\psi}_2$.

Calls for attention can be assumed to be a class of speech acts with the particular purpose to make the hearer aware of the fact that the speaker has the desire to communicate: In terms of goals, this means that a call for attention has the goal to make the hearer aware of the speaker's desire to perform further speech acts. This can be formulated as follows:

(5.68) CALL FOR ATTENTION A *Call for attention* is a speech act with which the speaker tries to make the fact that the speaker wants to perform further speech acts an element of the common ground:

$A\ B\ E\ S_E$
$\text{AGENT}(A),\ \text{AGENT}(B)$ $E : \text{ATTENTION CALL}(A, B) \Rightarrow S_E$ $S_E : \text{PURSUE}\Big(A, \boxed{\text{DES}(A, E' : \text{SPEECH ACT}(A, B, U, K))} \in \text{CG}\Big)$

The speaker performs a call for attention when she believes that the hearer is not engaged in communication. But what can be said about the hearer's assumptions on the speaker's goals in cases where a call for attention was performed? In principle, there are two cases for such a situation: 1. either the speaker is right and the hearer actually is ignorant about the speaker's first speech act *or* 2. the speaker's assumption about the hearer is wrong. In this case, the hearer actually perceived the original speech act but did not immediately react as expected. From the point of view of rhetorical relations and from the point of view of the hearer, the first case is not that interesting. The hearer was ignorant about the speaker's first speech act, which means that the hearer cannot establish any relation between the two given speech acts. After the utterance of a speech act like *HEY*, the addressee can just infer that the speaker wanted to get the attention of the hearer. More interesting is the case in which the hearer wasn't actually ignorant about the former speech act. In this case, the hearer is able to make assumptions on the speaker's plans, which means that he is able to recognize a rhetorical relation. In the following, these cases will be considered. Assume again that there is a speaker A and a hearer B and that A performed the two speech acts (5.67.a) and (5.67.b) above. Moreover, assume that B actually was *not* ignorant about E, i.e. assume that actually one of the propositions $\psi_1, \ldots, \psi_4$ is true. This situation can be presented as follows:

(5.69)

$E_1\ E_2\ S_A$
$E_1 \subset S_A$ $E_1 : \text{SPEECH ACT}(A, B, U_1, K_1)$ $E_2 : \text{ATTENTION CALL}(A, B)$ $E_1 \prec E_2$ $S_A : \text{BEL}\left(A, \neg\widehat{\psi}_1(E_1)\right)$ $\psi_i(E_1)$

Given that A called for attention, B can infer that A actually wants to communicate. However, B already knows that A wants to communicate because some ψ_i (for some $i = 1, 2, 3$) is actually true. Since both speech acts are unmarked and since E_1 precedes E_2, the support default can be applied and B can conclude that the call for attention in E_2 supports E_1. Therefore, the following presentation can be derived form (5.69).

(5.70)

$$
\begin{array}{|l|}
\hline
E_1\ E_2\ S_A\ T \\
\hline
E_1 \subset S_A \\
E_1 : \textsc{Speech Act}(A, B, U_1, K_1) \\
E_2 : \textsc{Attention Call}(A, B) \\
E_1 \prec E_2 \\
S_A : \textsc{bel}\ \left(A, \neg\widehat{\psi}_1(E_1)\right) \\
\psi_i(E_1) \\
E_2 \subset T, \\
\textsc{bel}\ \Big(B, \textsc{SubGoal}_A^T(G_{E_1}, G_{E_2})\Big) \\
G_{E_2} = \boxed{\textsc{des}\big(A, E' : \textsc{Speech Act}(A, B, U, K)\big) \in \textsc{cg}} \\
\hline
\end{array}
$$

Hence, B is in a situation where he already knows that A performed a speech act, where he knows that A called for attention, and where he knows that this call is a subgoal of a former goal of A. Therefore, B can infer that the speaker A must have assumed that E_1 is not in state 1:

(5.71) (5.70) $\Rightarrow \textsc{bel}\big(B, S_A : \textsc{bel}(A, \neg\widehat{\psi}_1(E_1))\big)$.

In this situation B is able to recognize A's whole plan: B can infer that A assumed that B wasn't aware of A's speech act which made A call for attention. In particular, B can now interpret this call for attention as a plan update and, hence, a repair move.

5.4.4 Repairs on the Content Level

Repair moves on the content level are speech acts which are performed in order to solve a possible grounding problem on stage 3, i.e. a problem which occurs on the level of the right interpretation of the given acoustic material. Of course, in general there are many reasons why a hearer is not able to come up with the right interpretation of a given utterance. The most basic way to repair a possible misunderstanding seems to be to *reformulate* the problematic part of the utterance. The following two examples are instances of reformulations produced in order to increase the hearer's ability to interpret a particular speech act.

(5.72) a. Did he leave?
b. I mean Bo.

(5.73) a. The piece begins with an anacrusis. (Jasinskaja and Karagjosova, 2015)
b. This is an unaccented note, which is not part of the first full bar.

In order to produce a *Reformulation* as a self-repair move, the speaker has to make assumptions about the problematic points in the speech act in question. In (5.72) the speaker seems to believe that the hearer is not able to conclude to which referent the third person pronoun is supposed to refer. In (5.73), on the other hand, the speaker must have assumed that the hearer does not know about the meaning of the word *anacrusis* and, therefore, the speaker gives a definition. Both cases have in common

that the first speech act is supposed to make sense for the hearer if the hearer replaces the problematic part of the first speech act with the reformulation. In example (5.72) the problematic part is the pronoun *he* and in (5.73) the problematic part is the noun *anacrusis*. Being the problematic part of an utterance in this case means that the hearer is ignorant about the meaning of this very part, For example, the hearer does not know to which referent it is supposed to be referring to. This can be formalized as follows: Assume that U' is a phonological or graphematical realization of some morphological material. Not being able to conclude what U' is supposed to be referring to can be characterized as follows: there is a referent or predicate in the content of the speech act which is supposed to be denoted by U' but where the hearer is ignorant about that denotation. Although predicates can of course also be the target of reformulations, the rest of this section is concerned with the reformulations of referents.

The property of U of being problematic will be denoted by $\mathrm{Y}(U')$. The definition in (5.74) formalizes the assumption on being a problematic part of a speech act. This definition uses the abbreviation $Univ(K)$ for the universe of the proposition K, i.e. the set of referents that occur in K:

(5.74) PROBLEMATIC PART: A subpart U' of an utterance U in a speech act E of speaker A towards hearer B is problematic if B is anticipated to or does not believe that this subpart U' is supposed to refer to a referent in the universe $Univ(K)$ of K:

$$\boxed{\begin{array}{|c|}\hline E,\ U' \\ \hline E : \textsc{Speech Act}(A, B, U, K) \\ U' \subset U \\ [\![U']\!] = x \\ \mathrm{Y}_B(U') :\Leftrightarrow \boxed{\begin{array}{|c|}\hline x \\ \hline E \subset S \\ x \in Univ(K) \\ \neg S : \textsc{bel}(B, [\![U']\!] = x) \end{array}} \\ \hline \end{array}}$$

Given this definition for Y, it is possible to state that a *Reformulation* becomes necessary if the speaker believes that $\mathrm{Y}_B(U')$ holds for some U' of a speech act E_1:

(5.75) $$\boxed{\begin{array}{|c|}\hline E \\ \hline E_1 : \textsc{Speech Act}(A, B, U, K) \\ U' \subset U \\ \textsc{bel}(A, \mathrm{Y}_B(U')) \\ \hline \end{array}} \Rightarrow \textsc{bel}(A, \widehat{\psi}_2(E))$$

In other words, if $\mathrm{Y}_B(U')$ holds for some $U' \subset U$, speaker A can conclude that $\widehat{\psi}_2(E)$ holds. In particular, this implies $\neg\widehat{\psi}_3(E)$. In this case, A can repair the subpart U' of U by reformulating it, i.e. by uttering a speech act which refers to the problematic referent of the first speech act. On that basis, the general structure of a reformulation can be described as follows:

(5.76) $$\textit{Reformulation}(E_1, E_2) = \left[\begin{array}{l} U' \\ \hline E_1 : \textsc{Speech Act}(A, B, U_1, K_1) \\ E_2 : \textsc{Speech Act}(A, B, U_2, K_2) \\ \textsc{SubGoal}_A(G_{E_1}, G_{E_2}) \\ U' \subset U_1 \\ K_2 = \left[\begin{array}{l} x \\ \hline [\![U']\!] = x \end{array}\right] \end{array}\right]$$

Given this characterization of *Reformulation*, it is now possible to describe the structure of its performance:

(5.77) $$\left[\begin{array}{c} E_1,\ T,\ T',\ S \\ \hline E_1 : \textsc{Speech Act}(A, B, U_1, K_1) \\ E_1 \subset T,\ T \prec T',\ T' \subset S, \\ S : \textsc{bel}\left(A, \left[\begin{array}{l} U' \\ \hline \Upsilon_B(U') \end{array}\right]\right) \Rightarrow \left[\begin{array}{l} E_2 \\ \hline E_1 \prec E_2 \\ E_2 : \textsc{Speech Act}(A, B, U_2, K_2) \\ \textit{Reformulation}(E_1, E_2) \end{array}\right] \end{array}\right]$$

So far, the structure of a reformulation was described from the point of view of the speaker. The next question is, how can a hearer recognize a *Reformulation* relation? In order to engage this question, assume that the hearer has a strong assumption about the part of the first speech act that is supposed to be problematic:

(5.78) $$\left[\begin{array}{c} E_1, S \\ \hline E_1 \subset S \\ E_1 : \textsc{Speech Act}(A, B, U_1, K_1) \\ E_2 : \textsc{Speech Act}(A, B, U_2, K_2) \\ \textsc{SubGoal}_A(G_{E_1}, G_{E_2}) \\ S : \textsc{bel}\left(B, \left[\begin{array}{l} U' \\ \hline U' \subset U_1 \\ \textsc{bel}(A, \Upsilon_B(U')) \end{array}\right]\right) \end{array}\right]$$

In a situation as described by (5.78), B can guess that E_2 is meant as a reformulation. However, if the hearer does not have any idea about which part of the first utterance is believed to be problematic this reasoning obviously does not work. It seems to be a general problem that the recognitions of a *Reformulation* only works if speaker and hearer have the same assumptions about the problematic part of the speech act. In cases where the hearer has no clue about the fact that the speaker believes that some part of the former speech act is problematic, the structure in (5.77) can simply not be applied. Take (5.79) as an example where a reformulation can fail in achieving its goal if the hearer does not know which part of the initial speech act is actually anticipated

to be problematic:

(5.79) He loves him. I mean Bob.

In principle it is possible that Bob is supposed to be a reformulation of *he* and of *him*. If the hearer does not have any strong assumptions about which part of the initial speech act is supposed to be problematic, this reformulation may fail. In other words, there are cases where the hearer already has to know which part of an utterance has to be reformulated in order for the reformulation to be successful. This gives the recognition of *Reformulation*s an unpleasant circular structure, from which one can only escape on the basis of the condition that it is common knowledge between hearer and speaker which part of an utterance has to be reformulated. This additional assumption is captured in the formula given in (5.78).

5.4.5 Support for Acceptance

The final stage for grounding is the acceptance of the given speech act. In the terminology of chapter 4, acceptance means that the hearer has adopted the speaker's goal of making the content of the speech act an element of the common ground. Problems on the acceptance level can be repaired in many different ways, for example by giving additional information which might increase the hearer's willingness to accept a speech act.

Repairs on the acceptance level are performed in cases where the speaker assumes that stage 4. of the extended list of grounding conditions in (5.62) is not achieved yet. In the following, some possible repair moves operating on the content level will be presented. All examples are instances of presentational relations.

(5.80) a. Eat your carrots!
b. Carrots are good for you. *Motivation*

(5.81) a. Could you pass me the hammer.
b. It's in the toolbox. *Enablement*

(5.82) a. Bill must have left.
b. His bike is gone. *Evidence*

Recall that the characterization of presentational relations includes the condition that the satellite increases the inclination to believe or to perform the content of the nucleus of the given relation. Translated in the terminology of this book this means that it helps *accept* the content of the nucleus.

If one uses the conditions $\widehat{\psi}$s as formal obstructions for grounding, the conditions that force a repair move on the acceptance level can be presented as follows:

(5.83)

$$\begin{array}{|c|}\hline A\ B\ T\ T'\ E\ S \\ \hline \textsc{agent}(A),\ \textsc{agent}(B) \\ E \subset T, T \prec T'\ T' \subset S \\ E : \textsc{Speech Act}(A, B, U, K) \\ S : \textsc{bel}\left(A, \widehat{\psi}_3(E)\right) \Rightarrow \begin{array}{|l|}\hline E' \\ \hline E \prec E' \\ E' : \textsc{Speech Act}(A, B, U, K) \\ \textsc{SubGoal}_A(G_E, G_{E'}) \\ \hline \end{array} \\ \hline \end{array}$$

This section investigates different realizations of self-repair moves on the acceptance level. In order to do this, it will be helpful to differentiate between two types of speech acts: speech acts that contain a future action of one of the interlocutors and cases of speech act which denote 'mere' propositions. Why should it be handy to differentiate between these types of speech acts? Some presentational relations, as for example *Motivation* or *Enablement*, are directed towards a future action of the hearer. In these cases, the first utterance of an utterance pair refers to a future goal-directed act of one of the agents A or B. Hence, the content K of the speech act $E : \textsc{Speech Act}(A, B, U, K)$ is of the following form:

(5.84)

$$K = \begin{array}{|l|}\hline C, E_G \\ \hline E \prec E_G \\ E_G : \textsc{goal-directed act}(C, G) \\ \hline \end{array}$$

The letter C in (5.84) denotes either A or B: $C \in \{A, B\}$. Such propositions will be referred to as *propositions of action type*. For a proposition of action type K, write $K : \textsc{action}$. If K is of action type with goal-directed act E_G and agent B, write $K = K(B, E_G)$. Propositions which are not of action type, i.e. descriptions of states of affairs in the world that do not make reference to future actions of one of the interlocutors, will simply be referred to as propositions. The distinction between mere propositions and propositions of action type will play a role in the discussion of an agent's ability to recognize certain rhetorical relations.

For the following considerations, assume that A performs an unmarked pair of speech acts (E_1, E_2). In this case, B can assume that E_2 is meant as a support for E_1. Moreover, assume that it is for some reason obvious for B that A believes that $\widehat{\psi}_3$ holds, i.e. B was already able to tell that E_2 is neither meant as a call for attention, a rearticulation, nor as a *Reformulation*. Therefore B can conclude that E_2 is a support on the acceptance level. The problem which is addressed in the rest of this section is the following: how is the hearer able to deduce which kind of relation connects E_1 and E_2?

Three different cases for the content of the first speech act K can be differentiated (without claiming that this list is exhaustive):

(5.85) 1. K is a mere proposition, i.e. does not make reference to a future action of one of the interlocutors.

