Human Sentence Processing: A Semantics-Oriented Parsing Approach

Inaugural-Dissertation zur Erlangung der Doktorwürde der Philosophischen Fakultäten der Albert-Ludwigs-Universität zu Freiburg i. Br.

> vorgelegt von

Lars Konieczny aus Bochum Sprecher des Gemeinsamen Ausschusses der Philosophischen Fakultäten: Prof. Dr. D. Mertens

Referenten: Prof. Dr. G. Strube, Prof. Dr. J. Dittmann und Prof. Dr. H. Spada

Datum der Promotion: 5. 7. 1996

Contents

1	Introduction	1			
	1.1 The Cognitive Science approach	2			
	1.2 The role of Syntax				
	1.3 The phenomena	5			
	1.4 Overview				
2	Previous work				
	2.1 The contribution of syntax to sentence understanding	12			
	2.1.1 Chomsky and Skinner	12			
	2.1.2 Competence and performance	13			
	2.1.3 Standard Theory (Chomsky 1965)	14			
	2.1.4 Derivational Theory of Complexity (Miller, 1962)	16			
	2.2 Perceptual strategies, Bever (1970)	17			
	2.3 Kimball's principles for surface structure parsing	19			
	2.4 The Sausage Machine	20			
	2.5 Augmented Transition Networks (ATNs)	22			
	2.6 Garden Path Theory	24			
	2.7 Lexical preferences: the Theory of Syntactic Closure	25			
	2.8 Abney's Licensing Structure Parser (1989)	31			
	2.9 Theta Attachment and the On-Line Locality Constraint (Pritchett, 1992)	33			
	2.10 Gibson's weighted parallel model ofhuman sentence processing	36			
	2.11 Multiple constraint models	41			
	2.12 CAPERS: A hybrid model of human parsing	43			
	2.13 Structural determinism: Gorrell's theory of syntax and parsing	46			
	2.14 Construal Theory	49			
	2.15 Tuning	52			
	2.16 Parametrized Head Attachment	53			

2.17 Conclusion			sion	56		
		2.17.1 S	erialism, parallelism, and underspecification	56		
		2.17.2 T	he status of lexical and higher order information	56		
		2.17.3 T	The competence base	56		
		2.17.4 I	ncrementality	57		
		2.17.5 V	Vhat are the constraints of the human parser?	58		
3	On	line tecl	nniques in parsing research	59		
	3.1	3.1 Eye-tracking				
	3.2	3.2 Regions				
	3.3 Standard measures					
		3.3.1 F	irst fixation durations	63		
		3.3.2 F	'irst pass reading times	64		
		3.3.3 T	otal reading times	66		
		3.3.4 A	A preliminary conclusion	67		
	3.4	3.4 New measures				
3.5 Accounting for lexical effects		ting for lexical effects	72			
		3.5.1 V	Vord length	72		
		3.5.2 N	Number of words in a region	76		
		3.5.3)pen questions	77		
	3.6	Conclus	sion	77		
4	General constraints on human parsers			79		
	4.1	The role	e of lexical heads	80		
		4.1.1 P	Preferences in German verb-second and verb-final clauses	80		
		4.1	.1.1 Introduction	80		
		4.1	.1.2 Experiment I	81		
		4.1	1.3 Experiment II	88		
		4.1 4 1	1.5 Ceneral Discussion	92		
		412 S	trictly linear parsing	100		
		4 1	2.1 Incremental attachment in verh-final clauses	100		
		4.1	.2.2 Self-paced reading experiments on subject-object asymmetries	100		
		4.1	.2.3 An eyetracking experiment	103		
		4.1.3 H	Iow lexical information is employed during parsing	111		
		4.1	.3.1 Subcategorization information	111		

		4.1.3.2 4.1.3.3	Thematic information Conclusion	122 125
	4.2	The architec	ture of the human sentence processing mechanism	128
		4.2.1 Serial,	, parallel, and minimal commitment models	128
		4.2.1.1	Predictions	128
		4.2.1.2	Evidence	130
		4.2.1.3	A preliminary conclusion	136
		4.2.2 Autor	nomous vs. interactive processing	137
		4.2.2.1	Attachment preferences in referentially biasing contexts	137
		4.2.2.2	Conclusion	150
	4.3	General disc	cussion	151
5	The SOUL system			
	5.1	The compete	ence base	154
		5.1.1 Head-	-driven Phrase Structure Grammar	
		(HPSO	G, Pollard & Sag, 1987, 1994)	154
		5.1.1.1	The formalism	156
		5.1.1.2	What do HPSG models represent?	159
		5.1.1.3	Phrase structure: the "skeleton"	161
		5.1.1.4	Head projection	164
		5.1.1.5 5.1.1.6	Content	166
		5117	Type deduction versus parsing	160
		5.1.2 HPSG	as a cognitive model of linguistic knowledge?	169
		513 Modif	fications and extensions in SOLU	100
		5131		171
		5.1.3.2	Complements and adjuncts	171
		5.1.3.3	Optional complements and pruning	173
		5.1.3.4	Lexical preferences	175
	5.2	Towards a th	neory of thematic salience	177
		5.2.1 Frequ	encv	177
		5.2.2 Instru	iments and agents	177
		5.2.3 Ontol	ogical necessity	178
		5.2.4 Conce	optual properties	179
		5.2.5 Conte	extual influences on salience	180
		5.2.6 Salien	ice at run or compile time?	181
	5.3	The processi	ing mechanism	182
		5.3.1 Mode	lling an adequate parser	182

	500		100	
	5.3.2	Signs as objects	183	
	5.3.3	Predicted and active attachment	184	
	5.3.4	States	185	
	5.3.5	Bottom up projection	185	
	5.3.6	The prediction stack	186	
	5.3.7	The treatment of optional complements	188	
	5.3.8	Overshadowing predictions	188	
	5.3.9	Ambiguity resolution	190	
	5.3.10	Incremental parsing with a head-driven grammar	199	
	5	5.3.10.1 SPEC and MOD in noun phrases	200	
	5	5.3.10.2 Inverted subcategorization information	204	
	5	5.3.10.3 Generalized sentence templates	205	
	5	5.3.10.4 Parsing with templates	207	
	Ę	5.3.10.5 Head-final constructions	207	
	5.3.11	Attachment preferences in verb-final constructions	212	
	5.3.12	Modulated lexical guidance	213	
	5.3.13	Revision of initial attachments	215	
5.4	Discu	ssion	218	
	5.4.1	Coverage of ambiguity resolution phenomena	218	
	5.4.2	SOUL and the Garden Path model	220	
	5.4.3	SOUL and the Licensing Structure Parser (Abney, 1987, 1989)	221	
	5	5.4.3.1 Verb attachment and low attachment	222	
	5	5.4.3.2 Head placement effects	223	
	5.4.4	Syntactic Preferences (Ford, Bresnan & Kaplan, 1982)	223	
	5	6.4.4.1 Modulated vs. unrestricted lexical guidance	225	
	5.4.5	Visibility and competition revisited	227	
5.5	Avoid	ing memory load: semantics-oriented parsing	227	
5.6	Sumn	nary	229	
Su	Summary and concluding remarks			

References

Acknowledgements

I am extremely grateful to Barbara Hemforth, Christoph Scheepers and Gerhard Strube for numerous fruitful discussions during the last five years in which we worked together so successfully. I would also like to thank Don Mitchell, Lyn Frazier, Chuck Clifton, Gerry Altmann, Gerard Kempen, Koenraad de Smedt, Steven Abney, Laurie Stowe, Jose Igoa, Simon Liversedge and many others for valuable comments and discussions about various aspects of this thesis. Extra special thanks go to Barbara Hemforth again, who kept many undesirable distractions and burdens away from me during the hot phase of writing this thesis. I owe a special debt of gratitude to Nick Ketley, who was there when I needed him to put a nice Australian accent into my English writings. I would also like to thank Thomas Mulack and Christoph Hölscher for their support in running the eye-tracking experiments. Thilo Weigel, Andreas Kleegraf, Ari Vollweiter, and Martin Ernst Kraus did a wonderful job in adjusting the eye-tracker software to our purposes. Thanks and apologies to all those important people I forgot because I'm currently too tired to remember.

1 Introduction

Understanding a sentence is easy, at least in general. How we comprehend a sentence and why we hear or read the majority of the sentences without a glitch, however, is far from clear. The ease of understanding a sentence stands in striking contrast to the complexity of the language we use.

(1) Time flies like an arrow.

Sentences in natural languages are usually highly ambiguous, i.e. they can be interpreted in many ways. Nevertheless, most people are aware of only one meaning shortly after they have read or heard a sentence. Sentence (1), for instance, has many possible interpretations¹. Nevertheless, most people read it metaphorically, roughly as "Time passes quickly". They might also have read it as an instruction type of sentence "Please, time these flies like you time an arrow!" or "Please, time these flies like an arrow times them!", though it might be harder to figure out a context in which such a sentence would actually make sense. Others might understand that instances of a particular species of flies, called *time flies*, are always keen on an arrow. Perhaps the word flies does not refer to an object of the insect kind, but to "the space over a theater stage where scenery and equipment can be hung" (Webster's 9th Student Collegiate Dictionary). The ambiguity of (1) partly arises from the fact that the words time, flies, and *like*, can be of more than one syntactic category. Is *like* a preposition, a verb, a noun, an adverb, an adjective, a conjunction? It might be each. It is *despite* the extraordinary ambiguousness of natural languages that sentence comprehension seems to be easy. Quite miraculously, the meaning we receive is quite often the meaning actually intended by the author or speaker.

^{1.}Actually, there are about 45 interpretations for the written version and more than 100 interpretations for the spoken version of the sentence (Altmann, personal communication)

1.1 The Cognitive Science approach

In Cognitive Science and its branches psychology, linguistics, computer science, philosophy and the neuro-sciences, among others, the approach taken on *cognitive* phenomena such as language comprehension is to understand them as *computational processes*, operating on *representations* that constitute (different kinds or *levels* of) our linguistic knowledge. For language processing in particular, a wide variety of knowledge sources come into play. In order to comprehend an utterance, we have to know what the particular words can mean and what *concepts* they are referring to, and, particularly, what the words can mean in the context of the other words in the utterance. In order to uncover the meaning of an entire sentence, one has to know how words can be combined *syntactically* in principle and how different combinations result in different meanings. The knowledge of how words can be syntactically combined with others entails the knowledge about their particular syntactic *category*, which is partly determined by their *morphological* structure, this in turn being determined by even finer grained information. Last but not least, the meaning of a sentence or parts of it may lie beyond the words, and quite often only the context in which a sentence is uttered determines what it might actually mean.

Sentence (1) is a striking example that illustrates the potential for ambiguity at almost all levels of representation. The question that arises is: given that linguistic knowledge includes lexical, morphological, phonological, syntactic, semantic, and pragmatic information, how are these sources of information put to use in a way that the right (intended) interpretation is chosen?

Of course, the intended interpretation is not always the one preferred by the perceiver. Consider the (real life) examples given in (2) and (3).

(2) Gary Barlow und Mark Owen (Take That) wollen laut "Bravo" Solokarrieren starten, Jason Orange werde auf Weltreise gehen und Howard Donald heiraten. (Badische Zeitung, 20. 2. 1996, Nr. 42, AAW 01)

Gary Barlow and Mark Owen (Take That), according to "Bravo", want to start solo-careers, Jason Orange will go on a journey around the world and

a. marry Howard Donald. (preferred)

b. Howard Donald will marry.

(3) ... würden keine zehn Pferde mich dazu kriegen, dahin zu gehen, wo all diese Verrückten sich umbringen, um Fotos zu machen.

Even ten horses would not get me to go to the place where all these crazy people kill themselves to take photos.

From "Frauenarzt Dr. Mertin", ZDF, 24. 2. 1996

The presumably intended interpretation of the latter example is roughly sketched in "I would not go there ... to take photos", whereas the interpretation that comes to mind initially is "... where they kill themselves to take photos." Examples like these illustrate that the human parser can very well be misguided by a sentence that requires an unpreferred interpretation. Such examples appear rather rarely in everyday conversation. Though people appear not to be led up the garden path very often in every day conversation, ambiguities of any kind appear in almost every sentence. Interestingly, we are simply not aware of them in the majority of the cases.

Not surprisingly, *ambiguity resolution* as a cognitive phenomenon has definitely been among the most prominent topics in the investigation of language processing in the last three decades. Its importance lies mainly in the potential to provide a deep insight into the human language system in general. If we understand how certain variations of ambiguous sentences or sentence-parts cause the reader or listener to adopt different readings or how certain variants induce measurable processing difficulties or even processing break-down, we will (hopefully) have learned something about the architecture or the underlying principles of the language system which determine what kinds of information can be used at different stages during processing.

The *Cognitive Science* approach to language, also referred to as the *psycholinguistic* approach, includes research strategies from several branches. Psychological experiments, mostly on *reading*, can tell us something about on-line preferences and the time course of language processing. Linguistics provides theories of linguistic knowledge. A *psycholinguistic model of human sentence processing*, then, integrates the facts from the different sources into a single system, proving the consistency of the formerly separately formulated assumptions. Ideally, the model is based on formal grounds to allow the different kinds of facts to be related to each other. Algorithms and representation techniques established in computer science and computational linguistics are employed in the task of modelling. When a model is set up, new questions will probably emerge entering the cycle of psycholinguistic research.

Metaphorically speaking, cognitive modelling amounts to solving a complex puzzle, unfortunately without knowing the complete picture. Moreover, many pieces are missing, leaving gaps to be filled. The pieces of evidence available allow for a variety of possible orderings, revealing different kinds of models. Fortunately, new pieces of evidence from psychological experiments, linguistic intuitions, neuro-linguistic findings etc. can be added, successively constraining the range of possibilities of how the pieces could be put together. Consequently, cognitive modelling entails the prediction of missing pieces thus simplifying the search, i.e. inspiring new experiments to be conducted.

1.2 The role of Syntax

Processing models developed to explain the observed preferences differ in the extent to which they consider grammatical knowledge relevant for language processing. On one side, models have been developed which totally leave aside linguistic analyses of the language input and instead use agrammatical heuristics (Bever, 1970; Schank, 1972; Herrmann, 1985). Theo Herrmann (1985), for example, claims that "in the normal case" a certain number of learned standard schemata about the canonical ordering of content words suffice to "structure" a sentence. It is only in those cases, where the available schemata are no longer applicable, that a syntactic analysis is called for, which, according to Herrmann's predictions, involves a conscious level of processing with a high degree of cognitive effort.

In the light of findings like those of Flores d'Arcais (1982, 1987, cf. Hemforth, 1993) the plausibility of these presumptions seems doubtful: Flores d'Arcais measured higher processing times for sentences with slight grammatical mistakes, for example, where prepositions were false or missing altogether (4), than with grammatically correct sentences, even though the ungrammaticality was not noticed by many of the subjects.

(4) The old lady sat ____ the white chair.

The non-awareness of syntax violations alone is compatible with syntax-free processors: because of the schematic ordering of the content words, no syntactic analysis occurs in which the grammatical errors could have been discovered. However, the increased processing times associated with syntactic violations remain unexplained. If one argues, however, that the increased processing time in (4) is due to a conscious and thorough syntax analysis, one must also predict, contrary to the facts, that the ungrammaticality is discovered in every case. Moreover, it remains unclear as to why such a syntax analysis would be needed to process sentence (4) in a syntax-as-lastresort account, since the content word's surface ordering should render it unnecessary.

Thus, findings like these cannot be explained consistently by language processing models which only take grammatical knowledge into account "as a last resort" (if at all). Actually, grammatical knowledge is always used, even in the analysis of simple sentences. The processing of this knowledge apparently occurs to a large extent automatically and thus unconsciously. On the basis of this knowledge alone, a correction of slight input errors is enabled which, although time consuming, is not necessarily noticed consciously (cf. Hemforth 1993).

Nowadays, it is widely agreed that linguistic competence is an indispensable part of realistically complex models of human language comprehension. Note that the primary function of a grammatical analysis is hardly that of distinguishing well-formed from non-well-formed utterances (cf. Hörmann, 1976). The parser structures the speech input so that its meaning can be recovered. Consider sentence (5).

(5) Den Jungen aß der Apfel. (Hemforth, 1993)

The boy_{acc} ate the apple_{nom}.

A model exclusively based on thematic/semantic roles would erroneously assume *the boy* to be the agent and *the apple* to be the patient of *eating* in this sentence. Only by virtue of syntactic constraints can language be used creatively, in that words can be combined in a way a thematic parser would have rejected as a seemingly implausible role-assignment (McCloskey, 1988, Hemforth, 1993).

Syntax clearly plays a crucial role in language comprehension. Nevertheless, one should always bear in mind that the goal of sentence processing is interpretation. Semantic and syntactic issues are hard to separate, both theoretically and empirically. A model of human sentence processing must be specific in its assumptions on the relationship between syntax and semantics, even if it is intended to focus on syntax processing exclusively.

There are a variety of possible scenarios with respect to the interplay of syntax and semantics. In chapter 4.2 I will discuss the issue of autonomous versus interacting processing modules. Evidence will be given suggesting that the syntax processor operates as an input system (Fodor, 1983), which is not irritated by higher level constraints in its first analysis.

1.3 The phenomena

Whether or not a processing mechanism can be regarded as cognitively adequate depends on whether it accounts for both linguistic and psychological phenomena. A parsing algorithm should for instance process the items in the same order as humans do. In German (as in English and many other languages), processing should thus follow the left to right ordering of words. Moreover, the parser should account for the fact that there is considerable evidence of interpretation starting before the end of the sentence is reached (Marslen-Wilson, 1973, 1975; Marslen-Wilson & Tyler, 1980; Hemforth et al., 1993).

A cognitive parser should also explain why the correct reading of some sentences, such as (6), is nearly irrecoverable, whereas sometimes, an increased processing load can only be observed in sophisticated experiments (7ab).

(6) Peter schenkte seiner Cousine ein Buch und einen Teddy, um niemanden zu übervorteilen, dem Bruder.²

Peter gave her cousin a book and a teddy, in order to treat both equally, to the brother.

In cases of ambiguities, the parser should initially pursue the alternative that people prefer. Evidence central to the processing model presented in this thesis comes from a series of self-paced reading and eye-movement experiments, which will be reported in detail in chapter 4.1 (see also Konieczny et al., 1994, 1995). In these experiments, subjects read German PP-attachment sentences, such as (7ab).

(7) Marion {a. beobachtete, b. erblickte} den Mann mit dem Fernglas.Marion {a. *watched*, b. *caught sight of*} the man with binoculars.

^{2.}The problem arises from a local ambiguity located at the words *und einen Teddy*: there seems to be a strong preference to coordinate the locally ambiguous *und einen Teddy* with the preceding NP *ein Buch*, which results in the problem that the last NP *dem Bruder* cannot be integrated. The correct coordination, of course, is the one which coordinates the constituent [*seiner Cousine ein Buch*] with [*einen Teddy seinem Bruder*].

(8) Ich habe gehört, daß Marion den Mann mit dem Fernglas {a. beobachtete, b. erblickte}.

I have heard, that Marion the man with the binoculars {a. watched, b. caught sight of}.

"I heard that Marion {a. watched, b. caught sight of} the man with the binoculars"

Both sentences are ambiguous in that the PP *with the binoculars* can be understood as an instrument of the action expressed by the verbs *watched* and *caught sight of*, or as an attribute of the object NP *the man*. Subjects preferentially attached the PP *with the binoculars* to the verb *watched* in (7a) but to the preceding NP *the man* in all other cases (7b, 8ab). Obviously, the placement of the verb, as well as its particular properties, influence human attachment preference here.

The model that I will propose here incorporates a tendency to attach an item to a constituent that has already been read. Besides the PP-attachment preferences above, this tendency will be shown to be also responsible for the coordination preference in (6) above and a variety of further attachment phenomena. However, in some cases, such as (7ab), the lexical bias of the particular head to which the item can be attached modulates the parsing preferences early on. Naturally, only if the lexical head of the attachment site is read *before* the location of the ambiguity can it modulate the parsing process.

In order to account for a broad variety of parsing phenomena, the *Parametrized Head Attachment* (PHA) principle (Konieczny et al., 1994) was formulated.

- (p1) Parametrized Head Attachment, PHA (Konieczny, Hemforth, Scheepers, & Strube, 1994)
 (Attempt to) apply Head Attachment (p1a) before Preferred Role Attachment (p1b) before Most Recent Head Attachment (p1c).
- (p1a) *Head Attachment* (Konieczny, Hemforth, & Strube, 1991) Prefer to attach an item to a phrasal unit whose lexical head has already been read.

The *Head Attachment* principle applies to all those cases, where an attachment ambiguity can be resolved by either attaching to a preceding head, or to a head yet to come.

In many other cases, such as the verb-second sentences in (7ab), there are two or more heads already read which are potential attachment sites for an ambiguous item. As we have seen above, the decision now depends on the *lexical bias* of either of the heads, as expressed in (p1b).

(p1b) Preferred-Role Attachment

Prefer to attach an item to a phrasal unit whose head *preferentially* subcategorizes for it.

Of course, there are cases where two (or more) preceding heads do not differ in their preference to bind a constituent, as in (9).

(9) El fisico dedujo las conclusiones del experimento. (Igoa, 1995)
 The physicist deduced the conclusions of (from) the experiment.

In these cases, a decision is supposed to be based upon *recency* (p1c).

(p1c) Most Recent Head Attachment

Prefer to attach an item to the head that was read most recently.

Most Recent Head Attachment is only applied if the other principles fail to provide a decision. This has been expressed in the unified *Parametrized Head Attachment* principle, *PHA* (p1), which furthermore serves the purpose of emphasizing the fact that attachment ambiguities are resolved on the basis of certain parameters of lexical heads, such as their *relative position* and *lexical preferences*.

According to PHA, and in particular due to *Head Attachment*, the parser initially builds a syntactic structure, which can be evaluated semantically as early as possible (*"semantics-oriented"* processing).

PHA is intentionally formulated at a meta-algorithmic level of a computational theory of language *processing*. Computational theory, in this respect, can be regarded as a variant of Marr's definition (see Marr, 1982) in that it is applied to procedural issues here. In that sense, a processing *principle*, such as *PHA*, specifies a mechanism, which does neither depend on representational assumptions nor on concrete algorithmic specifications. It must be shown, however, that it is possible to give an algorithm as a specification of the model that combines with independently (linguistically) motivated representational assumptions. This algorithm must therefore be shown to be applicable to a modern grammar theory, such as HPSG (Pollard & Sag, 1994), or the minimalist program (Chomsky, 1995)

Such a cognitive parsing algorithm must proceed in a way that corresponds to PHA. In the chapter that deals with the computational model I will present a serial algorithm, called the SOUL-processing mechanism (SOUL: Semantics-Oriented Unification-based Language processor), which works "as if" it was directed by PHA with respect to the initial attachment preferences.

1.3.1 Serialism, parallelism, guidance and filter

Although we seem to have access to human preferences when it comes to ambiguities, we still do not know when exactly during processing the preferred analysis is chosen. It might be the case that the parser calculates all syntactic alternatives at the location of an ambiguity in the first stage, but then decides to pursue only one of them, probably the *best* one, in the further analysis. That is to say that the parser uses certain *heuristics* in the second stage to filter out only very few or even only a single alternative. In that sense, *PHA* could be directly implemented as a second stage heuristic, filtering out the structural alternative favored by PHA. Note that the processes at both stages could apply at one and the same location in the sentence. Therefore, it will be very hard to distinguish *locally parallel* and *locally serial* models *empirically*. I will argue, however, that, among other things, the lack of an ambiguity effect in the ambiguous region (*der Professorin*) in constructions like (10ab) is most compatible with a serial approach.

(10) a. Daß der Student der Professorin die Klausur überreicht hat, …
That the student (to) the professor the test paper handed over has, …
That the student has handed over the test paper to the professor, …
b. Daß der Student der Professorin die Klausur bestanden hat , …
That the student (of) the professor the test paper passed has, …
That the student of the professor has passed the test paper, …

Among the currently most often discussed issues in this context is the question of what kind of information is used during the first stage of sentence processing. That is, should the first stage make use of all lexical features of the input, or should the parser ignore most lexical features (such as verb-frame information) and base its first analysis decisions only on categorial information (such as whether an item is a noun, a verb, or an adjective, etc.)? Note that this question is rather independent from the question of seriality: whether or not this information is used at the first stage does not restrict the parser to pursue only one or several paths of analysis. The question extends to higher level processes: is world knowledge or pragmatic information able to influence initial parsing decisions, or can it only have an effect after an initial decision has already been made on purely structural grounds. The evidence I present in this thesis strongly suggests that lexical information is fully employed during first analysis whereas higher level information only comes in at a later stage.

As a consequence of the empirical pre-conditions I will present a model of human sentence processing, called the SOUL (Semantics-Oriented Unification-based Language) processor. SOUL is a serial parser based on the modern unification based grammar theory HPSG (*Head-driven Phrase Structure Grammar*, Pollard & Sag, 1994). HPSG is a highly lexicalized grammar, which leaves only some very general mechanisms and principles outside the lexicon. It focuses on the representation of linguistic *signs*, in which the different levels (syntax, semantics, etc.) interact by virtue of non-derivational mechanisms. In SOUL, the information structure of signs is supplemented by a small set of (procedural) *methods* that enable signs to combine themselves with other signs. The object-oriented perspective will turn out to be well-suited for modelling an attachment behavior most compatible with the empirical findings. Attachment preferences will be accounted for in terms of the *visibility* of the potential attachment sites to the sign that attempts to attach itself to the current sentence structure.

As a consequence of lexicalization, it is important to analyze the amount of *transparency* (Berwick & Weinberg, 1985) HPSG allows so it can function as a competence base for a cognitive parser. Besides the issue of first stage usage of lexical information, this will most importantly include the question of incrementality, i.e whether syntactic structure can be built in the absence of lexical heads which provide necessary

information for structure formation. This topic is especially relevant when it comes to parsing the head-final constructions, such as (8), which are quite common in German.

1.3.2 A categorization of SOUL

Psycholinguistic models of human sentence processing can be roughly distinguished into two subclasses: those which follow the *transparency* appoach, and those which refer to the notion of *general cognitive architecture*. According to the *transparency* approach, the properties of the parser can be derived from the property of the underlying grammar. The second, more psychologically oriented direction considers properties of the cognitive architecture, such as limitations of the working memory and the constraints on the information flow between the sub-modules of the language system responsible for parsing phenomena while the relevance of the syntactic competence for performance phenomena varies between the instances of that direction.

The view taken by SOUL is somewhere in between, i.e the parser is assumed to be based on representations provided by a linguistically motivated grammar theory, namely, HPSG. On the other hand, the operations of the parsing mechanism do not automatically fall out of the structure of the grammar system, nor is the syntax processor, as a highly automatic or autonomous system, assumed to consume too many processing resources. What is assumed to be costly, on the other hand, is, roughly speaking, the maintainance of the entities of thought, namely, the discourse model, its entities and their interrelations established during the process of interpretation. Being conscious objects, they consume working memory resources, particularly those located in the central executive (Baddeley, 1986). Importantly however, economy in terms of properties of the discourse model will be demonstrated not to be able to drive the parsing decisions made at the initial stage of structure assembly. Instead, as will be shown, the HSPM (SOUL) is more or less incidentally suited to provide the basis for an economic on-line interpretation of the incoming words of a sentence. This might be called an evolutionary approach in that the syntax module is supposed to have developed as an efficient service module that maps an unstructured input onto an representation which can be semantically interpreted most efficiently. Once again, this is what makes up a *semantics-oriented* parsing approach.

1.4 Overview

The thesis is organized as follows: in chapter 2, I will present "state of the art" approaches in psycholinguistic research on human sentence processing, focussing on theories or models. Each model will be discussed in the light of general evidence on attachment preferences. The collection of models described includes *Parametrized Head Attachment* which has been motivated by evidence from self-paced reading studies on attachment preferences in verb-second and verb-final clauses. Since button pressing techniques suffer from several methodological weaknesses, as discussed in chapter 2, I will present data from eye-tracking studies on a similar set of materials as the self-paced reading studies in chapter 4.1.

In chapter 3, *eye-tracking* will be pointed out as being a sophisticated technique to investigate the moment-to-moment processes that occur during reading. New measures will be introduced that seem to reflect parsing processes more properly than conventional measures.

Empirical evidence about general constraints on human parsing will be presented in chapter 4. The role of lexical heads in parsing will be discussed in the light of empirical evidence from eye-tracking experiments on PP-attachment in isolated verb-second and verb-final sentences. The design includes a variation of the lexical bias of verbs with respect to their expectation of an instrument role.

Since reading isolated sentences may result in artificial reading strategies a further PP-attachment experiment was conducted where the sentences were presented in short contexts which were unbiased in not favoring one or the other interpretation.

PP-attachment ambiguities were disambiguated semantically at the end of the PP. To tap into the attachment process earlier, attachment preferences in verb-final sentences were extended to NP-attachment ambiguities such as (10), which allow for an early grammatical disambiguation through the case of the determiner (i.e. *dem / des Professors*) of the structurally ambiguous NP.

The evidence will be shown to strongly support the principle of *Parametrized Head Attachment*. It will be argued, however, that a certain class of parsing approaches, namely, *head licensing* models, being the strongest variant of *head projection* models, can also account for the data obtained. *Head licensing* will be shown to be incompatible with data on German subject-object asymmetries presented in section 4.1.2.

The notion of *head projection* is strongly related to the status of detailed lexical information in parsing. Chapter 4.1.3 presents a collection of empirical results from psycholinguistic literature, suggesting that detailed lexical information is nevertheless used immediately during initial structure assembly.

In chapter 4.2 I will analyze the predictions of models that assume alternative sentence structures to be built in parallel or serially, or a fully determinate decision to be postponed. The evidence given is supplemented by the findings on NP-attachment ambiguities presented in section 4.1.1. Finally, the question of whether the human parser has to be conceived of as an autonomous input module that cannot be distracted by higher level processes will be addressed. A modular approach will be substantiated by an experiment on PP-attachment in pragmatically biased contexts.

The pieces of evidence presented throughout the thesis will be put together in chapter 5. Based on *Head-driven Phrase Structure Grammar* (HPSG), the *Semantics-Oriented Unification-based Language (SOUL)* mechanism, introduced in this chapter, employs rich lexical representations to simultaneously assemble the structure of a sentence and compose its meaning. The principled behavior of the parser will result in the attachment preferences predicted by *Parametrized Head Attachment*. Predictions for a variety of parsing phenomena will be discussed in comparison to concurrent models.

2 Previous work

In this chapter, I will give an overview of important models of human sentence processing, focussing on models which ascribe processing complexity (at least partially) to the *human syntax processor*³. The overview cannot be exhaustive, of course. The models presented were not only chosen because of their explanatory power but also because of their typicality for a class of processing accounts. The discussion will start in the late fifties and early sixties, where modern psycholinguistics and many ideas which still have considerable impact have their roots, and conclude with models published only recently. The presentation is ordered fairly chronologically. Where possible and reasonable, the models will be classified according to various dimensions: is a *fully specified* phrase structure representation built on-line? Can several structures be pursued in *parallel* in cases of structural ambiguity? If not, do alternative structures at least *compete locally*? Is structure *predicted top-down* or can it only be *projected bottom-up* from the lexicon? Which kinds of grammatical information are used in the first analysis?

In chapter 4, I will discuss the classes of models resulting from the different dimensions on the basis of a broad range of empirical evidence, grounded on experiments presented in this thesis to an important degree. General implications of the different classes will be weighed in the light of the data. In this chapter, however, the models will mainly be discussed with respect to the question of whether or not they meet their own explicit goal of predicting parsing difficulties properly. These predictions

^{3.}Some important approaches may appear to be missing in this chapter. Crain's and Steedman's (1985) and Altmann's and Steedman's (1988) model, for instance, will not be presented here, because they do not attribute parsing complexity to the human syntax processor. They will be discussed, however, in the context of general architectural assumptions on modularity in chapter 4.

mostly result from model-specific parsing assumptions. Furthermore, inconsistent or implausible presumptions will also be pointed out here.

2.1 The contribution of syntax to sentence understanding

2.1.1 Chomsky and Skinner

The psychology of language in the early fifties was dominated by the view that the study of human language processing is in general independent of linguistic concerns. Bloomfield, for instance, a linguist whose structuralist approach (Bloomfield, 1933) introduced a new influential paradigm of language research, considered himself a *behaviorist* with respect to language *use*. *Behaviorism* was at its second zenith in the early fifties, and it was Skinner (1957), in particular, who offered a behaviorist approach to language in his book "Verbal Behavior". In claiming that verbal behavior is just a subtype of behavior in general, Skinner believed he could apply the same type of laws that had been shown to be successful in the explanation of the learning of certain behavioral (motoric) sequences in animals to the language domain. The organism's mental state, as an *intervening variable* between the observable input and output, was considered irrelevant. Instead, behavior should be explained in terms of input/output laws, relating the *stimulus*, i. e. what an organism perceives, to how it responds.

It was Chomsky (1959) who pointed out some severe weaknesses in the behaviorist approach to language. Firstly, he convincingly worked out the inherent flaws and inconsistencies of Skinner's framework, namely, the inconsistent use of the terms *stimulus* (which should be described in terms of its physical properties only) and *behavior* (which should be described in terms of its function rather than its form). In particular, stimuli were often identified in terms of the responses they produce rather than their physical characteristics. It is impossible, for instance, to define the properties of a physical object that causes someone to remark on its beauty, in terms of its physical attributes (c.f. Garnham, 1985). The claim that verbal behavior can be explained using the behavioristic concepts is therefore untestable, since in observing a response, it is always possible to name *some* stimulus that is somehow related to it.

Chomsky demonstrated that linguistic behavior is not determined by characteristics of the environment only. There is no way of explaining verbal behavior without reference to mental constructs, such as intentions, beliefs, and, most importantly, the *knowledge* a person must have in order to use a language. Chomsky identifies this knowledge with linguists' discoveries. With this claim, he takes a position which is diametrically opposed to Bloomfield's view. It is the renaissance of the 'rationalist' way of thinking initiated by Chomsky that can be considered as the birth of modern psycholinguistics and, to some substantial extent, the birth of cognitive science.

In a second, more abstract and mathematical type of argument, Chomsky (1957) falsified virtually all work in the psychology of language in the 1950s and before, which, inspired by *information theory* (Shannon & Weaver, 1949), Miller & Selfridge (1950), had described language processing as a sequence of transitions from one men-

tal state into another. Since the brain has only a finite set of neurons, is was assumed that the set of mental states must be finite in number as well. For each state, such as just having read the word "George", there are several possible transitions to the next state, such as reading (or speaking) the word "laughs" or the word "thinks", etc. The probability for a particular transition depends on the state it starts out from. Of course, the predictability of a transition increases with the number of previously passed states able to be taken into account.

Chomsky (1957) demonstrated that the mechanism, which is implicit in both Miller's (1951) and Skinner's (1957) models, has important limitations. Such a mechanism could be described as an abstract machine, a so-called *finite state automaton*. Using mathematical methods, Chomsky showed that a finite state automaton could never be set up to produce all the sentences of a natural language and, importantly, *only those*⁴. The problem can be illustrated with a series of so-called *center-embedded sentences*, such as (11b,c; cf. Garnham, 1985).

(11) a. The war ends the world

- b. The war that the general starts ends the world.
- c. The war that the general that the president ... appoints starts ends the world.

While (11b) adds a relative clause to the subject-NP in (11a), (11c) does much the same to the subject-NP in the relative clause in (11b). Although (11c) is somewhat harder to process, there is no reason to reject (11c) as an ungrammatical sentence, because the same grammatical rules apply to (11b) and (11c). It is obvious that there could be even more embeddings of the same type, in principle infinite in number. In a monolithic finite state automaton, however, each additional embedding would have to be specified via a distinct path through the transition network, requiring an *infinite* amount of states in the automaton for an infinite number of embeddings.

2.1.2 Competence and performance

An important presumption in this line of argument is Chomsky's distinction between the language *competence* of speakers/hearers, and their actual language *performance*. The language *competence* is the tacit grammatical knowledge shared among all native speakers of a language, which an *ideal speaker* would be able to use without any limitations. Real speakers, however, are handicapped by several weaknesses in the human cognitive system, such as limitations in the working memory capacity, etc. When real humans show difficulties in processing sentences like (11c), it is therefore not because their language competence forbids the generation of such a sentence, but because of certain weaknesses in their performance system.