2. $K = K(B, E_G)$ is of action type with an action of hearer B.
3. $K = K(A, E_G)$ is of action type with an action of speaker A.

In the rest of this section, these cases will be analyzed one by one. *First:* It was assumed that it is the goal of any speech act to make the content of the speech act an element of the common ground. In the case where the content of the speech act is a mere propositions, this implies that the hearer has to believe that very proposition. Therefore, supporting relations in which the first speech act presents a mere proposition have the purpose to ensure that the hearer actually believes the content of the first speech act. Examples for support relations where the first speech act presents a mere proposition are *Evidence*, *Elaboration* and *Explanation*. These relations target at different epistemic problems: *Evidence*, as one can tell by the name, provides further evidence for the truth of a given proposition. Example (5.82), in the following repeated as (5.86) is an example of *Evidence*.

(5.86) a. Bill must have left.
b. His bike is gone.

Evidence functions on the level of propositional attitudes: the speaker explains *why* he is in a certain belief state and hopes that this reason causes the hearer to be in a similar belief state.

In an *Elaboration*, the second speech act gives, by definition, basically the same information as the first speech act, but under a slightly different aspect or in greater detail. As an example for *Evidence*, take the following two examples from Asher and Lascarides, (2003).

(5.87) a. Max had a great dinner last night.
b. He had salmon and devoured cheese.

(5.88) a. Alexis did really well in school this year.
b. She got A's in every subject.

If an *Elaboration* is meant as a repair move, we may conclude that the hearer is assumed to believe the content of the second rather then the content of the first one even thought they basically state the same content.

The term *Explanation* as we use it here refers to a content level explanation, i.e. an explanation which presents causal relations on the level of states of affairs in the world. If it is used as a repair move, we may conclude that the hearer is willing to accept the reason of a state of affairs more easily than its effect. But if the hearer believes the cause, then she also accept its effect. In (5.89), we see an example of an *Explanation*:

(5.89) a. Max can't party today.
b. He is sick.

These relations have in common that the speaker believes that making the hearer believe the content of causes the hearer to believe the content of the first speech act as well. However, at this point no tools have been developed on the basis of which we

can model that the hearer is able to distinguish between these particular relations. Nevertheless, it will be possible to model the hearer's inference of the particular function of these relations. In order to do this, let Λ denote the following formula:

$$\Lambda(E_1, E_2) = \left| \begin{array}{l} E_1 : \textsc{Speech Act}(A, B, U_1, K_1) \\ E_2 : \textsc{Speech Act}(A, B, U_2, K_2) \\ \textsc{SubGoal}_A(G_{E_1}, G_{E_2}) \\ E_1 \subset S_1,\ E_1 \subset S_2 \\ S_1 : \textsc{bel}(A, \widehat{\psi}_3(E_1)) \\ S_2 : \textsc{bel}(A, \textsc{bel}(B, K_2) \twoheadrightarrow \textsc{bel}(B, K_1)) \end{array} \right| \tag{5.90}$$

In cases where K_1 is a mere proposition (i.e. K_1 is *not* of action type), Λ can either be an *Explanation*, an *Elaboration* or an *Evidence* relation.

$$\Lambda(E_1, E_2) \Rightarrow \left\{ \begin{array}{c} \textit{Elaboration} \\ \textit{Explanation} \\ \textit{Evidence} \end{array} \right\} (E_1, E_2) \tag{5.91}$$

Hence, in cases where the hearer knows that the speaker believes that $\widehat{\psi}_3(E_1)$, and where K_1 is mere proposition, the hearer can apply Λ in order to conclude that the pair (E_1, E_2) is either an *Explanation*, an *Elaboration* or an *Evidence*relation.
Second: Examples of speech acts with content of action type describing an action of the hearer are pleas, requests, or orders like for example in (5.80) and (5.81). Such speech acts contain future goal-directed acts or plans of the hearer, which means that they they are only successful if the hearer is willing to adopt a certain goal. A goal, on the other hand, is a certain mental state. This means that the speaker has to make sure that the hearer is at least willing to adopt these mental states. Recall that a goal was characterized as a combination of the following states of affairs:

(5.92) GOALS: A state of affairs G is a *goal* of an agent A if the following conditions hold:

- g_1: A believes that there are actions which can be performed by A and which cause G to hold.
- g_2: A desires G to hold in the future.
- g_3: A intends to perform some of the actions of g_1.

For the following considerations, these conditions will be called g_1, g_2 and g_3. In cases where it has to be pointed out that these conditions depend on the agent A, they will be called $g_i(A)$. Again, it is the case that a speaker who believes that one of these conditions do not hold has to do something in order to make this condition hold. On the basis of this assumption, one can classify rhetorical relations as follows: If A believes that g_1 does not hold, A has to give the hearer the information which is needed in order to perform a certain action. In terms of rhetorical relations, this configuration results in an *Enablement*. Example (5.81) is such a relation. The speaker does not believe that the hearer is able to perform the requested action because of a lack of information and,

therefore, provides the missing information that can be used to perform that action.

In cases where A believes that g_2 does not hold, A has to give B some reasons to desire the goal state. In cases where g_3 does not hold, A has to increase the intention to actually perform one of the proposed actions. Both cases result in some kind of *Motivation*. The last two cases are quite similar difficult to tear apart. Even though desires and the ability to bring about the desire with a certain action do not add up to an intention, desires and intentions are certainly intertwined (cp. Searle, 2005): A strong desire may increase the intention to act in a certain way. On the other hand, it is possible that one can intend to perform a certain action without actually having a strong desire towards the result of the action.

An example for a *Motivation* which is directed at increasing the desire towards the goal state is (5.93).

(5.93) a. You should definitely buy that necklace.
b. It makes you so beautiful.

The utterance given in (5.93.b) provides a reason for the hearer to possess the particular necklace. The event of buying the necklace itself is probably rather unpleasant.

An example for a *Motivation* which is directed at the hearer's intention to perform a particular action is the following:

(5.94) a. You should take the bike to go there.
b. The weather is nice and you won't get stuck in a traffic jam.

In the following, *Enablement*s and *Motivation*s will be classified on the basis of the speaker's assumption about the hearer's mental states. In order to do this, let Ψ_i denote the following formula:

$$\Psi_i(E_1, E_2) = \left[\begin{array}{l} E_1 : \textsc{Speech Act}(A, B, U_1, K_1(B, E_G)) \\ E_2 : \textsc{Speech Act}(A, B, U_2, K_2) \\ \textsc{SubGoal}_A(G_{E_1}, G_{E_2}) \\ \textsc{bel}(A, \widehat{\psi}_3(E_1)) \\ S : \textsc{bel}(A, \neg g_i(B)),\ E_1 \subset S \end{array}\right] \tag{5.95}$$

The former discussion showed that the following implications hold:

(5.96) 1. $\Psi_1(E_1, E_2) \Rightarrow \mathit{Enablement}(E_1, E_2)$.
2. $\Psi_i(E_1, E_2) \Rightarrow \mathit{Motivation}(E_1, E_2)$, for $i = 2, 3$.

However, it requires again a lot of world knowledge to guess which kind of information increases an agent's desires, an agent's intentions, or even an agent's ability to perform a certain action. The given tools make it, nevertheless, possible to model the hearer's ability to infer that a pair of speech acts performed under the conditions above is meant as a *Motivation* or an *Enablement*.

How can a hearer make guesses about the intended relation? Take a situation in

which the hearer was already able to tell that the speaker believes that there is a problem on the content level of a given speech act that has to be repaired. Given that the content of that speech act is of action type, the hearer can assume that the speaker wants to repair some conditions that have to hold for the hearer in order to perform a goal-directed act. This is summarized in the following presentation:

$$\text{(5.97)}\qquad \textsc{bel}\left(B,\left[\begin{array}{l} E_1 : \textsc{Speech Act}(A,B,U_1,K_1(B,E_G)) \\ E_2 : \textsc{Speech Act}(A,B,U_2,K_2) \\ S : \textsc{bel}(A,\widehat{\psi}_3)),\ E_1 \subset S \\ \textsc{SubGoal}_A(G_{E_1},G_{E_2}) \end{array}\right)\right) \Rightarrow \textsc{bel}\left(B,\left[\begin{array}{l} S' \\ \hline E_1 \subset S' \\ \textsc{bel}(A,\neg g_i(B)) \end{array}\right]\right)$$

On the basis of his assumptions in (5.97) and the conventions given in (5.96), the hearer is now able to conclude that the two speech acts are either meant as an *Enablement* or a *Motivation*.

Third: Typical examples of speech acts where the content of the speech act describes a future action of the speaker are promises and threats. Support relations occurring after such a speech act are similar to relations after a speech act that is supposed to make the hearer act. However, in these cases, the roles of the hearer and the speaker are mirrored. In a case where a speech act that contains a future action of the speaker is problematic, the speaker has to make sure that the hearer actually believes that the conditions for performing a goal-directed act hold. This means, that in this case the speaker has to give some evidence or explanation for the fact that the problematic conditions hold. The difference to the case, where the first speech act presented an action of the hearer is that the speaker of course knows about his own mental states and abilities and just needs to give explanations or evidence for his individual propositional attitudes. For the following considerations, let Φ_i denote the following formula:

$$\text{(5.98)}\qquad \Phi_i(E_1,E_2) = \left[\begin{array}{l} E_1 : \textsc{Speech Act}(A,B,U_1,K_1(A,E_G)) \\ E_2 : \textsc{Speech Act}(A,B,U_2,K_2) \\ \textsc{SubGoal}_A(G_{E_1},G_{E_2}) \\ \textsc{bel}(A,\widehat{\psi}_3(E_1) \\ S : \textsc{bel}\left(A,\left[\begin{array}{l} S' \\ \hline S \subset S' \\ S' : \textsc{bel}(B,\neg g_i(B)) \end{array}\right]\right),\ E_1 \subset S \end{array}\right]$$

A relation of type Φ_1 is caused by the fact that speaker A believes that hearer B questions the speaker's ability to perform the given action. An example of such a relation is the following:

(5.99) a. Tonight, I'll make coq au vin.
b. I learned that in the cooking class yesterday.

In a relation of type Φ_2, the hearer explains or gives evidence for why he has a certain desire. The following example shows a case where the speaker explains why he desires

the outcome of a given action.

(5.100) a. Tomorrow, I'll start exercising.

b. I want to fit in that suit again.

Relations of type Φ_3 explain why the speakers actually intends to perform a given action. In the following example, the speaker provides additional knowledge for the fact that it is for him no actual problem for him to perform a future action and therefore he is likely to perform it.

(5.101) a. I'll get you some milk.
b. I was going for the supermarket anyway.

Although the relations of type Φ_1, Φ_2m and Φ_3 are slightly different with respect to the question of which problem they want to address, they can all be classified as some kind of *Explanation*. At this point, the problem remains that so far there are no tools which make it possible to model the hearer's ability to differentiate between the different Φ_is. But given a Φ_i, it is easy to understand a hearer's recognition of an *Explanation*: Given that the hearer knows, that the speaker performs a repair move on the acceptance level, the hearer is able to conclude that the speaker believes that one one of the $g_i(A)$'s do not hold. Hence, the speech acts E_1 and E_2 are connected by one of the Φ_i, which means that they are related by an *Explanation*.

Summary of this section: This section was concerned with different types of self-repair moves and with the question under which circumstances what kind of self-repair moves can be predicted. It was assumed that self-repair moves occur if the speaker believes that there are some problems for the grounding process for a certain speech act and that problems at different stages of the grounding process produce different types of repair moves and, hence, different types of rhetorical relations. In particular, this section was interested in problems on the acceptance level, since such problems cause the realization of interesting rhetorical relations such as certain presentational relations. The methods an results of this section are rather speculative and are meant more as a possible road map for a more fine-grained investigation of rhetorical relations than an exhaustive analysis of all possible repair moves.

5.5 Summary

It was this chapter's general goal to integrate rhetorical relations into the theory of actions, plans, and speech acts which was developed in the preceedings chapters. In order to do this, notions of communicative plans of communicative subordination and coordination have been established. The notion of communicative subordination is based on the idea of subgoals which have been developed in chapter 3. The notion of communicative coordination on the other hand, applies to speech acts that can be

identified as belonging to the same communicative plan but which are not communicatively subordinated. After having established these definitions, it was discussed whether the notion of a communicative plan provides a satisfactory way to establish a notion of discourse coherence and it was proposed that the notion of a communicative plan provides at least a better understanding of coherence. On the basis of the common sense principles which have been developed so far, it was possible that the default case for a relation between unmarked utterances is the support relation. However, since not all relations are support relations, it was discussed which linguistic means can be used for overwriting this default. The idea developed in the relevant section was the following: a speaker can only presuppose content if she believes that this content is already an element of the common ground. This, however, means that the goal of a former speech act containing the presupposed content was already achieved. In this case, the speaker does not have to support the former speech act any more which means that the new speech act does not have to support its former speech. The final section was concerned with the realization of self-repair moves. Such moves have been defined as realizations of updates of communicative plans. It was also possible to show that in some cases certain rhetorical relations can be explained as such self-repair moves.

Chapter 6

Prominence and the Right Frontier

This final chapter is concerned with two major issues: introducing the notion of prominence and explaining the impact of discourse structure in terms of rhetorical relations on the prominence status of referents. The notion of prominence is defined by the criteria that it is a) relational, b) dynamic and c) that it attracts linguistic operations. This book, focuses on the notion of discourse prominence, which applies primarily to discourse referents, such as for example individuals, events, and states. The notion of prominence generalizes over several other notions which have been concerned with discourse structure, such as salience, topicality, givenness, nuclearity, etc. (Jasinskaja, Chiriacescu, et al., 2015).

This chapter is in particular interested in the impact of rhetorical relations on the prominence status of discourse referents. In section 6.1, the general notion of discourse prominence will be introduced and discussed. Moreover, this section will show that speech acts which pursue goals that are not achieved yet tend to be better anchor points for the attachment of new discourse material than speech acts which serve goals that are already achieved. Moreover, it will show that referents which occur in such speech acts are more prominent than referents which only occur in speech acts with already achieved goals. In that sense, a speech act that did not achieve its goal yet can be regarded as a prominent speech act since it attracts further linguistic operations. Section 6.2 on the other hand, will be concerned with the connection of prominence to the right frontier of a discourse graph. This section proves the claim that referents which occur in the right frontier of a discourse graph are more prominent (in the sense that they can easily be picked up by a third person pronoun) than other referents. Moreover, this section shows that the right frontier of a discourse graph consists exactly of the speech acts whose goals are not achieved yet.

6.1 Prominence in Language

This section is concerned with the introduction of the notion of prominence in language and with the question of how rhetorical relations change the prominence status

of discourse referents. In particular, the effect of the right frontier constraint (cp. section 5.1) on the prominence status of referents will be discussed. Moreover, it will be discussed whether the prominence status of a speech act depends on the status of its goals. Prominence refers to a structure-building principle that is characterized by the assumptions that it is a) relational, b) dynamic and c) that it attracts linguistic operations (von Heusinger and Schumacher, 2019).