Chomsky's competence/performance distinction started an on-going *epistemological* argument: since linguistic evidence, namely, judgments about the grammatical

^{4.}Of course, it is a trivial job to set up an automaton that generates any sequence of words from a language vocabulary, with the correct sentences of the language among them. It is, however, impossible to restrict a finite state automaton in a way that all non-sentences are rejected while all grammatically correct sentences can still be generated.

well-formedness of sentences, is necessarily based on intuitions of non-ideal speakers, it is in principle impossible to attribute the acceptability of a particular sentence to either *competence* or *performance* factors. Note that this problem arises only because Chomsky identifies grammar with the human syntactic knowledge of language. One could also argue that a grammar theory represents just an external formal description of what might be represented *mentally* in a completely different and less complete fashion, e.g. as a finite state grammar⁵.

2.1.3 Standard Theory (Chomsky 1965)

Chomsky provided an elaborate theoretical framework, which has prevailed in linguistic research and even the psychological research of language in the following decades. In order to illustrate the latter it is unavoidable to sketch his linguistic theory, at least briefly.

The grammar of a language is defined as the system of rules that generates all and only the well-formed sentences of the respective language. To be descriptively adequate, grammar not only has to generate the correct sentences but also provide adequate structural descriptions for them. Two grammars are considered weakly equivalent if they generate the same set of sentences. They are strongly equivalent if they generate the same structural descriptions. The goal of generative grammar theorists goes even one step further: an explanatorily adequate grammar theory has to be descriptively adequate, but it must also explain how the grammar of the respective language can be learned by children, given the linguistic input available to them. One of the central claims of Standard Theory and all its successors is that there is a universal biologically determined basis for the human capacity to learn language, the socalled language acquisition device (LAD). Grammar, according to this view, can only be explained if the principles underlying the language faculty, the mental organ for language learning, or the language instinct, as Steven Pinker (1994) calls it, are uncovered. In this section, I will only state some of the major assumptions of the early years of generative grammar, the Chomsky (1965) version in particular, which are basic for the following discussion. Some more details will be given when particular models of human sentence processing are discussed.

Example (11), presented earlier in this chapter, demonstrates the *recursive* nature of natural language grammars. Recursive grammars allow a certain type of phrasal unit, say an S (for sentence), to appear as a constituent of that phrasal unit itself. For example, one could put forth a rule (12) for English sentences, like (11), which says that a sentence (S) consists of a noun phrase (NP) and a verb phrase (VP).

(12) $S \rightarrow NP VP$

A noun phrase can combine with a relative clause (S[´]), expressed by rule (13)

^{5.}This is the position taken by other linguists, such as Katz (1981), Gazdar (1985), Montague (1974), and others, who consider themselves mathematicians (and therefore "Platonists", as opposed to "rationalists") rather than cognitive scientists.

(13) NP \rightarrow NP S^{\prime} (in itself recursive)

A relative clause, then, can consist of a relative pronoun (pron), and a sentence.

(14) $S' \rightarrow \text{pron } S$

Note that there is an S on the right hand side of rule (14), which could then be expanded by rule (12) again, starting a new cycle. A finite set of recursive grammar rules is thus capable of generating an infinite amount of constructions, in this way reflecting the productivity of natural languages.

Context free phrase structure rules as (12) to (14), however, cannot account for the commonalities of sentences like (15 a, b or 16 a, b).

- (15) a. I would like to read this book.
 - b. This book, I would like to read.
- (16) a. Peter gave the book to Mary.
 - b. Peter gave Mary the book.

Of course, rules can be found that generate (15a) or (16a) as well as (15b) or (16b), but the intuitive feeling that, in a deeper sense, (15a) and (15b) or (16a) and (16b), respectively, describe the same situation cannot be represented. Therefore, only weak generative capacity is assumed for context free grammars; i.e. context free grammars are considered weakly equivalent to the "real" grammar of natural language: they may be able to generate all and only syntactically well-formed sentences, but they do not provide adequate structural descriptions in all cases. To reach strong generative capacity of the grammar, two levels of structural representation were introduced in Standard Theory, the deep structural representation generated by context free phrase structure rules, and the surface representation, which is derived from the deep structure by a context sensitive component, namely, transformations. A context sensitive dative shift rule (17) could be assumed that derives (16b) from (16a), for instance.

(17) verb NP_i PP_k(to) \rightarrow verb NP_k NP_i

In Standard Theory, deep structure was the interface to semantic interpretation. The phonological form of the sentence, on the other hand, was generated from the sentence surface structure. However, in more recent variants of the Chomskian approach, e.g. Government and Binding Theory (Chomsky, 1981, 1986, as well as in earlier versions of the theory, e.g. Chomsky, 1957), this does not hold true anymore. Interpretational phenomena depending on the scope of quantifiers (e.g. *every* and *two*) on the surface structure showed that logical form cannot be derived from deep structure (18 a,b).

- (18) a. Every teacher at this school hates two pupils.
 - b. Two pupils, every teacher at this school hates.

Though both (18a) and (18b) are ambiguous with respect to quantifier scope, the preferred reading of (18a) is that for all teachers there are two pupils (which may be

different ones for each teacher) whom the teacher hates. The preferred reading of (18b), on the other hand, is that exactly two pupils exist who are hated by every teacher at this school. Examples like these show that surface structure must be taken into account for the derivation of the logical form of a sentence.

2.1.4 Derivational Theory of Complexity (Miller, 1962)

It was George Miller who introduced transformational grammar (TG) to psychology. Miller put forward the hypothesis that the language understanding system uses the grammar in a straightforward way, though he focused his attention solely on the transformational component of TG. In his theory, later referred to as the "Derivational Theory of Complexity" (DTC), he hypothesized that the difficulty in understanding a sentence was primarily determined by the length of derivation, i.e the number of transformation rules to be applied to the underlying kernel structure in order to achieve the correct surface string. The processing mechanism in DTC can briefly be sketched as follows: first, the surface phrase marker is computed from the sentence to be processed, though Miller is not very clear about how this should be done. Then, the transformational rules are applied in reverse to the surface phrase marker, i.e. the right hand side is matched against it, in order to recover the phrase marker of the underlying kernel string at the deep structure level. So if, for example, the phrase marker matches the *output* of a passive rule, the corresponding active form is derived by reversing the transformation. The resulting kernel structure is then passed to an interpretative component, together with tags indicating the applied transformations ("*kernel plus tag*" hypothesis)⁶.

Miller proposed that each step in derivation lasts a certain, though unspecified, amount of time. In addition, transformations were applied in a serial fashion. Thus, a sentence which requires additional transformations, compared to other sentences, should take more processing time than the latter. Indeed, early studies appeared to support this hypothesis. But the more sophisticated the experiments became, the more the results suggested that at least some transformations assumed by TG did not induce difficulties in understanding a sentence (see Fodor, Bever, Garrett, 1974). Movement of the particle in (19a), for instance, should have rendered this "transformed" structure more difficult than (19b), but no increased processing load could be established empirically. In general, this approach suffered from the problem of rapidly changing assumptions on what was considered a transformation in different versions of TG. Thus, psycholinguists had a hard time to keep pace with the latest theoretical developments in their empirical research.

- (19) a. John phoned the girl up
 - b. John phoned up the girl

The evidence from sentences like (19) has never seriously been taken to falsify TG itself. Although Chomsky was intrigued by DTC in the beginning (see Chomsky and

⁶.DTC was based on the 1957 version of transformational grammar, which assumed some semantic information at the surface structure level.

Miller, 1963), as soon as counter-evidence emerged, he denied that TG was intended to model actual language use. The notion of *generation* was only meant in an abstract mathematical (functional) sense without any actual procedural implications. In that sense, psycholinguistic evidence could in principle never falsify a linguistic theory of competence. Not surprisingly, many psycholinguists lost interest in TG as a foundation of a psychological model of language use, at least until the mid-seventies.

DTC only concentrated on the psychological reality of transformation, an approach that did not turn out to be very successful. Bever (1970) and Fodor, Bever, and Garrett (1974), on the other hand, claimed they had shown at least the syntactic deep structures to be psychologically real. In a famous series of click location experiments, they were able to demonstrate that subjects were sensitive to phrasal boundaries when reproducing the position where they had heard a "click" while listening to a sentence. After hearing the full sentence, subjects reported hearing the clicks closer to major syntactic boundaries then they actually were. Furthermore, the apparent displacement of the clicks was directed more strongly towards deep structure boundaries than towards surface structure boundaries. Though the evidence suggested that phrase boundaries may have played some role in these tasks, it was by no means clear whether the effects took place when the sentences were processed or only later when the positions of the clicks had to be reconstructed. Furthermore, like many other psycholinguistic results from that period, syntactic variations were indistinguishably confounded with semantic aspects, since major syntactic boundaries very often coincide with semantic boundaries.

Different accounts of competence and performance emerged during that period. Many psychologists as well as linguists returned to a more Platonist view of linguistics, claiming that linguistics is only in charge of providing *elegant* and parsimonious theories, independent of whether or not the rules and principles could ever be claimed to be mentally real. A somewhat stronger view was taken by many psychologists (and sometimes still is; e.g., Herrmann, 1985), claiming that there is no such thing as linguistic knowledge *besides* perceptual mechanisms.

2.2 Perceptual strategies, Bever (1970)

Since the arguments against the DTC as a theory of language use seemed convincing, it was apparent that a different approach was required to deal with the way in which the "internal representation" of a sentence is formed in the hearer's mind. Bever (1970) proposed some mechanisms independent of grammar labelled "perceptual strategies" in analogy with heuristics and mechanisms involved in visual perception. The empirical evidence from the *click location* experiments and others (see Garrett, Bever, & Fodor, 1966; Fodor, Bever, & Garrett, 1974; Chapin, Timothy, & Abrahamson, 1972) was taken as an indicator of the role of clauses as basic processing units. Bever (1970) offered the following strategies (among others).

(p2) In any sentence, the first clause with a ... Noun ... Verb ... (Noun) ... sequence is the main clause, unless it is marked as subordinate by a subordinating conjunction.

- (p3) Any Noun Verb Noun sequence is to be interpreted as actor-action-object.
- (p4) After a determiner, which signals the beginning of a Noun Phrase, the end of the NP is indicated by one of the following: (i) a plural morpheme, such as -s, (ii) a morpheme that indicates the beginning of a new phrase, such as the, that, will, may, or should, (iii) a word that is probably not a noun, for example, a verb that is only rarely used as a noun.

Perceptual strategies like these were supposed to explain the immense processing difficulty that consistently shows up in sentences like (20), where readers are *lead up the garden-path* in their first analysis.

(20) The horse raced past the barn fell.

The verb *raced* can either be analyzed as the main verb of a matrix clause (*The horse raced past the barn and fell.*) or as a past participle within a reduced relative clause (*The horse that was raced past the barn fell.*). Obviously, the main verb reading is strongly preferred (p2), rendering sentences like (20) nearly impossible to process.⁷ One of the major pitfalls of the perceptual strategy approach was the incredible amount of strategies that would have become necessary to explain preferences in ambiguous structures, the recovery of the unpreferred alternatives, and the whole spectrum of easily processable but still highly complex structures. It is obvious that building a consistent system of strategies independent of grammar able to handle all kinds of parsing problems is an unsolvable problem. The perceptual strategies would have had to duplicate the full system of the competence grammar, rendering one of the systems superfluous.

Since grammatical constraints are required anyway, the less costly way is to employ the knowledge specified by the grammar for the parsing process. Thus, psycholinguistics moved back to competence based models of human sentence processing in the early seventies.

However, most of the models of human sentence processing developed from the early seventies until today differ from former models in several important ways: firstly, there was a shift in the topic of research. Bever prepared the ground for a more processing oriented approach, centered around preferences in ambiguity resolution, instead of finding evidence for the psychological reality of the latest invention of generative grammar. Processing difficulty was not (at least not only) derived from the direct application of grammar rules but from a complex interaction of the representation of grammatical knowledge and independently motivated parsing principles generated by the architecture of the human parser. Kimball's seven surface structure parsing principles (Kimball, 1973) were the first step in a tradition of modern psycholinguistic research that aimed at finding general and universal principles of sentence processing.

^{7.}In fact, many native speakers consider (20) ungrammatical.

2.3 Kimball's principles for surface structure parsing

Kimball (1973) presents a two-stage model of parsing: in the first stage the surface structure of a sentence is computed on the basis of surface grammar rules (which need not fully overlap with the grammar rules assumed to hold for deep structure by a grammar theory). In the second stage, deep structure is derived from the surface structure.

Whether or not a particular sentence is easy to process depends on processing principles claiming, for instance, that sentences are parsed *top-down* or that no more than two sentences can be analyzed at the same time (thus accounting for differences in parsing deeply embedded clauses like 11c). At least three of Kimball's principles have been highly influential in successive work. Therefore, they will be presented in a little more detail:

(p5) *The principle of Right Association (Kimball, 1973)* Associate a terminal symbol to the lowest non-terminal node

In Kimball's model, the principle of Right Association is assumed to explain the preference for attaching the particle *out* to the verb of the embedded infinitival clause (*take*) in (21) as well as the weirdness of (22), where the initial analysis yields the implausible interpretation that the train is being taken out.

- (21) John figured that Susan wanted to take the cat out.
- (22) John figured that Susan wanted to take the train to New York out.

If attachment decisions have been made according to any of the proposed principles, revision of the structure is assumed to be computationally costly as noted in the Principle of Fixed Structure.

(p6) The Principle of Fixed Structure (Kimball, 1973)

When a phrase has been closed it is computationally costly to reorganize its constituents.

A preference for non-recursive structures is predicted by (p6, Kimball, 1975). By avoiding a duplication of nodes, the problem of infinite left-recursive loops in a top-down algorithm is circumvented.

(p7) If the parser can choose between a hierarchical structure of the form "A ... B ... C", where A dominates B and B dominates C, and the form "A ... B ... B ... C", the " canonical "A ... B ... C" is chosen.

Although Kimball alludes to the architecture of the human language processing device, the surface structure parsing principles are formulated rather independently from the processing architecture. Frazier and Fodor (1978) went one step further, both in the generality of the principles and in their foundation on architectural constraints.

2.4 The Sausage Machine

The Sausage Machine by Frazier & Fodor (1978, Fodor & Frazier, 1980) is the "classic" approach that aims at providing an explanation for parsing preferences based on the architecture of the human sentence processing device. The Sausage Machine is also a two-stage processor like Kimball's model, though the architecture is rather different: the first stage, the so-called *Preliminary Phrase Packager (PPP)* assigns lexical and phrasal nodes to strings of words within a narrow window of five or six words⁸, inspired by Miller's (1956) famous account on working-memory limitations.

When phrases have been packed, they are deleted from the PPP and handed over to the so-called Sentence Structure Supervisor (SSS) that combines the phrase packages into a phrase structure representation of the whole sentence. Revising structure that has been built by the PPP or the SSS, respectively, is assumed to be costly (Revision as Last Resort, see also Kimball's principle of Fixed Structure, p6). The "shortsightedness" of the PPP is claimed to be responsible for parsing preferences in sentences as (23 b), where the prepositional phrase for Susan can either be attached to the directly preceding verb (obtain) or to the verb of the matrix clause (bought). Obviously, there is a preference to attach the PP to the more recent verb. In the Sausage Machine, this is due to the fact that the verb of the matrix clause is not visible from within the PPP when the preposition for is encountered. Things look different, however, for sentences like (23 a), where the PP for Susan can either be attached to the VP of the sentence as an argument of the main verb (*bought*) or to the directly preceding object NP *the book*. Both attachment sites lie within the range of the PPP. In these cases, a principle that is assumed to hold for the PPP as well as the SSS applies, namely Minimal Attachment, yielding the least complex syntactic structure compatible with the input.⁹

(p8) Minimal Attachment

Each lexical item (or other node) is to be attached into the phrase maker with the fewest possible number of non-terminal nodes linking it with the nodes which are already present.

- (23) a. John bought the book for Susan.
 - b. John bought the book that I had been trying to obtain for Susan.

Frazier and Fodor claim that there is empirical evidence of a preference for VP attachment of the PP in sentences like (23 a). On the basis of a simple phrase structure grammar, it can be demonstrated by the syntactic alternatives for sentence (23a) that these findings meet the predictions of *Minimal Attachment*: when read from left to right, an attachment-conflict occurs on reading the article of the object-NP because it is possible either to attach the object-NP directly to the VP (24a) or to postulate a complex NP with an extra NP node (24b). Because of this extra NP, *Minimal Attachment*

^{8.}The proper unit of "items" in the PPP does not necessarily have to be words. Frazier and Fodor note that syllables or morphemes might turn out to be more appropriate.

⁹*Minimal Attachment* can be regarded as a generalized version of Kimball's simplicity principle (p6).

will hinder the construction of (24b) in favor of a direct integration of the object NP into the VP (24 a).

- (24) a. [S [NPJohn] [VP [bought [NP the book] [PP for Susan]]]
 - b. [S[NP John] [VP bought [NP [NP the book] [PP for Susan]]]]

Minimal Attachment is assumed to be the result of a race based account of phrase structure rule application: rules are expanded in parallel, and those rules which first arrive at preterminal nodes that can be matched against the input win the race. Thus, less complex phrase structures, requiring less expansions, are always preferred. Frazier and Fodor (1980) concede, however, that in the adult sentence processor, rules favored by Minimal Attachment may have acquired a "head start" over the others due to the frequency of successful application in the learning history. The staggered parallel search mechanism may only be adequate for the language learner, becoming more and more serial with experience.

The parsing preferences predicted by the limited capacity of the PPP overlap with predictions of Kimball's principle of *Right Association* to a large extent. But the overlap is not complete, the predictions differing in interesting ways. The PPP constraints, for example, also appear to explain the difference in the acceptability of sentences like (25) and (26). In (25), the PPP tries to combine the NPs *the rat, the cat,* and *the dog* into a single phrase package, yielding a preference for a conjoined reading of the three NPs which turns out to be incompatible with the following input. Lengthening the constituent as in (16), on the other hand, should facilitate the analysis because the constituents passed to the SSS can be integrated into the correct embedded structure more readily (the vertical lines indicate plausible packages of five or six words).

- (25) The rat the cat the dog bit chased ran away.
- (26) The beautiful young woman | the man the girl loved | met on a cruise ship in Maine | died of cholera in 1962.

The preference for *local association* resulting from the shortsightedness of the PPP was claimed to explain not only the well known right association preferences but also the preference to attach the adverb *yesterday* in (27) to the sentence to the right of it (*yesterday she announced* ...). The verb *claimed* was assumed to be too far from the adverb *yesterday* to be a good candidate for attachment compared to the matrix verb *announced*.

(27) Though Martha claimed that she will be the first woman president yesterday she announced that she d rather be an astronaut.

Ford, Bresnan, and Kaplan (1982), however, argue that no left association preference has to be assumed for (27), since complement clauses (or S' in general) have a strong tendency to occur in sentence final positions. Thus, early closure of the subclause (*Though Matha claimed ...president*) can easily be accounted for by a constituent ordering preference.

Wanner (1980) demonstrated that "local association" like phenomena occur even in sentences, such as (28) and (29), which should be parsable within the narrow window of the PPP. As Fodor and Frazier (1980) conceded, local association fails to make the correct right association predictions for sentences like (28) where all potential attachment sites lie within the limited window of the PPP. The adverb *yesterday* in (28) can either be attached to the matrix sentence (*Barbara said*) or to the embedded sentence (*Chris married*). Attachment to the lower embedded sentence is strongly preferred. Thus, a low right association preference has to be assumed independently from any local association preferences. Wanner (1980) presented an alternative approach to account for the observable preferences, as discussed in the next section.

(28) Barbara said Chris married yesterday.

(29) Women men girls love meet die.

In the years following this discussion, Frazier and colleagues have abandoned the idea of a "shortsighted" preprocessor. However, the core principles of the Sausage Machine, *Minimal Attachment* and *Right Association* found their place in the successive theory of Frazier and colleagues, the *Garden Path Model*, which will be discussed in section 2.6.

2.5 Augmented Transition Networks (ATNs)

Wanner (1980; Wanner & Maratsos, 1978) proposed a variant of an Augmented Transition Network (ATN, Woods, 1970), a one-stage top-down parsing architecture, to account for the evidence.

An ATN consists of a set of transition networks each of which usually specifies a particular type of phrase (S, VP, NP, PP, etc.). These transition networks consist of states and arcs which have to be traversed to get from one stage in processing to the next one. The arcs in these networks are labeled specifying which action is to be taken for permitting a transition. The types of arcs assumed in Wanners network are: CAT for arcs asking for a particular category (e.g. noun, verb, etc.); SEEK for arcs asking for phrases that are specified within a different subnetwok; SEND arcs terminate a network when it has been successfully passed through, thus rendering the goal of the previously encountered SEEK arc the new current state of the parsing process; JUMP arcs permit free transitions between states, expressing the optionality of certain sub-paths; WORD arcs can only be traversed if the current input is a particular word or morpheme (e.g. *to, of*).

According to Wanner, the predictions made by *Right Association* and *Minimal Attachment* can easily be accounted for by general arc ordering constraints in an ATN: for *Right Association*, all SEND arcs and all JUMP arcs have to be scheduled after every other arc type. For *Minimal Attachment*, all CAT and WORD arcs have to be scheduled before all SEEK arcs. Parsing preferences for sentences like (20), (21), (22), (28) and many others can be predicted by arc ordering.

Two major criticisms have been advanced against Wanner's ATNs by Frazier and Fodor (1980): firstly, the scheduling of arcs does not reflect general parsing principles, but appears to be a rather ad hoc approach yielding the correct descriptions of many parsing preferences. On the contrary, the principles of *Minimal Attachment* and *Local Association* emerge from the architectural properties of the *Sausage Machine*.

Secondly, there are parsing phenomena which are predicted by *Minimal Attachment* and *Right Association*, but not by Wanner's arc ordering. In those cases, where there is a conflict between two arcs of the same category, Wanner's ATN does not predict any differences. Crucial examples are the parsing preferences in sentences like (30 a, b), where the less complex attachment of the NP *the theory* as the direct object of the matrix verb *knew* (30 a) is supposed to be preferred. For both structural alternatives, a SEEK arc has to be traversed, a SEEK NP arc in the case of (30 a) and a SEEK S arc in the case of (30 b). Minimal Attachment correctly predicts a preference for the less complex object NP reading.

(30) a. The student knew the theory by heart.

b. The student knew the theory was wrong.

Both problems cannot be regarded as pitfalls of ATNs per se, since they can be overcome with some changes to the architecture of the model. One could, for example, order arcs with respect to their frequency of use. Since SEEK NP arcs are surely traversed more often than SEEK S arcs, the preference for (30 a) can be predicted¹⁰. Furthermore, Wanner suggests that if a conflict arises between two arcs of the same type (say SEEK), the decision is based upon e.g. the "bar level" in the terminology of X-bar theory¹¹ (Jackendoff, 1977; Chomsky, 1981; Stowell, 1981). Traversing arcs which point to networks with "lower" bar levels, i.e. which are closer to the lexical input, should be preferred.

From a technical point of view, the ATN approach was abandoned when it became clear that more realistically complex systems soon become intractable, partly due to the unfortunate combination of declarative and procedural kowledge in one and the same fairly over-powered formalism. From a theoretical point of view, linguists have only rarely been appealed by ATNs due to the almost complete lack of the capability to express linguistically motivated generalizations, such as the *principles* in more recent versions of TG.

i. [XP [YP [X' [X⁰ ZP]]]]

The maximal projection (XP) as well as the X^0 are obligatory X-bar levels, whereby there is some discussion as to the necessity of the intermediate X' level(s) in cases where they do not branch.

^{10.}In this sense, ATNs seem to be a natural competence base for frequency based parsing theories (e.g. Mitchell, 1994).

^{11.}Phrases are built up following the X-bar schema which states that complements (ZP) have to be realized as sisters of their lexical heads (X⁰), and specifiers (YP) as sisters to an intermediate X' level. The lexical head determines the type of phrase to be projected; i.e. nouns project to noun phrases (NPs), verbs to verb phrases (VPs) etc.

2.6 Garden Path Theory

In *Garden Path Theory*, the parser is presumed to work incrementally word-byword in a serial depth-first manner, such that in cases of structural ambiguity only one analysis will be pursued (*first analysis constraint*). If the first analysis later proves to be unacceptable according to syntactic, semantic, or discourse-pragmatic reasons, a measurably time consuming reanalysis will be induced (see Kimball's principle of Fixed Structure, p6, and *Revision As Last Resort, RALR*, Frazier and Fodor, 1978, see section 2.4). The principles of *Minimal Attachment* and *Right Association* are reformulated as (p9) and (p10), respectively (Frazier, 1987a).

(p9) Minimal Attachment (MA)

Do not postulate any potentially unneccessary nodes.

(p10) Late Closure (LC)

If grammatically permissible, attach new items into the clause or phrase currently being processed (i.e. the clause or phrase postulated most recently).

As in the Sausage Machine, MA ensures that in cases of structural ambiguity the parser builds the simplest possible structure during the first analysis. LC guarantees that new constituents are immediately attached to prior material, minimizing the chance of exceeding working memory limitations. In order to work properly, MA has to be given priority over LC in cases where both principles could apply.

Minimal Attachment and *Late Closure* are claimed to apply to parsing preferences in a wide range of structures, such as those in (31) and (32) taken from Frazier and Clifton (1996, pp. 11/12, the preferred reading is given first):

- (31) Minimal Attachment based preferences
 - a. Main clause / reduced relative
 [The horse raced past the barn] ?fell.
 [The horse [S' raced past the barn] fell].
 - b. NP versus S complement
 John [knew the answer to the physics problem] ?was wrong.
 John knew [s the answer to the physics problem was wrong].
 - c. NP conjunction versus S conjunction Jacob [kissed [Miriam and her sister]] ?laughed. [Jacob kissed Miriam] and [her sister laughed].
 - d. PP-attachment to VP/NP Sandra [wrote [a letter] to Mary]. Sandra wrote [a letter to Mary].
 - e. Complement / relative clause
 John told the girl [that Bill liked the story].
 John told [the girl that Bill liked] the story.

- f. Attachment of NP as second object / relative on first object
 Fred [gave the man the dog] ?bit the package.
 Fred gave [the man the dog bit] the package.
- (32) Late Closure based preferences
 - a. Direct object versus subject of S2
 While [Mary was mending the sock] ?fell off her lap.
 While Mary was mending [the sock fell off her lap].
 - b. Attachment of PP to lower clause / higher clauseI put [the book that you were reading in the library] ?.I put [the book that you were reading] in the library.
 - c. Attachment of S to lower clause / higher clause
 Fred will realize [that Mary left when the party ?starts].
 Fred [will realize [that Mary left] when the party starts].
 - d. Attachment of adverb to lower / higher clause
 We remembered [that the assignment will be due ?yesterday].
 We [remembered [that the assignment will be due] yesterday].

The Garden Path Theory (Frazier & Rayner, 1982; Frazier, 1987; 1990) proposes an autonomous syntax module which cannot be guided by higher-level processes, such as semantic or pragmatic interpretation. The latter is assumed to be accomplished in the *Thematic Processor* that proposes alternative interpretations if the structurally preferred analysis fails.

The status of lexical information is somewhat uncertain. According to the *lexical fil-ter hypothesis* (Mitchell, 1987, 1989), only the basic syntactic category of a word is considered during the first analysis. More detailed lexical information, e. g., about potential complements of a word, is proposed "to reject or to confirm whatever analysis has been constructed on purely syntactic information" (Frazier 1987b, pp. 524; but see Clifton, Frazier, & Connine, 1984, for evidence for the early use of verb-frame information).

To sum up, the Garden Path Theory appears to account for a wide variety of phenomena in a very parsimonious way.

Most of the parsing theories argue for their particular processing account referring to the Garden Path Model. A variety of criticisms with therefore be discussed in the context of alternative approaches.

2.7 Lexical preferences: the Theory of Syntactic Closure

The Theory of Syntactic Closure (Ford, Bresnan and Kaplan, 1982) involves a highly elaborate parsing model that, more than any preceding model, was set up to link procedural considerations to an elaborate linguistic grammar framework, namely *Lexical Functional Grammar* (LFG, Bresnan, 1978). In using a linguistic theory as a competence base in a cognitively motivated parser, Ford and colleagues believed they provided a theory of maximal simplicity, strength and explanatory power by offering a coherent and unified approach to a variety of seemingly unrelated phenomena, such as lexical expectation effects in gap-finding (Fodor, 1978), as well as the early closure and late closure phenomena presented in the previous sections.

The model was in large parts motivated by the observation that PP-attachment preferences, contrary to the predictions made by Minimal Attachment, seem to depend on the actual verbs used in the sentences. For instance, whereas the preferred reading of sentence (33a) is that the book was [*positioned on the rack*] by John, there is a preference to interpret [*the book on the rack*] in (33b) as that which John wants.

- (33) a. John positioned the book on the rack.
 - b. John *wanted* the book on the rack.

The theory of syntactic closure strongly influenced the model introduced in this thesis (although the actual differences will later turn out to be fairly strong, see chapter 5). I will therefore continue to discuss the theory in more detail.

In the theory of syntactic closure, it is presumed that rules of the competence grammar are used to construct the internal representations in sentence perception. A simplified set of rules for English is given in (34), with irrelevant LFG-specific function assignment annotations omitted (except some abbreviations in square brackets "[]").

- (34) grammar rules for English
 - a. $S \rightarrow NP VP$
 - b. VP \rightarrow (AUX) V (NP) ({NP, PP}) PP[*PCASE*]* PP[*ADJ*]* (S^{\prime})
 - c. NP \rightarrow (DET) ADJ* N
 - d. NP \rightarrow NP {PP, S'}
 - e. NP \rightarrow NP[*SUBJ*]'s VP
 - f. NP $\rightarrow e$

The parentheses "()" indicate optionality, such that "(NP)" is equivalent to "{NP, *e*}", meaning that either an NP is expected or the empty element *e*. Note also that a superscript asterisk on a category indicates that this category may be repeated any number of times, including not at all.

What gives rise to closure effects in the theory of syntactic closure is the order in which these grammatical rules are applied. This order is "jointly determined by strengths of alternative lexical forms, the strengths of alternative categories in the expansion of syntactic rules, and the sequence of hypotheses in the parsing process." (Ford, Bresnan, & Kaplan, 1982, pp. 741). Grammar rules are applied in a serial fashion: only one structure is initially obtained. With each application of a rule, both constituent and functional structure is built at the same time, such that syntactic decisions can be affected by functional as well as lexical information. Data-driven analysis procedures interact with rule driven (top-down, left-to-right) analysis procedures in the way outlined below.

Two types of memory are proposed: the *chart*, and the *agenda*. The *chart* saves all the well-formed substrings found at a given point in the analysis, such that the parser can easily recover them for reanalysis. The *agenda* contains a list of options that could have been taken at different positions in the string. A processor takes one option at a time and executes the operations appropriate for that option. Two types of options are distinguished: *hypothesizing* and *attaching* complete constituents. In more detail, closure effects result from a number of scheduling principles for options on the agenda, the most important of which is *Lexical Preference*, as shown in (35).

(35) Lexical Preference (determines hypothesizing)

"If a set of alternatives has been reached in the expansion of a phrase structure rule, give priority to the alternatives that are coherent with the strongest lexical form of the predicate." (pp. 747)

In order to let *Lexical Preference* have an influence on parsing, it is assumed that if a verb has more than one lexical form, the forms are ordered according to their *strength* or *salience*. (No attempt is made to give an account of *strength* other than the mere assumption that it could be determined by the frequency of the lexical form in every-day conversation.)

Let us consider a short example to illustrate the issue at hand. The sentence to be parsed is given in (36). Suppose the verb *position* has the two forms <(SUBJ), (OBJ), (PCOMP)> and <(SUBJ), (OBJ)>, the former being the stronger one.

(36) $_1$ John $_2$ positioned $_3$ the $_4$ book $_5$ on $_6$ the $_7$ rack $_8$.

Let us assume we step in after the direct object *the book* has been read. At that position, the VP-rule is expanded up to the point marked by " | ".

(37) $_2$ VP \rightarrow V NP $_5$ | ({NP, PP}) PP[*PCASE*]* PP[*ADJ*]* (S[^])

The next constituent expected by this rule is given by ({NP, PP}), which is either another object NP, a PP, or *e* (nothing). Since the only argument not already occupied in the strongest lexical form of *position* is a PCOMP, the PP will be hypothesized as a verb-argument, to which *on the rack* can be attached after it was read (*invoked attachment* (38))¹².

(38) Invoked Attachment (attach to where a constituent has been hypothesized, can lead to minimal attachment kind of effects)
"When there are alternative options for attaching a phrase into a structure, give (default) priority to the option for attaching the phrase to the partial constituent that caused the phrase to be hypothesized." (pp. 754)

In contrast, the verb in sentence (36) is *wanted*, which is presumed to have the 2-place form <(SUBJ), (OBJ)> as its strongest form.

 $^{^{12}\}mbox{-}$ For expository purposes I have omitted the fact that the attachment is delayed by *final argument*.

(39) $_1$ John $_2$ wanted $_3$ the $_4$ book $_5$ on $_6$ the $_7$ rack $_8$.

If we step in after the verb at position 3, the VP-rule is expanded up to the point marked by " | ".

(40) $_{2}VP \rightarrow V_{3}|$ (NP) ({NP, PP}) $PP[PCASE]^{*} PP[ADJ]^{*}$ (S[^])

At this position, a direct object could be hypothesized, which is confirmed by the strongest verb form of *wanted*. Therefore, when the NP *the book* is read, it could be (*invokedly*) attached to the VP at this position. However, since the direct object NP is the final argument in the preferred frame, the attachment is delayed according to (41):

(41) Final Argument (can delay attaching).

"Give low priority to attaching to a phrase the final argument of the strongest lexical form of that phrase and to attaching any elements subsequent to the final argument. Low priority is defined here with respect to other options that arise at the end position of the element whose attachment is to be delayed." (pp. 752)

The NP will therefore be given higher priority to be attached as the first constituent in rule (34d), resulting in the prediction of a PP or an S'. Since the PP is ranked higher (again due to assumed frequency reasons), it will be chosen as the next hypothesized constituent, according to principle (43).

(42) $_{3}NP \rightarrow NP_{5} | \{PP, S'\}$

(43) Syntactic Preference

"The default order of priority for alternative categories in the expansion of a phrase structure rule is the order of the strengths of the alternative categories" (pp. 749)

After the PP is read, it will be attached to the NP. Then, the complex NP could be attached as the direct object in rule (40), but again, *final argument* (41) prohibits the attachment. Only now can the actual attachment take place when it becomes clear that the sentence end is reached with all other options having failed.

Ford et al.'s model could thus be shown to account for lexical preference effects. As it stands, however, the proposal of four principles to explain a range of two sentences does not sound too elegant and simple. In fact, the model provides predictions for a much wider range of data, some of which will be discussed here.

In general, *Syntactic Preference* (43) is only employed when *Lexical Preference* (35) does not choose between alternative categories in the expansion of a rule. Consider sentence fragment (44):

(44) That silly old-fashioned ... (Fodor & Frazier, 1980)

Example (44) can either be continued as a subject NP, as in *That silly old-fashioned dress makes me nervous*, or as a sentential subject S', as in *That silly old-fashioned dresses*
make me nervous is well known. One might expect the lexical strength of the determiner form of *that* to determine the rule expansion in this example, but the model operates differently: before a lexical item is even considered, the parser operates top-down, i.e. it expands the S-rule it has to start with. Now, consider the S-rule given in (45).

$$(45) S \longrightarrow \{NP S'\} VP$$

For expanding an S, the parser has to choose one of the possible initial constituents given in the rule. Since the decision is based on *Syntactic Preference* (43), the NP is chosen, resulting in the obtained preference.

The syntactic closure model also accounts for the following collection of data. As they found in their questionnaire study, the PP *with the teacher* will preferentially be attached to *discussed* in (46a), but to *having* in (46b).

- (46) a. Sue discussed her daughter's difficulties with the teachers. (early)
 - b. Sue discussed the difficulties that her daughter was having with the teachers. (late)

Whereas the prediction of the former preference is straightforward, if one assumes a 3-place predicate for *discussed*, the latter is somewhat striking, since *having* does not take a *with*-PP preferentially. The explanation is as follows: the strongest form of *have* takes the two arguments <(SUBJ), (OBJ)>. After *having*, an NP-gap (*e*) co-indexed with *the difficulties* is postulated, which occupies the final argument of the predicate. Ford et al. claim that *Final Argument* can only delay its attachment very briefly, since there are no options in the grammar to postulate a complex NP from a gap¹³. Therefore, when the PP is read, it cannot be attached to an NP. However, after several cycles of applying *Syntactic Preference* for the different options in the lower VP, delaying with *Final Argument*, and backtracking due to the lack of further options, the first viable option the parser can backtrack to is attaching the PP into the lower VP as a WITH-COMP. In order to do so, however, a weaker lexical form including a WITH-COMP must replace the stronger 2-place form of *having*. Eventually, the strongest form of *discuss*, <(SUBJ), (OBJ), (PCOMP)>, is replaced by the 2-place form.

Thus, Ford et al. attribute the preference in (46b) to the mere existence of a second verb *(having)*, which overshadows the preferences of the first verb by employing another VP-rule, whose options must be retraced in order to arrive at options resulting from the higher VP.

It is obvious that the same attachment preferences are predicted in the sentences (47) and (48). In the latter, though globally well-formed, something is experienced as missing at the end, due to the strict requirement of a PCOMP argument of *put*.

(47) Joe carried the package that I included for Susan.

^{13.}It is not clear to me, though, exactly how the postulation of a complex [NP [*e* PP]] from the grammar they provide is ruled out.

(48) John put the book that Mary had been reading in the study. (Fodor & Frazier, 1980)

Note that in (49) the preference is different, although the verb properties appear to be identical.

(49) The woman positioned the dress that I wanted on the rack.

Ford et al. assume that the substitution of stronger lexical forms is only permitted, if the preferred form semantically *entails* the less preferred form. This is true for *include*, since "she included something for someone" *entails* "she included something", but this is not true for *want*: "She wanted the dress on the rack" does *not* entail "She wanted the dress."

A preliminary discussion

In general, the theory of syntactic closure gives a comparably detailed account for a variety of psycholinguistic phenomena previously ignored. Lexical preferences had never before been put into a serial model of human language processing.

On the other hand, the empirical findings on lexical preferences are somewhat weak. Ford et al. (1982) prepared booklets for subjects with one sentence per page each followed by a page with two unambiguous paraphrases of the possible alternatives, from which subjects had to choose the one they thought was the interpretation that came to mind first. This is pretty off-line, and emphasizing the ambiguity probably did not help much either. Using the frequency of interpretation choice as an indicator for *lexical* properties appears premature, since it is by no means clear whether the biases they obtained are not due to global semantic aspects of the entire sentence. Even if they exchanged only the verbs in their items, the expectedness of certain arguments may very well depend on the interpretation of both the verb and its other arguments, which may have a strong impact on the interpretation of the situation described (see Taraban & McClelland, 1988).

On the other hand, Ford et al.'s interpretation of the data receives some support from the fact that the same preferences were also found if the NPs were replaced by pronouns, as in (50).

- (50) a. They included everything for them. (NP attachment)
 - b. They carried everything for them. (VP attachment)

Since pronouns presumably do not carry any world knowledge information that could bias the attachment to the VP or to the NP, the obtained preferences strongly suggest that verbs start out with a lexical default bias, even though it could very well be modified by context during processing.

An important point that further weakens the approach is that *Syntactic Preference* in particular gives room for a lot of post-hoc parameter setting used in tuning the system to make it fit any data that could ever come up. As long as *Syntactic Preference* is

exclusively linked to the frequency of certain constructions in everyday speech, its explanatory power can in principle be questioned¹⁴.

Moreover, unless data from large corpora become available, *Syntactic Preference* as a principle which should *explain* experimental findings amounts to a circular argument, since the only way syntactic preferences can be estimated is by observation in psycholinguistic evidence. Even if *Syntactic Preference* were based on frequency counts in large corpora, the theoretical status of such evidence could still be disputed (see Gibson, Pearlmutter, Canseco-Gonzalez, & Hickock, in press).

Of course, this argument does to a certain extent also hold for *Lexical Preference*, as long as *Lexical Preference* lacks a theoretical foundation, and some psycholinguistic evidence proving its influence in on-line parsing. I will return to this issue several times throughout the following chapters. In chapter 5, following the discussion of the SOUL model, Ford et al.'s model will be discussed in more detail, and further problems will be addressed in an appropriate context.

All models described so far rely on phrase structure rules (or an equivalent ATN representation) to be accessible during parsing (Frazier, 1986, 1989). Syntactic structure can thus be proposed *top-down*, allowing for structure to be built even if lexical *heads* licensing the integration of current items are not yet available. Attachment does therefore not have to be delayed until phrasal structure can be projected from the respective lexical heads. This particular property, among others, distinguishes the theories discussed in the previous sections from most of the accounts based on *Government and Binding Theory* (Chomsky, 1981, 1986), which I will now turn to.

2.8 Abney's Licensing Structure Parser (1989)

A fairly radical hypothesis concerning the use of lexical information in syntactic processing is to be found in Abney's (1989) Licensing Structure Parser (LSP). The model is based on a highly lexicalized GB-like grammar with lexical heads providing *licensing relations* for their arguments (theta relations), adjuncts (modifier relations), and functional categories (functional selections). Characteristically, the attachment of constituents is restricted to circumstances where it is *licensed* by a lexical head.

The parser proceeds according to a shift-reduce (*shift-attach* in Abney's terms) algorithm which is based on a *stack* onto which (maximally projected) incoming items are shifted to be processed. Before a new item is shifted, however, the parser always attempts to attach the current top-most item (structure) to the partial structure in the second cell of the stack. In other words: *shift/attach* conflicts are resolved in favor of *attach*.

^{14.}Note, however, that the authors are attracted by Wanner's approach for arc-ordering in ATNs (Wanner, 1980, see section 2.5). Such an approach remains to be proven as applicable to the LFG approach.

In the case of categorial ambiguities (i.e. *shift/shift* conflicts), the strongest lexical form is chosen (the concept of strength is not very well specified as in Ford, Bresnan, & Kaplan's theory, see section 2.7).

Many of the attachment ambiguity phenomena dealt with in this thesis appear as *attach/attach* conflicts, i.e. as conflicts between two or more possible attachments to the same structure. In these cases, the parser operates according to the following three principles.

(p11) Theta Attachment

Prefer theta-attachment over non-theta-attachment.

(p12) Verb Attachment

Prefer attachment to verbs over attachment to non-verbs.

(p13) Low Attachment

Prefer attachment to the lowest attachment site.

The *Theta-Attachment* principle (p11) ensures that in processing sentences such as (51, 52, and 53) the same preferences will be forecast as were accounted for by *Mini-mal Attachment* in the *Garden Path Model*: the verbs (*watched* or *decorated* respectively) both provide a theta relation, whereas the nouns of the object-NP (*thief* or *cake*, respectively) only supply a modifying relation. The attachment of the PP to the verb will thus be preferred. In cases as (53), where neither the verb nor the object NP provide a theta role, the attachment to the VP is also preferred, however due to (p12).

In sentence (54), though structurally similar, the attachment preference seems to be different (at least intuitively): the PP *in the Volvo* it preferentially attached to the object-NP *his interest*. Note, however, that a theta-role is supplied by the noun (interest in the Volvo), but not by the verb (thought about... in the Volvo). In contrast to the *Garden Path Model*, this preference is predicted by the LSP, by virtue of *Theta-Attachment* (p11).

- (51) The actress watched the thief with binoculars.
- (52) Susan decorated the cake with strawberries.
- (53) Carl caught the cat in the car.
- (54) She thought about his interest in the Volvo.

Finally, "late closure" preference phenomena as in (55), are covered by the principle of *low attachment*.

(55) Barbara gave Chris who was eating the Black Forest cake.

Note that Clifton, Speer, and Abney (1991) claim to have shown that a preference for thematic attachment cannot override a verb attachment (*Minimal Attachment*) preference in sentences like (54). Since their findings pose a threat to Abney's as well

as the model presented in this thesis, I will discuss the validity of this evidence in some detail in chapter 4.2.

It is important to point out again that the LSP does not operate in a strictly incremental fashion, since incoming items may sometimes not be attached immediately. If a lexical head carrying the licensing information is not yet available while proceeding through the input string, no attachment can take place. Thus, Abney's model follows a strong *head licensing hypothesis*. The issue of incrementality will be discussed in further detail in chapter 4.1.2 on the basis of new evidence from parsing German subject/object asymmetries.

2.9 Theta Attachment and the On-Line Locality Constraint (Pritchett, 1992)

Pritchett (1988, 1992) aims at providing a strongly competence based model of human sentence processing. For this purpose, he builds his model on the linguistic framework of Government and Binding Theory (GB, Chomsky, 1981, 1986). In order to describe Pritchett's model, it is important to sketch the most relevant concepts of GB first.

In GB, the grammar is not longer specified as a set of explicit phrase structure rules, but as a small set of very general grammar principles that interact with each other to constrain the way in which well-formed sentences may be constructed. The basic construction schema for phrases is the X-bar schema (see footnote 11.). The derivation of the surface structure further depends on the principles which constrain or (enforce) the way in which constituents are moved into proper positions in the structure built from the X-bar schemes. The two most important grammar principles for Pritchett's model are the *theta criterion* (56) and *case theory* (57).

(56) Theta Criterion

Each argument bears one and only one theta-role (thematic role) and each theta-role is assigned to one and only one argument (Chomsky, 1981; p. 36).

(57) Case Theory

Every lexical NP must be assigned (abstract) case.

In GB, the assignment of theta roles and case is bound to certain structural positions. Verbs, for example, can directly assign theta-roles and case to their complements which are realized as sisters in the X-bar projection. Subjects receive their thematic roles from the verb, but their case is assigned by a functional head called INFL¹⁵ that carries the inflectional information of finite verbs. Thus, no nominative case can be assigned by infinite verbs (58).

¹⁵·INFL projects to IP, which is the representation of a sentence without a complementizer (formerly S). Complementized sentences are projections over "comp" (e.g. that) called CPs (formerly S').

(58) *Ich sah der Mann kommen. *I saw the_{nom} man coming.

Pritchett then proposes two highly abstract principles in his parsing model, *Generalized Theta Attachment* (GTA, p14) and the *On-line Locality Constraint* (OLLC, p15), which are supposed to give an explanation for first analysis as well as reanalysis processes and are able to cover a wide range of data.

(p14) Generalized Theta Attachment (GTA)

Every principle of the syntax attempts to be maximally satisfied at every point during processing." (Pritchett, 1992, p. 155)

(p15) On-Line Locality Constraint (OLLC)

The target position (if any) assumed by a constituent must be *governed* or *dominated* by its source position (if any), otherwise attachment is impossible for the automatic sentence processor." (Pritchett, 1992, p. 155)

Generalized Theta Attachment requires the parser to satisfy the principles of grammar, like the *theta criterion* (56) and *case theory* (57), at each point during sentence processing, disregarding *global* grammaticality of the whole sentence.

The OLLC determines the reanalysis processes in which a constituent is reattached within the automatic sentence processor. Whereas GTA exclusively evolves from theoretical notions of *Government and Binding Theory*, note that OLLC is stipulated independently of the competence grammar, in that nothing in the definition of *government* or *domination* motivates the necessity of a reanalysis constraint.

If locally permissible, phrases are attached to structural positions which are both *theta marked* and *case marked*. If the preferred reading fails to yield a grammatical structure for the whole sentence, the automatic sentence processor is only able to recover the ultimately correct analysis under the conditions specified by the OLLC. Reanalysis within the range of the OLLC is therefore not assumed to be costly. Whenever these constraints are violated, however, people can only "consciously" recover from the garden path.

According to Pritchett, easy reanalyses (within OLLC) are performed quite often during sentence processing if they lead to better fulfillment of GTA.¹⁶ A striking example for these costless reanalyses is shown in (59a).

- (59) a. Without her donations to the charity Bob failed to appear.
 - b. Without her donations to the charity failed to appear.

Following GTA, the pronoun *her* is initially attached as the object of the preposition *without*, since *without* can assign case to a pronoun-NP, but not to a pronoun in the specifier position of an incomplete NP. Then, when the noun *donations* is read, it would locally strand without theta-role or case if it was not attached as the head of

¹⁶.Obviously, this theory of reanalysis contrasts strongly with *reanalysis as last resort* (see section 2.4).

the object of "*without*". Fortunately, *her* can easily be reanalyzed as the specifier of the NP "*her donations*", because the source position of "*her*", the NP-argument of "*without*", clearly dominates the specifier position of a new NP in that position.

Note that without a reanalysis (59a) should be a severe garden path sentence, which it clearly is not. In contrast to that, (59b) is a "conscious" garden path sentence because after *her donations* is preferentially analyzed as the object of *without, donations* cannot be reanalyzed within the OLLC to become the subject of the matrix clause. The source position of *donations*, below the NP within the topicalized PP, neither dominates nor governs the target position (I-Spec).¹⁷

Consider the sentences (60 a, b).

- (60) a. I warned the werewolf was after Ron.¹⁸
 - b. I knew the werewolf was after Ron.

The NP *the werewolf* may be attached as the object of the preceding verb or as the subject of the following subclause. According to GTA, object attachment is preferred, since the structural object position of the verb *warned* is locally theta and case marked, whereas the licenser of the subject of the subclause is not yet locally accessible.¹⁹ Up to this point, GTA meets the predictions of *Minimal Attachment*. But neither MA nor GTA can account for the fact that in (60 a) the garden path effect appears to be stronger than in (60b). Actually, Pritchett assumes that there is no ("conscious") garden path at all in (60b).

What makes the difference? The verb of (60b), *knew*, only licenses *one* argument, which may (roughly) be realized either as a noun phrase or as a sentence. The initially misattached NP *the werewolf* has to be reanalyzed as the subject of the sentence, occupying the same structural position as its source position. The target position is obviously dominated by the source position. In (60a), however, *the werewolf* is initially attached as the *goal* of *warned* and has to be reanalyzed as the subject of its *theme* which takes a structurally different position. In this case, the target position is neither dominated nor governed by the source position thus leading to a severe garden path.

A preliminary discussion

Note that Pritchett does not provide a specified parsing algorithm, especially not a concrete reanalysis mechanism. He predicts performance phenomena solely on the basis of configurational *constraints* that have to be satisfied if the process is to be kept

^{17.}Government is defined in the following way:

government: α governs β iff α m-commands β and every γ dominating β dominates α , γ a maximal projection (Adapted from Chomsky, 1986a).

m-command: α m-commands β iff a does not dominate β and every γ that dominates α dominates β , γ a maximal projection (Adapted from Chomsky, 1986a).

¹⁸.What makes this example a little less convincing is that the use of "*warn*" without a goal is of borderline grammaticality according to native informants.

¹⁹ The automatic sentence processor works without any look-ahead or wait-and-see strate-

inexpensive. Though elegant at first glance, such an approach leaves the question open of how the correct analysis is found in case of a parsing failure.

Similarly to the *Licensing Structure Parser*, though less explicitly, Pritchett's account suggests a head licensing mechanism in which no attachment can take place before it is licensed by the respective heads. Some evidence contradicting this assumption has been presented by Hemforth (1993) and Hemforth, Konieczny, & Strube (1993). A detailed eye-tracking experiment showing evidence against head corner parsing will be presented in chapter 4.1.2.

Furthermore, no recency based principle like *Right Association* (see section 2.3, p5) or *Late Closure* (see section 2.6, p10) is assumed in Pritchetts model (as in the first version of the *Sausage Machine*, Frazier & Fodor, 1978). As a consequence, the theory presented in this section is subject to the same shortcomings as the Sausage Machine, as pointed out by Gibson (1991). The fulfillment of syntactic requirements, as forced by GTA, does not suffice to predict a preference to attach the adverb *yesterday* in (61a) or the particle *out* in (61b) to the lower clause.

- (61) a. Reinhard said he liked the TV show yesterday.
 - b. Anna figured that Klaus intended to take the dog out.

Gibson (1991) provides a different GB-based account inspired by Pritchett's, which is more explicit with respect to the algorithmic assumptions, and which includes a notion of recency.

2.10 Gibson's weighted parallel model of human sentence processing

Gibson's (1991) sophisticated account is based on a powerful incremental parser that generates all syntactically permissible structures in parallel at each point during processing. However, processing is assumed to consume *working memory* resources. The *processing load* is then computed for each analysis on the basis of certain *properties* of the structures derived from principles of GB. If an analysis differs too much from the others with respect to *processing load*, it will be abandoned and only the less costly analyses will be maintained. Like Pritchett, Gibson distinguishes structures which can be easily analyzed in the automatic parsing routines, even if they have not been preferred initially, from structures which can only be analyzed by special purpose recovery procedures. The mechanism will now be described in more detail.

The parser proceeds according to a *left-corner* algorithm (see Aho, Hopcroft, & Ulmann, 1974; Johnson-Laird, 1983), mixing *bottom-up* with *top-down* processing. In order to be able to predict structure *top-down*, the parser employs phrase structure rules compiled from a GB-based grammar. When an item matches the left-corner of the right hand side of a phrase structure rule, all categories are predicted which can or have to appear to the right of the current input word. The resulting *hypothesized* nodes are called *H-nodes*. All other nodes are called *confirmed* or *C-nodes*. Not only obligatory but also optional categories are predicted in this way. Lexically unfilled optional H-nodes can be pruned from the structure at the end of a parse.

For each structural alternative constructed by the left-corner parser, an abstract weight is calculated, the (integer) number of *processing load units* (PLUs). Processing load units are calculated from certain *properties* of the structures. The two most important properties are the property of *Thematic Reception* (TR) and the property of *Lexical Requirement* (LR). *Thematic Reception* (TR) is directly derived from the *Theta Criterion* in GB (Chomsky, 1981; see section 2.9, 56):

(p16) Property of Thematic Reception

Associate a load of x_{tr} PLUs to each constituent that is in a position to receive a theta role in some coexisting structure, but whose theta assigner is not unambiguously identifiable in the structure in question. (Gibson, 1991; p. 97)

The second property, the property of *Lexical Requirement*, can be derived from the *Projection Principle* in GB (Chomsky, 1981, lexical requirements must be satisfied at all levels of representation).

(p17) Property of Lexical Requirement

Associate a load of x_{lr} PLUs to each lexical requirement that is obligatory in some coexisting structure, but is satisfied by a H-node constituent containing no thematic element in the structure in question. (Gibson, 1991; p. 97)

Since *Thematic Reception* and *Lexical Requirement* are indirectly derived from the Principle of *Full Interpretation* (Chomsky, 1991), both are assumed to be based on the same variable ($x_{tr} = x_{lr} = x_{int}$), i.e. the load associated with local non-interpretability.

It is assumed that structures can be pursued in parallel as long as the number of PLUs associated with the structural alternatives do not differ too much. Gibson (1991) postulates a preference constant *P* that determines the maximally acceptable *distance* in processing load between two structural alternatives. As soon as this preference constant is exceeded, analyses which are too costly are abandoned. Processing breakdown (e.g. in a conscious garden path) is thus predicted when all continued analyses turn out to be incorrect at some point, while the correct one has already been abandoned earlier in the sentence.

Gibson's line of argument is formulated more mathematically than concurring approaches. Instead of deriving predictions about the parsing difficulty from principles or *properties* of the structures (or distance between alternative structures) and testing them empirically, he utilizes the empirical evidence (including intuitions) to *estimate* the upper limit (*P*) for the difference in PLUs that may not be exceeded²⁰. After proceeding in this way for a variety of sentences, these estimations are submitted to a system of inequalities. Only if a consistent solution can be found, the model can be said to be adequate.

Consider (62 a, b), for example. The NP *the evidence* can either be attached as the direct object of *knew*, or as the subject of a sentential complement of *knew*.

(62) a. [IP[NP The linguist] [VP knew [NP the evidence] ... by heart.

b. [IP[NP The linguist] [VP knew [IP [NP the evidence] ... was correct.

When the NP is analyzed as the direct object (62a), no NP in a theta position is left without an identifiable theta assigner and all lexical requirements are fulfilled, so there are $0x_{tr}$ PLUs. In (62b), on the other hand, the NP *the evidence* is supposed to be in a theta position²¹ but the theta assigner is not yet available, so there is $1x_{th}$ PLU. Though (62b) is considered slightly more difficult to process because of the increased processing load, there is no (intuitive) evidence for a processing breakdown. *P* must be greater than $1x_{th}$ PLU.

Consider now the examples given in (63a, b).

- (63) a. [IP [NP The student] [VP put [NP the book] [PP on ...] ... the table.
 - b. [$_{IP}$ [$_{NP2}$ The student] [$_{VP}$ put [$_{NP2}$ the [$_{N}$ book [$_{PP}$ on ...] ... the table in the shelf.

When the PP is constructed from the preposition *on*, it can either be attached to the VP or the preceding N'. $1x_{th}$ PLU has to be associated with (63a), because *on* requires an argument that is not yet lexically realized. On the other hand, $3x_{th}$ PLUs have to be associated with (63b) because the verb *put* as well as the preposition *on* are lacking an argument, and there is a coexisting structure (63a) that assigns a thematic role to the PP. The difference between the two structures is therefore $2x_{int}$ PLUs. In contrast to (62b), (63b) is considered a conscious garden-path. Thus, *P* must be between $1x_{int} < P < 2x_{int}$ PLUs.

"Late closure" preferences, as they can be observed in sentences like (64a, b), cannot be explained with *Thematic Reception* and *Lexical Requirement*. Adjuncts such as *last Sunday* in (64a) are neither lexically required nor do they take a theta role. In (64b), both preceding verbs, *put* and *thrown*, can assign a theta role to the PP *on the table*.

- (64) a. Chris said Yuki will visit him *last Saturday*.
 - b. Tom put the book he had thrown on the table.

The only difference between the potential structures of (64a, b) is that the matrix verb, *said* or *put*, respectively, is less recent than the verb of the complement clause, *visit or thrown*, respectively.

$$\sum_{i=1}^{n} A_{i} x_{i} - \sum_{i=1}^{n} B_{i} x_{i} > P$$

^{20.}The inequality for *P* is computed as follows:

P is the preference constant, x_i the number of processing load units (PLUs) associated with property *i*, n the number of properties that are associated with processing load, A_i the number of times property *i* appears in the unpreferred (more costly) structure, and B_i the number of times property *i* appears in the preferred structure.

^{21.}Contrasting other approaches, Gibson (1991) considers the subject position as a theta position.

To account for such "*late closure*" preferences, Gibson postulates the property of *Recency Preference* (p18).

(p18) Property of Recency Preference (RP)

The load associated with the structure resulting from attachment involving either a thematic or arbitrary H-node is equal to the number of more recent words that are associated with a corresponding thematic or arbitrary H-node.²²

In (64a), then, the processing load associated with high attachment of the PP *on Saturday* is 1 x_{rp} PLU. Similarly, thematic attachment to the verb *put* in (64b) is less recent than the favored thematic attachment to *thrown*. Since (64a) is considered a garden path sentence, $1x_{rp} > P$. (Note that x_{rp} is obviously different from x_{tr} and x_{lr})

An appealing aspect of Gibson's approach is that it appears to account not only for processing breakdown effects in ambiguous structures. Multiply center embedded relative clauses are another well known example of grammatical but unacceptable structures (see sections 2.1.1, 2.1.3). It is assumed that the capacity of the human sentence processor is exceeded in these cases. The constant *K* is assumed to constrain the total processing load that can be associated with structures without rendering them unacceptable. If the cumulated processing costs of a structure exceed *K*, the structure becomes unacceptable.²³

K is estimated by intuitions about processing overload. Consider the following example (65; example 351 in Gibson, 1991):

(65) The man that the woman that won the race likes eats fish.

What is the processing load associated with these structures? The maximum processing load is reached when the second complementizer *that* is processed. The structure Gibson assumes for (65) at that position is given in (66):

(66) [IP [NP the man_i [CP [NP O_i] that [IP [NP the woman_j [CP [NP O_j] that [IP]]]]]]

There are two lexical NPs *the man* and *the woman* and two non-lexical NPs, the relative clause operators O_i and O_j which do not have theta roles assigned, so there are $4x_{th}$ PLUs associated with (65). For the second complementizer *that*, the complement

$$\sum_{i=1}^{n} A_i x_i > K$$

^{22.} Note that a distinction is made between arbitrary and thematic H-nodes: any H-node is an arbitrary H-node, whereas only H-nodes which participate in thematic role assignment, i.e. which are for instance assigning or receiving a thematic role are thematic H-nodes.

^{23.}The inequality for *K* is computed as follows:

K is the maximum allowable processing load, x_i the number of PLUs associated with property *i*, *n*, the number of properties that are associated with processing load, and A_i the number of times property *i* appears in the structure.

(the IP) is missing, violating the Property of Lexical Requirement. The combined PLU associated with (65) is $4x_{th} + 1x_{lr} = 5x_{int}$.

Structures like (67; example 342 in Gibson, 1991), on the other hand, are regarded as acceptable.

(67) I saw the man that the woman that won the race likes.

In this example, there is no thematic role assigned for three NPs (one lexical NP, *the woman*, and two non-lexical operators) and the complementizer *that* is lacking its argument. The NP *the man*, however, has its thematic role assigned by the verb *saw*, in contrast to (65). So the processing load is only $3x_{th} + 1x_{lr} = 4x_{int}$ PLUs. Thus, the constant K can be estimated by the inequality $4x_{int} < K < 5x_{int}$.

A preliminary discussion

Generally, Gibson's account does not appear to be very parsimonious. All structural alternatives have to be computed at any given parse step. Fairly complex scanning procedures are necessary to calculate the processing loads associated with the different structures, each of which has to be constructed initially, thus enabling a decision as to whether or not they had better be abandoned (or even not constructed in the first place) because they are too complex.

Note further that the *sum* of PLUs of all structures maintained in parallel does not determine whether or not a sentence can be parsed easily. *K* can only be exceeded by a single sentence, regardless of how many structures have been constructed in parallel. *P* refers to the *difference* in PLUs, regardless of how complex the single alternative structures actually are. Thus, it is in principle possible to analyze an infinite amount of structures in parallel, as long as each of them does not exceed *K* or differ too much from the others. Note that Gibson does not consider structure building and deleting as well as computing and comparing the number of PLUs *per se* costly. All this, of course, presupposes a strange model of working memory. Moreover²⁴, it is absolutely not clear *why* structures should be weighted or ranked if the working memory is in principle capable of storing an infinite amount of them.

Besides this, the model seems to be empirically inadequate. Gibson does not predict a garden path in sentences like (68). Since the attachment of the PP *in the library* to the most recent verb [*was reading*] would be non-thematic, whereas the verb *put* could assign a thematic role to this PP, attachment to the verb of the matrix clause should be preferred.²⁵

(68) John put the book he was reading in the library²⁶.

^{24.}Thanks to Barbara Hemforth for this hint.

^{25.}Note, that no PLUs due to Recency Preference are associated with the thematic attachment, because only attachments which are matched for their thematic properties participate in the calculation.

According to most native speakers, at least a "surprise effect" can be noticed when reading (68), because something seems to be missing (Fodor and Frazier, 1980). Thus, contradicting Gibsons assumptions, the PP *in the library* seems to be preferentially attached to the more recent clause initially.

2.11 Multiple constraint models

The revival of connectionism in the past decade has brought interactive-activation models as theories of human parsing back on stage (e.g., Elman & McClelland, 1984). One of the most prominent accounts was presented in MacDonald, Pearlmutter, and Seidenberg (1994). Their multiple constraint satisfaction model aims at providing an integrated lexicalist framework that abandons the privileged status of syntactic processing outside the lexicon.

According to this theory, lexical entries come equipped with multiple kinds of constraints, including partial syntactic information, e.g. partial parse trees, based on the X-bar scheme. A rough idea of the information that is represented for a word like *John* is given in example (69; see MacDonald et al, 1994, p. 688).

(69) Partial representation of "JOHN"

semantics: *animate, human,* etc. Thematic roles to be filled by JOHN: *agent, experiencer, theme, goal* Argument structure (i.e. arguments taken by JOHN): *null* Lexical category: *noun* Syntactic structure: [N² [spec Y][N¹ [N⁰ JOHN] [complement Z]]]

The maximal projection of each word is stored in the lexical entry. Parsing proceeds by activating lexical entries with their partial parse trees and by *linking* them appropriately using all kinds of information given in the lexicon and the context of the utterance. The particular process of *linking*, however, remains fairly underspecified in MacDonald et al.'s paper.

All kinds of ambiguities reside in the lexicon. If a verb can be used transitively as well as intransitively, the two respective syntactic trees are represented in the lexical entry of the verb. Which of the two is activated depends on the strength of the different readings, mainly derived from the frequency of occurrence of this reading in the learning history of the individual. Exactly the same processes are supposed to underlie lexical as well as so-called structural ambiguity resolution: i.e. frequency based activation of lexical entries which then determine which kind of structure is preferentially built. All kinds of contextual information also come into play immediately.

^{26.}Gibson argues that indeed no "*late closure*" preference exists in sentences such as (68). His argument is based on the observation that the sentence "*Janet put the book Lyn was reading in the study in the rack*" appears to induce processing difficulties, which it should not if a late closure preference was present. See footnote 17. in section 1.4.7.3 for a discussion of this argument in the context of the model presented in this thesis.

They are not postponed until initial structure based analyses are given to a secondary checking stage. However, contextual information can only determine initial preferences if the lexical entries are equi-biased in frequency (bottom-up priority, Marslen-Wilson, 1975) and if the contextual constraints are strong enough.

MacDonald et al. claim to be able to explain a lot of the variation in the empirical evidence from psycholinguistic literature. Some studies find strong garden-path effects for particular constructions like, for example, the main-verb/reduced relative ambiguity (70), some others do not.

- (70) a. The thesis enjoyed by the professor ...
 - b. The thesis examined by the professor ...

Some find overriding context effects and some do not. According to the multiple constraint satisfaction model, the particular outcome of any experiment is due to the respective choice of verbs and the strength of the contextual information. If verbs with a strong bias for past tense readings, such as *enjoyed* in (70a), are used in main-verb/reduced relative ambiguity experiments, a strong garden path is to be expected there. In these cases, not even preceding thematic information facilitating the reduced relative reading, such as the non-animate NP *the thesis*, which is unlikely to be an agent, will override the lexical bias. However, if equi-biased verbs are used (70b), the context may determine the preferences.

The multiple constraint approach seems to gain support from a series of experiments. MacDonald (1994), for example, investigated the effect of argument structure, comparing transitive biased verbs like *push* (71a) with intransitive biased verbs *move* (71b) and unambiguous participles *driven* (71c). On the syntactically disambiguating part (*were afraid*), a garden path only showed up for intransitive biased verbs (71b). MacDonald interpreted her result as an indicator of the immediate influence of the lexical bias: because of the transitive bias of *pushed*, (71a) was already disambiguated in favor of a reduced relative at the preposition *into* succeeding the verb.²⁷ So no strong commitment for the main verb reading was established.

- (71) a. The rancher knew that the nervous cattle **pushed** into the crowded pen were afraid of the cowboys.
 - b. The rancher knew that the nervous cattle **moved** into the crowded pen were afraid of the cowboys.
 - c. The rancher knew that the nervous cattle **driven** into the crowded pen were afraid of the cowboys.

Evidence like this has been disputed extensively, but I want to postpone this discussion to chapter 4, where the empirical evidence on general architectural constraints of the human sentence processing device will be weighed up more thoroughly.

^{27.}Unfortunately it is unclear what happened at the preposition. Any current garden path model of sentence processing would have predicted an influence of subcategorization information several words after the verb is encountered.

A general problem with MacDonald et al. 's approach is that parsing operations are only very vaguely specified. There appear to be quite a lot of parameters that can be set to make the model fit the data. By only arguing with lexical frequency, they simply restate the problem of parsing preferences (cf. Stevenson, 1995). Different frequencies of verb-frames, for example, may be the result and not the cause of parsing preferences.

In the next section I will present the CAPERS model by Suzanne Stevenson, which is specified in much more detail than MacDonald et al.'s constraint satisfaction approach.

2.12 CAPERS: A hybrid model of human parsing

Suzanne Stevenson (1993, 1995) presents a GB-based connectionist model of human sentence processing that combines symbolic feature passing with numeric spreading activation. Thus, it is not a purely connectionist but a hybrid model. Grammatical constraints are established by the symbolic feature passing mechanism. Since we are mainly interested in the predictions on parsing preferences which are realized by spreading activation, I will concentrate on the dynamics of the connectionist part of the model in the following discussion.

The network that is built up during parsing directly represents a parse tree. When an input token is processed, it activates a group of *p*-*nodes* (phrase-nodes) representing the maximally projected x-bar structure headed by the lexical item. These are then connected to a number of *a*-*nodes* (attachment nodes) that are further connected to potential attachment sites along the right edge of the parse tree, as illustrated in (72), adapted from Stevenson (1995).





Note that attachments via a-nodes represent *sister* relations in the actual phrase marker.

The alternative a-nodes compete for activation. The amount of activation spread to an a-node depends on its compatibility with grammatical constraints, preferred subcategorization frames of lexical heads and the relative distance of a potential attachment site; i.e. the more recent a p-node the a-node is connected to, the more weight will be given to the connection. Preferentially transitive verbs give strong weights to a-nodes that connect them to direct objects, verbs that take a prepositional phrase give a strong weight to the respective a-node and so forth. The winner of the competition depends on the combined spreading activation processes from all these sources. All new attachment nodes except the winner are finally deleted.

Let us have a look at a particular attachment ambiguity, namely, PP-attachment: according to Stevenson (1995), the PP *with pickled cucumbers* in (73) can either be interpreted as a verb modifier or as a NP modifier²⁸.

(73) Chris decorated the sandwiches with pickled cucumbers.

When the preposition *with* is read, its *p-node* will be connected to the p-nodes representing both attachment sites via a-nodes. A decay function in the activation of the p-nodes postulated in the course of processing the sentence gives the a-node connected to the more recent object NP *the sandwiches* a slight advantage. On the other hand, since verbs like *decorated* strongly favor a PP-complement (an *ornative* in this particular case; i.e. something that something else is decorated with), the a-node connecting the PP to the VP is strengthened. No such predictive effect is assumed for the NP modification in this case. VP attachment will win in cases of strong lexical preferences, the a-node for NP-attachment will eventually be eliminated. If on the other hand the object NP showed a preference for a modifier, attachment preferences would change. Thus, for modifier / modifier -attachment ambiguities, slight changes in the respective lexical heads or even changes in the semantic / pragmatic context may determine the final attachment preference.

For sentences like (74), on the other hand, a strong preference for argument attachment of the first PP *on the table* to the verb *put* is predicted.

(74) Thomas put the questionnaire on the table into his bag.

This is assumed to be due to grammatical constraints on arguments which are reflected in the behavior of the competition model. Argument positions must be filled with exactly one argument (according to the theta criterion, 56). This necessitates the activation function in the connectionist network to exclude multiple attachments. Modifiers, on the other hand, may be connected in any number to the same node (i.e. zero to potentially infinite). If modifiers are defined by their potential for multiple attachments (see chapter 5 for a discussion in the HPSG-framework), a much less competitive activation function has to be assumed. Thus, for grammatical reasons,

^{28.}In contrast to work by Frazier (1987a) or Frazier and Clifton (1996), attachment of the instrumental PP is not regarded as argument attachment.

attachment to argument positions has to be more competitive than modifier attachment. Consequently, argument attachment will always win in the case of this kind of ambiguity.

The evidence Stevenson gives for multiple attachments can be illustrated by example (75).

(75) I went [PP with Barbara] [PP from Frankfurt] [PP to Hartford] [PP on Continental Airlines]

All the PPs in (75) are considered modifiers of the VP. One of the central questions is whether this is a viable description of the state of affairs. Definitely, not all the PPs considered here are equally optional. At least the goal *to Hartford* is strongly attracted by the verb *went*, so strongly, by the way, that it becomes very hard to tell it apart from an argument. Additionally, attachment of modifiers is not unconstrained. One cannot say something like *"I went with Barbara with Chrisi from Frankfurt …"*. A coordination (*with Barbara and Chrisi*) has to be used to attach multiple modifiers of a kind (e.g. co-agents), and that is highly similar to argument positions. If anything like thematic roles of non-arguments (e.g. instruments, locatives etc.) is represented in the network, modifiers have to compete as strongly for these thematic roles as arguments for theirs. Since lexical preferences of verbs for *modifiers* of a particular thematic type are supposed to influence the activation of a-nodes (see example 73 above), thematic roles of modifiers have to be represented. It is not too convincing then, that modifiers can cope without a winner-takes-all activation function.

Similarly to Gibson (1991, see section 2.10), Stevenson assumes that the winner takes all competition function overrides the recency preference in (76).

(76) The cleaning woman put the cups that Markus had used in the sink.

However, sentences like (76) seem to induce a slight garden-path due to a "late closure" preference of the PP *in the sink* (i.e. the PP is preferentially taken as a modifier of *used* such that the obligatory PP-complement of *put* is missing at the end of the parse).