In 6.1.1, the notion of prominence (and in particular the notion of discourse prominence) will be introduced in more detail and it will be shown that prominence is actually a generalization of former approaches to discourse. Moreover, this section discusses how prominence can be applied to make predictions about certain discourse phenomena, such as pronoun resolution and predictions about possible anchors for perspective taking. In 6.1.2, we will finally discuss the impact of rhetorical relations on the prominence status of referents. Moreover, this section proposes a connection between the prominence status of referents and the status of the goal of the speech act in which the referents occur. To be already more precise, it proposes that referents occurring in speech acts whose goals are not achieved yet are more prominent than referents who only occur in speech acts which already achieved their goals.

6.1.1 Discourse Prominence

This section wants to establish the notion of discourse prominence. Other linguistic phenomena, which can also be captured by the general term prominence in language, such as for example prosodic prominence, will not be approached in this section.

The notion of discourse prominence captures the intuition that at a given point in discourse, some referents (such as events or individuals) seem to "stand out" and are more important than other referents. It also captures the intuition that referents that seem to be more important than others are more likely to be picked up by anaphoric devices such as third person pronouns. In this case, discourse prominence targets the following problem: If the discourse contains more than one individual with the same grammatical gender, third person pronouns are in principle ambiguous. However, it is more likely that a third person pronoun resolves to the more prominent competitor. Speakers, therefore, choose a referential expression on the basis of the prominence status of the referent they want to refer to.

This section introduces several phenomena where the notion of prominence can be applied and, moreover, discusses some notions of linguistic theories which can be regarded as predecessors to prominence in language. In particular, givenness theory and centering theory can be regarded as theories which already contain the idea of prominence without making it explicit. The thematic role of a referent and how it influences its prominence status will be discussed. Finally, it will be shown how the prominence status of a referent changes its ability to serve as an anchor for perspective taking in narrative texts.

Prominence in discourse is assumed to be a structure-building principle on the basis of which interlocutors plan further discourse moves on the one hand, and, on the other hand form expectations about the content of further discourse moves. Moreover, the notion of prominence captures the dynamic nature of discourse, since it assumes

that the prominence status of discourse entities may shift in time. On the basis of this intuitive discussion, discourse prominence can be characterized by the following three criteria (Jasinskaja, Chiriacescu, et al., 2015; Himmelmann and Primus, 2013; von Heusinger and Schumacher, 2019):

(6.1) DISCOURSE PROMINENCE:

i. Discourse Prominence is a partial ranking of discourse elements of the same type.
ii. The ranking is updated as the discourse is progressing.
iii. Prominent units act as structural attractors for linguistic operations (such as for example the choice of referential expressions).

The notion of discourse prominence captures several approaches to language concerned with describing the fact that certain elements of the discourse are more in the center of attention, such as for example *Centering theory* or *Givenness theory* (Averintseva-Klisch, 2013, Chap. 8). In the following, the idea of these theories will be outlined and it will be shown that discourse prominence can be regarded as a generalization of these theories. Moreover, it will be shown that some disadvantages of these theories can be overcome on the basis of a notion of prominence in language.

Centering theory captures the intuition that some referents are more in the "center of attention" than other referents. These referents are more likely to be picked up with an short referential expression (such as for example a third person pronoun) than referents which do not stand as much in the center of attention. In its original version, centering theory, (Grosz, Joshi, and Weinstein, 1995), assumes that the set of referents in any utterance in a discourse constitutes an ordered set of elements which are potentially be picked up by a referential expression in the next utterance. The elements of this set are called the *forward-looking centers* of the utterance and it is assumed that the highest ranked forward-looking center is most likely to be referred to by a short referential expression in the next utterance. Grosz, Joshi, and Weinstein, (1995, p. 214) assume that the ranking in the forward-looking centers is given by their grammatical roles (subjects are higher ranked than objects which are higher ranked then other referents of the sentence). The ranking in the forward-looking center of an utterance provides an example for a prominence ranking since it is a ranking that shifts with the performance of any utterance and it attracts further linguistic operation in the sense that it makes predictions about the upcoming choice of referential expressions. However, as already explained in section 5.2.2, centering theory only makes claims about the local coherence of discourse and breaks down in cases where all chains of referential expressions break up, i.e. in cases where no element of the forward-looking centers occur in the new utterance. This disadvantage can be overcome by a prominence-based account to choices of referential expressions. Because of its dynamic character, prominence is not limited to the description of the local structure of discourse.

Givenness theory (Gundel, Hedberg, and Zacharski, 1993), on the other hand, assumes that the familiarity (or givenness) of a given referent influences the choice of a referential expression which is supposed to refer to that referent. The authors proposed a givenness hierarchy on the basis of which one can predict the choice of a particular referential expression. The following table shows this givenness hierarchy and which

referential expressions is predicted. It is ordered from the most given referent to the least given referent.

(6.2) GIVENNESS HIERARCHY: (Gundel, Hedberg, and Zacharski, 1993, p. 275)

TYPE OF GIVENNESS	REFERENTIAL EXPRESSION CHOICE
in focus	*he, she, it*
activated	*that, this, this N*
familiar	*that N*
uniquely identifiable	*the N*
referential	indefinite *this N*
type identifiable	*a N*

Givenness also captures the prominence criteria in the sense that it provides a prominence ranking of referents that influences the choice of referential expressions. However, in contrast to centering theory and in contrast to the characterization of prominence in (6.1), givenness does not depend on possible competitors (von Heusinger and Schumacher, 2019, p. 123). Therefore, givenness cam be regarded as a prominence-lending feature but does not capture the dynamic aspects of discourse.

So far, two theories concerned with descriptions of similar phenomena in discourse have been discussed. These theories try to model similar phenomena as prominence. However, both theories seem to fail at a given point: centering theory is limited to the description of local phenomena in the sense that it only takes adjacent sentences into account. Givenness theory, on the other hand, does not take into account that any familiarity feature of referents can only be described in their relation to other referents. However, both theories, implicitly, already proposed prominence-lending features. Centering theory proposes that grammatical role (subject $>$ object $>$ others) of the referring expression in a sentence has an impact on the prominence status of that referent: higher ranked referents are more likely to be talked about in the next sentence and are more likely to be picked up by a short referential expression. This type of prominence lending feature can be referred to as *syntactic prominence* (Jasinskaja, Chiriacescu, et al., 2015). Givenness tries to rank referents by familiarity. In that sense, more familiar referents are more prominent then less familiar referents.

Next to the syntactic role, there are other syntactic constructions, such as for example cleft constructions that increase the prominence status of the referent they are referring to. Foraker and McElree, (2007) for example, provided evidence for the claim that clefting increases the strength of the mental representation of the clefted referents. Givenness theory, on the other hand, licenses the possibilities to pick up referents with certain referential expressions based on the familiarity status of these referents. Hence, more given referents attract more linguistic operations.

Next to the givenness status of a referent, the thematic role of a referent as an argument of a predicate influences the prominence status of that referent. This phenomenon can be referred to as semantic prominence, since it focuses on the semantic role of a given referent in the argument structure of a given predicate. Another set of features contributing to the semantic prominence of referents are agentivity features

in the sense of Dowty, (1991). Dowty and, following him, Primus, (2010) pursued the idea that verbs select agentivity features and pass them over to their arguments. In a sentence like *John kicks Jones.*, John is highly agentive: he causes a kicking event, he does it volitional, he moves something with the kicking, event and so on. This makes John a *proto-agent* in Dowty's terminology. Jones, on the other hand, is highly involved in the kicking event, but he is just non-volitional involved which makes him much less agentive in this example. In order to characterize proto-agentivity, Dowty proposed the following list of agentivity features:

(6.3) CONTRIBUTING PROPERTIES FOR THE AGENT PROTO-ROLE:
- a. volitional involvement in the event or state
- b. sentience (and/or perception)
- c. causing an event or change of state in another participant
- d. movement (relative to the position of another participant)
- e. exists independently of the event named by the verb

A referent occurring as an argument of a verb that selects all of these features is a prototypical agent. These features also contribute to the semantic prominence status of a referent since it can be shown that referents who combine more agentivity features tend to be better antecedents for third person pronoun resolution than referents with less agentivity features, which means that they attract linguistic operations (Schumacher, Backhaus, and Dangl, 2015).

Next to the fact that third person pronouns are in principle very flexible but show a tendency to pick the most prominent referent, there are referential expressions which seem to have a tendency to pick up the least prominent referent. Examples of such referential expressions are German demonstrative pronouns *der, die, das*, and (more formally) *dieser, diese, dieses*. (Patil, Bosch, and Hinterwimmer, 2019; Patterson and Schumacher, 2020). Example (6.4), taken from Patil, Bosch, and Hinterwimmer, (2019), illustrates the prominence avoidance behavior of demonstratives and contrasts it with the more flexible third person pronoun. *Peter* is the most prominent referent in (6.4) in terms of syntactical and semantic prominence. Given the observation that the demonstrative *der* seems to be prominence-avoiding, it can be predicted that *der* cannot pick *Peter* as an antecedent.

(6.4) Peter$_i$ wollte mit Paul$_j$ joggen gehen, aber er$_{i/j}$/der$_j$ war
Peter$_i$ wanted with Paul$_j$ jogging go but he$_{i/j}$/der.DEM$_j$ was
erkältet.
catch a cold.
'Peter wanted to go jogging with Paul, but he had a cold.'

Another interesting discourse phenomenon which can be approached on the basis of the notion of prominent referents is perspective-taking in narrative texts. Hinterwimmer, (2019) investigated the question of which prominence factors make a protagonist available for serving as an anchor for perspective taking. In a story, there are local prominence features, as for example agentivity, and non-local factors, as for example being the topic of a short story or a paragraph. Hinterwimmer showed that both fac-

tors influence the choice of a referent who serves as anchor for perspective taking and, hence, for perspective speech such as *free indirect discourse*. In particular, he showed that not only local prominence factors can be taken into account for predicting which protagonist may serve as an anchor for free indirect discourse but also global factors. For an illustration, of how local and global prominence factors interact, consider examples (6.5) and (6.6):

(6.5) Mary went towards the entrance of the building. Suddenly a huge guy in a black coat came around the corner. He bumped into her head-on. Angrily, she smacked him in the face with her bag.
   a. #Ouch, how that hurt!
   b. That would teach the dumb jerk to watch out!

In the context of (6.5), Mary is both locally and globally prominent. She is locally prominent since she is the subject and the protagonist with the most agentivity features in the last sentence of the context of (6.5). Moreover, she is globally prominent because she is the discourse topic for the whole text. This makes *the guy* in context (6.5) unavailable as anchor for free indirect discourse. Even though it is reasonable to assume that the exclamative in (6.5.a) is a thought of the guy who got smacked in the face, the discourse is hard to interpret because *the guy* is simply not prominent enough to serve as perspective center.

However, it cannot just be the local prominence factors which can be taken as an explanation for this phenomenon. Consider (6.6) in contrast to (6.5). Similar to Mary in the last sentence of context (6.5), *the guy* combines the most local prominence-lending features in the last sentence of context (6.6). Therefore, one can assume that *the guy* is a potential perspective center in the further discourse. And indeed, (6.6.b) can without any problems be interpreted as thoughts of *the guy*. However, Mary can still serve as perspective center even though she is locally not prominent. This can be seen by the fact that it is unproblematic to interpret the exclamative in (6.6.a) as thoughts of Mary. This observation provides evidence for the fact that both local and global prominence cues influence a protagonist's ability to serve as a possible anchor for perspective taking.

(6.6) Mary went towards the entrance of the building. Suddenly a huge guy in a black coat came around the corner. She bumped into him head-on. Angrily, he smacked her in the face with his bag.
   a. Ouch, how that hurt!
   b. That would teach the old hag to watch out!

So far, the idea of prominence in language as a discourse structure building principle was introduced. It was shown that prominence can be regarded as a generalization of further attempts to analyze discourse structure (as for example, centering theory). Furthermore, it was discussed that the notion of prominence can be used to make predictions about the resolution of referential expressions and about the abilities of protagonists to serve as an anchor for perspective-taking. These are not the only possible applications of prominence. The flexibility of the notion of prominence makes it for

example also possible to apply it to temporal referents, such as events and states, and to model their ability to serve as temporal referent points (Becker and Egetenmeyer, 2018).

So far, the grammatical role of a referent, its agentivity features, and its familiarity status have been considered as prominence-lending features. 6.1.2 will, finally, investigate another prominence lending factor: rhetorical relations.

6.1.2 Discourse Prominence and Subordination

In the following, the influence of rhetorical relations on the prominence status of referents will be discussed. This discussion is based on the right frontier constraint (RFC). The RFC makes predictions about the ability of an utterance to serve as pivot for new discourse material in terms of discourse graphs (cp. section 5.1). According to the RFC, only nodes on the right frontier can serve as possible pivots for new discourse material. Given the further assumption that non-local antecedents of third person pronouns can only be found in the utterance to which the host of the pronoun is related to, one can conclude that there is a connection between rhetorical relations and the prominence status of referents. In (6.7), the assumptions which are needed to come this conclusion are collected:

(6.7) RIGHT FRONTIER CONSTRAINT

a. Only nodes at the right frontier of the discourse graph are accessible for attachment of new discourse material. (Jasinskaja and Karagjosova, 2015)

b. Non-local antecedents of third person pronouns in an utterance can only be in the pivot of that utterance. (Zeevat, 2011)

In order to illustrate how the RFC influences the prominence status of referents, consider again the following example:

(6.8) a. Max$_m$ painted the wall.

b. Chris$_c$ convinced him.

c. He$_{m/c}$ did a good job!

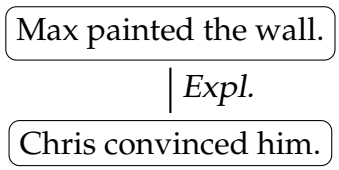


Figure 6.1: A discourse graph for (6.8.a) and (6.8.b).

The speech acts in (6.8.a) and (6.8.b) are related by an *Explanation*. Figure 6.1 shows a simple discourse graph for this discourse. Both nodes of this graph are on its right frontier and, hence, both speech acts can serve as a pivot for new discourse material. This means that the speech act in (6.8.c) doesn't show a strong bias to relate to one of its preceding utterances. In other words, the speech act in (6.8.c) can either be a comment on Max' action of painting the wall or it can be a comment on Chris' convincing Max to do so. On the basis of the further assumption given in (6.7.b) this means that in principle both referents, Max and Chris, can serve as possible antecedents for the third person pronoun *he* in (6.8.c).

This prediction contrasts with the prediction in the coordinated case: The first two speech acts in (6.9) are coordinated by a *Parallel* relation. Figure 6.2 shows a discourse graph for these utterances.