Let us sum up some of the major features of the model: it is a competitive multiple constraint model, and therefore corresponds to MacDonald et al.'s (see section 2.11). Similarly to MacDonald et al., grammatical and lexical constraints, along with higher order conceptual or contextual constraints, are assumed to initially influence attachment processes. In contrast to MacDonald et al. (1994), however, lexical frequency is not the major driving force of attachment preferences. The approach shares many predictions with the other competition based approaches. In particular, ambiguity should always lead to an increased processing load, since the network always needs more cycles to stabilize on a particular interpretation if competing alternatives exist. This particular prediction will be addressed in detail in chapter 4.

Moreover, since all, including even ungrammatical, attachments along the right edge of the phrase marker are created in a first attempt, it may be the case that attachment processes become increasingly costly the longer a sentence gets and, consequently, the more attachment sites become available. Every a-node joins the competition, so the more a-nodes are constructed initially, the more cycles it should take to stabilize the net.

A major advantage of this model is that it is fully formalized and implemented such that the predictions on new constructions can be tested straightforwardly in many cases.

2.13 Structural determinism: Gorrell's theory of syntax and parsing

In the preceding sections I presented models of human sentence processing which tried to account for parsing preferences by weighing structural alternatives constructed in parallel. Considering several alternatives for constituent attachment in parallel reduces the amount of work that has to be done if the preferred analysis fails. The approach discussed in this section also aims at reducing the processing load due to recovery of unpreferred structures. However, this goal is not approached by considering structural alternatives in parallel, but by leaving the description of phrase markers *underspecified*.

The grammatical representation of Gorrell's model is based on *d*-theory (description theory, Marcus, Hindle, and Fleck, 1983). In *d*-theory, syntactic structure is represented by a description of trees rather than trees themselves. During parsing, no parse tree is built. Instead of this, a description of the structural representation is constructed and then interpreted. A parse tree encodes three kinds of information (Partee et al., 1990): the *dominance* relations of nodes, their left to right (*precedence*) order and their grammatical type. Thus, (77b) is a description of (77a).

(77) a. [VP [v NP]]

b. dom(VP, v), dom(VP NP), prec(v, NP)

with *dom* standing for dominates and *prec* for precedes

Note, however, that (77b) is only a partial description of (77a), because in (77a) the VP-node immediately dominates the v-node and the NP-node, respectively. (77b) is underspecified in that only dominance, and not *immediate* dominance, is expressed. Dominance and precedence are constrained by the following conditions:

- (p19) For two nodes x and y, x dominates y iff the connection between x and y is composed exclusively of descending branches. (Gorrell, 1995, p. 11)
- (p20) In any well-formed tree, either prec(x, y) or prec(y, x) is true iff neither dom(x, y) nor dom(y, x) is true. (Partee et al., 1990, cf. Gorrell, 1995, p. 12)
- (p21)In any well-formed tree, if prec(x, y), then all nodes dominated by x precede all nodes dominated by y.

Gorrell (1995) distinguishes two sub-processes in the construction of syntactic representations. The first process is *structure building*, which is only based on *dominance* and *precedence* relations between constituents. Secondly, the structural description built by these "primary" relations is analyzed by the *structure interpreter* yielding socalled secondary relations based on GB concepts like *government*, *c-command*, *theta assignment*, *case assignment*, and *binding*. Structure building is constrained by the principle of *Simplicity* which is derived from the grammar principle of *Full Interpretation* (Chomsky, 1991).

(p22) Simplicity

No vacuous structure building. (Gorrell, 1995; p. 100)

Simplicity hinders the construction of non-branching nodes which are not required by the input.

In the tradition of Marcus (1980) and Marcus, Hindle, and Fleck (1983), Gorrell postulates a structurally *deterministic* model. Structures that have once been built cannot be revised by the automatic parser. If structure has to be revised, a conscious garden-path results. In contrast to Marcus (1980), structural determinism in Gorrell's model is restricted to representations constructed by the *structure builder*.

(p23) Structural determinism

The domain of determinism is limited to the primary structural relations, *dominance* and *precedence*.

Since dominance and precedence relations are assumed to be processed deterministically, revision of structural information causes a conscious garden path, whereas mere addition of structural information may lead to an increased processing load but not to processing break-down (p24).

(p24) Upon reanalysis, the reanalyzed constituent must participate in the same dominance and precedence relations as prior to reanalysis. (Gorrell, 1995, p. 117).

Let us have a look at examples like (78a, b) to see how Gorrell accounts for easily recoverable parsing preferences:

- (78) a. The girl knew the answer ... by heart. dom(IP, NP1), dom(IP, VP), dom(IP v), dom(IP, NP2), dom(VP, v), dom(VP, NP2)
 - b. The girl knew the answer ... was correct. dom(IP1, NP1), dom(IP1, VP), dom(IP1, v), dom(IP1, NP2), dom(VP, v), dom(VP, NP2), **dom(IP1, IP2), dom(VP, IP2), dom(IP2, NP)**

The dominance relations which are most relevant for this discussion are given below the examples. When the NP *the answer* is read, only the dominance relations given for (78a) should be postulated according to the principle of *Simplicity*. If, however, the sentence is later disambiguated, yielding structure (78b), dominance relations have to be added (e.g., those typed in boldface below 78b). All primary relations that have been postulated for (78a) still hold in (78b). Compatible with the psycholinguistic evidence, there may be some increased processing load measurable for structures like (78b) due to structure addition, but there is no conscious garden path. Consider now structures like (79a, b):

(79) a. Ian put the candy on the table ... last night.

dom(IP, NP1), dom(IP, VP), dom(IP, v), dom(IP, NP2), dom(IP, PP), dom(VP, v), dom(VP, NP), dom(VP, PP), prec(NP1, VP), prec(v, NP2), prec(v, PP), prec(NP2, PP)

b. Ian put the candy on the table ... into his mouth.

dom(IP, NP1), dom(IP, VP), dom(IP, v), dom(IP, NP2), dom(IP, PP), dom(VP, v), dom(VP, NP), dom(VP, PP),prec(NP1, VP), prec(v, NP2), prec(v, PP), prec(NP2, PP), dom(NP2, PP)

After reading the PP *on the table*, the structural description in (79a; i.e. VP attachment of the PP) should be preferred due to simplicity. However, if the sentence is continued with another PP that has to be interpreted as the argument of *put*, the PP *on the table* must be reanalyzed as being dominated by the preceding NP. Since the exclusivity constraint (p20) defines that constituents may *either* precede *or* dominate each other, the precedence relation prec(NP2, PP) has to be deleted in order to allow for the new dominance relation (the relation typed in boldface in 79b). Thus, (79b) is supposed to be a conscious garden-path.

One major problem of Gorrell's approach is that it overpredicts conscious garden paths (Sturt, 1996). If (79b) is a garden-path, then so should (80) where disambiguation is accomplished by a bias in plausibility.

(80) The spy watched the cop with the revolver.

Though it is well known that attachment of the PP to the preceding NP in sentences like (80) is a little more costly than attachment to the VP, it is by no means a conscious garden path. As in the preceding example, a precedence relation has to be deleted in order to allow the PP *with the revolver* to be dominated by the preceding NP.

Underspecification models in general will be discussed further in chapter 4.4, which addresses the predictions of serial, parallel, and minimal commitment approaches.

2.14 Construal Theory

Construal Theory (Frazier & Clifton, 1996) is the current successor of the *Garden-path Theory* (see section 2.6). The move from *Garden-path Theory* to *Construal Theory* was mainly motivated by a series of cross-linguistic experiments on relative clause attachment in sentences like (81).

(81) The daughter of the colonel who was standing on the balcony ...

The principle of *Late Closure* (p10) would clearly predict a low attachment preference of the relative clause *who was standing on the balcony* to the modifying NP *the colonel*. Indeed, the predicted preference seems to show up in English questionnaires

(Cuetos & Mitchell, 1988), though only inconsistently (Carreiras & Clifton, 1993). For Spanish, however, a very consistent preference for attaching the relative clause high to the NP "the daughter" was established in off-line questionnaire data as well as online self-paced reading experiments (Cuetos & Mitchell, 1988; Carreiras & Clifton, 1993), clearly contradicting the principle of *Late Closure*. A comparable high attachment preference at least for the final interpretation has since been shown to be viable in French (Zagar & Pynte, 1992), in Italian (de Vincenzi & Job, in press), in Dutch (Brysbaert & Mitchell, in press), and in German (Hemforth, Konieczny, & Scheepers, 1994).

To account for the effects found for relative clauses, Frazier and Clifton introduce a distinction between two kinds of syntactic relations, *primary relations* as defined in (p25) and non-primary or *secondary relations*, which are (roughly) modifiers.²⁹

(p25) Primary phrases and relations include

- a. the subject and main predicate of any (+ or -) finite clause
- b. complements and obligatory constituents of primary phrases.

(Frazier & Clifton, 1996; p. 41)

If a phrase can be attached either as a primary phrase or as a non-primary phrase, attachment via a primary relation will be preferred. If all alternatives are primary relations, the structural core principles of the *Garden Path Model*, namely, *Minimal Attachment* and *Late Closure*, still hold. In NP/S-ambiguities like (82), for instance, both attachment alternatives are primary. In this case, *Minimal Attachment* favors the simple NP reading, as required in (82a).

(82) a. John knew the answer to the physics problem very well.

b. John knew the answer to the physics problem was wrong.

In other constructions, the predictions may still meet those discussed for the *Garden-path Theory*, albeit for different reasons. For PP-attachment ambiguities like (83), for example, a preference for verb attachment is predicted.

(83) Joe bought the book for Susan.

The PP *for Susan* is considered to be a potential argument of the verb and a modifier of the object noun, thus, a preference for attachment via primary relations favors verb-attachment.

For exclusively non-primary alternatives, however, the *Construal Principle* (p26) is postulated.

^{29.}Note that despite of the overlapping terminology the definitions of primary and secondary relations differ considerably from Gorrell's (1995).

(p26) a. Construal Principle

i. Associate a phrase XP that cannot be analyzed as instantiating a primary relation into the current thematic processing domain.

ii. Interpret XP within that domain using structural and non-structural (interpretive) principles.

b. Current thematic processing domain

The current thematic processing domain is the extended maximal projection of the last theta assigner.

(Frazier & Clifton, 1996; p. 41f)

Thus, the attachment of non-primary phrases is initially underspecified in that the phrase is only *associated* to the *thematic domain* produced by the last theta assigner. The ultimate attachment site within the thematic domain is then determined by structural as well as higher-level (interpretive) principles involving semantic and pragmatic preferences.

For non-primary ambiguities, the predictions may have changed completely. Consider again sentence (81). Instead of purely structural preferences, attachment preferences in these structures appear to depend on the type of the potential host of the relative clause. (For Spanish and English, see Gilboy et al., 1995; for German, see Hemforth, Konieczny, & Scheepers, forthcoming). All kinds of pragmatic and semantic principles are claimed to influence the final attachment decision. In this particular case, *referentiality* of the host seems to play a role (i.e. relative clauses tend to attach to referential, definite, hosts, like *table* in 84), as well as the thematic processing domain closest to the relative clause (i.e. they prefer to stay within the most recent thematic domain, opened by the preposition *with* in 85).³⁰

(84) ... the table of wood that was from Galicia (Frazier & Clifton, 1996, p. 30)

(85) ... the girl with the hat that looked funny (op. cit., p. 31)

Language specific differences in attachment preferences are also supposed to be due to the interplay of semantic, pragmatic, and structural principles. Consider the NP *the daughter of the colonel* in (84). If the speaker/writer had intended an interpretation that requires the attachment of the following relative clause to the head noun *daughter*, she could have used an unambiguous construction in English, namely *the colonel's daughter*. According to Frazier and Clifton (1996), the listener/reader is aware of alternative unambiguous constructions. If one of the readings of a structure could have been expressed unambiguously, the alternative reading is preferentially pursued because the receiver presupposes that the message is as *clear* as possible. It is thus assumed that the Gricean maxim of clarity is responsible for a low attachment (or at least a less articulated high attachment) preference in English constructions like (84).

^{30.}The preposition *with* is assumed to assign a thematic role to its argument whereas semantically empty prepositions like *to* or *of* only transmit the thematic role assigned by the head noun (e.g. the student of physics) or the verb (e.g. He sent the book to her.).

A preliminary discussion

From a computational point of view, this example suggests that Construal Theory implies a structure *generator* which produces all structural variants of the currently parsed input in order to find out whether or not the input could have been formulated *unambiguously* in a different way. Such a generator would have to be intertwined with the parser. This does not occur to me as the most parsimonious assumption one can think of. In the light of the fact that *Minimal Attachment* and *Late Closure*, which are still assumed to be valid, have been motivated by considerations of structural *economy*, Construal Theory does not appear to be a very consistent model.

In general, many aspects in Construal Theory are only vaguely specified. No detailed specification of the notion of *association* is given and, as a consequence, no specification of the interplay of semantic, pragmatic, and structural principles is proposed either. Since all kinds of language related information may influence parsing decisions in structural association ambiguities, *Construal* can cope with a huge variation of empirical results. It is unclear, however, whether any evidence on non-primary relations can ever falsify the approach³¹.

Furthermore, many of the wide range of phenomena accounted for by the *Garden Path Model* (such as "Tom said Bill died yesterday") fall into the domain of *Construal* now, leaving only very few constructions (such as 82) to be covered by the *Garden Path Model*. Under these circumstances, the model appears somewhat overpowered.

The definition of *primary relations* plays the crucial role in determining whether or not the *Garden Path* principles apply. It turns out that what is a complement is fairly vague: in many cases it seems to be determined by the fact that a certain type of phrase can *in principle* be a complement, not that it really is. Consider (86).

(86) John put the book Bill was reading in the library.

One would usually assume that *in the library* can be a complement to *put*, but not to *reading*. If so, the latter would fail to be a *primary relation*. Then, however, high attachment would have to be predicted, due to the priority of primary relations. In order to predict the correct *Late Closure* preference, the PP must be able to be considered a primary relation, such that the principles of the *Garden Path Model* can apply. Indeed, the authors assume the PP to be a primary relation, because a *locative* PP could *in principle* be one, even if it turns out later not to be one. The consequences of such an assumption are far from clear yet.

Nevertheless, Hemforth et al. (1994, forthcoming) present evidence that cannot be covered by *Construal* easily. In questionnaires as well as in eye tracking experiments, they found that whereas *relative clauses* in structures like (87a) are preferentially attached high to the head noun, PPs (87b), which were matched in their content as closely as possible, show a preference for the attachment to the more recent noun.

(87) a. The daughter of the teacher who came from Germany ...

³¹. This criticism holds for any type of *multiple constraint model*, of course.

b. The daughter of the teacher from Germany ...

Note that all kinds of non-primary relations should be subject to the same set of principles which determines the final interpretation (whatever these principles are for a particular instance). PPs in sentences like (87b) have to be attached to the phrase marker via a non-primary relation. Nevertheless, in contrast to relative clauses, PPs clearly obey the predictions made by *Late Closure*. Hemforth et al. (forthcoming) propose that the differential attachment preferences are due to an interaction of structural and anaphoric processes, the latter being initiated by the relative pronoun. Language specific differences are assumed to be due to a differential sensitivity of the relative pronoun to anaphoric information: the anaphoric aspect of relative clause attachment might generally be reduced for languages such as English which allow for reduced relatives without an overt relative pronoun.

2.15 Tuning

Mitchell and Cuetos (1991; Mitchell, 1994; Mitchell & Brysbaert, in press) provide an alternative account of the language specific differences found for relative clauses. The *tuning hypothesis* is based on a variant of the *Garden Path Model* claiming that fully determinate phrase structures are built during parsing, but always only one at a time. There is neither any parallelism nor any underspecification. However, no general parsing principles like *Minimal Attachment* or *Late Closure* are assumed to determine parsing decisions in cases of ambiguity. Instead, the human parser initially chooses the structural alternative that has been successfully employed *most frequently* in the past. Interlingual or even interindividual differences in parsing preferences are due to different *learning histories*.

Mitchell (1994) substantiates this claim by corpus counts in Spanish, French, and English which predict the parsing preferences found in the particular languages, if only the structures are counted at the "correct" grain size (e.g. information at the level of *definiteness* of determiners appears to be too fine grained, according to Mitchell, Cuetos, Corley, and Brybaert, 1995).

Furthermore, preferences of children were shown to be strengthened by training sessions, indicating a biasing influence of the learning history.

Although Mitchell and colleagues have (to my knowledge) never committed themselves to a concrete computational approach, their account would fit a probabilistic ATN-approach (Wanner and Maratsos, 1978, see section 2.5) best, if one assumes arclabels at the right grain size (probably at the level of major category) and arcs that can be dynamically ordered according to the recent success history.

A preliminary discussion

It is rather unclear whether or not statistically based models do much more than restating the problem originally posed. It is highly plausible that preferences for particular structures are reflected in corpus analyses as well as in comprehension studies. Predicting parsing preferences by the frequency of occurrence of a particular structure shifts the burden of explanation to sentence *production*: why is it that some structural alternatives are produced more often than others?

Furthermore, the training effects presented by Mitchell (1994) may not only be due to influences of the learning history, but to a change in the sensitivity to the information provided by different parts of the relative clause (see Hemforth, in prep.).

2.16 Parametrized Head Attachment

In a series of self-paced reading studies (see Konieczny, Scheepers, Hemforth & Strube, 1994), PP-attachment preferences could be shown to vary depending on where the verb was placed in German main clauses and subclauses as in (a, b).

(88) a. Manfred beobachtete den Mann mit dem Fernglas.

"Manfred observed the man with binoculars."

b. Daß Manfred den Mann mit dem Fernglas beobachtete, ...

That Manfred the man with the binoculars observed, ...

"That Manfred observed the man with binoculars, ..."

In main clauses such as (a), subjects preferred to attach the PP *mit dem Fernglas* either to the VP or to the preceding NP depending on a verb-related lexical expectation of a particular instrument. However, subjects showed a reliable preference towards attaching the PP to the preceding NP in subclauses (b), where the verb is placed at the end.

It is not obvious how parsing models such as the *Garden Path Theory* (Frazier, 1987, section 2.6) or its current successor *Construal Theory* (Frazier and Clifton, 1996; see section 2.14) could account for these findings. Since the structure is assembled fairly top-down in these models (Frazier, 1987), the placement of the verb in the surface string should not be able to determine the preference for where a potential complement is to be attached.

In the same experiment, the preference changed in verb-second sentences with verbs that only posed a weak expectation³² of an instrument, such as *erblicken* (*catch sight of s.th.*) in (89).

(89) Manfred erblickte den Mann mit dem Fernglas.

"Manfred caught-sight-of the man with binoculars."

Subjects preferred to attach the PP *with binoculars* to the preceding NP in such sentences. Comparable on-line effects of *lexical preferences* (see section 2.7) have been established by others as well (Konieczny, 1989; Trueswell, Tanenhaus, & Garnsey, 1994).

^{32.}The "weak expectation" of an instrument had been established in a series of pre-tests.

Konieczny et al. (1994) proposed the *Parametrized Head Attachment* principle (PHA, p27) to provide a thorough account of the variety of ambiguity phenomena, including the verb-placement effect and the lexical preferences on PP-attachment. In chapter 4.1, I will present a series of eye tracking studies on PP- and NP-attachment ambiguities to further substantiate the validity of this principle.

(p27) Parametrized Head Attachment(PHA)

a. Head Attachment, HA (Konieczny, Hemforth & Strube, 1991)

If possible, attach a constituent g to a phrasal unit whose lexical head has already been read.

If more than one phrasal unit remains as a potential attachment site,

b. Preferred Role Attachment

attach the constituent g to a phrasal unit whose head provides a required (obligatory) or expected (optional) theta- or place/time- role for g.

If further attachment possibilities for *g* remain, then

c. Recent Head Attachment

attach the constituent g to the phrase whose lexical head has been read most recently.

In verb-final constructions such as (b), *Head Attachment* (p27a) predicts a preference for attaching incoming material to the preceding NP, since its lexical head, the noun "*Mann*", is the only one that has already been read. In verb-second constructions such as (a), however, the lexical heads of both the VP ("*beobachtete*") and the NP ("*Mann*") appear prior to the PP. Thus, attachment to either the VP or the NP is preferred according to the role expectations (p27b) of their respective heads, or *recency* of their occurrence (p27c).

The *PHA* model, like the *Garden Path Theory*, was proposed as a first analysis model that ignores semantic or pragmatic aspects of the discourse situation in its initial syntactic analysis. However, the *PHA* model regards lexical information, such as thematic features of lexical heads, and their particular order in the sentence as crucial for the first analysis. Only later can the initial analysis be revised by higher-order information such as world knowledge or pragmatic aspects introduced by the context (Konieczny, Hemforth and Völker, 1995, see also section 4.2).

The central idea in *Parametrized Head Attachment* is that that syntactic structure is proposed first which can be semantically evaluated earliest. Processing is thus assumed to be *semantics oriented*. The parser is forced to attach an ambiguous constituent preferably to a phrase whose lexical head allows the thematic integration of the phrase. This, of course, is only possible if the lexical head is already there. Furthermore, if more than one potential head is available, the parser will be guided by the role-taking properties of the respective heads. In general, the parser guided by PHA initially builds the structure that allows the maximum semantic integration, i.e. it is *semantics oriented*. I will not give any detailed examples for the application of PHA to various parsing preferences in this chapter, because the discussion will span most of the chapters to follow. In chapter 5, I will introduce a parser whose behavior covers the predictions of *Parametrized Head Attachment*. This will be illustrated on the basis of a wider range of examples.

A preliminary discussion

As a parsing principle suited for structural ambiguity resolution, PHA leaves a number of processing properties fairly vague. For that reason, PHA was assumed to be integrated in a parser which fulfills further requirements (see Konieczny et al., 1994, Hemforth et al., 1992), such as a left-corner algorithm that establishes a strictly linear attachment of words (i.e. each item is integrated into the structure as soon it is encountered, Frazier, 1987b). In contrast, the approach taken later in this thesis sets up a parsing mechanism from which the predictions of PHA emerge, instead of hard-coding PHA as a special purpose heuristic routine whose operating properties are independent of those of the parser.

Moreover, the principles assumed by PHA are vague with respect to whether or not attachment alternatives are assumed to locally *compete* with each other, as in the model of Stevenson (1995). This particular problem will be addressed in chapter 4.2, and as a result, a sequential mechanism will be outlined in chapter 5.

2.17 Conclusion

In this chapter, I have presented a series of models of human sentence processing which, in one way or the other, try to account for initial parsing preferences and partly for recovery from parsing failures. Apart from the detailed parsing principles assumed, general architectural claims are made more or less explicitly.

2.17.1 Serialism, parallelism, and underspecification

In cases of structural ambiguity, fully determinate structures are claimed to be built serially in some models as in *Garden Path Theory* (see section 2.6) and in Ford, Bresnan, and Kaplan's *Theory of Syntactic Closure* (see section 2.7). In other models all alternative structures are considered in parallel at least locally, as in Gibson's account (see section 2.10), the multiple constraint competition model put forward by Mac-Donald et al (1994; see section 2.11), or Stevenson's (1993, 1995) hybrid model of sentence processing (see section 2.12). Underspecified representations of phrase structure are assumed by Gorrell (1995; see section 2.13) and, at least partly, by Frazier and Clifton (1996; see section 2.14).

2.17.2 The status of lexical and higher order information

The models presented also differ with respect to the question of which kinds of information they allow to guide initial structure building. Parsing decisions may be mainly grounded on purely structural information as in *Garden Path Theory*. Ford,

Bresnan, and Kaplan, as well as Abney (see section 2.8), Pritchett (see section 2.9), Gibson (see section 2.10), MacDonald et al., Stevenson, and Konieczny et al. (see section 2.16) emphasize the guiding force of lexical information, in particular information on complements which can be predicted from lexical heads. Higher order information, as world knowledge or pragmatic conventions, are supposed to constrain initial attachment in MacDonald et al.'s and in Stevenson's model for all kinds of structures, and in *Construal Theory* for secondary relations.

2.17.3 The competence base

Most of the models presented in this chapter were based on one or the other variant of transformational grammar, the more recent ones mainly on the theory of Government and Binding (Chomsky, 1986). An exception is Ford, Bresnan, and Kaplan's account which is based on Lexical Functional Grammar (Bresnan & Kaplan, 1982). The models strongly differ with respect to the degree of responsibility for performance phenomena attributed to the principles or rules of the grammar. The strongest claim is made by Pritchett (1992), although his On-Line Locality Constraint does not automatically emerge from the principles of GB (cf. Gorrell, 1995). Gibson's approach is similar in that some of the guiding properties are derived from principles of GB. Nevertheless, he assumes the parser operates on phrase structure rules pre-compiled from the grammar, weakening its transparency (Berwick and Weinberg, 1985). Gorrell's approach, then, is based on representational assumptions (*d-theory*) which are completely independent of a particular grammar theory, whereas the major structure building principle (Simplicity) is derivable from GB and the Minimalist Program (Chomsky, 1995). The basic "parsing" operations in Stevenson's model, namely attaching and deactivation, are proposed at a level certainly lower than that of grammar-theoretical constructs. Nevertheless, the attaching mechanism is strongly determined by the *theta criterion* in GB.

A somewhat weaker stance is taken by Ford et al. (1982) in claiming that the rules of the grammar must correspond to the steps of structure generation, as formulated in the strong competence hypothesis (Bresnan, 1982). However, the parsing principles they assume do not emerge from any property of the grammar, but are additional stipulations that have nevertheless been shown to be seamlessly integrated into their LFG-based model. Note, however, that LFG is a rule-based theory, posing much weaker requirements on a parser than a principle-based theory like GB. The Garden Path Theory is slightly different, in that the principles Minimal Attachment and Late Closure are motivated by the more psychologically-oriented assumption of structural economy, and do not automatically follow from grammar-theoretical assumptions. Frazier (1986), while grounding her model on a variant of GB, claims that phrase structure rules are the units of mental representation, thus taking a weak notion of competence. PHA (Konieczny et al. 1994) does not require any commitment to grammar-theoretical assumptions and is exclusively motivated by a different kind of economy, which does not assume the structures to be the entities relevant to economy, but the number of unintegrated referential entities (see section 4.2). Note, however, that economy only *motivates* PHA, while not *determining* the parsing process directly.

Finally, MacDonald et al.'s model is completely independent of grammar theoretical assumptions. All preferences are assumed to be caused by the strength (frequency) of multiple constraints located in lexical items.

2.17.4 Incrementality

The time course of structure building is a further distinguishing feature of the various approaches. The amount of incrementality depends on the kind of grammar theory adopted. The transparent parsing strategy for principle based and lexicalized grammars is *head driven*. In *head licensing* models such as Abney's, Pritchett's, possibly Gorrell's, MacDonald et al.'s and Stevenson's, items cannot be attached to a phrase marker of the sentence before the respective lexical head licensing the attachment is encountered.³³ Fully incremental or *linear* structure building is explicitly assumed in Frazier's *Garden Path Theory* and Konieczny et al.'s model, as well as in Gibson's. Although the *Theory of Syntactic Closure* (Ford et al., 1982) assumes rule-based topdown processing, phrases are only attached when they are complete, thus diminishing the degree of incrementality assumed.

2.17.5 What are the constraints of the human parser?

The various constraints, either explicitly or implicitly assumed by the models presented so far, do have empirically testable consequences. Throughout this thesis I will discuss these assumptions in the light of psycholinguistic evidence, to a large degree based on experiments presented in the following chapters. I will argue in favor of a serial model that makes use of the availability of lexical heads and their respective properties, where fully determinate structure is built incrementally, and which is based on Head-driven Phase Structure Grammar (HPSG, Pollard & Sag, 1987, 1994).

Before I turn to the empirical evidence posing further constraints on parsing models, I will justify the use of eye tracking as an on-line method for investigating parsing preferences in the next chapter. Eye tracking will be discussed in comparison to other on-line measures, and eye tracking measures will be introduced that are claimed to be the most sensitive to an on-line processing load.

^{33.}Note that head projection implies a certain amount of underspecification, because no fully determinate structure is built locally in the absence of a licensing head.

3 On-line techniques in parsing research

In recent years, several sophisticated techniques became quite popular in the study of the moment-to-moment processes that occur during reading. Among the early techniques used in psycholinguistics are *phoneme-monitoring* for spoken language, and the *lexical decision task* at a certain point during *externally paced* reading (e.g. Clifton, Frazier, and Connine, 1984). All these techniques suffer from the fact that a secondary task has to be performed which is likely to interfere massively with normal reading.

Since the advent of cheap computer technology in the early 80s, *subject-paced presentation* techniques became increasingly popular. In self-paced reading, subjects read sentences displayed on a computer screen bit-by-bit such that each new segment of the text is only presented on demand; i.e. the subject pressing a button after having finished reading the current segment. The underlying logic is that the time between two button-presses reflects the time needed to read the section, i.e. to perform several kinds of linguistic processes, such as lexical access, syntactic and semantic integration, and so on. Presentation can be word-by-word or in larger segments, either segments of a fixed length, phrases or even entire clauses. Sometimes, an on-line secondary task, such as lexical decision, is combined with reading in order to force the subject finish some processing of a word before moving on to the succeeding word.

Though the self-paced reading technique could be demonstrated to be very useful in the investigation of a variety of issues, there are quite a lot severe problems, some associated with secondary tasks, others with each of the segmentation procedures and again others with button-pressing during reading in general (see Mitchell, 1984; Günther, 1989, for an overview). Even in its most natural form, i.e. self-paced reading without any on-line judgements, the button pressing itself generally interferes with normal reading, resulting in highly increased reading times (word by word reading times are about twice as long as subjects would have spent on words in normal reading), and, even more importantly, in "rhythmization" effects; i.e. subjects tend to equalize the word reading times along the sentence. Therefore, effects do often *spill over* into later sections, or show up only at the end of the sentence, if at all (Mitchell, 1984). When using a moving window technique, presentation can be cumulative or non-cumulative. With cumulative presentation, the link between button-press times and linguistic processing is even more loose: subjects learn quickly that they can proceed through large segments of the sentence quickly and only then re-read it more carefully (Just, Carpenter, & Woolley 1982).

Though most of these problems have been known for quite a long time, another problem often ignored in the literature must be pointed out specifically. It is associated with the popular *phrase-by-phrase* presentation mode. In studies investigating attachment ambiguities, one should avoid identifying segment-boundaries with phrase-boundaries, because it has been demonstrated recently that subjects use segment boundaries as syntactic cues for closing phrases (Scheepers et al. 1994), which they would not have done without phrase-wise segmentation.

Summing up, self-paced reading data must be interpreted with much caution. A number of indispensable problems weaken the explanatory power of such results, some of which do not only simply reduce the sensitivity of the technique, but bear the potential to produce considerable artifacts. Note that even simply slowing down the reading speed may *produce* effects which would not have occurred in normal reading. Some higher level context effects, for example, are known to influence reading only after a certain amount of time, such that slow readers show context effects whereas others do not (Perfetti and Lesgold, 1977).

A very different and very sophisticated technique has emerged lately, namely the evocation of event related potentials (ERPs). The standard procedure is as follows: subjects read or listen to single words which are presented externally-paced on a computer screen or over speakers, respectively, while an EEG is taken from a number of sensors placed in certain positions on the subject's skull. It has be demonstrated that certain sorts of semantic violations in the material produce a negative potential shift about 400 ms after the on-set of the word introducing the violation, compared to the base-line obtained from a violation-free control sentence (Kutas & Hillyard, 1980, 1983, 1989). Furthermore, more recently, some research groups have discovered a positive shift about 300 to 600 ms after the onset of a word that introduced certain kinds of syntactic violations (Osterhout et Holcomb, 1992; Hagoort, Brown, & Groothusen, 1993; Mecklinger et al, 1995). These interpretations are, however, highly debatable and constantly changing, and the most recent interpretation of the syntactic positive shift is that it reflects reanalysis effort. Nevertheless, ERPs seem to provide data which allow researchers to distinguish qualitative patterns of syntactic and higher-level processes.

Although ERPs bear the potential to provide a deeper insight into the architecture of the human sentence processing mechanisms, the current techniques suffer from a number of severe problems. First, for technical reasons, the words are presented externally paced, quite often at a very slow constant pace of about 600 milliseconds for each word, allowing for the occurrence of undesired side-effects, and rendering the analysis of early parsing preferences difficult.

Second, and more importantly, in order to eliminate the immense amount of noise in the data, subjects have to perform a huge number of trials, leaving insufficient space for filler sentences which could distract from the problem under investigation. Moreover, in quite a number of studies using ERPs, this lead to poorly balanced material, such that subjects could have easily developed strategies to minimize parsing effort temporarily.

Third, as the direction of the potential shift is relative to the base-line, sometimes it is not clear whether some effects are due to the intended variation or to certain differences on the prior word (or whatever has been used as the base-line) that have produced an opposite potential change in the base-line (Steinhauer, personal communication)

To conclude, though the ERPs provide a promising tool for the investigation of online processes, the technique is still far from mature, and a whole lot of intrinsic problems still lack even the idea of a solution. So long, for the issues at hand, ERPs are currently of little help.

In the sections following, I will describe in more detail the technique of *eye movement recording*, which I used for the experiments reported in chapter 4.

3.1 Eye-tracking

Recording the eye-movements of subjects engaged in the task of reading emerged as one of the most informative techniques in psycholinguistic research. It has been shown that the duration that subjects spend on a word or segment reflects the processing load induced by that particular piece of material (Just and Carpenter, 1980; Rayner & Pollatsek, 1987). In this section I will briefly discuss some aspects of this technique relevant to this thesis, discuss standard procedures for data analysis and, as a result, introduce some new procedures, which will then permit us to draw more accurate conclusions from the data obtained. For a more detailed overview of the basic facts about eye-movements in reading, the reader is referred to Rayner and Pollatsek (1989), Rayner et al. (1989), and Just and Carpenter (1987).

The most obvious advantage of eye-tracking over other techniques is that subjects can read the materials in a rather "natural" way without having to perform an unnatural secondary task. A pitfall of eye-tracking, on the other hand, is that most equipments require subjects to have their head fixed in a way that they cannot move too much, which can be quite uncomfortable. However, subjects get used to the situation very quickly, usually after a few training items.

Although very informative, eye movement data are not perfect reflections of the processes associated with language comprehension. Several factors distort the signal,

some of which are not associated with any kind of language processing at all. Lowlevel perceptual factors and even purely motoric aspects can influence the fixation duration and certain properties of the saccades (O'Regan & Levy-Schoen, 1987; Vitu, 1993).

The *eye-mind span* (Just and Carpenter, 1980) is crucial for the use of eye-movement data in psycholinguistic research. If eye-movement patterns turned out to be completely independent of linguistic processes, they would obviously be of no use at all for the present purpose. There are two reasons at least why eye-movement records have to be treated with caution. First, there is the problem of the *perceptual span*. Readers extract useful information not only from the very spot the eyes fixate, but also from the *parafoveal* area, which ranges from 4 characters to the left to up to 15 characters to the right of the fixation point (McConkie & Rayner, 1975). Therefore, words are often previewed before they are actually fixated. Rayner et al. (1982), however, provide evidence suggesting that the main portion of information is extracted from the word fixated plus two or tree characters of the next (parafoveal) word.

Second, spillover effects observable on a word constant across conditions but succeeding the word that was actually manipulated have occasionally been demonstrated (Morrison, 1984). It is, however, by no means clear whether such effects are due to post-lexical processes, or whether the successor word is just deprived of an effective preview so that its lexical access is slowed, if the prior word was difficult to read (Henderson & Ferreira, 1990).

Summing up, extraction of information is closely, though not perfectly, related to where the eye fixates³⁴ and the evidence available in general suggests that eye movements do provide a good reflection of cognitive processes in reading.

3.2 Regions

A reasonable way to cope with the problems associated with the eye-mind span is to treat groups of several words as a single segment or *region* in the data analysis (not in the presentation of the material), such that all words in a region contribute to only one score. For each word in a region, the complete processing load is then accumu-

³⁴.Short words, mostly closed-class words, such as determiners, prepositions and conjunctions, are often skipped. "Null-fixations" can be handled in a number of ways, all of which are associated with both advantages and disadvantages. Firstly, they can be treated either as "zeroduration" fixations (*unconditionalized* analysis) and included in the data analysis, or, secondly, as missing values and excluded (*conditionalized* analysis) from data analyses. It has frequently been argued that including them as zero-duration fixations is rather implausible (Rayner & Pollatsek, 1989), since skipped words are usually not left unprocessed, but are most often previewed on the preceding word. If null-fixations were included, they would cause the average reading time to be artificially low. Excluding them, however, is inevitably associated with the loss of valuable information. Thirdly, some researchers replace null fixations with either the mean duration of the condition under consideration, or with the low cut-off value in that condition. However, the standard deviation will be artificially decreased in either of the cases, resulting in higher F-values and thus in higher probabilities to yield significant effects. Thus, the most accurate treatment of null fixation durations appears to be the exclusion from the data analyses (*conditionalized* analysis).

lated in the score, no matter whether it was actually fixated or only previewed (except for the first word, of course).

Region boundaries are usually identified with phrase boundaries. However, it is not always useful to do so. If, for example, the material has been designed to reveal an effect very early within a phrase, like, for instance , on the determiner of a nounphrase, it might be useful to treat the word preceding the critical word plus the critical word itself as the critical region. The reading time scores on this region will now include both the time spent on the critical word and the time spent for pre-processing the critical word, which may only be noticeable on the preceding word in cases where the critical word was skipped. However, it is important to keep the word prior to the critical word constant across the experimental conditions.

Since each word in a region adds a certain amount of variance to the data, a region should be as short as possible.

3.3 Standard measures

Once the range of a region for the data analysis has been defined, the duration the eyes spent in the region can be computed. Differently from the self-paced reading technique, which provides just one duration score for each region, the detailed information about every single fixation and saccade might allow the experimenter to distinguish different stages of processing. Readers do not only often fixate several times on a region, they also look back to earlier regions in the sentence, re-read portions of the sentence, fixate on the region again, sometimes go back again, and so on. The eye movement patterns thus provide quite a lot of information valuable for drawing a more complete picture of what actually might have happened when subjects read a certain part of a text.

The most common measures, the "classical" measures are *first fixation durations*, *first pass reading times* (*gaze durations*), and *total reading times*. In the early papers on eye-tracking experiments, only these measures were given in addition to the frequencies of regressions from the critical region. More recently, some new measures have evolved, which are suited to overcome certain pitfalls in the classical measures. It is important to be very clear about what kind of processing is reflected in each of these measures. I will therefore discuss each of them in detail in the following sections.

3.3.1 First fixation durations

If readers fixate only once on a word before they move on to the next word, the duration of the single fixation would reflect the underlying processes fairly well. Unfortunately, words, especially longer ones, are often fixated more than once. The duration of the *first fixation*, although often provided in psycholinguistic literature, is thus one of the most disputable reading time scores. First fixation durations (FFD) in single word regions are sometimes claimed to reflect processes of lexical access, which is supported by a correlation of FFDs with several lexical variables, such as word length (Inhoff, 1984). Some researchers even claim that FFDs reflect very early

higher level processes, such as initial analysis processes in the parser or plausibility effects (Murray & Rowan, 1996). However, it can be argued with good reason that FFDs are of low value with respect to the analysis sentence processing mechanisms.

First of all, even if the two aforementioned claims were true for single word regions, FFDs in longer regions, as provided by many researchers even in more recent papers, are quite often complete nonsense. FFDs cover only one fixation, which is necessarily a fixation on one word only. Since it is generally quite uncertain which word in a region is looked at first, the FFDs will necessarily cumulate scores from different words, including rather irrelevant ones. In particular, when the words of interest are positioned near the right boundary of the region, as, for example, in semantically biased PPs, the larger portion of data is probably collected from irrelevant material at the beginning of the region, such as a preposition and a determiner in a PP.

Furthermore, even in single word regions, it is completely unclear whether or not FFDs reflect lexical or even higher level processes. A lot of variance is obviously due to lowest-level aspects of eye-movements, such as oculumotoric strategies (see Vitu, 1993). If, for example, the first fixation is in a rather unfavorable position within the new word, e.g. on the last letter, the probability for a within-word saccade increases strongly, resulting in a decreased first fixation duration. The data currently available strongly suggest that FFDs do reflect lexical and higher level aspects to some extent, but only because some words are indeed fixated *only once* before the eyes move on to another word. In these and only these cases, the fixation durations will actually reflect processes at many, if not all levels of sentence processing.

Generally, however, since it is in general impossible to restrict readers to fixate words only once, FFDs will be contaminated by events of false saccade planning, i.e. fixations on unfortunate locations in a word. If these cases were excluded from the data, FFDs could nevertheless represent linguistic processing quite accurately. One would then be forced to only use short words (about 4 characters), in order to avoid too many rejected cases, which is not always feasible. An apparently more natural way to measure the reading time in a region is therefore to amass all fixations in a region into one score, as described in the next section.

3.3.2 First pass reading times

The first pass reading time (FPRT) accumulates the durations of all fixations on a region starting when the region is entered from the left, until there is a saccade to another region. For single word regions, first pass reading times are also called *gaze durations*. The FPRT is the most commonly used measure in psycholinguistic research. FPRTs are usually said to reflect the processing load induced by the portion of text in that particular region rather directly, since subjects will not go on to another region until the region is sufficiently processed. It is, however, not absolutely clear, what "sufficiently" actually stands for in this context: is it sufficient to syntactically attach an item to the structure, or must there be at least some portion of semantic, pragmatic and contextual interpretation? Just and Carpenter (1987) claim, and this claim is supported by the eye-tracking data they provide, that gaze durations even
reflect processes of semantic interpretation. In their *"immediacy hypothesis"* they propose that processes at every stage of processing start as soon as a word is read, and that as much processes as (logically) possible have to be finished before the eye moves on to the successor word.

In research on human sentence processing, researchers are often interested in the question of which analysis the parser prefers *initially* in the case of an ambiguity, in which more than one analysis is lexically and syntactically permissible. Parsing preferences are experimentally investigated using variations of sentences which are ambiguous at a certain location in the sentence. In order to get an insight into the initial parsing preferences, the sentences provide a certain section, often subsequent to the ambiguous region, that forces the parser to adopt a certain analysis in one of the variants, and another analysis in another variant. Presuming that the parser settles on a single analysis in the ambiguous region and pursues only the preferred analysis, one would expect processing difficulties when this section, the *disambiguating region*, is read, if it turns out to be inconsistent with the preferred analysis, either syntactically or semantically.

Most researchers agree that the reanalysis effort shows up in FPRTs, assuming that subjects fixate on a region until a consistent analysis has been found. This assumption, however, implies that FPRTs collect the processing durations of both the first and the second (nth) analysis in the parser. Since such FPRT-effects are observed even in experiments in which disambiguation is accomplished through the semantic bias of the ambiguous phrase itself, FPRTs can be claimed to reflect not only syntactic, but also semantic processes, sometimes even at the level of world knowledge.

It is important to emphasize that FPRTs do not (necessarily) inform us about the duration of certain processes. Some semantic evaluation processes, for example, might only be started within the time range of the FPRTs in a region where the relevant information becomes available, such that certain match or mismatch violations are also detected within the range of FPRTs, at least in a number of cases sufficient to produce a significant effect. Since the duration itself covers processes of different kinds, which may or may not proceed in a sequential fashion and some of which start in the preview phase, no conclusion about absolute durations of particular processes can be drawn (Rayner & Pollatsek, 1989). Furthermore, some effects, especially those induced by higher level violations, sometimes spill over into subsequent regions.

Are FPRTs a good measure of parsing preferences? Not necessarily. Unfortunately, subjects do not always stay in a region until they find a consistent analysis. It happens quite often (up to about 40% of the trials) that subjects look back to earlier places in the sentence in order to read some portions of the sentence again. In these cases, FPRTs will not reflect all processes induced by a mismatch, namely pursuing a second analysis and, if it works, its semantic integration and evaluation. Thus, the FPRT itself will probably be considerably short, because it only represents the time needed to detect the mismatch and to initiate a saccade to prior words in the sentence relevant in the reanalysis process. In the light of the fact that usually more regressive saccades can be observed in the mismatch condition, the average FPRTs are likely to be contaminated by the trials with regressions, so that, paradoxically, shorter average FPRTs

might be observed in regions with material more difficult to process than in those with consistent material.

FPRTs would therefore be treated more properly, if trials in which regressive saccades from the critical region after first pass reading were treated separately from trials without regressions after first pass reading, as suggested by Altmann et al. (1992). In the so called *regression contingent analysis*, the set of all trials is separated into two subsets with (henceforth FPRT+) and without a regressive saccade (henceforth FPRT-) out of the critical region after first pass reading. The rationale is that FPRTs will represent more comparable events within each of the subsets. It is obvious that FPRT- are considerably sensitive to garden path effects. Altmann et al. (1992) provide data that seems to suggest that the FPRTs corresponding to fixations followed by a regressive saccade (i.e. FPRT+), though considerably smaller, are as well sensitive to some garden path phenomena, as are FPRTs not followed by a regression (i.e. FPRT-). The result of our experiments provided later in this thesis will further support this assumption. Note, however, that FPRT+, corresponding to fixations followed by a regressive saccade might reflect different aspects of processing than FPRT-, corresponding to fixations followed by a progressive saccade. Whereas FPRT+ will probably not reflect complete (higher level) integration processes, since re-reading of text outside of the region has been done prior to that, FPRT- may very well do so. One could therefore argue that initial analysis preferences were much better mirrored in FPRT+. Note, however, that arguing about parsing preferences on the basis of FPRT+ means comparing a condition, in which regressions are initiated after a parsing failure as a consequence of a misguiding parsing preference, to another condition where no such failure would be expected and where the regression must have occurred rather accidentally. What FPRT+ reflect in the latter case is completely unclear.

Note also that the post-hoc division into two groups of trials poses several statistical problems. First of all, since the FPRT+ cases are excluded from the FPRT- data, and vice versa, both suffer from a huge number of missing data. Since the separation is done on a post-hoc basis, the exclusion of data may be pretty unbalanced over conditions, potentially resulting in empty cells in the design. This is in particular true for FPRT+, since readers perform regressive eye-movements in only 5 to 40% of the trials, and definitely more often in the garden-path condition than in the control condition. Although there are ways of coping with this problem, regression contingent data must therefore be treated with caution.

Conceptually, the *regression contingent analysis* appears to be a partly consequential step in the right direction. As an important pitfall, it ignores valuable information in the data, namely the re-reading of the prior text itself, which is more likely to happen the "stronger" the actual garden path effect is .

3.3.3 Total reading times

The third "classical" measure is the sum of all fixations in a region, regardless of whether or not the fixations occur during first pass reading or during a re-reading period. The *total reading times* (TRTs) are claimed to reflect processes at later stages in processing. However, the claim that the TRTs will necessarily reflect *preferences* at a

later stage of processing, but not at early stages, is definitely too strong a claim, as will become clear shortly.

The argument is a consequence of that which has been said in the section about FPRTs above. If the re-reading phase follows a regressive saccade from that region, the second pass reading of a region is within what I will henceforth call the "*regression path*", meaning the set of *all fixations following a regressive saccade from a region until the region is fixated again or skipped*. The *regression path* is thus the set of all fixations on the portion of text that is re-read after something has caused a regressive saccade from the critical region. In the cases where the TRTs result mainly from re-readings *within* the regression path, the TRTs may actually reflect *first analysis preferences*, and they will do so even better than FPRTs in these cases (see last section), because the complete re-reading process, which is a direct result of a failure of the *initial analysis*, is included in the TRT and not in the FPRT.

If, however, the re-reading mainly results from a regressive saccade from a subsequent region, it is completely unclear what the TRTs actually represent. Nothing in the TRT tells the researcher which kind of re-reading contributes most to it, and at what location the re-reading has actually been initiated. Therefore, we consider TRTs usable only as a rough estimation of what might have happened in a particular region in comparison with FPRTs, but not as a serious measure of processes at any stage of processing.

This criticism extends to another measure, sometimes misleadingly called *second pass reading time*, which is simply computed from the *total reading time* minus the *first pass reading time*. It should be obvious here that this measure is a far from perfect estimation of second pass processes in a region, though sometimes provided as such in parsing literature.

3.3.4 A preliminary conclusion

I have argued against the standard assumption of the classical measures *first fixation duration, first pass reading times,* and *total reading times,* namely that they reflect processes at different stages of processing, in particular lexical and sub-lexical processes, early syntactic processes, and processes of the thematic processor, respectively. Each of the measures may or may not reflect processes at every level of processing, although they have been demonstrated to be *correlated* with variables at different stages to different extents (Rayner & Sereno, 1994). It should be clear from what has been said above, however, that neither are the measures clean in a sense that they only reflect certain processes and not others, nor do they *completely* (sufficiently?) reflect the processes they should reflect, in the sense that some portions of these processes actually take place in regions other than the critical one.

It must be emphasized that even though eye-tracking will presumably not provide a direct access to the *processes* at different levels, eye-movement data certainly can give an insight into the human parsing architecture. The chain of inference is rather indirect in that it has to pass evidence about parsing *preferences*³⁵ resulting in parsing *difficulties* in experimental designs set up appropriately. The criterion of quality can thus be reduced to the question of to what extent a measure reflects processes of *reanalysis*.

All three classic measures, then, suffer from one drawback: they focus on one region only. The big advantage of the eye-tracking technique, however, is that researchers can not only keep track of the fixations in a single region, but also of the saccades to, and fixations in different regions. Such data provide indispensable information about what is going on during reading.

3.4 New measures

3.4.1 Regression-path durations

Regression path durations (RPDs), first proposed in Konieczny et al. (1994) were developed to overcome the above-mentioned shortcomings. RPDs accumulate the durations of all fixations in the *regression path* (see definition in section 3.3.3) and add them to the first pass reading time. Thus, the measurement of the RPD starts as soon

^{35.}I will ignore here the fact that observations of reading latencies can also be accounted for differently, especially in multiple constraint type of models (MacDonald et al., 1994; see section 2.13). The point I want to make here is that architectural interpretations of the data can only be made on the basis of the experimental design, not from different types of eye-movement measures.

as a region is entered with a progressive saccade, and it ends when the region is left with a progressive saccade, as illustrated in figure 1^{36} .



FIGURE 1. Regression path durations

Because RPDs include fixations in regions prior to the critical region following a regressive saccade from the critical region, they reflect the processing load induced in that region more realistically.

Note also that RPDs are, more than any other eye-tracking measure, comparable to self-paced reading times (without its shortcomings, though), since both reflect the time the reader needs to shift attention to the next word or region.

^{36.}RPDs have been used under different names recently, such as *cumulative region reading times* (Brysbart and Mitchell, 1995), *go past* (Crocker et al., personal communication), *right boundary* (Traxler and Pickering, personal communication), and the *Liversedge first pass measure* (Liversedge, 1994). A similar measure has been developed very early by Kennedy et al. (1989), called *total pass per word*. Leaving aside the rather unclear definition of *total pass* (do all fixations have to result from regressions directly?), this measure suffers from rather implausible presumptions. For reasons not obvious to the author, the total pass is *divided* by the number of word*s in the critical region*. See section 3.5 for a detailed discussion of methods for length adjustment.

I will adhere to the term *regression path duration*, because it will allow me to focus on different aspects of the *regression path* needed for the definitions of further measures below.

RPDs in early regions will generally be lower than RPDs in later regions, simply because there is more "space" for longer regression paths in the latter case. Therefore, only RPDs of the conditions should be compared, in which the critical regions are equally distant from the beginning of the sentence.

Problems

RPDs, using the current definition, even collect multiple re-readings of the preceding text. Therefore, RPDs will possibly reflect not only first-pass, but also second pass, and in general n-pass preferences. In order to reduce the impact of preferences in later stages, however, one can restrict the RPDs to the *first* regression path, as proposed by Konieczny et al. (1995, submitted). *First RPDs* incorporate all fixations including the fixations in the critical region when it is read again (for the second time), but excluding fixation following further regressive saccades. First RPDs are particularly useful in regions at clause boundaries. Whereas total RPDs must be treated with caution at clause boundaries, because they will incorporate wrap-up processes which probably reflect processes at all levels and preferences at many stages of processing in addition to mere re-checking, first RPDs can reduce the problem to some extent (though not necessarily). The best thing to do in clause boundary regions, however, is to stay away from RPDs at all.

A second possible problem with RPDs might be caused by the fact that RPDs in different conditions may result from qualitatively different regression paths, i.e. from fixations on different portions of the text. It may very well occur, for example, that in the condition in which readers garden-path more often, they produce shorter regression paths, which only include fixations on material relatively close to the disambiguating region, whereas the regression paths in the non-garden pathing condition are comparably long both spatially and in duration, though regressions will occur less frequent. On average, then, RPDs may give a reasonably reversed picture of actual parsing difficulties. Note, however, that the RPD variance would be considerably high in the latter condition, since the RPDs would include both the low values from the most frequent trials with no regression after the first pass reading, and the extremely high values caused by rare, but extremely long regression paths. No effect is likely to be observed in such a situation.

There are (at least) two sensible solutions to this problem, both of which are currently under development but not yet implemented in the analysis procedure used in this thesis. Nevertheless, they should be mentioned here in order to point out directions of future research.

3.4.1.1 Selective RPDs

Standard RPDs do not distinguish between qualitatively different regression paths. It can be of particular interest, however, to distinguish re-reading of portions of the text relevant to the reanalysis from the mere re-reading of the entire sentence or only some irrelevant portions of the text. Furthermore, it may also be sensible to distinguish different types of regression paths which all incorporate re-readings of relevant passages.

For example, imagine a PP attachment ambiguity in which the PP can either be attached to the VP or to the NP, such as in (90).

(90) The spy saw the cop with binoculars.

Depending on which attachment is preferred by the parser, one could expect different regression paths, if the disambiguating information rules out the preferred NPattachment or the preferred VP-attachment, respectively. Under the hypothesis that the preferred but false attachment site will be re-fixated mostly³⁷, the regression path would range only to the NP (*the*) *cop* in the former case, but also to the verb *saw* in the latter case, resulting in a considerably longer regression path here (provided the words in between are fixated as well). Separate data analyses for both cases would help to confirm or falsify such an assumption.

3.4.1.2 Regression contingent RPDs

The standard RPD measures include both trials with and without regressive saccades following first pass reading. For those trials without a regressive saccade, RPDs by definition do not differ from FPRT(-). It might be informative, however, to look at the RPDs particularly in those cases in which a regression from the critical region actually occurs. Since the re-reading is likely to last much longer than simple firstpass reading times, the distribution of RPDs can be bi-modal, indicating different types of events. If this pattern shows up in the data, it would be sensible to distinguish cases with (RPD+) and without a regressive saccade (FPRT-) after first pass reading.

3.4.1.3 Load contribution

Along the regression path, some regions contribute more to the RPD than others. Especially when reanalysis processes are to be investigated, it is of particular interest which regions are most intensively re-processed after the initial analysis failed. For example, researchers might come up with the hypothesis that the ambiguous region is most important in reanalysis, or the potential attachment sites, or only thematic licensers, and so on. I propose the following measures, which, I believe, will obtain relevant information: first, the frequency of fixations in these regions *within the regression path* of the critical region, and second, the sum of the corresponding fixation durations, i.e. the *total reading times* within a regression path. I will call the latter *load contribution*.

In contrast to the standard measures and the RPDs, *load contribution* (as well as regression path dependent fixation frequencies) is a two-place function $lc(reg_j, reg_k)$, namely the total reading time in a region (reg_j) within the regression path of another region $(reg_k)^{38}$.

^{37.}Of course, one could come up with the opposite hypothesis equally well.

As an alternative to the selective RPDs described above, one could also cumulate the load contributions of the regions regarded as relevant for the reanalysis (or in other words, exclude the load contribution of irrelevant regions from the RPD).

In general, I strongly believe that load contribution measures will help us to uncover important properties of the human sentence processing mechanism at many of its levels.

3.5 Accounting for lexical effects

It has been demonstrated that the time readers spend on a single word strongly depends upon lexical factors, many of which are irrelevant for models of human sentence processing at the level of investigation in this work. *Word length* (number of characters) and *word frequency*, in particular, are known to strongly influence reading time, and have therefore to be controlled in the material. The best way to do so is to only use words of about the same length and frequency in the critical regions. Since it is very often not feasible to do so, lexical factors have to be eliminated statistically in the analysis of the data as well as possible. Researchers have proposed a number of procedures to account for lexical factors, some of which became very predominant in recent parsing literature.

3.5.1 Word length

It is known from eye-movement studies that words are not read character by character. Nevertheless, the duration of reading a single word is a linear function of the number of characters of the word (Just and Carpenter, 1987), each character consuming about 30ms *additional* time to read³⁹.

For reasons not obvious to the author, the most widely accepted length adjustment procedure is *reading time per character*. There are a number of severe problems with this procedure, as pointed out by Trueswell, Tanenhaus, and Garnsey (1994). To summarize, by dividing reading time scores by the number of characters it is implicitly presumed that all processes that contribute to the reading time score are influenced by the word length. That is, not only word recognition, but also other lexical processes, such as retrieving syntactic and semantic information associated with a word, are determined by its length. In other words, if one would exclude the time needed to recognize a word, which is presumably influenced by its length, all further processes performed in the lexicon would also take longer, the longer the word actually is. However, we know from regression analysis results (Just and Carpenter 1987, Mitchell 1984), which indicate that the function of word length on reading time has not

^{38.}Note that the first pass reading time measure suggested by Liversedge (1994) is a special case of load contribution, namely the case in which both regions are identical (i.e. $lc(reg_j, reg_j)$). Note also that what could reasonably be called *second pass reading time* (although regression contingent) can easily be computed from the *load contribution* of the critical region minus its *first pass reading time*.

^{39.}In our own studies we got about 20ms per character for our German subjects, most of which were students (undergraduates).

only a constant slope, but also a certain constant intercept of about 150-200 ms, that there is a constant amount of processing load *not* influenced by word length.

To clarify the problem, let us presume a linear function f of word length l, yielding the estimated reading time score. Let i be a constant intercept, and s a constant slope. The estimation function, then, f(l) is given in (91) below:

$$(91) f(l) = ls + i$$

A sensible length adjustment procedure applied to the estimated reading time score should yield a value independent of any word length given. If we adjust the predicted score by dividing by the number of characters, however, we yield (92):

(92)
$$f_2(l) = \frac{f(l)}{l} = \frac{ls+i}{l} = s + \frac{i}{l}$$

The adjustment procedure f_2 only yields a constant adjusted value, if *i* (*intercept*) equals 0, since the deviation amounts to *i*/*l*. Therefore, the greater the intercept or the shorter the word, the greater becomes the distortion. Note also that the deviation function $f_d(l) = i/l$ is a non-linear (hyperbolic) function of word length. Therefore, the same difference in word length causes a much larger difference in deviation for shorter words than for longer words. This is particularly important in the light of the fact that it is often argued that the actual length adjustment procedure is only marginally important when it can be shown that the mean word lengths do not substantially differ between the conditions. Given the distortion function above, however, not only the average word length is important, but also its variance and even the distribution of word lengths within each condition⁴⁰.

One could be tempted to add the intercept to adjustment function *f2*. The obvious way to do this is to subtract the intercept from the reading time score and only then divide by the number of characters. As we can see in (93), the adjustment procedure amounts to a hypothetical value (s) independent of word length.

(93)
$$f2b(l) = \frac{f(l) - i}{l} = \frac{ls + i - i}{l} = \frac{ls}{l} = s$$

^{40.} In order to illustrate this issue, let us consider a short example. Let us assume that in each of two conditions there are three measures of three words. In one condition, the words are 2, 3, and 7 characters long, in the other condition 1, 5, and 6 characters. The average word length, then, is 4 letters with a variance of 14 in both conditions. The average deviation from the hypothetical base value is ((1/2 + 1/3 + 1/7) / 3) * i ms/l for the first condition, and ((1/1 + 1/5 + 1/6) / 3) * i ms/l for the second condition, which amounts to about 0.325 *i ms/l* in the first, and 0.45 *i* ms/l for the second condition. If we substitute a realistic value of 180 ms for *i*, we obtain about 58.6 ms/l deviation in the first, and 82 ms/l in the second condition, resulting in a predicted difference of about 23.4 ms/l between the conditions. In other words, although average word length and standard deviation in word lengths do not differ between conditions, we expect the average reading time per character to be about 23 ms smaller in the first condition only due to the difference in the distribution of word lengths.

Although this looks quite perfect at first glance, we should give it some further consideration. Note that the transformation only yields a linear (constant) result, if the processes to be investigated contribute to the slope but not to the intercept. In other words, it is presumed that the intercept does not differ between the experimental conditions, whereas the slope does. This amounts to saying that the effects we are interested in are related to the word length; i.e. that the processes under investigation will take the longer, the longer the word or region in which they take place actually is. This assumption, however, appears rather unreasonable, since we are looking at (reanalysis) processes presumably associated with a level of processing higher than lexical access; processes which presumably do not co-vary with shallow lexical factors, such as word-length.

Consequently, the experimental variation can be assumed to be reflected rather in differences in the intercept than in the slope. If so, the intercept has two components, a lexical one, which is constant over the conditions (i_c), and another one (i_v), which reflects condition-specific processing load and which therefore varies between the conditions. Since we are not interested in eliminating the design specific variance in the intercept, we will only subtract the subject *average* intercept from the score before dividing it. Therefore, there will always be a residual intercept left in each condition, which will then be divided by the number of characters as before, again resulting in a non-linear distortion function.

(94)
$$f_{3}(l) = \frac{f(l) - \bar{i}}{l} = \frac{ls + i_{c} + i_{v} - \bar{i}}{l} = s + \frac{i_{c} + i_{v} - \bar{i}}{l}$$

Although the distortion will be considerably smaller here, the adjustment is quite imperfect, again due to the division by character transformation.

As a preliminary conclusion, dividing the reading time score by the number of characters, whether or not the intercept is subtracted before, will never completely eliminate the influence of the region length. Instead, the result will be a distortion function with a negative slope and a non-linear component with the potential to produce considerable artifacts (see Trueswell et al., 1994, for an empirical validation).

Two further procedures, fairly similar to each other, have been suggested, which do not suffer from this problem. Both have in common that the influence of the component reflected by the slope is eliminated. Konieczny et al. (1994) suggest computing the product of the number of characters in the region and the slope, and subtract it from the actual reading time score. The rationale is that only the additional amount of time due to lexical processes should be eliminated. The result, the condition specific intercept $(i_c + i_v)$, reflects both the subject specific base rate (i_c) and the condition (and subject) specific amount of time due to processes initiated in that particular experimental condition (i_v) . Since the adjusted score is submitted to a *dependent* analysis of variance, there is no need to account for the subject specific base rate in the dependent variable.

(95)
$$f4(l) = f(l) - ls = ls + i_c + i_v - ls = i_c + i_v$$

·----

The second transformation procedure, suggested by Ferreira and Clifton (1986) for self-paced reading data, and adopted by Trueswell et al. (1994) for eye-tracking data, is very similar, except that the intercept is also included in the formula. In this procedure, the predicted score resulting from the application of the subject specific regression function f(l) is subtracted from the actually measured reading time score. This resulting residual value is then submitted to statistical analysis.

(96)
$$f5(l) = f(l) - (ls + i) = ls + i_c + i_v - (ls + i) = i_c + i_v - i_v$$

Both transformations therefore result in a value independent of the number of characters.

Unfortunately, both transformations suffer from a problem apparently absent in the *per character* transformation. Since the product of slope and number of characters is subtracted from the score, the error produced by a falsely predicted slope increases linearly with each additional character in the region $((s_e - s_p)^{*1})$. The longer the region, the larger the error will be. It is therefore extremely important to achieve the most accurate estimation of the slope.

The accuracy of the slope estimation, however, depends on a number of factors which have mostly been ignored in current literature. First of all, it appears to be important to estimate the slope for each subject separately, since there seems to be a fair amount of variance between subjects (see Trueswell et al., 1994). Nevertheless, some aspects seem to be even more important⁴¹.

Although it might appear reasonable to establish the regression coefficients for each region separately and apply them to the corresponding reading time score, this is probably exactly the wrong way to adjust the data. Note that in the critical region, at least one condition will supposedly produce parsing difficulties reflected in increased reading time data. In these cases, the variance in the data is strongly determined by the across-condition variance. At very best, then, the region length is not, or only very weakly, confounded with the experimental factors. If it is, however, the overall regression coefficients are likely to be strongly distorted in that region⁴².

Consequently, the regression coefficients used in the length adjustment should *never* be computed in the regions that show considerable parsing differences between the experimental conditions. Optimally, the subject-specific coefficients should be estimated in a region which is not likely to produce any processing difficulty in any condition. Only in such regions can one be sure to yield fairly sensible regression coefficients.

⁴¹.Since Trueswell et al. (1994) do not provide reliability analyses for the coefficients within each subject, it is not clear whether or not the between-subject variance is actually due to estimation errors caused by one of the factors described below, most importantly the *between condition* variance.

Another source of distortion is associated with the fact that in quite a lot of cases not all words in a region are actually looked at during first-pass reading. On average, these cases lead to an underestimation of the length effect, if the reading time score is correlated with the length of the entire region. The best way to counter this problem would be to compute the sum of the lengths of the word that were actually fixated during first-pass reading (*"first-pass region length"*). At the moment, however, there is no analysis program I know of (not even my own), that provides varying region lengths depending on which words were fixated or not.

3.5.2 Number of words in a region

Note that each word in a region adds a certain amount of time to the score, which stems solely from lexical access processes reflected in the intercept (in an unambiguous position). When the critical region differs in the number of words between the conditions, it is absolutely necessary that the adjustment procedure also accounts for this difference. The most reasonable way to do so is to subtract the number of words in the region multiplied with the subject-specific intercept from the reading time score. Note that the intercept must have been established in a regression analysis using word-by-word reading time scores (in a region in the sentence where no parsing difficulties can be expected).

Some researchers suggested dividing the reading time score by the number of words in the region (e.g. Kennedy et al., 1989). It should be clear at this point that we would run into a similar kind of problem as with the *by character* account described above. With this transformation it is assumed that the processes we are interested in do show up more strongly, the longer the region is. While garden-path phenomena may indeed be stronger, the longer the distance between the ambiguity and the disambiguating region actually is (see Ferreira & Henderson, 1991; Hemforth et al. 1994), they will presumably not interact with aspects of merely recognizing the words in the region. Again, the most appropriate way to account for the number of words, then, is to subtract a certain amount of time (the intercept) for each word in the region⁴³.

^{42.}In order to illustrate this, imagine two groups of items corresponding to distinct conditions in the experimental design, both of which show reasonable regression coefficients within each condition, say: an intercept of 190 ms and a slope of 20ms. Suppose that *i*.) in one condition a rather strong garden path is detected in the critical region while no garden path is detected in the other condition, and *ii*.) that the average region length in the first condition is *shorter* than the length in the latter. In this particular constellation, there will probably be a *negative* correlation between region length and reading times score, resulting in an absurdly increased intercept, and more importantly in this context, in a bizarre *negative* slope. But even if one can find a positive correlation between region length and reading time score, the contaminated variance makes the regression coefficients extremely likely to be at least weakly distorted. Note further that in a 15-character region, for example, even a slope-estimation error of three milliseconds results in 45 ms error with the potential of producing considerable artifacts.

^{43.}It might be important to consider a number-of-word correction even if the critical region is constant in length (number of words) across the conditions and even over the cases, because some words might not have been fixated during first pass reading. We have as yet no empirical data to support or falsify this claim.

To conclude, the best way to account for a lexical word length effect is clearly to control the lengths of the relevant regions in the material in a way that they do not substantially differ between the conditions⁴⁴.

3.5.3 Open questions

Whereas it seems quite reasonable to eliminate lexical and sub-lexical factors from the first pass reading time score, it is not so clear whether or not such factors also influence reading time scores when words are re-read, e.g. during a regression path. Evidence in favor of or against the inclusion of length effects during re-reading could stem from a correlation analysis of *load contribution* scores of re-read words with their corresponding lengths. Unfortunately, such data are not yet available. As long as no such evidence is available, we will ignore the word lengths of re-read portions of the text, since the lexical access has presumably already taken place before the regression path was entered.

It must also be mentioned that lexical factors other than *word length* contribute to reading time, especially *frequency*, *concreteness*, and many more. Whereas the frequency of words (in large corpora) can be taken from several linguistic data-bases, statistics about the other factors are not equally easy to access. We have therefore suggested (see Konieczny et al. 1994) measuring *lexical decision times* of all relevant words and include these as a covariate in the analysis of variance. Lexical decision time (LDT) data must, however, be established within a separate procedure of its own, which is a rather costly enterprise. Furthermore, it is often quite hard or not possible at all to get both types of data, i. e., an LDT study and the eye-tracking experiment from one and the same subject. Then, mean standardized LDTs (z-standardized for each subject) must be used as a covariate. From an empirical point of view, LDTs are surely advantageous over mere region lengths, but since the second most important lexical factor, *frequency*, is strongly correlated with the word length, it is not clear whether or not the outcome justifies the effort of running a separate lexical decision study of its own.

3.6 Conclusion

Eye movement records have emerged as an extremely useful source of information in the enterprise of parsing research. While other techniques currently seem to be restricted to quite unnatural presentation modes, subjects can read text during eyetracking in much the same way they are used to do. Eye movement records represent information of many kinds, including, among others, fixation durations, the sequence of fixations, and regressions, thus providing a reflection of cognitive processes at a level of detail still unmatched by other techniques. Even though the enthusiasm of the early years about the potential of standard eye-tracking measures has calmed

^{44.}Note, however, that even if the critical region contains identical material in all conditions, there might still be a length distortion if there is variance in length between the items, due to the fact that the distribution of missing values may vary between the conditions (see Konieczny, Hemforth and Völker, 1994).

down a little, it was demonstrated that there is much more information in the records if only the narrowed perspective on single regions is abandoned. Especially the notion of the *regression path* provides the invention of a series of new measures, in particular the several variants of *(first) regression path duration* (RPD), and *load contribution* (LC), both of which will probably tell us some interesting stories about the human sentence processing mechanism in the future.

4 General constraints on human parsers

In this chapter, I will discuss the empirical evidence in detail which constitutes the constraints a model of human parsing must satisfy to be adequate. The role of lexical heads in parsing will be the topic of the first part of this chapter, more general architectural constraints that of the second part. The general constraints and specific parsing principles (highlighted as best suited to the evidence) will serve as the empirical foundation of the parsing mechanism to be specified in chapter 5.

Data from two eye-tracking experiments on PP-attachment (in isolated sentences as well as in neutral contexts) and one on NP-attachment will provide evidence for the relevance of the availability of lexical heads and their respective properties for initial attachment decisions. The evidence will be shown to strongly support the principle of *Parametrized Head Attachment*.

It will be argued, however, that a certain class of lexicalist parsing approaches, namely, *head licensing* models, can also account for the data obtained. Data from an eye-tracking experiment on German subject-object asymmetries will be shown to be incompatible with *Head Licensing* in section 4.1.2.

The immediate use of detailed lexical information as it is assumed in the principle of *Parametrized Head Attachment* has been questioned quite often in psycholinguistic literature. In section 4.1.3 I will discuss empirical results which have been claimed to either support or disconfirm a guiding influence of lexical information in structure assembly.

In the second part of this chapter, general constraints on parsing will be discussed. The predictions of different classes of models will be analyzed which assume alternative sentence structures are built in parallel, serially, or that a fully determinate decision is postponed. The evidence given is supplemented by the findings on NP- attachment ambiguities presented in section 4.1.1. Finally, the question of whether the human parser has to be conceived of as an autonomous input module that cannot be distracted by higher level processes will be addressed. A modular approach will be substantiated by an eye-tracking experiment on PP-attachment in pragmatically biased contexts.

4.1 The role of lexical heads

4.1.1 Preferences in German verb-second and verb-final clauses

4.1.1.1 Introduction

There are two major features of German that make it particularly interesting for the investigation of parsing processes: firstly, many syntactic constraints are expressed in the morphology of lexical items, which allows for a higher degree of variability in constituent ordering. Secondly, and consequently, the variability in ordering heads and their complements permits us to vary the availability of prominent lexical information during the course of on-line sentence processing. In German main clauses, for instance, the verb has to appear as the second constituent in the sentence (97, 98), whereas in subclauses it usually follows its arguments and modifiers (99, 100).

- (97) Susanne verzierte den Kuchen mit dem frischen Obst. Susan decorated the cake with the fresh fruit.
- (98) Gestern gab der Freund der Tochter einen Kuß.Yesterday gave the friend (to) the daughter a kiss."Yesterday the friend gave the daughter a kiss."
- (99) Daß Susanne den Kuchen mit dem frischen Obst verzierte, …
 That Susan the cake with the fresh fruit decorated, …
 "That Susan decorated the cake with the fresh fruit,…"
- (100) Daß der Freund der Tochter einen Kuß gab, …That the friend (to) the daughter a kiss gave, …

"That the friend gave a kiss to the daughter ..."

In this chapter, I will present on-line data from eye-tracking experiments which show that attachment preferences of NPs and PPs in constructions like (97) to (100) are dependent on the availability of lexical heads providing an attachment site as an argument or modifier, respectively⁴⁵.