(6.9) a. Max$_m$ repaired the roof.
b. Chris$_c$ painted the wall.
c. He$_{*m/c}$ did a good job!

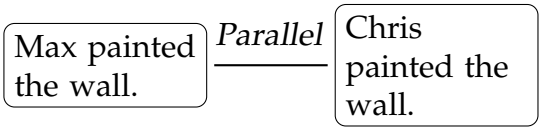


Figure 6.2: A discourse graph for (6.9.a) and (6.9.b).

In figure 6.2, only the node which represents (6.9.b) is on the right frontier of the graph. According to the RFC, only this utterance can serve as a pivot for new discourse material and, therefore, the utterance in (6.9.c) can only relate to (6.9.b). Given the assumption in (6.7.b), this means that it is the only possibility for the pronoun *he* in (6.9.c) to be resolved to Chris.

But what does this mean in terms of prominence of referents? It appears that in the subordinated case (6.8), the older discourse referent is longer available as possible antecedent for a third person pronoun than in the coordinated case(6.9). If the speaker wants to talk about Max in the coordinated case, the speaker has to choose a stronger referential expressions (such as for example a proper name). The subordinated case, on the other hand, allows that more referential expressions refer to Max (even though they are potentially ambiguous). In other words, referents which appear in the first utterance of a subordinating relation attract more linguistic operations and, hence, stay prominent, whereas referents which occur in the first utterance of a coordination (and not in the second utterance anymore) have a tendency to become less prominent. In the terminology of the right frontier, this means that the prominence status of referents decreases once the utterances they appear in are not part of the right frontier anymore. This last formulation also captures the dynamic character of prominence, since the right frontier of a discourse graph changes with every new utterance inf a discourse. This will be made precise in section 6.2.

In the analysis of the examples (6.8) and (6.9), only the right frontier of the discourse graph was used. But it was not taken into account that the discourse graph just consists of two nodes. Therefore, the observations which were made so far are easily generalizable to all utterances of a discourse graph:

(6.10) PROMINENCE AND THE RIGHT FRONTIER: Referents which occur only in utterances that are not at the right frontier of a discourse graph are less likely to be picked up in new discourse material. Therefore, the prominence status of referents decreases once the utterances they occur in are not part of the right frontier anymore.

This means that the right frontier and, hence, rhetorical relations have a direct impact on the prominence status of discourse referents. Of course, rhetorical relations are not the only prominence lending factor, as was shown in the previous section, but they

are an important factor among others. Section 6.2 will basically be concerned with the proof of the claim of the RFC and with the proof of the claim given in (6.10).

However, in this book rhetorical relations are regarded from the point of view of goals of speech acts. Based on the terminology of goals and subgoals of speech acts, section 6.2 shows that the right frontier of a discourse graph consists exactly of the speech acts where the respective goals are not believed to be achieved yet. On the basis of this assumption, one can rephrase (6.10) as follows:

(6.11) PROMINENCE AND OPEN GOALS: Referents which occur in speech acts whose goals are not achieved yet are more prominent than referents which occur only in speech acts whose goals are already achieved.

A possible interpretation of (6.11) is the following. A referent who only occurs in a speech act whose goal is already achieved had certainly a purpose when it was mentioned. However, the goal of that speech act is achieved and the referent is not mentioned anymore. Therefore, one can assume that the referent was maybe important in the context of that speech act but not in the context of the current discourse. Referents that occur in a speech act whose goal is not achieved yet can still be important for the proceeding of the discourse and, hence, are prominent. This claim is also in accordance with classical findings from psychology, such as the *Zeigarnik effect* (Zeigarnik, 1927). According to the Zeigarnik effect, human agents tend to recall activities better that have been interrupted then activities that have been completed. However, if an activity is interrupted then the goal of the activity was not achieved yet and, therefore, non-achieved goals seem to be easier to be recalled then achieved goals. This finding fits with the assumption that referents that occur in a speech act whose goal is not achieved yet are more prominent than others.

Summary of this section: This section was a general discussion of discourse prominence, and of the question of which factors contribute to discourse prominence and how discourse prominence can be applied to make predictions about certain discourse phenomena such as pronoun resolution (local and non-local) and perspective-taking in narrative texts. Moreover, the impact of different types of rhetorical relations on the prominence status of discourse referents has been discussed. Finally, it was proposed that referents which occur in a speech act where the goal of a speech act is anticipated to be not achieved yet have a tendency to stay more prominent. Section 6.2 will be concerned with the question of how this last feature can be translated in a claim about speech acts on the right frontier of discourse trees. Moreover, it is concerned with the derivation of the right frontier constraint on the basis of speech acts whose goals are not achieved yet.

6.2 The Right Frontier

Finally, this section wants to prove the following claim:

(6.12) Discourse Referents which occur in speech act on the right frontier of a discourse tree are more prominent then other discourse referents.

Moreover, it wants to derive the *Right Frontier Constraint* (6.7) on the basis of the theory of goals of speech acts established in this book.

The ingredients for deriving these claims will be provided in this section: §6.2.1 will show why speech acts where it is not obvious that there goals are achieved, still serve as possible anchors for attaching new discourse material. This will be done using the metaphor of the communicative *Table*, where all given speech acts are subject of negotiation but disappear from the table once their goals are achieved. §6.2.2 shows how a discourse tree can be constructed from the material on the table. This part emphasizes the point of view that a discourse graph is basically a reconstructed version of the speaker's communicative plan. The reconstruction process as it is described here has the form of an algorithm that takes speech acts as input and returns a discourse graph. §6.2.3 shows a possible algorithm that takes a discourse graph and returns the right frontier of that discourse graph. Introducing such an algorithm seemed to be necessary, since no description of the right frontier has been provided that goes beyond the intuitive visual impression so far. Finally, 6.2.4 shows that the elements on the right frontier of a discourse tree are exactly the speech acts which are still under negotiation.

On the basis of the insights of these considerations, the right frontier and the claim in (6.12) can be derived as follows:

1. Speech acts with open goals are still accessible as pivots of new speech acts 6.2.1.
2. The right frontier of a discourse tree are exactly the speech acts that pursue goals that are not achieved yet (6.2.2, 6.2.4)
3. Therefore, speech acts in the right frontier are still accessible as pivot for new speech acts (RFC)
4. Since non-local antecedents of third person pronouns can only occur in the pivot of their containing speech act, only referents occurring in the right frontier can be picked up by third person pronouns. (claim (6.12))

Since prominent entities serve as structural attractors, The last point states that referents in the right frontier are more prominent than other referents. This was the claim of (6.10).

6.2.1 Open Goals and Open Problems

This section wants to explain why older speech acts are still accessible for the attachment of new discourse material in the case where it was not made public yet that the goal of these speech acts was already achieved. A speech act where it is not clear among the interlocutors that it achieved its goal is called a *speech act under negotiation*. Given this terminology, one can say that this section aims at explaining the claim that speech acts under negotiation can serve as pivots for new discourse material.

In order to do this, the notion of supporting speech acts and the notion of coordination of communicative goals as they have been introduced in section 5.2 will be used. Recall that a goal of a speech act E is communicatively coordinated to the goal of a speech act E' if E does not support E' and if both speech act belong to the same communicative plan. As an example that contains both coordinated and subordinated

communicative goals, consider the following discourse, taken from Scalzis science fiction novel *Old Man's War*:

(6.13) a. I did two things on my seventy-fifth birthday. E_1
b. I visited my wife's grave. E_2
c. Then I joined the army. E_3

In this example, one can assume that the speech act (6.13.c) is coordinated to the speech act in (6.13.b) in the sense that it does not support it and that they belong to the same communicative plan. However, both speech acts together form an *Elaboration* for the speech act in (6.13.a) and, in particular, support that speech act in the sense that the goal of (6.13.a) can only be achieved if the goals of the two following speech acts are achieved. This assumption can be illustrated with the help of a discourse graph as in figure 6.3:

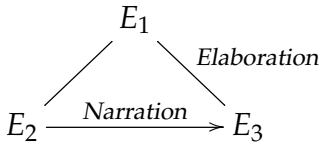


Figure 6.3: A discourse graph for (6.13).

The speech act given in (6.13.a) overtly expresses the communicative plan drawing the two following speech acts together. This makes it rather easy to apply the definition of communicative coordination to the following speech acts (6.13.b) and (6.13.c). The fact that (6.13.c) is attached to (6.13.b) by communicative coordination is due to the fact that it still supports (6.13.a) but does not support (6.13.b). The discussion of this example shows that coordinated attachment can be derived by saying that two speech acts are coordinated if they are adjacent in the discourse graph and if they directly support the same communicative goal. This means that even in the communicatively coordinated case, the new speech act supports another communicative goal to which it can be attached. This idea is crucial for understanding the ideas for discourse attachment that will be presented in the rest of this section. However, the idea above has the following problem: there are cases of speech acts where the highest ranked communicative goal is not overtly expressed, but where speech acts still support this communicative goal. In such cases, the highest ranked communicative goal has to be reconstructed by the interlocutors in order to make coordinated attachment on the highest overtly expressed level possible. This was for example the case in a narrative text as it was discussed in 5.2.2.

After this initial discussion, the question of why speech acts under negotiation can serve as a pivot for new discourse material will be approached in more detail. Recall that a speech act is under negotiation if it is not signaled that the goal of the speech act is achieved. In this case, the speaker still has to be able to support that speech act. Why is this so? In cases where the speaker still believes that the goals of a speech act are not achieved yet, the speaker still wants to achieve the goal of the given speech

act. In other words, acts under negotiation have to be accessible for further support moves. In particular, the most recent speech act in a discourse is always a possible attachment point for further support moves since it has not been signaled that the goal of this speech act is achieved. This observation explains the fact that the speaker can always attach to the earliest speech act of which he believes that it needs support.

However, so far this assumption does not explain how it is possible that one can attach to a speech act in a non-supportive (i.e. in a coordinative) way. For explaining discourse attachment in a coordinative sense, one needs the assumption that any speech act belongs to a communicative plan. Moreover, the definition of communicative coordination given in section 5.2 and the results of the discussion above are needed. 5.2 defined coordinated attachment as an epiphenomenon of attachment in the supporting case. Recall that a speech act E is communicatively coordinated to its predecessor E' if both speech acts directly support a common goal but if E does not support E'. The following considerations will make use of the notions of goal stacks which were developed in section 3.1. Later on, the metaphor of a communicative *Table* in the sense of Farkas and Bruce, (2009) will also be introduced. The *Table* delivers another pictorial way for illustrating this idea of attachment.

The idea of the table as a metaphor for communication will be illustrated by means of the following two examples. But before this idea can be discussed, these examples will be modeled as goal stacks.

(6.14) a. Max repaired the roof.
b. Chris convinced him.

(6.15) a. Max repaired the roof.
b. And Chris painted the wall.

The example in (6.14) shows a support relation; (6.15), on the other hand, shows a coordinating relation. Again, it will be assumed that both examples serve a common goal, which can in this case be paraphrased as Inform.about.construction.side. Based on this assumption, one can construct a goal stack for (6.14) as follows:

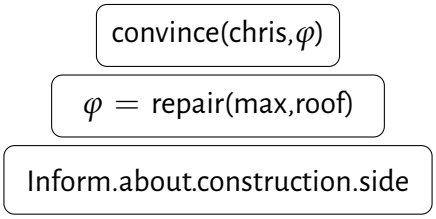


Figure 6.4: A goal stack for (6.14).

This goal stack is a stack of subgoals. In particular, both speech acts are under negotiation. The goal stack for (6.15) given in figure 6.5 looks different. The second speech act in (6.15) explicitly signals that the speaker believes that the goal of the first speech act is already achieved. Therefore, this goal can pop of the goal stack and the new goal can be approached. In particular, after repair(max,roof) popped of the goal stack, it is not

under negotiation anymore and, hence, not available for further attachment of further speech acts.

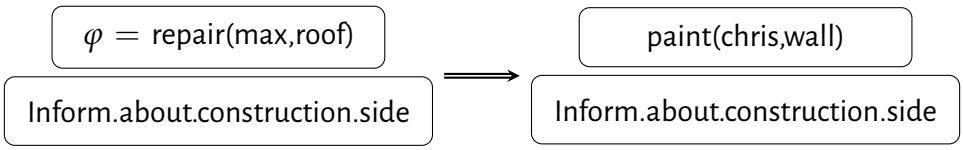


Figure 6.5: Developement of the goal stack for (6.15).

In the following, the *Table* metaphor from Farkas and Bruce, (2009) will be used in oroder to illustrate this phenomenon. Technically, the table is the same as a QUD stack in the sense of Ginzburg, (2012) and Roberts, (1996). The table contains any proposals for a common ground update, which are not yet elements of the common ground, and hence, can still be rejected or accepted. In order to use the terminology of this book, the table simply contains all elements that are under negotiation. The table is interesting because it is a very good metaphor for the assumption that both interlocutors have to decide which content may become an element of the common ground. Literally speaking, the interlocutors are sitting at a table and all speech acts are placed on this table and they have to decide which speech act may achieve their goals. It therefore emphasizes that both interlocutors can decide upon which facts "fall off the table" and enter the common ground.
Remark If one pushes the table metaphor too far, one gets the problem of seemingly having several different levels of communication: A "content" level of communication, which consists of the speech acts that enter the table, and a "negotiation" level where the two interlocutors negotiate about the content of the table. If this is the case, the question arises whether the negotiation level constitutes again its own table. It this is the case, one would have to add a third order table, where the negotiations are negotiated and so on. This would end up in infinite levels of tables. Although this observation poses some theoretical problems (as in many cases where one wants to model phenomena as joint activity or joint attention), it captures some empirical facts, as for example situations where interlocutors communicate about communication. Moreover, it makes it possible to obtain a new view on some special speech acts as for example acceptance moves. These can be regarded as a speech act of one of negotiators who sits at the table and who is willing to accept a speech act which lies *on* the table. However, in the following, problems which may occur if one takes the table metaphor too serious will be avoided.

In the following, the table metaphor will be a little bit adjusted in order to fit the approach of this book. It will be assumed that the table contains all proposals for a common ground update, and at the same time the table represents the overarching communicative goal of the discourse. In other words, it will be assumed that the interlocutors already have agreed upon the particular use of the table in the situation of the conversation. This can be illustrated by stressing the table metaphor a little further (but in another direction as in the remark above). One and the same table can be used

as a dining table, an office table, or as table for playing cards. It has just to be agreed upon its particular use. This means that one and the same table can represent a dining table, an office table, or a playing table.

The same is the case for a conversational table: it can be used to make guest list, to negotiate what one is going to have for supper, or to inform about a construction site. It can be used for all these different purposes, the interlocutors just have to agree upon the fact for what purpose the table is used. If this is done, the conversational table can represent a making-a-guest-list table, a deciding-on-supper table, or a inform-about-construction-site table.. In the examples above, the interlocutors already agreed upon the fact that the table is used to negotiate information concerning a particular construction side. This means that this conversational table represents an Inform-about-construction-site.

In order to be able to illustrate whats happening on the table, the convention will be used that speech acts are presented as *cards* which lie on the table. Moreover, the convention will be used that a card lies on top of another card if the respective speech act supports the other speech act. On the basis of these conventions, one can represent the tables of the discourse in (6.14) after the first two utterances as follows:

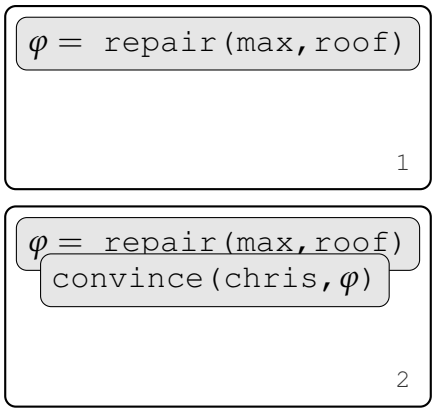


Figure 6.6: The Table after the first utterance of (6.14.a)and after the utterance of (6.14.b).

The coordinated case in (6.15) can also be approached on the basis of the table metaphor. With the performance of (6.15.b), the speaker signals that he believes that the goal of the former speech act is achieved. After the utterance of (6.15.b) the speech act in (6.15.a) is therefore not under negotiation anymore. In other words, it "fell off" the table. This means that, after the utterance of (6.15.b), the table only contains the proposal to update the common ground with the content of that speech act, i.e. it looks as shown in figure 6.7. The table after the utterance of (6.15.a) is not explicitly shown here, as it looks exactly as the first table in figure 6.6.

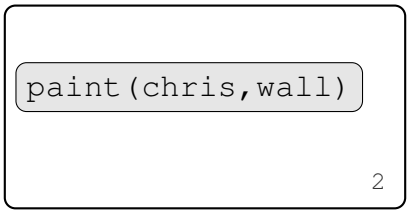


Figure 6.7: The table after the utterance of (6.15.b).

The table is a useful way to illustrate the material under negotiation. However, one has to keep in mind that the material that has left the table is not under negotiation any more but does not necessarily become an element of the common ground. The speaker can signal, that he *believes* that some material has become an element of the common ground which means that the content of this speech act falls off the table. However, with a little bit of effort, both interlocutors still have the possibility to bring the material which already left the table back on the table. For example, the hearer of the discourse given in (6.15) still has the opportunity to bring the claim that Max repaired the roof back on the table by saying something like:

(6.15) c. Wait a minute...Max never repairs roofs. He has acrophobia.

However, note that in this example Max is picked up by his full name. The fact that no utterance which contained Max as a referent was on the table any more makes it harder for Max to be picked up by a third person pronoun.

It was claimed that speech acts which are under negotiation (i.e. on the table) serve as possible attaching points for further discourse material, either because they still have to be supported or because their superordinated communicative goal has to be supported. In that case, the attachment of new discourse material is coordinated.

(6.16) ATTACHMENT IN DISCOURSE: Only a speech act E which is under negotiation can serve as attaching point for new discourse material E'. There are two cases which can be differentiated:

- E' supports E.
- E' signals that the speaker believes that the goal of E' is achieved. In this case E' supports the goal to which E is subordinated and is therefore communicatively coordinated to E.

Given this characterization for the class of speech acts, which can serve as attaching points for new discourse material, this section places its focus on cases where it is actually signaled that the goal of one of the utterances is believed to be achieved. Assume that it is signaled that the goal of a speech act which has got some support is achieved. This automatically means that all its supporting speech acts' goals are also achieved. As a result, if a speech act falls off the table all its supporting speech acts also fall off the table. Take for example the following discourse:

(6.17) a. Max fell.
b. Bill pushed him.

c. They were having a fight.
d. He was sent to the hospital.

In the following, it will be assumed that (6.17.c) explains (6.17.b) and that (6.17.b) explains (6.17.a). Therefore both speech acts (6.17.b) and (6.17.c) support (6.17.a). However, it is reasonable to assume that (6.17.d) attaches to (6.17.a) by a *Narration* or *Result* relation and, hence, that it does not support (6.17.a). But this means that the speaker believes that the goal of (6.17.a) is already achieved, which means that all the subgoals of that goal are also achieved. In other words, they fall off the table. This observation can be generalized to the statement that if it is signaled that a goal of a speech act is achieved, then this speech act and all speech acts which support this goal are not under negotiation anymore. In the following, this claim will be formalized: assume that there is a temporally ordered sequence of speech acts $E_1 \prec \cdots \prec E_n \prec E$ which are under negotiation. Moreover, assume that the speaker signals with the most recent speech act E that he believes that the goal of some former speech act – lets say E_i – is achieved. This, however, means that the speaker believes that all goals of the speech acts $E_i, \ldots, E_n$ are also achieved. Therefore, $E_i, \ldots, E_n$ can be dropped from the table and E can either be meant as a support for one of the speech acts $E_1, \ldots, E_{i-1}$ or it can be communicatively coordinated to E_i. The claim in (6.18) summarizes the observation above. (6.18) uses the convention that E_0 is the empty speech act. In this case E is coordinated to E_1. The assumption, that E and E_1 belong to the same communicative plan is left implicit in in (6.18).

(6.18) LEAVING THE TABLE: Let A be the speaker and B be the hearer of a temporally ordered sequence of speech acts $E_1 \prec \cdots \prec E_n \prec E$. Assume that there is a $i \in \{0, \ldots, n\}$ such that E supports E_i (i.e. $\text{BEL}(B, \text{SUBGOAL}_A(G_{E_i}, G_E))$) and that E does not support E_j for all $j \in \{i+1, \ldots, n\}$ (i.e. $\text{BEL}(B, \neg\text{SUBGOAL}(G_{E_j}, G_E))$). Then the speech acts $E_{i+1}, \ldots, E_n$ are not under negotiation anymore.

The claim of (6.18) is illustrated in figure 6.8. It is assumed in this figure that speech act E signals that the goal of E_2 is achieved and, therefore, that it supports E_1. As a result, E_2 and all speech acts supporting E_2 can fall off the table.

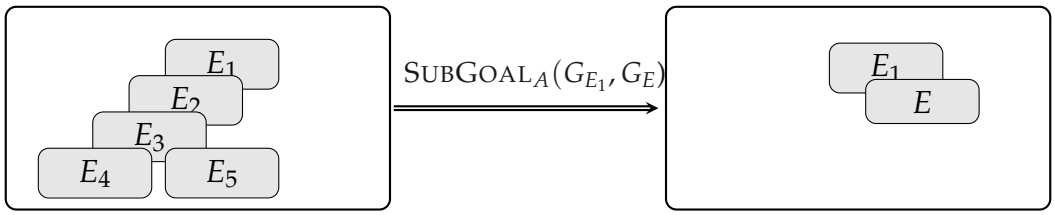


Figure 6.8: A table before and after the utterance of a speech act E, which signals that the goal of E_2 is achieved. E_2 and all its supporting speech act fall off the table at once.

The counterpart of the rule given in (6.18) in Ginzburg's theory is called qud-*downdate* (Ginzburg, 2012, p. 103). The idea of QUD-downdate is the following: once a QUD is answered, this QUD and all QUDs to which the answer is entailed by the former answer can be popped of from the QUD-stack.

So far this section was concerned with a characterization of the set of speech acts which still serve as a possible anchor points for further communication. It was claimed that this set consists exactly of the speech acts which are still under negotiation. Moreover, a criterion was given on the basis of which it is possible to say which speech acts are still under negotiation. 6.2.2elaborates on this criterion. In particular, it will be used to show how one constructs a discourse graph given the knowledge about all the discourse material that is still under negotiation.

6.2.2 From the Table to the Tree

This section is concerned with a construction of discourse graphs on the basis of the terminology of communicative goals. To be more precise, an algorithm will be proposed which models an update of a discourse graph with new discourse material using assumption about the goals of the new speech act. It will be assumed that all discourse moves actually belong to the same communicative plan in order to avoid examples which fall in the category of relations of the type *inform-accident-mention-fruit* (cp. section 5.2.2). This is also in accordance with the discussion of section 6.2.1, where it was assumed that the interlocutors' communicative table already restricts the content which can be negotiated on that table.

A discourse graph can be interpreted as the hearer's reconstruction of the speaker's communicative plan. Somehow, the hearer has to deal with the incoming discourse material and to interpret and integrate it in what he believes is the speaker's strategy. In an ideal case where there is no communicative confusion or misunderstanding, the structure of the discourse tree and the structure of the communicative plan are similar. The discourse tree has to be constructed from the material which was actually under negotiation, i.e. from the material on the table. The hearer has the task to integrate new speech acts in the structure what he believes is the speaker's communicative plan. Therefore, the process of constructing a discourse tree can be divided into at least three phases:

(6.19) PHASES OF A DISCOURSE TREE:

1. The speaker pursues a communicative plan. This plan is realized with the performance of speech acts.
2. As a result, any speech act enters the table after its performance.
3. The hearer has to reconstruct the communicative plan from the material on the table. In this reconstruction process, the speaker has to use cues which have been provided by the speaker and rules such as the rule given in (6.18).

Figure 6.9 is an attempt to illustrate these phases of communication. It is assumed that both interlocutors have the same impression about the content of the table. Moreover, this figure already uses the convention that a subgoal relation is indicated by a vertical line. This convention will also be used for the general construction of discourse trees.

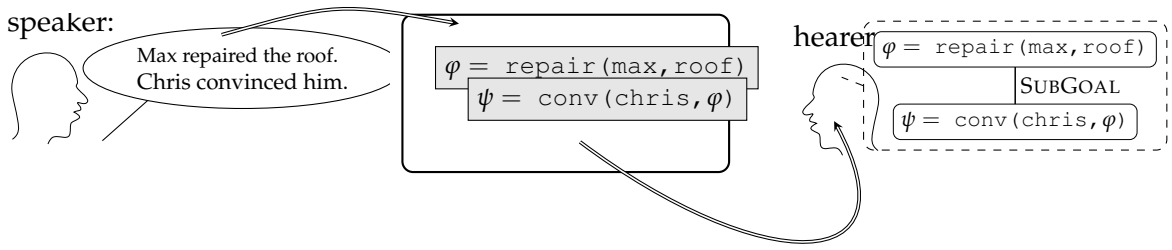


Figure 6.9: The three phases of communication. The hearer realizes some communicative plan, the single speech acts enter the table, and the hearer reconstructs the communicative plan from the material on the table

In an ideal case, the discourse tree and the communicative plan of a conversation are two sides of the same coin. Nevertheless, they differ of course in several aspects. The construction of the communicative plan is a top-down process: the speaker pursues some communicative goal and constructs a plan in order to achieve this goal. The construction of the discourse tree, on the other hand, is a bottom-up process. The hearer can only interpret and integrate step by step new material in the discourse tree, since the speaker has to realize his communicative plan only step by step.

In the following, it will be shown in an informal way how the hearer's interpretation process results in a discourse tree. This interpretation process depends primarily on the hearer's assumptions on causalities but can be guided by the linguistic cues with which the speaker tries to make his plan transparent, such as for example the use of certain discourse connectors. The hearer's construction of a discourse tree is explicated on the basis of (6.20), which is basically the same example as (6.9) but which does not contain anaphoric ambiguity.

(6.20) a. Max repaired the roof.
b. Chris had convinced him.
c. Then they had supper.

Assume that the hearer of (6.20) is at the point after the utterance of (6.20.b). In this situation the speech-acts (6.20.a) and (6.20.b) are still on the table and from this information the hearer manages to construct a discourse tree for these speech-acts:

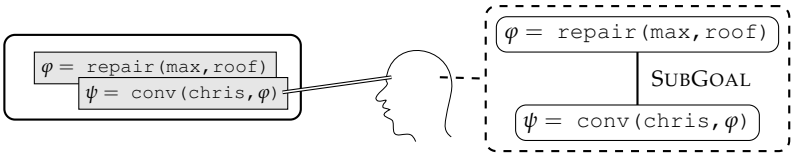


Figure 6.10: The situation after (6.20.b). Both speech acts are on the table and build a stack. This is reflected by the shape of the discourse tree.

Now the new discourse material *Then they had supper.* enters the *Table*. As already explained, this speech act has two functions:

1. it signals that the speaker believes that some former goal is achieved
2. it contributes new information for negotiation.

At this point, the hearer has to guess of which of the former speech-acts the hearer believes that its goal is achieved. It is reasonable to assume that the new speech act rather attaches to (6.20.a) than to (6.20.b). The temporal structure of the example above also supports this interpretation (however, at this point this assumption is stipulated). In the terminology of goals and subgoals, this means that the hearer believes that the new speech act is not performed in order to support (6.20.a) or (6.20.b) and hence can only be attached as a communicatively coordinated speech act to (6.20.a). Figure 6.11 shows the tree which can be constructed from the material given in (6.20).

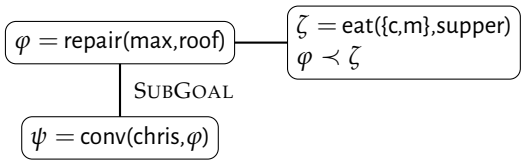


Figure 6.11: The situation after (6.20.c). The new discourse material is coordinated to (6.20.a).

Note, that at the point where the last speech act is integrated in the discourse tree, this speech act is still under negotiation and it is as under negotiation as long as the speaker does not signal that he believe its goal to be achieved.

The rest of this section is devoted to a formal presentation of the algorithm which was used in this example. This algorithm takes an incoming speech act and an already constructed discourse tree as an input and returns a discourse tree which contains the new discourse material. The algorithm depends on the hearer of the discourse and his assumptions about the speaker's communicative goals and world knowledge. In the formulation of this algorithm, this assumption is represented by the fact that it checks for the hearer's beliefs.

In order to be able to represent the fact that one speech act supports another speech act in a tree, discourse trees are assumed to be *directed*. Using direction, a support relation can then be represented by the following convention: if E_1 and E_2 are two adjacent speech acts such that E_2 supports E_1, then the edge between E_1 and E_2 is directed towards E_1. This means, that in cases where E_2 supports E_1 the graph $E_2 \rightarrow E_1$ is a subgraph of the discourse tree. Based on the definition that edges are pairs of nodes, this can also be formulated by saying that the pair (E_2, E_1) is an edge of the discourse graph. In cases of adjacent speech acts where the second speech act does not support the first speech act, the graph will be represented by a subgraph where the edge is directed in the other direction. In this case the graph $E_1 \rightarrow E_2$ is a subgraph of the discourse tree.