^{45.}Experiments I and III were conducted in cooperation with Barbara Hemforth, Christoph Scheepers and Gerhard Strube. The experiments have been carried out within the context of the project SOUL (Deutsche Forschungsgemeinschaft, DFG, Str 301 / 4-1,2,3).

In the first experiment, the influence of the availability of lexical information as well as the properties of lexical heads (subcategorization preferences) on PP-attachment preferences in isolated sentences are investigated. In experiment II, I will present data on PP-attachment in *verb-second* and *verb-final* sentences in neutral contexts, thus only varying the availability of the verbal subcategorization information. The evidence on attachment preferences in *verb-final* sentences will be extended to NP-attachment preferences in experiment III.

4.1.1.2 Experiment I

In the first experiment, two sets of verbs were used, which differed with respect to their bias towards taking a PP as a complement: one set of verbs, like *beobachten (to watch)* and *schlagen (to hit)* showed a strong preference for a with-PP, whereas the others, like *bemerken (to notice)* and *erblicken (to catch sight of)*, did not, which had been checked in extensive pre-studies (Strube, Hemforth, & Wrobel, 1990).⁴⁶

Secondly, the PP was varied so that its content forced or strongly biased either an attachment to the VP (101), (103) or to the NP (102), (104). The materials were carefully selected according to pretests⁴⁷ such that the PPs fit equally well to the verbs and to the nouns used in the direct object NPs, respectively.

- (101) Marion beobachtete das Pferd *mit dem neuen Fernglas*. Marion watched the horse *with the new binoculars*.
- (102) Marion beobachtete das Pferd *mit dem weißen Fleck*. Marion watched the horse *with the white patch*.
- (103) Martina erblickte die Schlange *mit dem starken Teleobjektiv.* Martina caught-sight-of the snake *with the strong telephoto lens.*
- (104) Martina erblickte die Schlange *mit dem spitzen Giftzahn*. Martina caught-sight-of the snake *with the sharp poisonous fang*.

Additionally, I took advantage of the variation of the *verb-placement* in German clauses: whereas the verb appears in the clause-final position in certain subclauses, such as (105) to (108), it appears as the second constituent in main clauses, such as (101) to (104). In the subclauses, the ambiguous PPs are thus read in the absence of the verb for which they might be an argument or modifier.

(105) Ich habe gehört, daß Marion das Pferd mit dem neuen Fernglas beobachtete.

I have heard that Marion the horse *with the new binoculars* watched. "I have heard that Marion watched the horse *with the new binoculars.*"

⁴⁶.In the pre-studies, subjects had to complete sentence fragments like "Peter watched the man with ...". The relative frequency of noun versus verb-modification was taken as an indicator for the verb-bias.

^{47.}Subjects had to rate the plausibility of syntactically disambiguated versions of the sentences.

(106) Ich habe gehört, daß Marion das Pferd mit dem weißen Fleck beobachtete,.

I have heard that Marion the horse *with the white patch* watched. "I have heard that Marion watched the horse *with the white patch*."

- (107) Jemand sagte, daß Ute die Schlange *mit dem starken Teleobjektiv* erblickte. Somebody said that Ute the snake *with the strong telephoto lens* caught-sight-of. "Somebody said that Ute caught-sight-of the snake *with the strong telephoto lens*."
- (108) Jemand sagte, daß Ute die Schlange mit dem spitzen Giftzahn erblickte. Someone said that Ute the snake with the sharp poisonous fang caught-sight-of. "Someone said that Ute caught-sight-of the snake with the sharp poisonous fang."

Methods

Subjects. Twenty undergraduate students (native speakers of German) from the University of Freiburg were paid to participate in the study. All of them had normal, uncorrected vision and they were all naive concerning the purpose of the study. During an experimental session of about 40 minutes, each of the subjects had to read 45 isolated sentences and yes-no questions while their eye movements were monitored by a Dual Purkinje Image Eyetracker. The results of two subjects had to be excluded because of too many track losses.

Materials and Design. The experimental sentences were manipulated according to a 2*2*2 within-subjects design with the factors *verb-placement* (final vs. initial), *lexical preference* (3-argument verb vs. 2-argument verb), and *semantic bias* (VP- vs. NP- attachment biased). The order of the sentences was randomized. For each experimental condition, three sentences were presented, resulting in 24 target sentences per subject. The materials were rotated such that every sentence from every set was presented to an equal number of subjects. The exclusion of two subjects led to an imbalance, however. Excluding two more subjects, quasi-randomly chosen to obtain a balanced distribution of the material, did not change any of the effects described in the following sections. One sentence had to be excluded because of too many missing data. Additionally, there were 29 filler sentences.

Procedure. Prior to the experiment, the subject was fitted to a head-rest to prevent head movements during reading. This was followed by a brief calibration procedure and a warming-up block of five successive filler sentences. The experiment was then made up of four blocks. Each block was initiated by a brief calibration procedure and contained ten sentences - two filler sentences followed by eight randomly mixed target sentences or filler sentences, respectively. Before a sentence was presented, the subject had to fixate a cross-marking on the screen which indicated the start-position of the sentence-string. When the subjects had finished reading the sentence, they pressed a button that erased the sentence from the screen. Each sentence of the experiment was then followed by a simple yes/no-question which the subject was to answer by pressing one of two buttons (left-hand button: "yes", right-hand button: "no"). Subjects were told to read normally. They answered with a high degree of accuracy (93%) which did not vary across conditions.

Apparatus. The subjects' eye movements were monitored by a Generation 5.5 Dual Purkinje Image Eyetracker. Viewing was binocular, but eye movements were recorded only from the right eye. The eyetracker was connected to an AT 386 computer which controlled the stimulus-presentation and stored the output from the eyetracker. The sampling rate for data collection was 1 KHz. The sentences were presented on a 20-inch color monitor, beginning at the 6th column of the character matrix. The subject was seated 83 cm from the face of the screen, so that 3 letters equaled 1 degree of visual angle. External distractions and light reflections were screened off by a black tube and the room was slightly darkened.

Dependent Variables and Data Analyses. The eye-movement data were summarized with respect to word positions and processing stages. The data were summarized for each word resulting in four dependent variables, namely first fixation duration, first pass reading times, total reading time, and first as well as total regression path durations (see chapter 3 for a detailed description of the eye-tracking measures). Reading times lower than 100 milliseconds were excluded from data analyses. Consequently, the data reported in the following section represent conditionalized fixation durations (see Rayner et al., 1989).

To control for word-length effects, z-scored *lexical decision times* for each word were used as a covariate in the statistical analyses of this experiment. The lexical decision times were taken from a pre-study conducted with a different group of subjects. In experiment II in this paper, for which I did not have access to lexical decision times, I calculated statistics for length adjusted reading times.

Hypotheses⁴⁸

Parametrized head attachment predicts different PP-attachment preferences, depending on the position of the verb and the lexical properties of the potential attachment sites. The critical constituent where a syntactic attachment preference can first be detected is the noun within the PP itself because at this point the plausibility of the first analysis of the parser can be determined.

Presuming the sentence is read from left to right, the head of the VP, the verb, is still not available on the first reading of the PP in *verb-final* sentences. However, the head of the object-NP will already have been read. Thus, rather than delaying the attachment to the end of the sentence, the PP should initially be attached to the object-NP according to *Head Attachment*. If the attachment is semantically implausible, as in the case of a VP-attachment biased PP, a *reanalysis* is started, during which the alternative VP-modifying reading is generated. This reanalysis should require some additional time. Therefore, the reading times on the noun of a VP-attachment biased PP.

In *verb-2* sentences the lexical heads of both VP and object-NP have already been processed when the PP becomes available. Consequently, *Preferred Role Attachment*

⁴⁸.In the hypotheses, I will only look at models from which clear predictions for the constructions under consideration can be derived.

causes the PP to be attached initially to the VP if the verb bears a lexical expectation of it. If the PP-object is an implausible verb argument, however, longer reading times should be observable which result from the initiation of reanalysis. When the preferred reading of the verb does not expect a prepositional argument, the PP is initially attached to the most recent head, i.e. the object noun. This causes the processor to initiate reanalysis when a VP-attachment biased PP is read. Seen as a whole, PHA predicts an interaction between *lexical preference* and *semantic bias* for verb-2 sentences in that reading times on the noun of the PP should be longer if *semantic bias* contradicts either of the verb-dependent PP-attachment preferences.

Furthermore, an overall interaction of the factors *verb-placement, lexical preference,* and *semantic bias* can be predicted, which results from the differing PP-attachment preferences in the different conditions of *verb-placement*.

Predictions of other models. The *Garden Path Model* (see section 2.6) or its successor *Construal Theory* (see section 2.14) do not predict different attachment preferences for different verb positions or verb frame information. Depending on the assumption as to the structure of German VPs, either the structurally less complex (presumably VP-) attachment is assumed or, if the attachment possibilities do not differ in complexity, late closure predicts attachment to the most recently postulated NP. Within *Construal Theory*, VP-attachment via a primary relation should be preferred in any case.

Predictions from head licensing approaches (e.g., Abney's *Licensing Structure Parser*, Prichett's account, or the *Multiple Constraint* approach put forward by Mac-Donald et al., 1994; see the respective sections in chapter 2), may be comparable to those from *PHA*. Since structure building is restricted to lexical projection, only the object-NP is available as an attachment site for the PP in *verb-final* sentences. Assuming a phrase is preferably attached rather than left dangling, predictions are comparable to PHA. Of course, this would amount to including a principle like "*attach if possible*".

In verb-2 sentences, the particular lexical properties of the verb as well as the object-NP can influence attachment preferences in Ford et al.'s, Stevenson's, Gibson's and MacDonald's models, but not in Frazier's and Frazier and Clifton's. In Abney's and Gorrell's models, lexical properties may play a role in as much as complement versus adjunct attachment is involved. Since the PPs have to be attached to the NPs as modifiers, verb-attachment should be predicted in all cases, irrespective of *lexical preferences* of the verb.

Whether or not the *Tuning* approach (Mitchell et al., 1995, see section 2.15) has any predictions to offer depends on whether or not the *lexical preference*s are considered to be too finely grained information. In the current version, this seems to be the case. Therefore, no *lexical preference* effect is predicted.

In Gorrell's approach (see section 2.13), VP attachment should be easier because of simplicity, irrespective of *verb-placement*. In *verb-second* sentences, a strong garden path should be predicted if the PP has to be reanalyzed as part of the complex NP. If a precedence relation is established in *verb-final* sentences for the object NP and the PP, a conscious garden path should also result in these cases. If no such commitment is

made in the absence of the verb, the structure addition necessary in the case of complex NP readings may increase reading times slightly.

For Stevenson's connectionist model, attachment to a preceding head should be preferred, although the reason why depends on additional assumptions which have to be further specified. If no attachment can take place before a licensing lexical head is read, the predictions should be identical to those of the other head licensing approaches mentioned above. If, on the other hand, attachment is possible, it could be assumed that a lexicalized head can spread more activation to an attachment node than a hypothesized but yet unfilled node. Consequently, attachment of the PP to the preceding object NP should win the competition. Attachment in *verb-final* sentences should depend on the lexical / semantic properties of the verb and the object noun.

Results

All reading time measures were submitted to a full factorial 2*2*2 analysis of variance for repeated measures including the factors *verb-placement*, *lexical preference*, and *semantic bias*. Since *lexical preference* was varied between items, this factor had to be realized as a between-items factor in F2-analyses. For word-by-word analyses, the z-scored lexical decision times per word were included as an additional covariate.

TABLE 1. shows the first fixation duration and first pass reading time for the prepositional object by levels of *verb-placement* and *semantic bias*. A significant interaction between *verb-placement* and *semantic bias* was found for the first fixation duration on the noun of the prepositional phrase. This interaction effect is due to simple effects of *semantic bias* in verb-2 and *verb-final* sentences: in verb-2 sentences, the first fixation on the noun of the PP takes longer if *semantic bias* demands an NP-attached interpretation (F1_{1,16}=8.09; p<.02; F2_{1,21}=6.87; p<.02). Comparing the *semantic bias* conditions in *verb-final* sentences revealed slightly but not reliably higher reading times if the PP was semantically biased for VP-attachment (F1_{1,16}=2.97; p<.11; F2_{1,21}=3.26; p<.09).

	J X X I	5			
	first fixation dura	tion			
	verb-final	verb-2			
VP- biased	271	259	df= 1,16	F1= 10.17	p<.01
NP- biased	239	337	df= 1,22	F2= 13.64	p<.01
	first pass reading	time			
	verb-final	verb-2			
VP- biased	380	463	df= 1,16	F1= 1.73	ns (p =.20)
NP- biased	323	480	df= 1,22	F2 = 2.21	ns (p =.16)

TABLE 1. First fixation duration and first pass reading time (milliseconds) for the noun of the PP
by levels of verb-placement (final vs. second) and semantic bias (VP-biased vs. NP-biased). The
inferential statistics refer to the two-way interaction effects resulting from subject-analyses (F1)
and item-analyses (F2), respectively.

With respect to the *first pass reading times* on the noun of the PP, the interaction of *verb-placement* * *semantic bias* did not show up reliably. The linear contrasts for the two conditions of *verb-placement* resulted in a simple effect of *semantic bias* in *verb-final* sentences: VP-attachment biased PP-nouns showed reliably longer *first pass reading times*

than the NP-attachment biased condition (F1_{1,16}=6.31; P<.03; F2_{1,21}=3.39; p<.08). No reliable difference between the *semantic bias* conditions was found in verb-2 sentences.

TABLE 2. Total reading times and total regression path durations (milliseconds) for the noun of the PP by levels of verb-placement (final vs. second) and semantic bias (VP-biased vs. NP-biased). The inferential statistics refer to the two-way interaction effects resulting from subject-analyses (F1) and item-analyses (F2), respectively.

	total reading times				
	verb-final	verb-2			
VP- biased	655	694	df= 1,16	F1= 3.55	p<.08
NP- biased	528	711	df= 1,22	F2= 2.77	p<.12
	total regression path	durations			
	verb-final	verb-2			
VP- biased	472	1635	df= 1,16	F1= 3.68	p <.09
NP- biased	389	1733	df= 1,22	F2< 1	ns
	first regression path	durations			
	verb-final	verb-2			
VP- biased	457	1353	df = 1, 16	F1 = 2.73	p <.12
NP- biased	385	1424	(df=1, 22	F2 < 1.5	ns)

A marginal interaction could be established for first (FRPDs) and *total regression path durations* (TRPDs). FRPDs and TRPDs were increased in the VP-attachment biased conditions in *verb-final* sentences (FRPDs: $F1_{1,16} = 5.50$, p < 0.04; $F2_{1,21} = 2.69$, p < .13; TRPDs: $F1_{1,16} = 6.03$, p < 0.03; $F2_{1,21} = 5.03$, p < .04), whereas TRPDs for NP- and VP-attachment biased PP-objects in v2-sentences did not differ reliably.

For *total reading times*, the interaction between *semantic bias* and verb position was not reliable (see TABLE 2.). Nevertheless, for *verb-final* sentences, *total reading times* were longer if the *semantic bias* opted for a verb-argument reading (F1_{1,16} = 3.97, p < 0.07; F2_{1,21} = 5.45, p < .04).

A three way interaction *verb-placement* * *lexical preference* * *semantic bias* in FRPDs and TRPDs was found, though only for subject analyses (FRPDs: $F1_{1,16} = 6.32$, p < 0.03; TRPDs: $F1_{1,16} = 5.00$, p < 0.04). This three way interaction is due to a two way interaction *lexical preference* * *semantic bias* that only showed up in verb-2 sentences (see TABLE 3).

TABLE 3. *First* and *total regression path durations* (milliseconds) for the noun of the PP in verb-2 sentences by levels of *lexical preference* (2-place vs. 3-place) and *semantic bias* (VP-biased vs. NP-biased). The inferential statistics refer to the two-way interaction effects resulting from subject-analyses (F1) and item-analyses (F2), respectively.

	first regression path	durations			
	2-place	3-place			
VP- biased	1537	1168	df= 1,16	F1= 6.12	p<.03
NP- biased	1257	1591	(df= 1,22	F2= 2.85	p<.10)
	total regression path	durations			
	2-place	3-place			
VP- biased	1789	1422	df= 1,16	F1= 3.44	p<.09
NP- biased	1497	1819	(df= 1,22	F2 = 1.42	ns (p<.25)

No reliable effects of *semantic bias* or *lexical preference* showed up at the verb for any of the dependent variables, either for *verb-second* or for *verb-final* sentences (see TABLE 4).

TABLE 4. First fixations, first pass reading time, first and total regression path durations and total reading times (milliseconds) on the verb by levels of verb-placement (final vs. initial), lexical preference (3-arguments vs. 2-arguments), and world knowledge (instrumental vs. attributive).

		3-argument verb		2-argume	ent verb
		instrumental	attributive	instrumental	attributive
first fixations	verb-final	287	324	274	296
	verb-2	245	244	236	203
first pass reading	verb-final	537	575	499	527
times	verb-2	313	347	376	305
first regression path	verb-final	1364	1434	1456	1429
durations	verb-2	424	428	476	407
total regression path	verb-final	1772	2045	1996	1864
durations	verb-2	442	445	476	407
total reading times	verb-final	775	824	830	819
	verb-2	570	747	636	568

Discussion

The general pattern of the data confirms the predictions of *PHA*. The most important measures to be considered here are the *first pass reading times* and the *first regression path durations* on the noun of the PP, since the *first fixation duration* reflects only some of the processes which become relevant at that word with respect to the attachment preferences (see section 3.2). Regarding the *first pass reading time* on the noun of the PP, a reliable overall interaction of the factors *verb-placement, lexical preference,* and *semantic bias* was found which had been predicted by different preference principles in different *verb-placement* conditions. For verb-2 sentences, an interaction between *lexical preference* and *semantic bias* was found, as was predicted by *preferred role attachment*: if the verb yielded an argument role for the PP, the *first pass reading time* was significantly longer for an NP-attachment biased PP than for a VP-attachment biased PP, as predicted by *preferred role attachment*. In *verb-final* sentences, the *first pass reading time* on the noun of the PP was shorter in the semantically NP-attachment biased condition than in the VP-attachment biased condition, as was predicted by *head attachment*.

To sum up, the data clearly contradict approaches, such as the Garden-Path Theory, Construal, or Gorrell's account, which predict a VP attachment preference irrespective of *verb-placement*. They are compatible with PHA, but also with the various *head licensing* accounts and Stevenson's competition model.

4.1.1.3 Experiment II

One of the problems with the materials used in the previously described experiment is that the critical region (the noun of the PP) of the sentences differed between the experimental conditions. This is true for *lexical preference* as well as the *semantic bias* conditions. Since eye movements are particularly sensitive to superficial differences between strings, it could be argued that the differences obtained may at least be partially due to these non-minimal contrasts. Therefore, in experiment II the critical region was held constant in all conditions. Different biases were induced by the sentential contexts preceding the critical region (109a-d). *Lexical preferences* were not varied in this experiment. Most of the verbs used were biased to expect a PP.

Another criticism that may come up regarding experiment I is the fact that the sentences were presented in isolation without any context. Reading strategies for isolated sentences might differ from normal reading in context; e.g. reading times are usually slower. Since parsing preferences are very sensitive to the speed of incoming information, it might be the case that different effects show up if more time is spent on a region. To cope with this problem, two neutral context sentences were presented with the experimental material, one preceding and one following the experimental sentence.

(109) Context

Sarah hatte vergeblich versucht, einen wirklich schönen Liebesbrief zu verfassen.

Sarah had tried in vain to write a really beautiful love letter.

a. Sarah / entzündete / das Papier / mit / der Gasflamme,/ bevor sie / den Brief in den Kamin legte, was ihr selbst etwas kitschig vorkam./
Sarah lit the paper with the gas flame before she put the letter in the chimney, which seemed kitschy even to herself.

- b. Daß Sarah / das Papier / mit / der Gasflamme / entzündete,/ bevor sie / den Brief in den Kamin legte, kam ihr selbst etwas kitschig vor./ That Sarah the paper with the gas flame lit before she put the letter in the chimney seem kitschy even to herself. That Sarah lit the paper with the gas flame, ...
 c. Sarah / löschte / die Lampe / mit / der Gasflamme,/ bevor sie / den Brief in den Kamin legte, was ihr selbst etwas kitschig vorkam./ Sarah put out the lamp with the gas flame before she put the letter in the chimney which seemed kitschy even to herself.
- d. Daß Sarah / die Lampe / mit / der Gasflamme / löschte,/ bevor sie / den Brief in den Kamin legte, kam ihr selbst etwas kitschig vor./ That Sarah put out the lamp with the gas flame before she put the letter in the chimney seemed kitschy even to herself.

Sie gab sich gern einer melancholischen Stimmung hin.

She liked to surrender herself to a melancholic mood.

Methods

Materials and Design. The experimental sentences were manipulated according to a 2*2 within-subjects design with the factors *verb-placement (final vs. second)* and *semantic bias* (VP- vs. NP-attachment biased). All sentences were presented in neutral contexts with one sentence preceding and one sentence following the target sentence. The order of the sentences was randomized. For each experimental condition, four texts were presented, resulting in sixteen target sentences per subject. These sixteen sentences were taken from a pool of sixteen sets of sentences differing in content. The materials were rotated such that every sentence from every set was presented to an equal number of subjects. Additionally, there were 32 filler texts.

Subjects. Sixteen undergraduate students (native speakers of German) from the University of Freiburg were paid to participate in the study. All of them had normal, uncorrected vision and they were all unfamiliar with the purpose of the study.

Procedure. The experiment was made up of six blocks plus a warming-up block consisting of two training texts. Each block was initiated by a brief calibration procedure and contained eight texts—one filler text followed by seven randomly mixed target or filler texts. During an experimental session of 45 minutes, each of the subjects had to read 50 texts while their eye movements were monitored by a Dual Purkinje Image Eyetracker.

The fixation duration measures were summarized for each region of a sentence as indicated by the slashes in example (109), and were submitted to a full factorial 2*2 analysis of variance for repeated measures including the factors *semantic bias* (VP-biased vs. NP-biased), and *verb position* (verb-final vs. verb-second). The slope of the regression of word length on *first pass reading times* was calculated on a non-critical region (the NP preceding the structurally ambiguous PP) and the product of slope and word length was subtracted from all reading time measures (see chapter 3 for a discussion). All of the reported effects in the next sections will thus be adjusted with respect to word-length. I will concentrate the description of the data on the *first path reading times* and the *first* and *total regression path durations* at the critical sentence-positions, i.e. the PP-object (det + noun), and the main verb.

Results

A reliable interaction of the factors *verb-placement* and *semantic bias* was established for *first* and *total regression path durations* as well as for *total reading times* at the PPobject. For all three measures, reading times were reliably increased if NP-attachment was semantically biased in *verb-second* sentences (FRPD: F1_{1, 15} = 12,49, p <.01; F2_{1,15} = 10.87; p < 0.01; TRPD: F1_{1, 15} = 8.25, p <.02; F2_{1, 15}=14.48; p < 0.01; TRT: F1_{1, 15} = 11.51, p <.01; F2_{1, 15}= 10.41; p < 0.01). For *verb-final* sentences on the other hand, *total regression path durations* were marginally increased if VP-attachment was biased (TRPD: F1_{1,15} = 4.26, p < 0.06; F2_{1, 15}=2.89; p < 0.11).

	first pass reading times				
	verb-final	verb-2			
VP- biased	192	282	df= 1,16	F1< 1	ns
NP- biased	162	243	df= 1,16	F2< 1	ns
	first regression path	durations			
	verb-final	verb-2			
VP- biased	337	493	df= 1,16	F1= 10.55	p < 0.01
NP- biased	260	757	df= 1,16	F2=10.69	p < 0.01
	total regression path	h durations			
	verb-final	verb-2			
VP- biased	417	504	df= 1,16	F1= 10.00	p < 0.01
NP- biased	271	899	df= 1,16	F2= 15.82	p < 0.01
	total reading times				
	verb-final	verb-2			
VP- biased	572	542	df= 1,16	F1= 1.94	ns (p <.19)
NP- biased	600	827	df= 1,22	F2= 5.78	p < 0.04

TABLE 5. First pass reading times, first and total regression path durations and total reading times (milliseconds) for the PP-object by levels of *verb-placement* (final vs. second) and *semantic bias* (VP-biased vs. NP-biased). The inferential statistics refer to the two-way interaction effects resulting from subject-analyses (F1) and item-analyses (F2), respectively.

At the verb of *verb-final* sentences, *first* and *total regression path durations* as well as *total reading times* were increased, though not reliably so, if NP-attachment was semantically biased.

Subject anal	yses (11) and item		spectrely.	
	first pass reading	times		
	VP-biased	NP-biased		
verb-final	200	185	F1 _{1, 15} < 1, ns	F2 _{1, 15} < 1,ns
verb-second	273	279	F1 _{1, 15} < 1.21, ns	F2 _{1, 15} < 1, ns
	first regression pa	ath durations		
	VP-biased	NP-biased		
verb-final	365	460	F1 _{1,15} =2.93, p<.11	F2 _{1, 15} < 1, ns
verb-second	273	292	F1 _{1, 15} < 2.37, ns (p<.15)	F2 _{1, 15} < 1, ns
	total regression p	oath durations		
	VP-biased	NP-biased		
verb-final	411	588	F1 _{1,15} =3.75, p<.08	F2 _{1, 15} = 1.51, ns (p<.24)
verb-second	293	311	F1 _{1, 15} < 1.03, ns	F2 _{1, 15} < 1, ns
	total reading time	es		
	VP-biased	NP-biased		
verb-final	319	459	F1 _{1, 15} = 4.96, p<.05	F2 _{1, 15} = 3.73, p<.08
verb-second	415	567	F1 _{1, 15} < 1, ns	F2 _{1, 15} < 1, ns

TABLE 6. First pass reading times, first and total regression path durations and total reading times (milliseconds) at the verb in verb-second and verb-final sentences by levels of semantic bias (VP-biased vs. NP-biased). The inferential statistics refer to the semantic bias effect resulting from subject-analyses (F1) and item-analyses (F2), respectively.

Discussion

The data from Experiment II support the evidence on the influence of *verb-placement* found in Experiment I. Even if the materials appear in fairly neutral contexts, the interaction between attachment preferences in *verb-second* and *verb-final* sentences shows up at the PP-object. In this experiment, it cannot be due to superficial differences in the critical regions because these were identical for the bias and *verb-placement* conditions.

The shorter reading times on the clause-final verb in VP-attachment biased conditions can easily be explained since the pre-verbal PP, to be interpreted in these cases as a verbal argument, already restricts the interpretation of the verb more than a preverbal complex NP does. Additionally, only a very cautious interpretation of reading times at the verb is possible anyway, because the verbs were different for the two bias conditions. As can be seen in TABLE 6., even in *verb-second* sentences, verbs in NPattachment biased sentences were read slightly more slowly than verbs in VP-attachment biased sentences, though not reliably so.

4.1.1.4 Experiment III

The validity of a parsing principle like *Parametrized Head Attachment* should not be substantiated by experiments on only a single structural phenomenon like PP-attachment. PHA is supposed to be valid for the whole range of structural ambiguities it can be applied to. In this section, I will focus on processing local NP-attachment ambiguities in German verb-final constructions.⁴⁹ As is demonstrated in (110) and (111), these constructions can induce a local conflict associated with a morphological case marking ambiguity:

(110) Daß der Arzt der Sängerin ein Medikament gegeben hat, wußte niemand.

That the doctor [the singer_[fem, {gen/dat}]] a remedy given has, knew nobody.

Nobody knew that the doctor has given a remedy to the singer.

(111) Daß der Arzt der Sängerin ein Medikament entdeckt hat, wußte niemand.

That the doctor [the singer $_{\rm [fem, \{gen/dat\}]}$ a remedy discovered has, knew nobody.

Nobody knew that the doctor of the singer has discovered a remedy.

Locally, the NP *der Sängerin* can either be interpreted as an indirect object (i.e. a dative-marked argument) of the verb (as in 110), or as a genitival modifier of the subject NP *der Arzt* (as in 111), since it is morphologically ambiguous with respect to case marking. The structural ambiguity is resolved when the verbal participle at the end of the subordinate clause is encountered: *geben* (to give) in (110) requires a dative complement, and thus demands the critical NP to be part of the VP (112).

(112) [_{S'} Daß

[_S [_{NP} der Arzt]

[_{VP} [_{NP} der Sängerin] [_{NP} ein Medikament] [_V gegeben] [_{aux} hat]]] ...]

Entdecken (to discover), on the other hand, is a simple transitive verb. Therefore, the final structure of (111) is (113) in which the critical NP is part of a complex subject-NP.

(113) [_{S'} Daß

[_S [_{NP} [_{NP} der Arzt] [_{NP} der Sängerin]]

[VP [NP ein Medikament] [V entdeckt] [aux hat]]]...]

The same structures apply for sentences like (114) and (115), where the critical NP is unambiguously case-marked as *dative* (114), or *genitive* (115), respectively.

(114) Daß der Arzt dem Sänger ein Medikament gegeben hat, wußte niemand.

That the doctor [the singer_[masc, dat]] a remedy given has, knew nobody.

Nobody knew that the doctor has given a remedy to the singer.

⁴⁹.Structures like these were first investigated by Bader (1990). See also Scheepers, Hemforth, & Konieczny, 1994).

(115) Daß der Arzt des Sängers ein Medikament entdeckt hat, wußte niemand.

That the doctor [the singer_[masc, gen]] a remedy discovered has, knew nobody.

Nobody knew that the doctor of the singer has discovered a remedy.

Predictions

In this section, I will discuss predictions on the processing of NP-attachment ambiguities like (110, 111, 114, 115) which can be derived from the models discussed in chapter 2. As will be shown, the models differ with respect to the positions in the sentences where processing difficulties are to be expected, as well as in the prediction of the particular conditions under which these difficulties should appear. At first, I will sketch the different predictions regarding the influence of morphological case marking, and then I will turn to the assumptions concerning the ambiguity-resolution at the VP-participle.

The morphological case marking of the critical NP

Predictions of the PHA-Model. According to the *PHA-Model,* the parser initially tries to attach the structurally ambiguous NP to the already processed nominative NP. This is a consequence of the *head attachment* principle: since the head of the subject NP, but not the head of the VP, is already available when the critical NP is read, the integration of this critical NP into a complex subject-NP is preferred. If the case marking of the second NP is definitely incompatible with the preferred NP-modifying reading, reanalysis should be induced. This is the case if the second NP is unambiguously dative marked, since dative-marked NP-modifiers are not permitted in German. The earliest point in the sentence where such a conflict can appear is at the determiner of the critical NP. Thus, when an unambiguous dative-case determiner is read, there should be an increased processing load due to reanalysis processes beginning.

Predictions of the Garden Path Model/Construal Theory. When the parser processes the determiner of the subject-NP, it can either directly attach the subject-NP to the Snode, or it can generate additional NP-nodes for a more complex NP (cf. 112 and 113). According to Minimal Attachment, the parser should avoid the latter in the first analysis, because potentially unnecessary nodes would have to be postulated. Hence, the parser initially generates a simple subject-NP. When the critical second NP der Sängerin is perceived, it can easily be integrated into a VP (112), whereas interpreting it as a modifier of a complex subject-NP would require reanalysis. Consequently, the structurally ambiguous NP der Sängerin will be preferably attached to the VP as a complement. These predictions should hold true not only in the Garden-Path Model, but also in *Construal Theory*, though maybe for different reasons. If attachment to the preceding NP is considered a primary relation, Minimal Attachment and Late Closure apply. If not, verb attachment should be chosen because attachment via primary relations is preferred to non-primary attachment. However, with regard to the morphological case marking of the second NP, no early effects can be predicted within the Garden-Path Model/Construal Theory-framework. There are a few verbs in German e.g. gedenken (roughly to remember)-which require a genitival complement, as in $(116)^{50}$.

(116) Als der Arzt des Sängers gedachte, ...

When the doctor [the singer_[masc, gen]] remembered, ...

When the doctor remembered the singer, ...

Thus, even if the parser accounts for the case marking of the critical NP as soon as possible (i.e. when the determiner of the critical NP is read), an unambiguous genitive case marking cannot explicitly hinder VP-attachment. Consequently, effects due to the case marking of the second NP are very unlikely to appear, unless a disambiguating verb becomes available⁵¹.

Predictions of competition models. In contrast to the PP-attachment experiments presented in experiments I and II, predictions from PHA and competition models with incremental attachment (non-head-licensing models, see the discussion of predictions from Stevenson's model in section 4.1.1.2) differ for NP-attachment ambiguities. Since in competition models structural alternatives are considered in parallel, an ambiguity effect should show up at the ambiguous feminine NP *der Sängerin*; i.e. it should take longer to process ambiguous NPs in comparison to the unambiguous masculine NPs (*des Sängers* or *dem Sänger*, respectively).

Predictions of head licensing approaches. If attachments can only take place when they are licensed by a lexical head, only attachment to the preceding NP is possible. Then the predictions are identical to PHA.

Subcategorization requirements of the VP-participle

Predictions of the PHA-Model. If the sentence is grammatical, and if it contains no genitive-complement verb, no conflict will occur at the VP-participle when the preceding critical NP has already been disambiguated by its morphological case marking. Therefore, processing delay resulting from reanalysis at the participle is only to be expected in sentences with an ambiguous case marking of the critical NP. The structurally ambiguous NP is initially attached to the subject NP, according to *head attachment.* Thus, there should be no conflict with a simple transitive participle which requires only a direct object in addition to the subject. However, when a ditransitive (or preferably ditransitive) VP-participle is read, the preferred analysis has to be revised because the indirect object position remains unsaturated. Hence, a time consuming reanalysis should be detectable at a ditransitive participle.

^{50.}The fact that structures associated with such verbs are very unusual is not relevant for the current considerations, since the strictly principle-based approach of the Garden-Path Model/Construal Theory does not include any effects of the relative frequency of a given lexical form. I have to thank Christoph Scheepers for this important hint.

⁵¹.Note that even if genitival verb-arguments did not exist in German, it is not clear how the case could hinder VP-attachment in the absence of the verb. Case is certainly information that is considered to be assigned by the head at a later stage. Though case marking has not been explicitly dealt with in the framework of Garden Path Theory, it would be extremely surprising if they considered case information to be compiled into the phrase structure rule system, while subcategorization information is excluded.

Predictions of the Garden-Path Model/Construal Theory. The VP-participle determines the final decision of the parser. If the subcategorization requirements of the VP-participle are compatible with the initial parse, no reanalysis is required. Since *minimal attachment* forces the parser to postulate an indirect object in the dative or genitive case, respectively, only a verb which provides such an argument position will be easily processed. Otherwise, there should be measurable processing cost due to reanalysis. In any case, a simple transitive verb (or the region immediately following it) should therefore take longer to read.

Predictions of competition models. If the structural alternatives for the ambiguous NP have been computed in parallel and if attachment has been resolved by a competition process, reading times on the participle should be increased in any of the conditions with an ambiguous second NP because reanalysis will always be necessary in a proportion of the sentences read (see Mitchell, 1994; Frazier, submitted).

Predictions of head licensing models. If the NP-attachment to the subject-NP was the only one possible according to the *head licensing* approach, verbs that force VP-attachment should be more difficult to process, since the NP previously attached to the subject-NP has to be reanalyzed as a verb argument. So far the prediction of head licensing models are fairly similar to PHA. In the cases where NP-attachment to the subject-NP was ruled out by the dative case marking of the second NP (*dem ...*), however, integrating the additional dangling NP may slightly increase reading times.

Predictions of Gorrell's model. In sentences with an ambiguous second NP, no complex NP should be assumed due to *Simplicity.* As long as no commitment is made as to precedence relations between the subject NP and the PP in the absence of the verb, the structure addition that proves necessary in case of strictly transitive verbs may lead to slightly increased reading times. If, however, a commitment is made (such that NP1 precedes NP2), a conscious garden path should result at the transitive verb.

Method

Subjects. Twenty-four undergraduate students (native speakers of German) from the University of Freiburg were paid to participate in the study. All of them had normal, uncorrected vision and they were all unfamiliar with the purpose of the study. During an experimental session of 45 minutes, each of the subjects had to read 64 isolated sentences while their eye movements were monitored by a Dual Purkinje Image Eyetracker.

Materials. The experimental sentences in this study were manipulated according to a 2*2 within-subjects design with the factors *verb requirements* (dative complement *required vs. ruled out*), and *case ambiguity of NP2* (ambiguous vs. unambiguous). Unambiguously case marked NPs always fit the subcategorization requirements of the verb, i.e. no ungrammatical case assignments were presented. For each experimental condition, six sentences were presented, resulting in 24 sentences per subject. Each subject was presented with a different set of materials. Additionally, there were 40 filler sentences, structurally different from the target sentences. The order of the sentences was randomized. The target sentences were presented in two lines on the

screen, separated by two empty lines after the comma. Consequently, the entire subclause containing both critical regions was presented in one line.