The idea of the algorithm is rather easy: if the new speech act does support its preceding speech act, the hearer has to look for the earliest speech act it is subordinated to. In this situation, there are two cases for that speech act: first, the new speech act supports its direct predecessor. In that case it can directly be integrated in the given tree with a "subgoal edge". Or second, the new speech act supports some former speech

act E'. In this case the hearer can conclude that the goal of any speech act after E' is believed to be achieved, and the speech act can be coordinated to the oldest speech act that supports E'. (6.21) shows an informal version of this algorithm:

(6.21) DISCOURSE TREE UPDATE - INFORMAL:

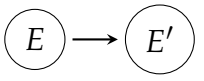
- Input: An ordered list of speech-acts $E_1, \ldots, E_n$ from which the hearer already constructed a discourse tree T.
  A new speech act E which has to be integrated to T.
- Output: A discourse tree T' which contains T as subtree and the new speech act E as a vertex.
- Integrate the new speech E act as follows:
  (i) Check whether the hearer believes that $\text{SUBGOAL}(G_{E_n}, G_E)$. If this is the case, add the vertex E and the edge (E, E_n) to the discourse tree.
  (ii) Otherwise, look for the earliest speech act E_i such that the hearer believes that $\text{SUBGOAL}(G_{E_{i-1}}, G_E)$ and add (E_i, E) to the discourse tree.

Recall that this construction is based on the assumption that a speaker may support his speech acts as long as he believes that their goals are not achieved yet. This means in particular the following: given the assumption that $\text{SUBGOAL}(G_{E_{i-1}}, G_E)$, the hearer can conclude that the speaker believes that the goals of the speech act E_i is already achieved and, hence, by the assumption given in (6.18), that the goals of all speech acts $E_i \ldots, E_n$ are already achieved. Therefore, the new speech act does not have to support E_i anymore, which means that it can be assumed to be communicatively coordinated to E_i.

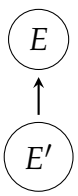
Basically all formats for the visual realization of discourse trees use the convention that nodes of older speech acts are higher and more on the left side of the discourse tree. This convention can be reproduced for the output of the construction above on the basis of the following convention.

(6.22) Let E and E' be two speech act with $E \prec E'$.

1. If E' is coordinated to E, draw the edge (E, E') as a horizontal line.
2. If E' supports E, draw the edge (E', E) as avertical line.



Convention (6.22.1)



Convention (6.22.2)

Before the construction given in (6.21) is formalized, it will be illustrated on a the basis of a further example taken from Asher and Lascarides, (2003):

(6.23)
a. Max had a great evening last night. E_1
b. He had a great meal. E_2
c. He ate salmon. E_3
d. He then won a dancing competition. E_4

Assume that the hearer of (6.23) already constructed a discourse tree for E_1, E_2, E_3 and now has to deal with E_4. The speech act E_4 contains the discourse connective *then* and therefore signals that its goal is *not* a subgoal of the preceding speech act. Therefore, the hearer has to consider older speech acts in order to find a speech act which can be supported by E_4. The hearer has to check whether E_4 could be performed as a support for E_2 first. Given a little bit of world knowledge, he can conclude that this is not the case since, in contrast to having salmon, a dancing competition is normally not a sub event of great meal. Winning a dancing competition, however, can be regarded as evidence for having a nice evening and hence the subgoal relation holds between the two speech acts E_4 and E_2. According to (6.21) this means that the edge $E_4 \rightarrow E_2$ has to be added to the already constructed discourse tree.

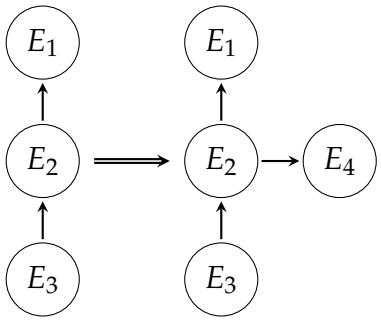


Figure 6.12: The attachment of the new node E_4 to the former discourse tree. The upward pointing arrows indicate a subgoal relation. The rightward pointing arrows indicate communicative coordination.

The algorithm in (6.24) formalizes the algorithm in (6.21). It uses the convention that the set of nodes of a graph T is denoted with $v(T)$ (v for *vertex*) and that the set of edges of T is denoted with $e(T)$.

(6.24) DISCOURSE TREE UPDATE:

Data: New speech-Act E
Discourse tree $T = (e(T), v(T))$ such that
the set of vertices $v(T) = \{E_1, \ldots, E_n\}$ consists of former speech-acts with
$\forall i = 1, \ldots, n-1 : E_i \prec E_{i+1}$
Result: Discourse tree T' with $v(T') = v(T) \cup \{E\}$.

$T' = \emptyset$
if $\text{bel}(B, \text{SubGoal}(G_{E_n}, G_E))$ **then**
  $v(T') := v(T) \cup \{E\}$ and $e(T') := e(T) \cup \{(E, E_n)\}$
else
  $i = n - 1$
  while $\text{bel}(B, \neg\text{SubGoal}_A(G_{E_i}, G_E))$ **do**
    $i = i - 1$
  end
  else $v(T') := v(T) \cup \{E\}$ and $e(T') := e(T) \cup \{(E_{i+1}, E)\}$
end

This section was concerned with the reconstruction of discourse trees from the material on the table. A formal construction rule was presented on the basis of which new discourse material can be integrated in a given discourse tree. The next section, finally, approaches the right frontier of a discourse tree. In particular, it proposes a formal way to obtain the set of the elements of the right frontier of a tree and shows that these elements correspond exactly to the speech acts of the discourse which are still under negotiation. Given the assumption that speech acts which are still under negotiation may serve as attachment points for further communication, this means that this section, finally, derives the right frontier constraint.

6.2.3 The Nodes at the Right Frontier

§6.2.2 developed almost all tools for deriving the right frontier constraint. However, there is still no clear characterization of the elements of the right frontier fo a discourse tree. This gap is filled by the following section which aims at giving a precise description of the elements of the right frontier of discourse trees. It proposes an algorithm which takes a given discourse tree as input and returns the set of the elements of the right frontier of that tree. In 6.2.4, it will finally be shown that the set returned by that algorithm in cases of discourse trees is exactly the set of speech acts which are still under negotiation.

The idea of the algorithm proposed in this section is quite simple. The algorithm uses a discourse tree's additional property that its nodes are also ordered by the time of their appearance. This means, in particular, that any discourse tree T has a "natural" root E, its oldest node that corresponds to the earliest speech act of the discourse. At least in a heuristic way, the algorithm makes also use of the convention that the nodes of supporting speech acts are drawn beneath the supported node. On the basis of these conventions, the algorithm roughly works as follows. It starts at the oldest node E and goes to the right (from the perspective of the reader) as long as possible. Going to the

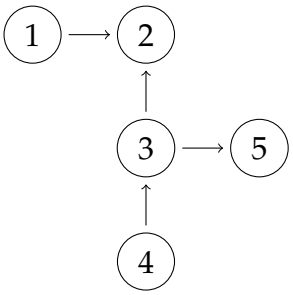
right means that one passes speech acts which occurred later in the discourse and are coordinated to E. If it is not possible to go to the right anymore, add the element where one has to change the direction to RF, the set of elements of the right frontier and go down as long it is not possible to go right and add any element to RF. Do so until a point is reached where one can go right again. Then repeat the whole process:

(6.25) CALCULATING THE RIGHT FRONTIER - INFORMAL:

- Input: A discourse Tree T.
- Output: The set RF of the nodes of the right frontier of T.
- Calculate RF as follows by starting at the oldest node of the tree and do the following:
  - (i) If you are at the youngest node, add it to RF.
  - (ii) As long as you are not at the youngest node, do:
    1. If you can go right, go right.
    2. If not, add the node to RF and go down.

Before a formal version of (6.25) will be given, consider an example to see how this algorithm works. Take the discourse graph given in (6.26) as an input for (6.25) and assume that the nodes of that graph are temporally ordered by their labels.

(6.26)



According to the construction rule in (6.25), this means that one has to start at node 1. and has to go to the right until it is not possible to go any further to the right. In the case of the graph given in (6.26), this means that one has to go to 2., and that one has to add this node to the set RF of elements of the right frontier. Afterwards, one has to go down until one reaches a node where one can turn right again. In the case of (6.26), such a node is already reached at 3. Then one has to go right again for as long as possible. In the case of (6.26), the algorithm already arives at 5. after one step. 5. is already the youngest node, which means that it is included to RF and that this is the node where the algorithm terminates. The three steps are illustrated in figure 6.13. The gray path in this figure indicates the way of the algorithm. The gray colored nodes belong to RF.

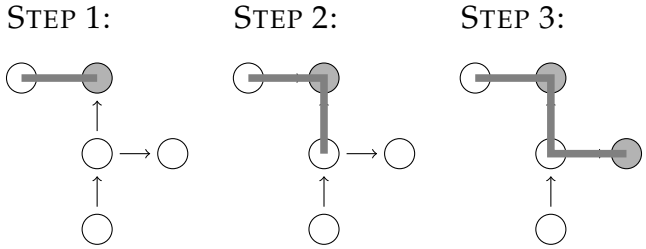


Figure 6.13: Three steps for calculating the right frontier for the graph (6.26).

The algorithm in (6.27) is a formal version of (6.25).

(6.27) CALCULATING THE RIGHT FRONTIER:
Data: A discourse graph $T = (v(T), e(T))$ with $v(T) = \{E_1, ..., E_n\}$ and $E_1 \prec ... \prec E_n$.
Result: The set $RF(T)$ of nodes at the right frontier of T.

```
1  RF(T) = ∅
2  i = 1
3  while i ≤ n do
4  |  if ∃k > i with (E_i, E_k) ∈ e(T) then
5  |  |  i = k;
6  |  else
7  |  |  RF(T) = RF(T) ∪ {E_i}
8  |  |  i = i + 1
9  |  end
10 end
```

This section was devoted to a presentation of the construction rule for the calculation of the right frontier RF of a given discourse graph. §6.2.4, finally shows that the set RF of nodes on the right frontier of a discourse graph consists exactly of the speech acts which are still under negotiation.

6.2.4 Towards the Right Frontier

The description of the form of the right frontier, given in the last section, is the last tool to finally derive the right frontier constraint of a discourse graph. What is left to do is to connect the discourse trees which are constructed by algorithm (6.27) to the material that is still under negotiation, i.e. it has to be shown that the set of elements on the right frontier and the speech acts under negotiation coincide. This section is therefore devoted to the proof of the following claim:

(6.28) Let T be a discourse tree constructed from the material on the table as in (6.24) and let $RF(T)$ be the set of nodes on the right frontier of T. Then $RF(T)$ coincides with the material which is still on the table.

The claim that the nodes on the right frontier and speech acts which are under negotiation coincide will be proven by by induction on the size of the discourse tree. The proof uses the following notational convention: let $E_1 \prec \cdots \prec E_n$ be the speech acts of the discourse. Then T_{E_i} refers to the discourse tree which consist exactly of the speech acts $E_1, \ldots, E_i$. Moreover, as in the algorithm above, $RF(T)$ denotes the set of the elements on the right frontier of the tree T.

Trivial Case: Assume that E is the first and only utterance in a discourse. Then it is under negotiation (since it was not signaled that the speaker believes its goal to be achieved) and the discourse tree $T_E = \{E\}$ consists of just one node. Since this node is the youngest node of the discourse tree it is also an element of the right frontier according to algorithm (6.27),. Therefore, the claim holds true for trees with just one node.
Induction Step: Now assume that the claim in (6.28) holds true for all discourse trees with n nodes, where $1 \leq n$. Assume that there is a sequence $E_1 \prec \cdots \prec E_n \prec E$ of $n+1$ speech acts, and assume that T_{E_n} is already a discourse tree for the speech acts $E_1, \ldots, E_n$. Let $E_{i_1} \prec \cdots \prec E_{i_k}$ be the set of speech acts which were under negotiation before the performance of E. Since T_{E_n} is a tree of size n, the induction hypothesis hold for T_{E_n} and, hence, $RF(T_{E_n})$ coincides with the set $\{E_{i_1}, \ldots, E_{i_k}\}$ of the speech acts which have been on the table before the performance of E. Moreover, the youngest speech act in $RF(T_{E_n})$ coincides with E_n, i.e. $E_{i_k} = E_n$. In this situation there are basically two possibilities for E to be attached to T_{E_n}: first, the hearer assumes that E supports E_n or, second, the hearer assumes that E supports a former speech act. Therefore, both cases have to be considered. In the following, B refers again to the hearer and A to the speaker of the discourse.
First Case: Assume that B believes that E supports E_n, i.e. $\text{BEL}(B, \text{SUBGOAL}_A(G_{E_n}, G_E))$. In this case, all the elements of the right frontier of T_{E_n} (i.e. the speech acts $E_{i_1} \prec \cdots \prec E_{i_k}$) are still under negotiation after E was performed (since they have not fall off the table yet). This is because E does not signal that the goal of some former speech act is achieved and, therefore, all speech acts on the table stay on the table after the utterance of E (cp. claim (6.18)). Therefore, the set of speech acts on the table after the performance of E is the set $RF(T_{E_n}) \cup \{E\}$. Hence, it is left to show that this set coincides with set on the right frontier of the new tree T_E. Before E was performed, E_n was the last speech act of the discourse and, hence, also the speech act where algorithm (6.27) terminated. However, after E has been integrated in T_{E_n}, E sits below the node E_n since E supports E_n. This, however, means that algorithm (6.27) has to take the same way in T as in T_{E_n} since in T there is no earlier opportunity to turn to the right. This situation is sketched in figure 6.14. The gray path indicates the way of the algorithm when it parsed the tree before it was updated with E and the gray bullet indicates where the algorithm terminated in T_{E_n}.

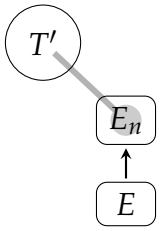


Figure 6.14: A sketch of the case for the termination of algorithm (6.27) in the case where E supports E_n.

Since algorithm (6.27) also passes node E_n in the updated tree T_E, it has to terminate in E. However, until it arrives at E_n, it only selects nodes of $RF(T_{E_n})$ for the right frontier of $RF(T_E)$. Therefore the right frontier of T_E consists exactly of $RF(T) \cup \{E\}$, which was to be shown.

Second Case: In this case, the hearer assumes that the new speech act supports an older speech act. E signaled somehow that the speaker believes that the goal of E_n is achieved, which means that $\text{BEL}(B, \text{SUBGOAL}_A(G_{E_i}, G_E))$ for $i < n$ and $\text{BEL}(B, \neg\text{SUBGOAL}_A(G_{E_j}, G_E)$ for all $i < j < n$: By the description of the material which falls off the table given in (6.18), this means in particular that the speech acts $E_{i+1}, ..., E_{n-1}$ are not under negotiation anymore. Therefore, after E was uttered, only the following speech acts are left under negotiation: $(RF(T_{E_n}) \setminus \{E_{i+1}, ..., E_{n-1}\}) \cup \{E\}$. However, the construction algorithm for discourse trees forces the new node E to be attached right to the older node E_{i+1}. This situation is sketched in figure 6.15.

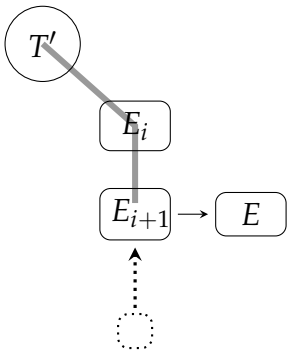


Figure 6.15: A sketch of the case where algorithm (6.27) passes E_{i+1}. Since E is coordinated to E_{i+1}, the algorithm can turn right and terminates at E. The dotted line indicates the speech acts $E_{i+2}, \ldots, E_n$

In the discourse tree T_E, algorithm (6.27) still passes E_i and E_{i+1} but has to turn to the right after E_{i+1} where it terminates in E. Therefore, the right frontier of T_E consists exactly of the speech acts $(RF(T_{E_n}) \setminus \{E_{i+1}, ..., E_{n-1}\}) \cup \{E\}$, which are all the speech acts that are still under negotiation. But this is exactly what was left to show.

Therefore, in both possible cases for attaching E to T, the material still under negotiation coincides with $RF(T_E)$ and this is exactly the claim of (6.28). Note, that the proof also shows that algorithm (6.27) always terminates.

The statement given in (6.28) claims that the right frontier of a discourse tree and the set of elements on the table coincide. Assumption (6.16), on the other hand stated that only speech acts which are under negotiation can serve as attachment points for new speech acts. But the combination of these two results is exactly the statement of the RFC. Moreover, assumption (6.11) of section 6.1.2 stated that discourse referents which occur in speech acts that are still on the table are more prominent than referents that already left the table. On the basis of these two results, the following conclusion can be formulated.

(6.29) THE RIGHT FRONTIER AS PROMINENCE FACTOR:
Discourse Referents which occur in speech acts on the right frontier of a discourse tree are more prominent than other discourse referents.

Summary of this section: This section was concerned with the analysis of the right frontier of a discourse tree. It introduced an algorithm that constructs a discourse tree from the material on the table and an algorithm that constructs the set of elements on the right frontier of a discourse graph. Moreover, it showed that the set of elements of the right frontier of the discourse graph consists exactly of the set of speech acts which are still under negotiation. This means that referents occurring in speech act that lie on the right frontier of a discourse graph are more prominent than referents which do not occur in such speech acts.

6.3 Summary

This chapter was concerned with the notion of discourse prominence, attachment in discourse, and with the question ofhow rhetorical relations can be considered as prominence lending cues. It was proposed that only speech acts which pursue goals that have not been achieved yet may serve as possible pivots for new discourse material. Given the additional assumption that non-local third person pronouns can only be resolved to referents in the pivot of the utterance in which they occur, non-local pronouns are more likely to be resolved to referents that occur in speech acts which pursue goals that have not been achieved yet, i.e. in speech acts which are still under negotiation. In terms of discourse prominence, this means in particular that the prominence status of referents which only occur in speech acts that are not under negotiation anymore decreases in comparison to referents which occur in speech acts that are still under negotiation. Moreover, this chapter investigated discourse trees and how one can construct discourse trees from the material that is still under negotiation. Finally, it investigated the right frontier of the discourse tree. It proposed a construction method for the elements on the right frontier of a given discourse tree and showed that these elements coincide with the set of speech acts that are still under negotiation. On the basis of the assumption that only speech acts which are under negotiation may serve as possible attachment points for new discourse material, this also proves the right frontier con-

straint and the claim that referents on the right frontier of a given discourse graph are more prominent than others.

Chapter 7

Conclusion

This book tried to provide a better understanding for rhetorical relations, subordination in discourse, and some particular discourse phenomena that accompany subordinating rhetorical relations. In order to do that, it approached rhetorical relations from the point of view of the function of a speech act in discourse, rather than from a semantic point of view. It was assumed that human agency is in general built on plans. Performing speech acts is a particular appearance of human agency and, hence, speech acts are organized in plans as well. Given this assumption, the recognition of a rhetorical relation is nothing but a plan recognition task. It was another major issue of this book to provide an explanation for the right frontier constraint (RFC). The RFCformulates the observation that discourse subordination relations work in some way against the recency effect and leave older discourse material longer available for the attachment of new discourse units. This observation is exemplified in (7.1):

(7.1) a. Max_m painted the wall.
b. Chris_c convinced him.
c. $\text{He}_{m/c}$ did a good job!

The two utterances (7.1.a) and (7.1.b) are related by an *Explanation*, i.e. a subordinating rhetorical relation in which case the RFC predicts that the new discourse material in (7.1.c) can relate to both utterances in the former discourse. Given the further assumption that non-local antecedents of third person pronouns can only be found in the pivot of the speech act in which the pronoun occurs, one can see that Max and Chris can serve as non-local antecedents for the third person pronoun *he*. Coordinating rhetorical relations, on the other hand, do not show this behavior, as exemplified by (7.2):

(7.2) a. Max_m repaired the roof.
b. Chris_c painted the wall.
c. $\text{He}_{*m/c}$ did a good job!

The utterances (7.2.a) and (7.2.b) are related by a *Parallel* and, hence, are coordinated. Therefore, the RFC predicts that (7.2.c) can only be related to the most recent speech act, i.e. to (7.2.b). Again, this can be seen by the fact that *he* has a strong tendency to only be able to resolve to Chris.

The main idea that was developed in this book in order to obtain an explanation for this phenomenon was the following: rhetorical relations relate speech acts and speech acts are a special type of goal-directed acts. On the basis of this assumption, it is possible to apply similar assumptions on coherent behavior of agents to discourse, as for example the following: once an agent pursues a goal, one can assume that the agent's action will be directed to that goal. This assumption turns distinct actions of an agent into a coherent behavior from the perspective of an observer. As an example, take again an agent who wants to pick some apples. If an observer is already aware of the fact that this agent wants to pick some apples, the observer can make sense out of the agent's action of, for example, getting a ladder. Embedding discourse structure into a general theory of action allows to apply a similar principles to model a hearer's interpretation of distinct speech acts and to model the hearer's ability to construct a coherent discourse from distinct speech acts. For example, take a hearer who already knows that the speaker wants the hearer to close the window (for some reason). Now assume that the speaker performs a speech act as (7.3):

(7.3) It's cold in here.

In this case, the hearer is able to conclude that the speaker wants to persuade the hearer from the fact that it is actually a good idea to close the window by saying something as (7.3), because he already knew about the speaker's general plan. If the hearer obtained the knowledge about the speaker's goal to make the hearer close the window because of a former utterance of the speaker, like for example in (7.4), the hearer is actually able to relate two distinct speech acts on the basis of the speaker's goals. In other words, the hearer is able to construct a rhetorical relation:

(7.4) a. Could you please close the window!?
b. It's cold in here.

This shows that a hearer who relates two distinct speech acts fulfills a similar task as the observer of an agent who tries to make sense out of the agent's distinct actions.

However, the fact that the target of a speech act is most of the time another agent (the hearer) provides additional difficulties for the speaker. In general, one can assume that it is the goal of the speaker to influence his audience by means of the performance of a certain speech act. In the terminology of this book, this assumption can be expressed by saying that the speaker wants to make the hearer *committed* to a certain proposition. However, making the hearer committed to a proposition means that the speaker has to make assumptions about the hearer's mental state, i.e. the speaker has to guess whether the hearer is willing to update his commitments with the content of the speech act. Since the speaker cannot directly look in the mind of the hearer, the speaker needs evidence for the success of his speech act. On the basis of this assumption, it is possible to make two predictions about discourse structure in oral conversation: First,

one can deduce the fact that hearers feel the need to produce acceptance moves (such as nodding the head or uttering something as *mhm* or *yes*) as overt signals for the hearer's acceptance of the speech act's content and, second, one can deduce the prediction that two speech acts are by default related by a*support relation*. The term support relation refers to a relation where the goal of the second speech act is a subgoal of the goal of the first speech act.

The argument for the necessity of acceptance moves goes roughly as follows: the hearer knows that the speaker needs evidence for a change in the hearer's attitudes towards the content of the speech act, since by default, agents do not believe that state of affairs change if they don't have evidence for a change in the given state of affairs. Therefore, in order to signal that the speaker's goal is achieved, the hearer has overtly accept the speech act with an acceptance move. The assumption that human agents do not change their mind if they don't have a reason to change their mind is called *inertia*.

Deriving the support relation is a little bit more elaborated than deriving the function of acceptance moves. It needs similar assumptions (in particular, inertia is needed) but is more based on the perspective of the speaker: in the absence of an acceptance move, the speaker may assume that the hearer does not accept a given speech act, i.e. the speaker may assume that this speech act was unsuccessful. In that case, the speaker has to give support for the speech act in question in order to achieve the goal of the speech act. In other words, in the absence of an acceptance move, speakers produce by default a speech act which supports the former speech act. These two results can be regarded as two major results of this book.

However, this book originally wanted to approach the RFC. As already explained, the RFC makes predictions about the ability of discourse units to serve as possible anchors for new discourse material. It generalizes over the observation that coordinating rhetorical relations make it more difficult for new discourse material to directly attach to older speech acts than subordinating relations. These two example (7.1) and (7.2) exemplified this observation. these examples show that the RFC, and hence the question whether a rhetorical relation is subordinating or coordinating is a prominence-lending factor. The term *prominence* refers to the concept of discourse prominence which describes the observation that entities of the same type are ranked, that this ranking shifts in time, and that speakers use this ranking for choosing a referential expression for referring to an entity of the discourse (Jasinskaja, Chiriacescu, et al., 2015). A major effect of being prominent is that it is easier to use a short referential expression (such as for example a third person pronoun) to pick up a more prominent referent. Less prominent referents tend to need longer referential expressions to be referred to. Since the RFC makes predictions about the ability of discourse units to serve as pivot for new discourse material, it has a direct impact on the prominence status of the referents which occur in a given utterance. Coordinating rhetorical relations decrease the prominence status of older discourse referents, whereas subordinating or, in the case of this book, supporting relations sustain the prominence status of older referents. This is again exemplified by example (7.2), which shows that the coordinating relation prevents the third person pronoun *he* from being resolved to Max, i.e. a referent mentioned in an older discourse unit. Therefore, the type of rhetorical relation connecting two utterances has a direct impact on the prominence status of the referents in the pivot of the

relation and, therefore, is a further prominence lending factor.

This book aimed at providing a better understanding of this factor, i.e. it aimed at providing a satisfactory explanation of the RFC. The given explanation is based on the assumption that only speech acts that pursue goals which have not been achieved yet are still open for discourse attachment. This assumption is based on the intuition that speakers have to be able to talk about speech acts if their goals have not been achieved yet. On that basis and in the case of just two utterances, the explanation that is proposed in this book can be summarized as follows: Assume that (E_1, E_2) is a pair of speech acts and assume that E_2 supports E_1. This means that the speaker believes that the performance of E_2 helps to achieve E_1 and, hence, that the goal of E_1 has not been achieved yet. Since the goals of both utterances have not been officially achieved yet, the assumption above implies that new discourse material can attach to both utterances, which is what the RFC predicts in this case. If, on the other hand, E_2 does not support E_1, the speaker believes that the goal of E_1 has already been achieved and, therefore, the speaker does not have to talk about E_1 anymore, which is also predicted by the RFC. This simple idea was outlined in the last chapter, where it was adjusted for larger discourse segments. On the basis of an algorithm, that determined the nodes on the right frontier of a discourse graph, it was possible to show that these nodes correspond to the speech acts which are still under negotiation, i.e. where the goals have not been achieved yet.

In this book, discourse structure was embedded into a more general framework of assumptions about goals and actions of rational agents by regarding speech acts as goal-directed acts and by showing how common sense principles for rational agent's behavior can be applied in order to construct a coherent discourse from distinct speech acts. Moreover, it was possible to show that (sometimes) acceptance has to be marked and that in the absence of an acceptance move a new speech act supports its preceding speech act. Finally, it showed how one can derive the RFC as a prominence-lending feature on the basis of the assumption that speakers pursue goals with speech acts and that speech acts in which the goals have not been achieved yet have to stay available for further communication.

At the end of this conclusive remark, I want to mention some open questions which can (perhaps) be approached with the methods which are provided in this book. The framework of questions under discussion (QUD) was mentioned in the proceedings of this book several times. QUDs are assumed to be overtly expressed or implicit questions which have to be answered by the interlocutors. Therefore, QUDs can be regarded as a structuring principle of any discourse. In this book, however, it was proposed that each speech act is performed because the speaker pursues a particular goal with that speech act. I think that a framework which is based on goals of speech acts is in principle more flexible than a QUD-based framework, since such a framework is also able to treat speech acts which aren't assertions or questions. But if this is actually the case, it should be possible to embed a QUD-based discourse model into a goal-based discourse model. In other words, it should be possible to construct the goals of a speaker from the QUDs of the discourse. However, this problem is not explicitly adressed in this book and remains a problem for further research.

Another interesting phenomenon which can perhaps be approached with the meth-

ods of this work is the difference between non-at-issue and at-issue content (Potts, 2005). Non-at-issue content can be defined by saying that it is the content of a speech act that does not directly answer a given QUD. Another possible definition of non-at-issue content is that this content cannot directly be negated by negation of the whole speech act. For example, in (7.5) the apposition *an astronomer* is non-at-issue content since it cannot directly be negated if one negates the whole speech act, i.e. the fact that Kepler was an astronomer survives the negation of the claim that Kepler died in misery.

(7.5) Kepler, an astronomer, died in misery.

It is now a further interesting problem whether a goal-based approach to discourse can handle non-at-issue content such as appositions. Potts treated non-at-issue content on the basis of a two-dimensional semantics which models the fact that non-at-issue content does not seem to directly contribute to the truth value of the hosting sentence (since it cannot be targeted by negations). Jasinskaja, (2016) already proposed to treat non-at-issue content not as genuinely different to at-issue content but simply as an answer to an intermediate QUD. Given the assumption that communicative goals are a more general than QUDs, it should also be possible to deal with non-at-issue content on the basis of goals of speech acts. For example, non-at-issue content could be treated as a certain type of repair move which repairs some possible misunderstanding. However, in general it is not clear yet how one could model non-at-issue content and derive its typical features on the basis of goals of speech acts.

Bibliography

Anscombe, G. E. M., 1957, Intention. Harvard University Press.

Asher, N. & A. Lascarides, 2009, *Agreement, Disputes and Commitments in Dialogue*. In: Journal of Semantics, 26: 109–158.

Asher, Nicholas, 2014, *Selectional restrictions, types and categories*. In: Journal of Applied Logic 12.1. Logic Categories Semantics, pp. 75–87. ISSN: 1570-8683.

Asher, Nicholas & Alex Lascarides, 2003, Logics of Conversation. Vol. 9. Studies in Natural Language Processing. Cambridge University Press, pp. 183–221. ISBN: 9780521650588.