Procedure. The experiment was made up of seven blocks plus a warming-up block consisting of five successive filler sentences. Each block was initiated by a brief calibration procedure and contained 9 sentences—one filler sentence followed by eight randomly mixed target sentences or filler sentences, respectively. An exception was the last experimental block which was made up of one filler sentence and four test/ filler sentences.

The fixation duration measures were summarized for each word of a sentence⁵², and were submitted to a full factorial 2*2 analysis of variance for repeated measures including the factors *verb requirements* (dative complement *required* vs. *ruled out*) and *case ambiguity of NP2* (ambiguous vs. unambiguous). To account for word-length and other lexical effects, the z-scored lexical decision times for each word were included as covariates in the statistical analyses. I will concentrate on the *first pass reading times* and the *regression path durations* at the critical sentence-positions, i.e. the determiner of the second NP, and the VP-participle, because—as I have noted earlier—these measures are assumed to be the most sensitive ones in detecting the points of the first analysis where a syntactic reanalysis is induced. However, regression path durations have to be handled with some care in clause final positions because they may reflect sentence wrap up or rechecking processes in these positions.

Results

Determiner of NP_{2} . At the determiner of the second NP, neither first pass reading times, nor regression path durations, nor even first fixation durations showed reliable differences across the experimental conditions (all Fs < 1.6). However, since determiners are always difficult positions to look at, it seemed reasonable to inspect the data more carefully.

It is known from several eye-tracking studies that particularly short words are frequently skipped during reading (e.g. Rayner & Pollatsek, 1989; Just & Carpenter, 1987). This also became apparent in the experiment presented here: the critical determiner, which has a string length of only three characters, was skipped in nearly 55% of the cases. This resulted in too few fixations for a valid data analysis. However, the fact that short words are often not fixated does not imply that they are not processed at all. Sometimes the succeeding word falls into the *identification span* of eight character spaces to the right of the fixation (see Rayner & Pollatsek, 1989). Rayner et al. (1982) provide evidence suggesting that the main portion of information is extracted from the word fixated plus two or three characters of the next (parafoveal) word. Since the determiner is only three characters long, it is very likely to be pre-processed

^{52.}Fixation durations at a given word greater than the 97th percentile were treated as extreme values, and thus were excluded from data analysis. This was done for each experimental condition, in order to keep the design balanced. Reading times lower than 100 milliseconds were eliminated from the analyses such that the data express *conditionalized fixation durations*.

on the preceding word (see section 3.2), namely, the noun of the subject NP. Therefore, it is very plausible to assume that in cases where the determiner of the second NP is skipped, it has been processed (at least partially) during fixations on the preceding subject noun.

To cope with this, additional data analyses with a *summed first pass reading time* (SFPRT) measure for the region containing the subject noun and the determiner of the second NP were carried out. This means that the *first pass reading times* (including zero *first pass reading times*) of the particular words of that region were simply summed up.⁵³ All cases in which the *summed first pass reading time* measure added up to zero were treated as missing values in the following analyses.

Note that the verb has not yet been processed in that position. The factor *verb*requirements simply reflects *case-marking* for unambiguously case-marked NPs. As TABLE 7 illustrates, a marginal two-way interaction of the factors *verb* requirements and *case ambiguity of NP2* was found at the region containing the subject noun and the second determiner. This interaction is reducible to a marginal difference between the two *verb* requirements conditions in the case of *unambiguous case marking of NP2*: if case marking demands an attachment to the VP (*dative* case), the SFPRTs tend to be higher than in the NP-attachment condition (*genitive* case). Not surprisingly, the effect of the *verb* requirements is far from being significant (F1, F2 < 0.5) in the *ambiguous case marking* condition. Testing the relative influence of the *case ambiguity of NP2* in each of the *verb* requirements conditions revealed no reliable effect (all Fs < 1.5).

TABLE 7. SFPRTs (ms) for the region containing the subject noun and the second determiner (average region length = 10 characters in each condition). The inference statistics (F1 for subject analyses, [F2] for item analyses) refer to the respective effect of verb requirements in each condition of the factor case ambiguity of NP_2 (ambiguous vs. unambiguous). The statistics of the overall two-way interaction are shown in the last row.

dative verb-complement					
required ruled out $F1_{1,22}$ [$F2_{1,23}$]					
ambig. case of NP ₂	376	370	n.s.		
unambig. case of NP_2	401	363	F= 3.47 [<i>3.90</i>]; p<.08 [=.06]		
Interaction "verb requirements BY case ambiguity of NP ₂ ": F= 3.24 [3.09]; p<.09 [=.09]					

VP-participle. TABLE 8. shows the average *first pass reading times* on the VP-participle in the respective conditions of *verb requirements* and *case ambiguity of NP2.* As can be seen, a significant two-way interaction of the factors could be established. If the morphological case marking of the second NP is ambiguous, longer *first pass reading times* result for a participle which requires a dative-case complement (VP-attachment) than for a simple transitive participle (NP-attachment). If the second NP is unambiguously case marked, the *first pass reading times* on the participle do not differ across the *verb requirements* conditions. Additional analyses revealed a reliable difference between the two *case ambiguity* conditions if the participle forces VP-attachment: the *first pass reading con-*

^{53.}Consequently, the *first pass reading time* at the subject noun were taken as the dependent variable whenever the determiner of NP_2 was skipped.

dition (F1_{1,22}=7.07; p<.02; $F2_{1,23}=2.06$; p=.16). If the verb forces NP-attachment, however, the case ambiguity of the second NP shows no effect on the *first pass reading times* at the participle (*F*1, *F*2 < 1).

TABLE 8. First pass reading times (ms) on the VP-participle. The inference statistics (F1 for subjects analyses, [F2] for item-analyses) refer to the respective effect of verb requirements in each condition of the factor case ambiguity of NP_2 (ambiguous vs. unambiguous). The statistics of the overall two-way interaction are shown in the last row.

	dative verb-				
	required	ruled out	F1 _{1,22} [F2 _{1,22}]		
ambig. case of NP ₂	472	380	F= 5.29 [4.48]; p<.04 [.05]		
unambig. case of NP_2	383	382	n.s.		
Interaction "verb requirements BY case ambiguity of NP_2 ": F= 5.81 [3.98]; p<.03 [.06]					

The mean regression path durations (RPDs) for the VP-participle are listed in TABLE 9. Descriptively, the overall pattern of RPDs is comparable to the *first pass reading times* pattern (TABLE 8.) since the highest are observable for a dative-complement participle (VP-attachment) in the case of an ambiguous case marking of the NP₂. However, the two-way interaction of the factors *verb requirements* and *case ambiguity of NP2* did not reach significance (F1_{1,22}=1.27; p=.26; *F2*_{1,22}=1.75; p=.20). Instead of that, case ambiguity of the second NP showed a reliable main effect (but only in the subject analysis): the RPDs on the participle are prolonged in the ambiguous case marking condition (656 ms) compared to the unambiguous case marking condition (606 ms; F1_{1,22}=4.49; p<.05; *F2*_{1,23}=1.19; p=.29).

	dative verb-complement		
	required	ruled out	
ambiguous case of NP_2	669	643	
unambig. case of NP_2	609	603	

TABLE 9. Total regression path durations (ms) for the VP-participle.

Discussion

The experiment revealed evidence for the construction of a complex subject-NP during the first structural analysis of the sentences considered: the summed *first pass reading times* of the region containing the first noun and the (presumably parafoveally processed) second determiner tended to be longer when an unambiguous dative case-marking of the second NP hindered its attachment to the preceding subject-NP. Thus, there is support for an immediate influence of morphological case marking in the direction predicted by *Parametrized Head Attachment*.

A significant effect was observed at the VP-participle: here, the *first pass reading times* were prolonged when a ditransitive subcategorization-frame required the critical NP to be attached to the VP, given that no early disambiguation had already taken place due to an unambiguous case marking of the critical NP. This result is consistent with *Head Attachment* as defined in PHA. *Head licensing* approaches such as Abney's

Licensing Structure Parser (see chapter 2, section 2.8) would also predict the NPattachment preference correctly. However, in the case where the unambiguous determiner *dem* has marked the critical NP dative, the NP must be left dangling. In this case, one additional argument has to be integrated into the verb-frame when the verb is read, which might bring about additional processing load. However, the processing load for unambiguous dative NPs measured on the verb was not longer than for unambiguous genitival NPs.

The reported data on NP-attachment ambiguities in German *verb-final* sentences provide clear evidence against *Minimal Attachment*, since NP, not VP-attachment, was preferred initially. The data do not seem to be compatible with Gorrell's account either: no increase in processing load was observed when the complex NP reading was forced, not to mention the complete lack of a conscious garden path.

Competition-based accounts also do not correctly predict the data. In the region including the first noun and the determiner of the second NP, summed *first pass reading times* for the ambiguous determiner *der* are not longer than *first pass reading times* on the "easier" unambiguous condition *des*. Comparably, *first pass reading times* on the verb are not longer in the condition with ambiguous second NPs than in the "easier" transitive condition. An ambiguity effect only shows up in *total regression path durations*, probably reflecting somewhat increased re-checking or wrap-up processes for cases in which the second NP was ambiguous (note that the participle is very close to the end of the clause).

4.1.1.5 General Discussion

The data presented so far give a fairly consistent picture of attachment processes in verb-second and verb-final clauses. Phrases which can either be attached to a preceding noun phrase or to the verb phrase tend to prefer attachment to the preceding noun, if the verb is not yet lexicalized. If the verb has already been read, attachment preferences depend on the detailed subcategorization preferences of the respective heads. This is fully compatible with the predictions of *Parametrized Head Attachment*. Taken together, the data strongly confirm PHA, whereas they pose several problems to any other type of model.

With respect to the *Garden Path Model*, all experiments presented in this chapter show evidence against *Minimal Attachment*. At first glance, therefore, it seems tempting to assume that *Minimal Attachment* does not apply to the structures under consideration whereby *Late Closure* would then become responsible for the obtained preferences. In fact, the known data about processing German *verb-final* constructions, like those considered here and in Konieczny et al. (1994), can be correctly predicted by *LC*. Note, however, that the predictions of *Late Closure* alone would be incompatible with the phenomena observed in *verb-second* sentences, like (97) or (98).

So far, the data can also be generally interpreted as confirming the attachment preference predictions of *head-licensing* models. The lack of an *integration* effect for an additional unattached complement will certainly not suffice to exclude such models as inappropriate. In the next section, I will therefore deal in more detail with *head*- *licensing* approaches, whose attachment predictions will be discussed in the light of an eye-tracking experiment on German subject-object asymmetries.

In chapter 5, PHA will be integrated into a more fully specified model of sentence processing. The principle as it has been formulated so far is however underspecified with respect to several questions concerning the architecture of a model of sentence processing: although I always discussed it as part of a serial approach, the way it has been described so far is fully compatible with a locally parallel (competition) approach where one of the structural alternatives considered in parallel is chosen on the basis of the availability and internal structure of the lexical heads. I will discuss the evidence on architectural constraints concerning serial, parallel, and underspecified approaches in more detail in section 4.2.1.

The *lexical preference* effects in verb-2 sentences in Experiment I were discussed as evidence supporting the principle of *Preferred Role Attachment* that models a guiding effect of lexical subcategorization information. Unfortunately, though, the data are not fully decisive for initial analysis because the processing load was measured a few words after the verb. In section 4.1.3 the evidence will be weighed up more thoroughly in the context of a variety of data discussed in the literature to arrive at a more convincing conclusion.

4.1.2 Strictly linear parsing

4.1.2.1 Incremental attachment in verb-final clauses

Languages permitting head final constructions, like Dutch, German, or Japanese, are particularly suited to dissociate predictions of strictly linear parsing (i.e. fully incremental attachment of each item to the phrase marker of the sentence as soon as it is read) from head licensing. Since the head which licenses attachment is only read after its arguments, any evidence about constituents being attached to the phrase marker of the sentence counts against head licensing.

In Dutch (Frazier, 1987b) and German (Bader, 1990; Mecklinger et al., 1995; Hemforth et al., in prep.), subject (117a) and object relative clauses (117b) were investigated. Both the relative pronoun as well as the NP within the relative clause were ambiguous with respect to their grammatical function (subject vs. direct object). Only the verb at the end of the relative clause provided the relevant disambiguation by its number feature that only agreed with either the relative pronoun or the full NP in the relative clause rendering the agreeing phrase the subject of the sentence. Using online tasks like self-paced reading and ERP-responses, Mecklinger et al. (1995) found evidence for an increased processing load at the verb in the relative clause if it forced an object relative reading (117b). This is interpreted as evidence for an early commitment to the subject reading of the relative pronoun.
(117) a. Das sind die Professorinnen, die die Studentin gesucht haben.

These are the professors_[fem, plur] who the student_[fem, sing] sought have. These are the professors who have sought the students.

b. Das ist die Professorin, die die Studentinnen gesucht haben.
 This is the professor_[fem, sing] who the students_[fem, plur] sought have.
 These are the professors who the students have sought .

Highly comparable results from self-paced reading studies are reported for equivalent constructions in Dutch (Frazier, 1987b).

Another line of evidence comes from German complement clauses. Bader and Lasser (1994) investigated constructions like (118a,b):

(118) a. ..., daß [sie_{nom} ... [zu fragen] erlaubt hat].

..., that she ... to ask permitted has.

"that she has given permission to ask ..."

b. ..., daß [[sie_{acc} ... zu fragen] erlaubt worden ist].

..., that her ... to ask permitted been is.

"that permission has been given to ask her ..."

Note that there are two licensing verbs to which the ambiguous pronoun *sie* can be attached. The "matrix"-verb of the subclause [... erlaubt ...] succeeds the infinitival form of the verb fragen in the embedded complement. If - as it can be assumed - subjects read from left to right, according to head licensing parsing, (118b) should be easier to process since *fragen* precedes *erlaubt* as a possible licenser of *sie*. When the attachment of *sie* is delayed until a licenser is read, the first possible and therefore preferred attachment should be as in (118b). Unfortunately, (118b) is one of the rare real German garden-path constructions, whereas (118a) is not, which totally contradicts the predictions of head licensing parsing. If, however, we assume the pronoun to be immediately interpreted as the subject of a coming active-voiced sentence, the attachment to fragen is not even considered during processing, whereas erlaubt hat succeeds in fitting this prediction perfectly. On the other hand, when ... worden ist marks the sentence as passive-voiced, the pronoun must have been misinterpreted, leading readers up the garden path. Thus, to sum up the results of these investigations, there is strong evidence that complements and (potential) subjects are in fact interpreted as such even before licensing information from the verb is available.

More evidence comes from Japanese, where Inoue (1991) reports "surprise effects" constructions like (119) on the transitive verb *eat*, where it becomes obvious that the preceding NPs *Bob*, *Mary*, and *apple* cannot be part of the same phrase marker, to which, as Inoue argues, they are initially attached.

(119) Bob ga Mary ni $[t_{nom/i}$ ringo wo tabeta] inu_i wo ageta.

 $Bob_{nom} Mary_{dat} [t_{nom/i} apple_{acc} eat_{past}] dog_{acc/i} give_{past}.$ Bob gave Mary the dog which ate the apple. (Inoue, 1991, after Sturt and Crocker, 1995)

The experimental data presented so far can be counted as evidence against head licensing parsing but in favor of strictly incremental (linear) parsing. But there are some problems that may yield the data a little less convincing than they appear to be at first sight. Firstly, garden path effects were only measured on clause final verbs. Therefore, it can be the case - as was argued by Pickering and Barry (1991) - that to account for the data no attachment must have taken place before the verb licensing attachment is read. Increased reading times on the verb in some constructions may reflect increased effort in integrating of the preceding constituents.

An ERP-study by Steinhauer and Friederici (1995), however, gives evidence for attachment processes before the verb. In sentences like (120), an increased P600 was found time locked to the unambiguously accusative NP *den Professor*, which is usually taken to reflect syntactic reanalysis processes.

(120) Daß den Professor der Student gesucht hat, …
That the professor (acc) the student (nom) sought has, …
That the student has sought the professor, …
(Steinhauer and Friederici, 1995)

Though this data is hardly compatible with head licensing parsing, it is still open for several alternative interpretations. It is still possible that parsing does not proceed fully incrementally but that smaller constituents like NPs or PPs are parsed in chunks, which are then attached to the phrase marker of the sentence (Perfetti, 1990).

4.1.2.2 Self-paced reading experiments on subject-object asymmetries

Most of the problems in the experiments described so far can be circumvented by looking at subject-object asymmetries in verb-second clauses. It is well known that for verb-second as well as for verb-final sentences, the first NP is preferentially interpreted as the subject of the sentence, though variable ordering of constituents is possible. The question at hand is that of when a violation of this preference is first noticed, leading to increased processing times.

In self-paced reading studies, Hemforth (1993; Hemforth et al., 1993) showed that subject- and object-NPs fronted to the *Vorfeld* position in verb-second clauses such as (6a,b,c) are attached to a sentence structure even before the subsequent verb occurs, since an unambiguous non-nominative case marking of an NP, which is fronted to the 'usually subject' *Vorfeld* position, led to higher processing times when the noun of the first NP was read. If incompatibilities with the subject-first preference showed up at the second NP (121c), reading times were increased as soon as the determiner was read.⁵⁴

- (121) a. Der gute Schauspieler_{nom} bewunderte die kluge Frau_{nom or acc}.
 - "The good actor admired the smart woman."
 - b. Den guten Schauspieler $_{\rm acc}$ bewunderte die kluge ${\rm Frau}_{\rm nom\ or\ acc}.$ "The smart woman admired the good actor."
 - c. Die kluge Frau_{nom or acc} bewunderte der gute Schauspieler_{nom}.
 "The good actor admired the smart woman."

The results of this experiment were taken as evidence for a variant of a left corner parsing strategy (see Johnson-Laird,1983; Earley, 1970; Aho, Hopcroft, & Ullmann, 1974). According to Johnson-Laird (1983), the human sentence processing system "... parses the left-hand corner of each tree (or subtree) from the bottom up and the rest of the tree (or subtree) top down" (p. 298). According to this left-corner algorithm, attachments for sentences like those under investigation in our experiments would proceed as follows:

The determiner is recognized as the left-hand corner of a NP-rule. A N' will be predicted and the adjective, as well as the noun, can easily be attached to the NP-node. No phrase marker of the sentence has been constructed yet. Only when the NP is completed is it recognized as a potential left corner of a sentence. According to the preference for subject-first sentences, an attachment as the subject of the sentence is attempted. If this attachment fails, in the case of an accusative-NP, a time-consuming reanalysis will be necessary. This variant of left-corner paring has also been called *arcstandard left corner parsing* (e.g., Abney & Johnson, 1991).

It was argued that the evidence we found in our self-paced reading experiments may have been misleading due to the nature of the experimental task. Word by word presentation in self-paced reading almost necessarily results in a decrease in reading speed compared to normal reading. It might be the case that slowing down the reading process leads to artificial reading strategies. So, subjects possibly do more work on words or regions preceding the licensing lexical head of a constituent than they would normally do, just because they have more time to do so. Obviously, the results have to be validated with a less interfering technique.

4.1.2.3 An eyetracking experiment

Materials and design

The experimental sentences were manipulated according to a 2*2 within-subjects design with the factors *ambiguity of NP1* (ambiguous vs. unambiguous) and *order of constituents* (subject-verb-object vs. object-verb-subject). The order of the sentences was randomized. For each experimental condition, four sentences were presented,

^{54.}Note, that though the determiner "der" is not unambiguously nominative in German (it could be feminine/singular/ dative or plural/genitive), none of the non-nominative readings is possible with the transitive verb "bewunderte" (admired).

resulting in 16 target sentences per subject. The material was rotated so that every sentence from every set was presented to an equal number of subjects. Additionally, there were 60 filler sentences.

- (122) a. Die hungrige Füchsin bemerkte den fetten Hahn. The hungry vixen_{nom or acc} noticed the fat rooster_{acc}.
 - b. Der hungrige Fuchs bemerkte den fetten Hahn. The hungry fox_{nom} noticed the fat rooster_{acc}.
 - c. Die hungrige Füchsin bemerkte der fette Hahn. The hungry vixen_{nom or acc} noticed the fat rooster_{nom}.
 - d. Den hungrigen Fuchs bemerkte der fette Hahn.

The hungry fox_{acc} noticed the fat rooster_{nom}.

TABLE 10. Experimental design: subject/object asymmetries. The numbers refer to the examples given in the text.

	NP1-amb	NP1-unamb
SO	(122a)	(122b)
OS	(122c)	(122c)

Subjects

24 undergraduate students (native speakers of German) from the University of Freiburg participated in the experiment. All subjects were paid to participate in the study. They had normal, uncorrected vision and they were all naive concerning the purpose of the study. During an experimental session of ca. 45 minutes, each of the subjects had to read ca. 76 sentences while their eye movements were monitored.

Procedure and apparatus

Procedure and apparatus were for the most part identical to that of the experiments described in the previous sections. The phase of warming-up trials consisted of ten sentences. The experiment itself was built up over 6 test sections, each starting with a filler sentence followed by 10 sentences, which were randomly taken from the set of filler sentences or the test sentences. Between the test sections, a brief calibration procedure was run. Each sentence was followed by a simple yes/no-question, which the subject had to answer by pressing one of two buttons (left-hand button: "yes", right-hand button: "no").

Data analyses

Eye-movements were recorded in order to provide an on-line measure of processing complexity due to the predicted preferences. For the statistical analyses, the data were summarized for each word yielding two dependent variables, namely *first pass reading times* and *regression path durations*, RPDs, (see chapter 3).

Hypotheses

Presuming a subject-before-object preference, increased reading times should be found from the *beginning* of an unambiguously accusative NP1 if words are attached to the phrase marker of the sentence as soon as they are read (*left-corner, arc-eager*). Since determiners are usually fixated only very briefly, or even skipped very often, increased reading times are expected on the adjective.

If the attachment of NP1 to the phrase marker of the sentence is delayed until it is completed (*left-corner, arc standard*), higher reading times are not to be expected before the noun.

If attachment is only possible after the lexical head which provides the respective licensing relation has been read (*head-licensing*), no parsing difficulties can be predicted before the verb is read. Whether or not a first NP that is unambiguously accusative leads to increased processing times depends on which integration processes are assumed at the verb.

From the beginning of the second NP, higher reading times should show up for both left-corner variants if an ambiguous first NP is followed by an unambiguous nominative NP which has to be the subject of the sentence. Again, effects are expected on the adjective of the second NP because of the high skipping rate of determiners.

For head-licensing, on the other hand, an increased processing load should only be measurable at a (nominative) head noun.

Results

For the first NP, reading time measures were submitted to an analysis of variance for repeated measures including only one factor, *case of the first NP (nominative, accusative, and ambiguous)*. The two ambiguous conditions were collapsed because they were identical up to the verb.

For the second NP, all reading time measures were submitted to a full factorial 2*2 analysis of variance for repeated measures including the factors *constituent order (SO vs. OS)* and *ambiguity of the first NP (ambiguous vs. unambiguous.* For word-by-word analyses, reading times were adjusted for word length subtracting 20 msec per character.

First NP. First pass reading times and total regression path durations of the determiner, adjective, and noun of the first NP are presented in TABLE 11. At the *determiner* of the first NP, no effects of case marking could be established for *first pass reading times* (or for *regression path durations*, which are identical at the first word). For *adjectives* in accusative NPs, *total regression path durations* (TRPDs) were reliably increased in comparison to adjectives in nominative NPs (F1_{1, 22} = 16.01, p < 0.01, F2_{1, 14} = 16.96, p < 0.01) as well as to adjectives in ambiguous NPs (F1_{1, 21} = 9.55, p < 0.01; F2_{1, 14} = 5.12, p < 0.05). No reliable differences showed up in *first pass reading times* though.

Surprisingly, at the *noun* of the first NP, *first pass reading times* for accusative nouns appear to be shorter than FPRTs for nominative (F1_{1, 21} = 3.87, p < 0.07; F2_{1, 13} < 1, ns) or ambiguous nouns (F1_{1, 21} = 13.65, p < 0.01; F2_{1, 12} = 2.33, p < 0.16). The same pattern showed up in *total regression path durations* (acc vs. nom: F1_{1, 22} = 8.16, p < 0.01; F2_{1, 13} = 4.27, p < 0.06; acc vs. amb: F1_{1, 22} = 3.40, p < 0.08; F2_{1, 13} < 1, ns). This effect may be due to the especially long reading times on the preceding adjective. Since part of the noun lay within the perceptual span of the preceding word, more preprocessing of the noun may have occurred on accusative adjectives. Furthermore, the semantics of the adjective constrains that of the following noun. When more time was spent on the adjective, higher order conceptual information may have had a chance to semantically prime potentially following nouns.

For total reading times, increased processing loads showed up on determiners of accusative NPs in comparison to nominative (F1_{1, 23} = 6.10, p<0.03; F2_{1, 15} = 6.44, p<.03) and ambiguous NPs (F1_{1, 23} = 12.00, p<0.01; F2_{1, 15} = 9.40, p<0.01). On adjectives, total reading times were increased for accusative NPs in comparison to nominative NPs (F1_{1, 22} = 5.81, p<0.03; F2_{1, 14} = 4.96, p<0.05).

TABLE 11. First pass reading times, total regression path durations, and total reading times at the determiner, the adjective and the noun of the first NP by levels of *case marking of NP1* (unambiguously nominative or accusative, or ambiguous). Inferential statistics refer to the main effect.

	first pass reading times						
	acc	nom	amb	subject analysis	item analysis		
det	219	224	220	F1 ₂ , ₄₆ < 1 _, ns	F2 _{2, 30} < 1, ns		
adj	335	352	362	F1 ₂ , ₄₃ < 1, ns	F2 _{2, 29} < 1, ns		
noun	301	326	367	F1 ₂ , ₄₃ = 5.66, p<.01	F2 _{2, 25} = 1.11, ns		
	total regression path durations						
	acc	nom	amb	subject analysis	item analysis		
det	222	225	222	F1 ₂ , ₄₃ < 1, ns	F2 _{2, 30} < 1, ns		
adj	587	390	458	F1 ₂ , ₄₃ = 9.18, p<.001	F2 _{2, 29} = 10.21, p<.001		
noun	348	432	408	F1 _{2, 45} = 5.93, p<.01	F2 _{2, 25} = 2.49, p<.11		
	total reading times						
	acc	nom	amb	subject analysis	item analysis		
det	432	334	307	F1 _{2, 46} = 7.02, p<.01	F2 _{2, 30} = 6.12, p<.01		
adj	1111	879	946	F1 _{2, 45} = 2.80, p<.08	F2 _{2, 29} = 3.74, p<.04		
noun	638	679	720	F1 _{2, 45} < 1, ns	F2 _{2, 25} < 1, ns		

Verb and second NP. No data will be presented for the determiner of the second NP, because it was skipped in 68 percent of the cases.

TABLE 12. First pass reading times, total regression path durations, and total reading times at the verb and at the adjective and noun of the second NP^a by levels of constituent order (SVO vs. OVS) and ambiguity of first NP (unambiguous vs. ambiguous). Inferential statistics refer to the interaction of the factors constituent order and ambiguity of first NP.^b

		First pass reading times				
		OVS	SVO	subject analysis	item analysis	
verb	unamb.	289	293	$F1_{1, 20} = 1.66$, ns	F2 _{1, 15} = 4.22, p<.06	
	ambig.	319	274			
adj	unamb.	262	296	F1 _{1, 21} < 1, ns	F2 _{1,15} < 1, ns	
	ambig.	290	365			
noun	unamb.	374	388	F1 _{1, 18} < 1, ns	F2 _{1,13} < 1, ns	
	ambig.	360	332			
		Total regression path duration				
		OVS	SVO	subject analysis	item analysis	
verb	unamb.	345	316	F1 _{1, 21} < 1, ns	F2 _{1, 15} = 2.83, p<.12	
	ambig.	347	282			
adj	unamb.	630	452	F1 _{1, 21} < 1, ns	F1 _{1, 15} < 1, ns	
	ambig.	697	560			
noun	unamb.	619	544	$F1_{1, 19} = 2.31$, ns	F2 _{1, 13} = 3.05, p<.11	
	amb.	873	588			
		Total reading times				
		OVS	SVO	subject analysis	item analysis	
verb	unamb.	892	695	F1 _{1, 22} < 1, ns	F1 _{1, 15} < 1, ns	
	ambig.	1001	833			
adj	unamb.	763	845	F1 _{1, 21} = 1.97, p<.18	F2 _{1, 15} = 20.94, p<.001	
	ambig.	1033	857			
noun	unamb.	737	765	$F1_{1, 19} = 1.62$, ns	F2 _{1, 13} = 2.51, p<.14	
	ambig.	748	561			

a. Determiner 2 was skipped in 68 percent of the cases, so it could not be analyzed properly.

b. Differing degrees of freedom result from missing values.

Verb. At the verb, a reliable main effect of constituent order was established in first pass reading times (F1_{1, 20} = 1.60, ns; F2_{1, 15} = 7.07, p < 0.02), total regression path durations (F1_{1, 21} = 10.84, p < 0.03; F2_{1, 15} = 16.48, p < 0.01) and total reading times (F1_{1, 22} = 7.11, p < 0.02; F2_{1, 15} = 16.18, p < 0.01). As simple effects show for *first pass reading times* and *total regression path durations* in particular, this main effect is mainly due to increased reading times for those cases where an unambiguously nominative NP2 follows an ambiguous first NP and the verb (which also shows up in the marginal interactions in F2-analyses presented in TABLE 12). In object-verb-subject-sentences with an ambiguous first NP, first pass reading times, total regression path

durations and total reading times are at least marginally increased compared to subject-verb-object sentences with an ambiguous first NP (FPRT: $F1_{1, 20} = 1.60$, ns; $F2_{1, 15}$ = 7.07, p < 0.02; TRPD: F1_{1, 21} = 10.84, p < 0.03; F2_{1, 15} = 16.48, p < 0.01; total reading times: $F1_{1, 22} = 7.11$, p < 0.02; $F2_{1, 15} = 16.18$, p < 0.01). Whereas the increase in total reading times can be due to second pass (i.e. reanalysis) processes, this is not viable for first pass reading times and total regression path durations. Obviously some preprocessing of the determiner of NP₂ has taken place on the verb. Though the determiner "der" alone is not unambiguously nominative (see footnote 54.), no grammatically permissible analysis is available on the transitive verb plus non-accusative determiner if the first NP has already been attached as the subject of the sentence. Reanalysis to the correct OVS-order or repair processes (see Konieczny, Hemforth, Scheepers, & Strube, in press) show up in an increased processing load. More time is also spent for re-reading the verb in OVS-sentences compared to SVOsentences (total reading times - first-pass reading times: 603 msec vs. 402 msec). This difference is reflected in the total reading times in these structures (see TABLE 12, F11, ₂₂ = 7.11, p < 0.02; F2_{1, 15} = 16.18, p < 0.01). A marginal main effect of ambiguity at the first NP was also established for total reading times in subject analyses (F1_{1, 22} = 3.44, p < 0.08; $F2_{1,15} < 1$, ns).

Adjective. First pass reading times at the adjective are shorter in object first sentences than in subject first sentences (OVS: 276 vs. SVO: 331; 5.82, p < 0.03; F2_{1, 14} = 2.54, p < 0.14). This counter-intuitive effect becomes plausible when total regression path durations are taken into account: as expected, TRPDs for object first sentences were longer than for subject first sentences (OVS: 664 vs. SVO: 506; F1_{1, 21} = 5.76, p < 0.03; F2_{1, 14} = 9.91, p < 0.01). Obviously, the increased processing load of object first sentences is detected very quickly and shows up in more, or longer regressions, not in longer first pass reading times at the critical region (see chapter 3 for a discussion on the different measures).

In those cases when an unambiguous nominative or accusative NP succeeds an ambiguous first NP, first pass reading times, regression path durations, and total reading times are increased (FPRT: $F1_{1, 21} = 7.16$, p < 0.02; $F2_{1, 15} = 4.18$, p < 0.06; TRPD: $F1_{1, 21} = 4.11$, p < 0.06; $F2_{1, 15} = 1.93$, p < 0.19; total reading times: $F1_{1, 21} = 2.96$, p < 0.11; $F2_{1, 15} = 4.33$, p < 0.03).

An increased processing load was expected for sentences where an unambiguous nominative NP followed an ambiguous NP1 because subjects were supposed to have committed themselves to the subject reading of the first NP which then had to be revised. For *first pass reading times*, however, no reliable simple effect between object first and subject first sentences with an ambiguous first NP could be established. If anything, the means in these two conditions tend towards the opposite direction than predicted. Again, first pass reading times are misleading because, as shows up in *total regression path durations*, the garden-path is obviously detected very fast and leads to more re-reading of earlier passages of the sentence (ambiguous NP1, OVS vs. ambiguous NP1, SVO: F1_{1, 21} = 1.63, ns; F2_{1, 14} = 7.05, p < 0.02). This difference is not fully reliable, probably due to an ambiguity effect for SVO sentences (ambiguous SVO vs. unambiguous SVO: F1_{1, 21} = 4.57, p < 0.05; F2_{1, 15} = 1.51, ns). Object first sentences with an unambiguous accusative NP1 led to increased *first pass reading times* and *total*

regression path durations compared to sentences with unambiguous nominative NP1s (FPRT: F1_{1, 21} = 5.53, p < 0.03; F2_{1, 14} = 3.33, p < 0.09; TRPD: 1_{1, 21} = 8.07, p < 0.02; F2_{1, 14} = 7.18, p < 0.02). This may be due to either spill over effects or to repair processes adjusting the accusative NP to the nominative subject expected at that position.

Noun.Major effects of constituent order (F1_{1, 19} = 7.02, p < 0.02; F2_{1, 12} = 3.21, p < 0.10) as well as ambiguity of NP₁ (F1_{1, 19} = 4.64, p < 0.05; F2_{1, 13} = 5.56, p < 0.04) were found for total regression path durations at the noun of NP₂. Both these effects are mainly due to increased reading times for nouns in object first sentences with an ambiguous NP1 in comparison to object first sentences with an unambiguous NP₁ (F1_{1, 20} = 5.18, p < 0.04; F2_{1, 14} = 4.83, p < 0.05) and subject first sentences with an ambiguous NP₁ (F1_{1, 20} = 5.18, p < 0.04; F2_{1, 14} = 4.20, p < 0.06; F2_{1, 12} = 3.58, p < 0.09).⁵⁵

Discussion

The data show clearly that unambiguously accusative NPs in sentence initial positions are more difficult to process than unambiguous nominative or ambiguous NPs which can be interpreted as nominative. This effect clearly shows up before the end of the first NP, namely, at the adjective. Given that usually not more than three characters of a succeeding word are within the perceptual span, this early effect cannot be due to the preprocessing of the noun, since in German the case is marked by word final suffixes. Thus, the results draw a clear picture: a sentence phrase marker is built and incoming items are attached to it before the verb is read. *Head licensing*, but also *arc-standard left corner* parsing, which was assumed to explain the data from the selfpaced reading experiment in Hemforth (1993; Hemforth et al., 1993) can be excluded as viable parsing strategies, whereas an arc-eager left-corner parsing algorithm is most compatible with the data.

However, two alternative interpretations of the data have to be considered. Firstly, it could be the case that adjectives and nouns in accusative NP1s are generally more difficult to access in the mental lexicon. This could explain increased reading times on the respective words in the first NP. The lexical access explanation does not seem to be valid, however, because on the *second* NP, accusative adjectives and nouns are processed *faster* than nominative ones.

The second alternative has been proposed by Gorrell (in press) and is associated with the difference in complexity of the two structures to be built. Based on an approach by Travis (1991), Gorrell assumes a difference in the amount of structure building that is necessary when the first NP is nominative or accusative, respectively (see Hemforth, 1993, for a similar line of argument). The subject NP is supposed to stay in SpecIP (123a) whereas the topicalized object NP is supposed to be moved to SpecCP (123b).

^{55.}A further effect that is rather complicated to explain and will not be discussed further showed up for total reading times: total reading times for subject first sentences with an ambiguous first NP were shorter than for subject first sentences with an unambiguous first NP (F1_{1,20}=4.12, p < 0.06; F2_{1,12} = 2.41, p < 0.15).

(123) a. [IP Die Frau [I' sah₁ [VP die Mutter t_1]]]

b. [_{CP} Die Frau [_{C'} sah₂ [_{IP} die Mutter [_{I'} t_2 [VP t_1 t_2]]]]]

Given these representational assumptions, it can easily be seen that for subject NPs only an IP has to be postulated whereas a more complex CP has to be constructed for topicalized object NPs. For ambiguous first NPs the simpler structure should be adopted first leading to an increased processing load if the second NP is incompatible with this analysis. Therefore, Gorrell claims, increased reading times for accusative NP1s are not supposed to reflect reanalysis but increased local effort in structure building.