Asher, Nicholas & Alex Lascarides, 2013, *Strategic conversation*. In: Semantics and Pragmatics 6.2, pp. 1–62.

Asher, Nicholas & Laure Vieu, 2005, *Subordinating and coordinating discourse relations*. In: Lingua 115, pp. 591–610.

Austin, John Langshaw, 1962, How to do things with words. William James Lectures. Oxford University Press.

Averintseva-Klisch, M., 2013, Textkohärenz. Kurze Einführungen in die germanistische Linguistik. Universitätsverlag Winter. ISBN: 9783825361563.

Beaver, David Ian, 1997, *Presupposition*. In: Handbook of Logic and Language.

Becker, Martin & Jakob Egetenmeyer, 2018, *A prominence-based account of temporal discourse structure*. In: Lingua 214, pp. 28–58. ISSN: 0024-3841.

Benz, Anton & Katja Jasinskaja, 2017, *Questions Under Discussion: From Sentence to Discourse*. In: Discourse Processes 54, pp. 177–186.

Bratman, Michael, 1987, Intention, Plans, and Practical Reason. Center for the Study of Language and Information.

Bratman, Michael, 2018, Planning, Time, and Self Governance: Essays in Practical Rationality. Oxford University Press. ISBN: 9780190867898.

Bratman, Michael E., 1990, *What is Intention*. In: Intentions in Communication. MIT Press, pp. 15–32.

Clark, Herbert H., 1996, Using Language. Cambridge University Press.

Clark, Herbert H. & Susan E. Brennan, 1991, *Grounding in Communication*. In: Perspectives on socially shared cognition, pp. 127–149.

Clark, Herbert H. & Edward F. Schaefer, 1989, *Contributing to Discourse*. In: Cognitive Sience 13, pp. 259–294.

Cohen, Philip R. & Hector J. Levesque, 1990, *Intention is Choice with Commitment*. In: Artif. Intell. 42.2-3, pp. 213–261. ISSN: 0004-3702.

Cohen, Philip & Hector J. Levesque, 1988, *Rational Interaction as the Basis for Communication*. In: Pattern Recognition - PR, p. 39.

Condoravdi, Cleo & Sven Lauer, 2011, *Performative Verbs and Performative Acts*. In: Proceedings of Sinn und Bedeutung 15, pp. 149–164.

Condoravdi, Cleo & Sven Lauer, 2012, *Imperatives: meaning and illocutionary force*. In: Empirical Issues in Syntax and Semantics 9, pp. 37–58.

Davidson, Donald, 1980, Essays on Actions and Events. Oxford University Press.

Demey, Lorenz & Joshua Sack, 2015, *Epistemic Propabalistic Logic*. In: Handbook of Epistemic Logic. Ed. by Hans van Ditmarsch et al. College Publications.

Dowty, David, 1991, *Thematic Proto-Roles and Argument Selection*. In: Language 67.3, pp. 547–619. ISSN: 00978507, 15350665.

Einstein, A. & L. Infeld, 1960, The Evolution of Physics from Early Concepts to Relativity and Quanta. Simon and Schuster. ISBN: 9780671201562.

Fagin, Ronald, 2003, Reasoning About Knowledge. Vol. 1st MIT Press pbk. ed. A Bradford Book. ISBN: 9780262562003.

Farkas, Donka F. & Kim B. Bruce, 2009, *On Reacting to Assertions and Polar Questions*. In: Journal of Semantics 27.1, pp. 81–118. ISSN: 0167-5133. eprint: https://academic.oup.com/jos/article-pdf/27/1/81/4601470/ffp010.pdf.

Foraker, Stephani & Brian McElree, 2007, *The role of prominence in pronoun resolution: Active versus passive representations*. In: Journal of Memory and Language 56.3, pp. 357–383. ISSN: 0749-596X.

Frege, Gottlob, 1892, *Über Sinn und Bedeutung*. In: Zeitschrift für Philosophie und philosophische Kritik 100, pp. 25–50.

Geurts, Bart, 2018, "Communication as commitment sharing: speech acts, implicatures, common ground".

Ginzburg, Jonathan, 2012, The Interactive Stance: Meaning for Conversation. Oxford University Press.

Grice, H. P., 1968, *Utterer's Meaning, Sentence-Meaning, and Word-Meaning*. In: Foundations of Language 4.3, pp. 225–242.

Grice, Paul, 1975, *Logic and Conversation*. In: Syntax and Semantics 3, pp. 41–58.

Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein, 1995, *Centering: A Framework for Modeling the Local Coherence of Discourse*. In: Computational Linguistics 21.2, pp. 203–225.

Grosz, Barbara J. & Candace L. Sidner, 1990, *Plans for Discourse*. In: Intention In Communication. Ed. by Philip R. Cohen, Jerry Morgan & Martha E. Pollack. A Bradford Book - The MIT Press, pp. 418–444.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski, 1993, *Cognitive Status and the Form of Referring Expressions in Discourse*. In: Language 69.2, pp. 274–307. ISSN: 00978507, 15350665.

Harras, G., 1983, Handlungssprache und Sprechhandlung: eine Einführung in die theoretischen Grundlagen. De Gruyter Studienbuch. De Gruyter. ISBN: 9783110176773.

Heim, Irene, 1992, *Presupposition Projection and the Semantics of Attitude Verbs*. In: Journal for Semantics 9, pp. 183–221.

Himmelmann, Nikolaus P. & Beatrice Primus, 2013, "Prominence Beyond Prosody – A First Approximation". In: pS-prominenceS: Prominences in Linguistics. Proceedings of the International Conference. Ed. by Amedeo De Dominicis.

Hinterwimmer, Stefan, 2019, *Prominent protagonists*. In: Journal of Pragmatics 154, pp. 79–91.

Hintikka, Jaakko, 1962, Knowledge and Belief - An Introduction to the Logic of the Two Notions. College Publications. ISBN: 978-1904987086.

Hobbs, Jerry R., 1979, *Coherence and coreference*. In: Cognitive Science 3.1, pp. 67–90. ISSN: 0364-0213.

Hobbs, Jerry R., 2005, *Toward a Useful Concept of Causality for Lexical Semantics*. In: Journal of Semantics 22, pp. 181–209.

Jasinskaja, Katja, 2010, *Modelling Discourse Relations by Topic and Implicature: The Elaboration Default*. In: Constraints in Discourse 2. Ed. by Candy Sidner, Anton Benz & Peter Kühnlein, pp. 61–79.

Jasinskaja, Katja, 2016, "Not at issue any more". In: 17th Szklarska Poręba Workshop.

Jasinskaja, Katja, Sofiana Chiriacescu et al., 2015, *Prominence in Discourse*. In: Prominences in Linguistics: Proceedings of the pS-prominenceS International Conference, pp. 134–153.

Jasinskaja, Katja & Elena Karagjosova, 2015, *Rhetorical Relations*. In: Blackwell Companion to Semantics. Ed. by Lisa Matthewson et al.

Kamp, Hans, 2015, *Using Proper Names as Intermediaries Between Labelled Entity Representations*. In: Erkenntnis 80.

Kamp, Hans, Josef van Genabith & Uwe Reyle, 2005, *Discourse Representation Theory*. In: Handbook of Philosophical Logic. Ed. by Dov Gabbay & Franz Guenthner.

Kamp, Hans & Uwe Reyle, 1993, From Discourse to Logic: Introduction to Model-theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory. Studies in Linguistics and Philosophy. Dordrecht: Springer. ISBN: 978-0-7923-1028-0.

Knott, Alistair & Robert Dale, 1994, *Using linguistic phenomena to motivate a set of coherence relations*. In: Discourse Processes 18, pp. 35–62.

Krifka, Manfred, 2015, *Bias in commitment space semantics: Declarative questions, negated quetions, and question tags*. In: Semantics and Linguistic Theory 25, p. 328.

Lambalgen, Michiel van & Fritz Hamm, 2005, The Proper Treatments of Events. Blackwell Publishing.

Lauer, David, 2014, *What Is It to Know Someone?* In: Philosophical Topics 42.1, pp. 321–344.

Lewis, David K., 1973, *Causation*. In: Journal of Philosophy 70.17, pp. 556–567.

Luhmann, N., 2017, Die Kontrolle von Intransparenz. Suhrkamp Verlag. ISBN: 9783518754528.

Maier, Emar, 2017, *Fictional Names in Psychologistic Semantics*. In: Theoretical Linguistics 43, pp. 1–45.

Mann, William C. & Sandra A. Thompson, 1988, *Rhetorical Structure Theory: Towards a functional theory of text organization*. In: Text, pp. 243–281.

McCready, Eric, 2014, Reliability in Pragmatics. Oxford University Press.

McCullagh, C. Behan, 1975, *Causal Theories of Action*. In: Philosophical Studies: An International Journal for Philosophy in the AnalyticTradition 27.3, pp. 201–209.

Meyer, John-Jules Ch., Jan Broersen & Andreas Herzig, 2015, *BDI Logic*. In: Handbook of Epistemic Logic. Ed. by Hans van Ditmarsch et al. College Publications, pp. 489–542.

Patil, Umesh, Peter Bosch & Stefan Hinterwimmer, 2019, *Constraints on German diese demonstratives: language formality and subject-avoidance*. In: Glossa: a journal of general linguistics 5, pp. 1–22.

Patterson, Clare & Petra Schumacher, 2020, *The timing of prominence information during the resolution of German personal and demonstrative pronouns*. In: Dialogue and Discourse 11.1, pp. 1–39.

Pearl, J., 2009, Causality. Causality: Models, Reasoning, and Inference. Cambridge University Press. ISBN: 0521773628.

Poesio, Massimo & David R. Traum, 1997, *Conversational Actions and Discourse Situations*. In: Computational Intelligence 13.3, pp. 309–347. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/0824-7935.00042.

Polanyi, Livia, 1988, *A formal model of the structure of discourse*. In: Journal of Pragmatics 12.5, pp. 601–638. ISSN: 0378-2166.

Pollack, Martha E., 1990, *Plans as Complex Mental Attitudes*. In: Intentions in Communication. MIT Press, pp. 77–104.

Potts, Christopher, 2005, The Logic of Conventional Implicatures. Ed. by Hagit Borer David Adger. Oxford University Press.

Primus, B., 2010, Cases and Thematic Roles: Ergative, Accusative and Active. Linguistische Arbeiten. De Gruyter. ISBN: 9783110912463.

Rasmussen, Mattias Skipper, 2015, *Dynamic Epistemic Logic And Logical Omniscience*. In: Logic and Logical Philosophy.

Roberts, Craige, 1996, *Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics*. In: OSU Working Papers in Linguistics 49, pp. 91–136.

Roberts, Craige, 2011, *Topics*. In: Semantics: An International Handbook of Natural Language Meaning. Ed. by Claudia Maienborn, Klaus von Heusinger & Paul H. Portner. 2nd ed. Vol. 2. de Gruyter, pp. 1908–1934.

Roberts, Craige, 2017, "Speech Acts in Discourse Context". To appear in: New Work on Speech Acts.

Rödl, Sebastian, 2018, *Joint Action and Plural Self-Consciousness*. In: Journal of Social Philosophy 49.1, pp. 124–136. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/josp.12226.

Russell, Bertrand, 2005, *On Denoting*. In: Mind 114.456, pp. 873–887.

Scalzi, John, 2005, Old Man's War. Tor Books. ISBN: 0-7653-0940-8.

Schank, Roger & Robert Abelson, 1977, Scripts, Plans, Goals and Understanding. Ed. by Roger Schank. The Artificial Intelligence Series. John Wiley and Sons.

Schumacher, Petra B., Jana Backhaus & Manuel Dangl, 2015, *Backward- and Forward-Looking Potential of Anaphors*. In: Frontiers in Psychology 6.

Searle, John R., 1969, Speech Acts: An Essay in the Philosophy of Language. Cambridge, London: Cambridge University Press.

Searle, John R., 1983, Intentionality: An Essay in the Philosophy of Mind. Cambridge paperback library. Cambridge University Press. ISBN: 9780521273022.

Searle, John R., 2005, *Desire, Deliberation and Action*. In: Logic, Thought and Action. Ed. by Daniel Vanderveken. Vol. 2. Logic, Epistemology, and the Unitiy of Science. Springer.

Searle, John R. & Daniel Vanderveken, 1985, Foundations Of Illocutionary Logic. Cambridge University Press.

Shanahan, Murray, 1997, Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia. Cambridge, MA, USA: MIT Press. ISBN: 0-262-19384-1.

Shohan, Yoav, 1990, *Nonmonotonic Reasoning and Causation*. In: Cognitive Science 14, pp. 213–252.

Simons, Mandy, 2013, *Presupposing*. In: ed. by Sbisa Marina & Ken Turner. Vol. 2. Hanbooks of Pragmatics. DeGruyter.

Stalnaker, Rober, 1978, *Assertions*. In: Syntax and Semantics 9, pp. 315–332.

Stalnaker, Robert C., 1973, *Pragmatic Presuppositions*. In: ed. by Martin Davies et al. Mindreading. Oxford University Press.

Sweetser, Eve, 1990, From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure. Cambridge Studies in Linguistics. Cambridge University Press.

Txurruka, Isabel Gómez, 2003, *The Natural Language Conjunction And*. In: Linguistics and Philosophy 26.3, pp. 255–285.

van Bockstaele, Bram et al., 2012, *Learning to Lie: Effects of Practice on the Cognitive Cost of Lying*. In: Frontiers in psychology 3, p. 526.

Veer, Anna, Mariëlle Stel & Ilja van Beest, 2013, *Limited Capacity to Lie: Cognitive Load Interferes with Being Dishonest.* In: SSRN Electronic Journal 9, pp. 199–206.

von Heusinger, Klaus & Petra Schumacher, 2019, *Discourse prominence: Definition and application*. In: Journal of Pragmatics 154, pp. 117–127.

Vrij, Aldert et al., 2006, *Detecting deception by manipulating cognitive load*. In: Trends in cognitive sciences 10, pp. 141–142.

Wittgenstein, Ludwig, 1953, Philosophische Untersuchungen. Frankfurt am Main: Suhrkamp Verlag.

Zeevat, Henk, 2004, *Particles: Presupposition triggers, context markers or speech act markers.* In: Optimality theory and pragmatics. Ed. by Reinhard Blutner & Henk Zeevat. Palgrave Macmillan, pp. 91–111.

Zeevat, Henk, 2011, *Why Study Rhetorical Relations*. In: Semantics. Ed. by Paul Portner Claudia Maienborn Klaus von Heusinger. Vol. 1. DeGruyter, pp. 946–969.

Zeigarnik, Bluma, 1927, *Das Behalten erledigter und unerledigter Handlungen*. In: Untersuchungen zur Handlungs- und Affektpsychologie. Ed. by Kurt Lewin.