How does this explanation fit the data? Increased structure building effort should show up in locally increased reading times, i. e. the *first pass reading times* on the words inducing additional structure building, because either of the structures would have been built within the *initial* path of analysis. What we find, however, is an increase in *total regression path durations* but nothing at all in *first pass reading times*. Obviously, subjects re-read the determiner in the case of accusative NP1s, indicating a reanalysis process induced by the fact that the top-down prediction of a *subject*-NP could not be verified. Additional structure building alone clearly fails to account for this pattern of results.

From this experiment, it can be concluded that the parsing strategy which is most natural for strongly lexicalized grammars, *head licensing* or *head-driven parsing*, is empirically not viable. Does that mean that we should abandon lexicalized accounts completely? The lexical preference effect established for PP-attachment in section 4.1.1.2 suggests that lexical heads do have a guiding influence on parsing decisions if they are only available. In the following section, I will discuss the role of lexical information in more detail.

4.1.3 How lexical information is employed during parsing

It is obvious that people use various types of lexical information to establish an interpretation of a sentence and to figure out the most plausible interpretation when a sentence is ambiguous. It is much less obvious, however, at which point in the course of sentence comprehension this variety of lexical information is put to use. Parsing research focuses mostly on subcategorization information, i. e. the information that constrains the combinatorial capabilities of lexical heads and their complements, and more recently on *thematic* information, which, being crucial for interpretation, determines the thematic argument roles the complements can (plausibly) take. The questions addressed in this section will be relevant for two reasons: first, evidence on the use of detailed lexical information will determine how this information must be represented in the competence base of a cognitive model of parsing and, second, it may provide some important insights into the internal architecture of the language faculty.

The line of argument will proceed as follows: I will firstly discuss the question of whether subcategorization information is capable of modulating structural parsing preferences and if so, how early lexical influences can be established in the course of sentence processing using highly sensitive experimental techniques. Secondly, as it will become clear that although the evidence provided substantiates the early use of lexical information, it remains to be shown whether or not this information is used to guide the *initial* structure assembly. Finally, I will briefly comment on the similarities and differences between the use of subcategorization and thematic information in general.

4.1.3.1 Subcategorization information

Verbs fall into distinct categories depending upon whether they combine with only a subject, as in "Don sneezed.", with an additional object, such as "Peter loves Mary.", with two objects, as in "John gave Paul the record.", or perhaps with a PP-object like "Paul put the record on the rack.", and so on. In standard grammar approaches, the verb's subcategory is represented separately from the major category, such as N, V, ADj, etc., usually by providing a list of "subcategorized" complements (<NP, NP, PP>), or functional arguments (<SUBJ, OBJ, PCOMP>), which will not be distinguished henceforth for expository reasons. Such a representational separation⁵⁶ raises the question of whether or not subcategory information is used immediately when a syntactic structure has to be proposed. In a top-down depth-first approach, for example, subcategorization information cannot help in choosing one of several possible grammar rules available to expand a phrasal node. It may therefore be proposed that if a word matches the major category of a predicted terminal node, subcategory information may initially be ignored such that the word can be integrated immediately. In such an account, subcategorization information is used only at a later stage to check the structure and to rule out incompatible ones. If so, an important fol-

^{56.}Note that there are different accounts, such as *categorial grammar* (Ades & Steedman, 1982), in which major category and subcategory are not distinctly represented.

low-up question is whether such a check is only performed at the end of sentence, at clause or phrase boundaries, or even earlier, e.g., word-by-word.

At the opposite extreme, some recent accounts claim that syntactic structures cannot even be built without the detailed information provided by lexical heads. In recent grammar frameworks, such as *Head-driven Phrase Structure Grammar* (HPSG, Pollard & Sag, 1987, 1994), the "rules" or constraints outside the lexicon specify neither categorial nor subcategorization information so that the assembly of structure would be completely unconstrained if no detailed lexical information were available. It seems reasonable, then, to restrict structure building to circumstances in which lexical heads provide necessary information (constraints). Clearly, lexical information is crucial in a "head-driven" account.

Verbs may carry multiple subcategorization lists, e. g. due to the fact that some complements need not necessarily occur with the verb, as in "*Peter read.*", opposed to "*Peter read a book*". In some early approaches, such as the *lexical analysis strategy* put forth by Fodor, Garrett and Bever (1968), it was assumed that in the case of a structural ambiguity all lines of analysis consistent with any of the verb's subcategorization restrictions were explored in parallel. The parallel approach however was abandoned quickly in the early seventies, and only recently have parallel accounts evolved again (e.g. Gibson, 1991; Stevenson, 1995; see chapter 2.10 and 2.12). This issue will be discussed in detail in section 4.2.1. In this section, however, I will focus on serial models of parsing, which may or may not make use of lexical information, or even completely rely on detailed lexical information in order to be able to do anything sensible at all.

Lexical preferences and attachment ambiguities

In between these extremes, models have been proposed in which grammar-rule driven top-down processes interact with the bottom-up processing of detailed lexical information. One of these accounts, the theory of syntactic closure (Ford, Bresnan & Kaplan, 1982) is discussed in detail in chapter 2.7. The authors assume that multiple subcategorization lists (*lexical forms*) are ordered according to individual lexical preferences which may determine whether or not a certain complement is to be expected during parsing. The preferred lexical forms of the verbs *carry* and *include*, for example, cause the PP *for Susan* in (124a) to be attached to the VP, and to the NP *the present* in (124b), respectively.

- (124) a. John carried the present for Susan.
 - b. John included the present for Susan.

As already mentioned in chapter 2.7, Ford et al. (1982) ran a questionnaire study to confirm their claims, with sentences such as (124a,b) and a variety of others. Immediately after reading each sentence, subjects had to mark one of two unambiguously rephrased versions of the sentence to indicate which of the interpretations had occurred to them first. The results confirmed their assumption that the lexical properties of the verbs are capable of modifying structural preferences. Unfortunately, this finding is considerably off-line: what "appears" to the subjects as their first interpre-

tation can very well be the result of a cumulative influence from various sources of information settling on one interpretation after several steps of reanalysis. However, the data provided by Strube et al. (1990) and Konieczny (1989) from self-paced reading studies on similar constructions, as well as the data from the experiment presented in section 4.1.1.2, all using PP attachment ambiguities similar to those in (124ab), strongly support the assumption that lexical preferences influence parsing in these constructions, even before the end of the sentence.

Early influences of detailed lexical information have also been found in a study conducted by Clifton, Frazier, and Connine (1984). They had subjects read sentences such as (125) and (126).

(125) The baby-sitter {a. read, b. sang} the @ story to the sick child.

(126) The baby-sitter {a. read, b. sang} to @ the sick child.

Only *optionally* transitive verbs were used; i.e. verbs after which the direct object NP may be omitted. For verbs like *read*, however, the *transitive* reading is preferred, whereas verbs like *sing* are *preferentially intransitive* (according to the experimenters' judgements). If it were assumed that subcategorization preferences are used immediately to build an appropriate structure (one that is most likely to meet the expectations), one would expect that a determiner following a preferentially intransitive verb (*sang*) and, conversely, a preposition following a preferentially transitive verb would result in processing difficulties.

The sentences were presented externally paced (300 ms + 50 ms break) in a stationary window. After the first word following the verb (indicated by "@"), the presentation was discontinued until the subjects performed a lexical decision task on an unrelated word, which was presented at a separate location on the screen.

The lexical decision times indicated that preferences had a very early impact: subjects performed faster when the word following the verb matched its preferred argument frame (908 ms and 877 ms mean secondary task reaction time for 125a and 126b, respectively) than when it mismatched the frame (1000 ms and 1008 ms mean secondary task reaction time for 125b and 126a, respectively).

The aspect to be emphasized is that the effect occurred before any information about the semantic content of the verbal complement became available because subjects had only seen either the determiner of the NP complement or the preposition before the lexical decision target was presented. The experiment thus provided clear support that lexical information about preferred subcategorization frames is used very quickly. (This study will be further discussed in chapter 5.5, where important implications of this finding will be related to the detailed predictions of the SOUL mechanism.)

NP vs. sentential complement expectation

The evidence on a different type of ambiguity, however, is far less clear. In sentences like (127a,b), both verbs, *saw* and *doubted*, can take either an NP object or a sen-

tential complement. Thus, the ambiguity does not result from the potential omission of a complement, but from alternative realizations of a complement. Verbs appear to impose certain preferences on the type of complement:

- (127) a. The reporter saw her friend was not succeeding.
 - b. The candidate doubted his sincerity would be appreciated.

In these examples, *saw* prefers to combine with an NP, whereas *doubted* expects a sentential complement preferentially. The NP following the verb could thus be interpreted as the direct object, which would result in a garden path when the second verb is read, or as the subject of the (reduced) sentential complement. If the verb-preferences on the complement type were able to influence parsing immediately, one would expect subjects to be led up the garden path only in (127a), as the verb in (127b) would direct the parser to pursue the proper analysis.

Few of the studies set up to investigate this issue established decisive evidence, even though some claim so (Holmes, 1987; Holmes et al., 1987; Kennedy et al., 1989; Ferreira & Henderson, 1990). However, all of these studies have substantial flaws, either on the level of the presentation technique used, data extraction, or simply at the level of material and design construction. Ferreira and Henderson (1990), for example, conducted an eye movement study which appeared to disconfirm the claim that subcategorization information is employed immediately during parsing. They used sentences such as (127ab) mixed with unambiguous counterparts (unreduced sentential complements), such as (128ab).

(128) a. The reporter saw that her friend was not succeeding.

b. The candidate doubted that his sincerity would be appreciated.

The results suggested that the subjects were garden-pathed to quite the same extent in both ambiguous conditions, regardless of the verb's preferences. A closer look, however, reveals that the authors based their arguments only on *first fixation durations* and *total reading times*. The regression pattern on the first post-disambiguating region shows an increased regression frequency for verbs with an NP-complement preference, which the authors interpret as evidence that the (structurally disfavored S-complement) preferences of the verb might have guided the parser to perform an efficient *reanalysis*. My impression, however, is that the more straightforward interpretation would suggest that the subjects were simply garden-pathed more strongly when they read an NP-preference verb. The increased number of regressions indicate increased parsing difficulties, which probably would have shown up in the *regression path durations* (see chapter 3), had they been computed⁵⁷. If so, the data would confirm the lexical guidance assumptions even in these kinds of constructions.

Moreover, Trueswell et al. (1993) discovered that some of the verbs Ferreira and Henderson used preferred neither the sentential complement, nor the NP comple-

^{57.}The occurrence in the post-disambiguating region, however, might suggest that the preference may have been used as a filter. On the other hand, first pass preference effects may very well be slightly delayed to the subsequent word (see chapter 3).

ment continuation. Verbs like *pray* and *agree*, for instance, appear more frequently in constructions differing from those under consideration here, such as an intransitive construction (*John prayed every night*) or with an infinitival complement (*John agreed to wait in line*). Therefore, the "ambiguity effect" might possibly be due to disfavored continuations in both alternatives. Furthermore, Trueswell, et al. (1994) found in a fragment completion study that the extent to which the optional complementizer *that* was included in the completion strongly depended on the verb, ranging from 100% to less than 20%. Juliano and Tanenhaus (1993) report that the *that* preference is strongly related to lexical frequency: the lower the frequency, the more often the verb occurs with a *that*. Trueswell et al. further pointed out that reading times of post-verbal NPs were correlated with the verbs' *that*-preferences. Summing up, quite a number of alternative accounts exist for Ferreira and Henderson's data. The possibility that lexical preference influence the structure building process in NP/S-complement ambiguities can therefore not be ultimately rejected.

Lexical guidance vs. filter

Today, there is a broad consensus that some lexical information does influence the parsing process very shortly after a word is processed. However, it is still an open question whether this information is used to guide the initial structure-building process (*lexical guidance, lexically directed assembly*), or whether it is only used to monitor or evaluate the independently built structures (*lexical filter, structure checking models*), and to rule out inconsistent ones. All studies presented above fail to provide a clear distinction, due to two reasons. First, in most studies, the disambiguation point follows the ambiguous region, such that by the time the effect is expected, the parser may have already quickly revised its initial analysis towards the lexically preferred one (see Mitchell, 1989). Given the possibility of fast revisions, many techniques used, such as sentence-by-sentence self-paced reading, are not sufficiently sensitive to the guidance vs. filtering distinction. If we consider increments to be passed from the structure assembler to the structure checker as fine as simple phrases or even words, it becomes very hard to find distinctive measures at all.

Nevertheless, there is evidence from an early study suggesting that even strict subcategorization restrictions could *not* prevent the parser from ungrammatical *gap-filling*. Frazier et al. (1983) investigated sentences like:

- (129) Everyone liked the woman_j who the little child_? {a. begged, b. forced} (__j) [__j to sing those stupid French songs].
- (130) Everyone liked the woman_j who the little child_k {a. begged, b. started} [$__k$ to sing those stupid French songs for $__i$].

In sentence (129), two gaps coindexed with the same filler have to be posited right after the second verb (*begged* or *forced*), whereas (130) opens another independent gap position at the end of the sentence following the preposition *for*. Since the relative pronoun (*who*) or the coindexed NP (*the woman*) must as a genuine filler occupy some gap, the only permitted reading in (129) is the one indicated by the subscript index (*j*), meaning *the woman* is both the object of *begged* and *forced*, and the subject of *sing*. In

(130), however, *the woman* must be the filler of the final gap following *for*, such that the subject-gap for *sing* is filled with the NP *the little child*.

The important property of these sentences is that when the second verb is read, the parser does not know how the gap to be posited here should be ultimately filled. Frazier et al. (1983) claim that people handle this ambiguity by employing a general strategy, namely the *most recent filler strategy*, in both kinds of sentences. Following the *most recent filler strategy*, the parser fills a gap, as soon as it is encountered, with the most recent (potential) filler, i. e. the NP closest to the left of the gap. They thus predicted sentences like (130) to be read faster than those like (129), which is exactly what they found in their reading study, even though the sentences like (130) were actually longer.

Interestingly, the *control* properties of verbs like *forced* in (129) and *started* in (130), in contrast to those of *begged*, place different constraints on the subject of the embedded infinitival construction, such that neither type of sentence remains ambiguous with these two verbs *(force* requires the subject of the embedded verb to be co-indexed with its direct object *the woman*, whereas *started* identifies its subject *the child* with the subject of the embedded verb). If this information were used at the time the gap-filler binding was established, then the *most recent filler* effect, observable with the ambiguous verb *begged*, should have been eliminated.

However, Frazier et al. (1983) found that the *most recent filler* effect was actually unaffected by lexically unambiguous verbs. They thus claim that the processor initially assigns filler-gap dependencies using the *most recent filler* strategy and only later checks the grammatical constraints on permissible filler-gap assignments imposed by the verbs⁵⁸.

Of course, it can be argued that the technique and measure they used in the experiments (sentence reading time) were not sensitive enough to distinguish their interpretation from alternatives. Ford and Dalrymple (1988), in particular, have questioned Frazier et al.'s interpretation. They propose that the observed preference may have been caused by the manner in which the parser employs lexical control information in the parsing process. When the subject gap of the embedded verb following a control verb is encountered, it must be coindexed with one of the control verb's arguments. It may well be that in order to assign a dependency with the embedded subject, the parser searches through the control verb's list of complements starting with the subject, and only if the control information prohibits the assignment, the next complement is consulted, and so on. Thus, the subject-control strategy could also account for the finding that sentence (130) is read faster than (129), since the verb started permits its subject to be co-indexed, whereas forced requires its direct object to be identified with the embedded subject. The empirical result at hand may therefore not distinguish between the two explanations, and only future research might provide distinctive results.

^{58.}The fact that the application of the *recent filler strategy* may lead to an ungrammatical filler-gap assignment has important implications on the competence - performance relationship that must be assumed in such a model.

What makes Frazier et al. 's result so interesting, however, is that if they are right in assuming that the parser initially ignores detailed restrictive lexical information of verbs, such that even ungrammatical *filler-gap assignments* can be established initially, similar phenomena might also be observable in processing *attachment* ambiguities.

Following Mitchell (1987, 1989, Mitchell et al., 1992), a research logic different from the one underlying previous studies is thus advised: instead of establishing influences of lexical preferences on structure assembly, one should rather investigate whether the parser facing strict subcategorization restrictions will nevertheless follow independent preferences *despite* these restrictions. If one could demonstrate that the parser ignores restrictive information during initial structure assembly, it would seem unlikely that considerably less restrictive subcategorization *preferences* are able to *guide* the initial structure assembly.

Mitchell (1987) has conducted a study which appears to provide evidence against the lexical guidance hypothesis. Subjects read sentences like (131) and (132).

(131) After the child visited the doctor / prescribed a course of injections.

(132) After the child sneezed the doctor / prescribed a course of injections.

The sentences were presented in two sections, indicated by the slash ("/"). Presentation of the two sections was subject-paced and non-cumulative. If the human parser uses the subcategorization information of the verb to directly build the appropriate structure, as Mitchell (1987, 1989) claims, no differences in the processing times between the first sections of the sentences should have been observed.

However, the subjects took longer to read the first part of sentence (132) than the first part of sentence (131). Thus, Mitchell interpreted this fact as support for the *two-stage structure-checking* approach, in which a syntactic structure is built in the initial stage on the basis of only major category information before more detailed lexical information, such as the subcategorization information, is used to rule out illegal attachments. Accordingly, subjects first attached *the doctor* to *sneezed* (e. g. due to *Minimal Attachment*), despite the intransitiveness of the verb. Only after this initial attachment does the subcategory of *sneezed* rule out such an attachment and initiate a reanalysis of *the doctor*, resulting in increased reading times of the first section in (132).

A couple of objections, most of which are associated with the segmentation technique used, can be held up against this interpretation⁵⁹. However, Mitchell's inter-

^{59.}Subjects use some segmentation borders as a "chunking" cue if the segment exceeds the one-word size (see Cromer 1972). If a segmentation boundary meets a subclause boundary, as in (131), everything is fine. If, however, the segmentation boundary is somewhere within a subclause, as in (132), some wrap-up like chunking procedures might have been initiated where they usually would not occur. Leaving aside the unnaturalness, this would add an additional chunking procedure compared to the other case. Moreover, the memorizing effort might be bigger, since *the doctor* cannot be integrated into the sentence, but may remain a chunk of its own until it can be integrated into the second part of the sentence. Thus, even if the parser had carried out the right analysis initially, i.e. it never had attempted to attach *the doctor* to *sneezed*, the pattern of results remains explainable.

pretation seems to gain support from several follow-up studies conducted by Stowe (1989), and most recently by Adams and Mitchell (1994) and Adams, Clifton and Mitchell (forthcoming). The latter two are replications of the (1987) study with a subject-paced word-by-word reading and an eye-tracking technique, and their results seem to suggest that the effect in the earlier study is not due to segmentation artifacts.

However, a closer look at the material they used uncovers the fact that about half of their verbs, such as *talk* and *yawn*, have rare but permissible transitive lexical forms⁶⁰. Does this weaken the expressiveness of the study in any way? Since these forms are hardly preferred ones, a parser initially guided by lexical preferences, such as Ford et al.'s (1982), would still have problems in simulating Mitchell's results. On the other hand, as will become clear in chapter 5.5, the results provide no evidence against the immediate use of lexical information in a model that gives highest priority to *head attachment*, such that the parser initially attempts to attach an item to the structure with its lexical head already processed. The possibility recovering a transitive form, even a rare one, that permits the attachment of the NP as a direct object in the first analysis, allows for an interpretation fully compatible with a single-staged, modulated lexically guided account, such as *parametrized head attachment* and the SOUL mechanism outlined later in this thesis⁶¹ (see chapter 5).

Mitchell (1989) claims that his interpretation of the data is still valid despite the apparently contradicting evidence from a series of studies conducted by Stowe (1989). In a word-by-word subject-paced reading experiment she presented sentences like (133). The crucial point in these sentences is that there is a strong preference to assume the intransitive passive-ergative reading of *the truck stopped*, if the subject is *inanimate*, as in the case of *truck* (133b), whereas in the case of an *animate* subject, like *police* in (133a), the transitive reading is preferred.

(133) Before the {a. police, b. truck} stopped the driver was already getting nervous.

Stowe found longer reading times at the verb (*was*) following the ambiguous NP when the first noun was *animate* (*police*) than when it was *inanimate* (*truck*). Although these findings suggest that lexical information has been used immediately during parsing, the data again did not sufficiently prove the lexical guidance hypothesis, as Mitchell (1989) pointed out. Since the increased reading times were measured after the ambiguous position in the sentence, it is still possible that the information was only used to filter out the correct reading shortly after the initial stage had assembled the transitive reading.

⁶⁰ according to Webster's Ninth Collegiate Dictionary.

^{61.}There are still further options as to how Mitchell's and Adams et al.'s results could be explained. Firstly, as it is permissible in English, subjects might have expected a comma following an intransitive verb. It could thus be the lack of the comma that made them read the NP as a direct object. Secondly, even with a strictly intransitive verb, an NP can be a valid continuation as an adjunct, as in "After the child sneezed these days,...". Nevertheless, I share Mitchell's and Adams et al.'s view that the direct object attachment is at least attempted.

Stowe conducted a follow-up study, in which she only presented sentences with an inanimate subject like (133b). The potential direct object NP was however varied with respect to its plausibility as an object (134).

(134) When the truck stopped the {a. driver, b. silence} became very frightening.

Subjects spent reliably longer reading the NP following the first verb if it was implausible as a direct object (134b), than in (134a) where the NP can be interpreted as a direct object. This implausibility-effect, in accordance with Mitchell (1989), should not have occurred if the initial analysis had been guided by the verb's lexical preference not to expect an object if the subject is inanimate. Thus, the parser seems to perform the initial attachment of the object despite the verb's preference which is then shown to be used only in a second stage of processing, though very early.

However, this interpretation is once again not completely compelling. As before, the operations of a single-staged, modulated lexical guidance approach, such as *parametrized head attachment* and SOUL (chapter 5), also correspond to Stowe's findings: despite the preferentially intransitive form of *stopped*, the attachment of the NP following the verb would still be attempted in both cases such that the disfavored (transitive) lexical form might eventually be recovered to let the NP be integrated as the direct object. If so, the plausibility of the noun as a direct object may in fact produce an effect⁶².

Complement and adjuncts

The question of the immediate use of subcategorization information is naturally related to the question of whether complements are treated differently from adjuncts during the initial stage of parsing. If subcategorization information does not determine the initial parse, complements should not be initially distinguishable from adjuncts. On the other hand, many of the models presented in chapter 2 postulate a preference for complement over adjunct attachment (e.g., Gibson, 2.10, Stevenson, 2.12, Abney, 2.9).

Clifton, Speer, and Abney (1991) conducted a series of reading studies. In a selfpaced, segment-by segment reading experiment, as well as in an eye-monitoring experiment, subjects read sentences like (135-136).

(135) The saleswoman tried to interest the man {a. *in a wallet*, b. *in his fifties*} during the storewide sale at Steigers.

^{62.}Again, there is still another explanation of Stowe's findings, as pointed out by Strube (personal communication): since the passive/ergative reading of the *truck stopped* is only preferred, but by no means strictly required, the subjects surely had chosen the preferred reading in most of the cases, but probably not in all cases. If the subjects had chosen the disfavored reading in, say, about 30% of the cases, hence, there is a sufficient amount of cases for the plausibility effect to evolve. The data presented by Stowe are thus completely consistent with the assumption that the parsing process has been guided by lexical information.

(136) The man expressed his interest {a. *in a hurry*, b. *in a wallet*} during the storewide sale at Steigers.

In all of these sentences, the PP was semantically biased so that it could either be plausibly attached to the preceding verb (*in a wallet* to *interest* in 135, and *in a hurry* to *expressed*, in 136), or to the preceding NP (*in his fifties* to *the man* in 135 and *in a wallet* to *interest* in 136). Note that in (135), the PP is taken as an argument by the verb (*interest*) and as an adjunct by the NP (*the man*). In (136), however, the situation is reversed: the noun (also *interest*) takes the argument here, whereas the PP must be an adjunct if it is attached to the verb.

Whereas the *Garden-Path Model*, and in particular *Minimal Attachment*, predicts that in both (135) and (136) the attachment to the verb (verb-attachment) is initially preferred (135a and 136a), the licensing structure parser (Abney, 1987, 1989) predicts that the PP is preferentially attached as an argument (argument-attachment) initially (135a and 136b). Note that the prediction for the initial preferences only differ for sentence (136ab), because only the noun *interest*, and not the verb *express*, can take a PP as an argument.

Clifton et al. obtained increased first pass reading times at the PP when it was biased to force the NP attachment in both (135) and (136)⁶³. Only in the region following the PP, however, did first pass reading times show an advantage of argument over adjunct attachment. The same effect showed up in the total reading times even on the PP. The authors took the results as support for a two-staged garden-path model, in which the initial structure is built following *minimal attachment* and in which the *thematic processor* then guides the reanalysis in order to establish the best theta-role assignment.

This interpretation seems convincing only at first glance. As I argued in chapter 3, it is far from clear which kind of process the first pass reading times actually reflect. If one admits that reanalysis processes can sometimes cause regressions to earlier regions, lower first pass reading times in one condition might only reflect that regressions were initiated earlier. The processing load induced by the failure to attach an item in a way proposed by the preferences of the parser would then be better mirrored if the durations of all fixations following a regressive saccade from that region until the region is read again or skipped were included in the measure. This is essentially what *regression path durations* (RPDs) are. Unfortunately, the authors do not provide RPD data. However, they provide the frequencies of regressive saccades from the PP: the probability of such saccades was significantly higher for forced verbattachment (135a, 136a), and approximately higher for forced adjunct attachment (136b), which did

⁶³ I will focus my attention on the eye-movement data, which can be considered much more valid than the self-paced reading data. The latter suffer from extremely long reading times due to the artificial secondary task, and especially from the phrase-by-phrase presentation method, which has been demonstrated to produce considerable artifacts several times (Scheepers et.al, 1994; Günther, 1989).

not differ too much from the remaining conditions (6% for verb-arguments, and 4% for noun-adjuncts⁶⁴). Note that (136) was the sentence for which the predictions of a structure-based initial analysis model like the GP-model and lexical frame based models, like Abney's Licensing Parser, differ. If we take the regression probability as an additional indicator for occurrences of reanalysis processes, the data seem to support an interpretation opposite to what the authors suggest. Although the absolute frequency of regressions was rather small (4-14%), it seems absolutely likely that they would have found a *first pass* effect of argument attachment vs. adjunct attachment in the PP, if only they had analyzed the RPDs.

Note that the first pass data given are likely to be compromised by the "regression effect": it is quite possible that short first pass reading times are sometimes due to an early initialization of regressive eye-movements caused by the initiation of reanalysis. This seems even more likely in the light of the fact that the condition with the shortest average first pass reading time, namely verb-attachment, is also the condition with the highest frequency of regressions. It would definitely have been more accurate if the cases with regressive saccades following first pass reading had been excluded from the analysis of the first pass reading measure (as it is done in the *contingent analysis*, suggested by Altmann et al. 1992). Although the exclusion of those data is associated with the loss of valuable information, one could then claim that the data reflect the real processing load in that region (though not for all cases).

Furthermore, the total reading time data provided support the hypothesis that RPDs would have uncovered a first pass argument vs. adjunct effect. The total reading times incorporate the durations of additional fixations stemming from re-readings of the region both *within* its regression path, i.e before the region is passed for the first time, and outside its regression path (within the regression path of a later region), i.e after the region has already been passed. Depending on the proportion of the two, the total reading times may very well reflect *first pass* preferences, if they result mostly from *within* regression path re-fixations (see also Liversedge, 1994). Clifton et al. found significantly longer total reading times in the PP, if the PP had to be attached as an adjunct. It is by no means clear, as the authors suggest, that this result reflects second stage processes. In the light of the regression probability data presented above, the result is more likely to be due to reanalysis processes caused by the initial parsing failure in the PP. If so, the total reading times reflect more properly initial parsing preferences than the first pass reading times.

There are further indicators of an argument over adjunct advantage in the data. Since we cannot take the first pass reading data in the PP as a tempting indicator of first pass preferences, the data obtained in the following region might perfectly reflect

^{64.}The low frequency of regressions for noun-adjuncts is only superficially puzzling. Note that the first pass reading times for noun adjuncts was highest, such that the RPDs would probably still be high. Note also that the first pass reading times as well as the total reading times in the region following the noun-adjuncts showed a huge increase compared to all other conditions. The data thus suggests that in many cases noun-adjuncts, such as *in his fifties*, were initially attached and interpreted as verb arguments ("*interest in his fifties*") and that the initiation of a reanalysis is often delayed to the next region, possibly due to comparably long-lasting inferences needed to rule out the verb argument interpretation.

first pass *preferences* as well, since they probably cover those cases in which the semantic disambiguation in the PP showed up only one or a few words later. Clifton et al. reported a significant advantage of argument attachment over adjunct attachment in both first pass reading times and total reading times in the region following the PP. Again, neither the fact that only the total reading times in the prior region, the PP, nor the fact that the region after the PP follows the disambiguating region in place and time necessarily implies that the effects reflect processes at a stage following the initial analysis. In particular for semantically disambiguated material, we know that effects are sometimes delayed, at least in a certain portion of the cases.

Taken together, the results do suggest an interpretation very different from what the authors claim: argument attachment *is* preferred over adjunct attachment in the initial analysis. As a final proof of this hypothesis, however, further data analyses, in particular RPDs, would be required⁶⁵. As they stand, the data confirms both the predictions of *Licensing Structure Parser* and those of *Parametrized Head Attachment*.

4.1.3.2 Thematic information

In recent years, interest in subcategorization information as a guiding factor in parsing decreased in favor of thematic information. While thematic information in all its variants originated in the early attempts of linguistics and computer science to diminish the role of syntax in language processing (e.g. in *case theory*, Fillmore, 1968, and Schank's conceptual dependency account, 1972), it was later integrated into grammar theories (e.g. Chomsky, 1981), as it became clear that thematic grids are a powerful and indispensable source of information in syntax. Although subcategorization information and thematic information are clearly distinct (Pollard & Sag, 1994), there are nevertheless strong interdependencies (thematic arguments must map onto subcategorized complements), such that most of the psycholinguistic evidence presented earlier could easily be reformulated as evidence on thematic information in processing.

Many of studies have been conducted in the search of thematic guidance effects in gap-filling (Tanenhaus et al. 1989) and structure building ambiguities (Trueswell, Tanenhaus, & Garnsey, 1994; Trueswell, Tanenhaus, & Kello, 1993). There has been an ongoing debate about whether there is a structural (Frazier, 1979, 1987a) or perceptual (Bever, 1970) preference towards the main verb analysis, or whether the analysis is guided by *a priori plausibility* (Crain & Steedman, 1985), or thematic role restrictions (Tanenhaus et al. 1989). The argument is quite analogous to the debate on subcategorization information in that the proponents of the *Garden Path Model* claimed the initial syntactic analysis was clearly pursued exclusively on structural grounds (i. e.

^{65.}There is another pitfall in the design, as pointed out to me by Barbara Hemforth (personal communication): whereas noun and verb arguments were identical in most of the cases (*in a wallet* in (135) and (136), noun and verb adjuncts always differ (*in his fifties* vs. *in a hurry* in 135 and 136). Although the authors report that the mean lengths and word frequencies of the items do not differ substantially within each position, we can not ultimately exclude the possibility that the content words in verb-adjunct phrases are *lexically* easier to access than those of nounadjuncts, thus pushing a supposed "verb attachment effect" in the first pass reading time data.

guided by *Minimal Attachment*, Frazier 1987a). Higher level information is analyzed by the *thematic processor*, which is capable of instructing the reanalysis process if the initial analysis turns out to be semantically flawed.

Attempts have been made to support the guiding influence of lexical thematic aspects (Tanenhaus et al, 1989). Whenever such influences have been observed, apparently even eliminating a residual minimal attachment effect, it could have been argued that the considerably weak sensitivity of the methods used allowed for an interpretation consistent with the garden path model⁶⁶, since the semantically guided reanalysis initiated by the thematic processor can easily account for semantic effects, though only at a later stage.

On the other hand, residual effects of "structural biases" have been observed in studies using an eye-tracking technique (e.g. Ferreira & Henderson, 1990). As before, none of these studies are free of substantial experimental flaws, either in design, material presentation (line breaks within the critical zone in only one condition), or data analysis.

As one of the most typical examples on thematic information in parsing, I will only briefly discuss a recent study by Trueswell et al. (1994). They had subjects read "classical" garden path sentences beginning with an NP and a lexically (morphologically) ambiguous verb with identical simple past and past participle forms, such as *loved* in (137).

(137) The teacher loved ...

Fragment (137) could thus be continued as in (138a), using the simple past form, or as in (138b), in which the verb starts a passive-voiced reduced relative clause.

(138) a. The teacher loved to talk about geography.

b. The teacher loved by the class was very easy to understand.

Trueswell et al. (1994) used only sentences of the latter kind (138b), in which the first verb has to be read as a past participle verb in a passive voiced relative clause, because it is later disambiguated through the passive continuation *by the class*. However, these sentences were contrasted to others which only differed in that the first noun was *inanimate*, as in

(139) The textbook loved by the class was very easy to understand.

^{66.}This is surely true for self-paced reading with sentence-by-sentence presentation and even with smaller segments, such as phrases or words, since the button-pushing in self-paced reading slows down reading enormously (see chapter 3). The preferred method of Tanenhaus and his group is the *stop making sense* technique, in which the subjects have to push a certain button from the point where they thought the sentence was semantically spoilt. As some researchers pointed out (e.g. Fodor, 1990), the technique seems likely to induce special perceptive strategies in which semantic aspects may be emphasized too much to permit the conclusion that no structural biases were active in parsing.

If the readers are initially guided by structural means only, they could possibly be garden-pathed on the inconsistent PP *by the class*. On a multiple constraint approach (see chapter 2.11), the prediction depends upon the type of noun used in the first NP. In a main clause, the NP is the subject of the verb and as such, it will typically take an *agent* role. In a reduced relative clause, the NP becomes the object of the verb, which will typically take the role of *theme* or *patient*. Crucially then, *inanimate* nouns can hardly be (intentional) *agents* but are generally good *themes*. For ambiguous verbs the alternative sets of thematic roles for each possible reading will be activated. The semantic fit between the NP and its possible thematic roles on the one hand, and the alternative argument structures for the past tense and participial reading on the other hand, can determine the initially preferred reading.

In sentence (138b), then, with an *animate* noun, the multiple constraint model predicts a garden path on the PP *by the class,* just as the *Garden-Path Model* does. With *inanimate* nouns, however, predictions (might) differ: in the *Garden-Path Model*, one might expect at least a residual effect on the verb or the "*by*"-PP, compared to an unambiguous sentence like (140) (unreduced relative clause) and (141) (morphologically unambiguous verb) if the evaluation process in the thematic processor is delayed to a certain extent.

(140) The textbook that was loved by the class was very easy to understand.

(141) The poster drawn by the illustrator was used for a magazine cover.

The multiple constraint model, however, predicts that the residual effect can be eliminated⁶⁷. The first pass reading times on the verb and the "by"-PP appear to confirm this prediction. However, in the second pass reading times (i.e. the time of all rereadings of the region) Trueswell et al. found increased reading times for inanimate nouns suggesting that if *regression path durations* (see chapter 3) had been provided, which reflect processing preferences more realistically than first pass reading times, they would probably have obtained a residual effect on the verb for inanimate unreduced relatives.

Moreover, as pointed out by Frazier (in press), there are a number of fundamental shortcomings in almost every study that was set up to confirm the lexicalist constraint satisfaction approach. All studies (with one exception) have investigated the comprehension of sentences with structurally disfavored syntactic analyses, such as reduced relative and sentential complement constructions, and have ignored the structurally favored sentence counterparts in the experiment, such as (138a) Leaving aside the counter-balancing problem in the materials, this could have completely eliminated the potential to observe residual structural effects in the disambiguating region. Since the garden path model attributes thematic effects to reanalysis processes guided by the thematic processor, and since all investigated structures would require reanalysis according to the Garden Path model, these studies were therefore in princi-

^{67.}However, only if the verb itself is equi-biased. Due to the multiplicity of constraints immediately interacting in such models, they are generally hard to be falsified.

ple not capable of distinguishing between the two types of models (at least unless it is possible to distinguish the effects of first analysis and those of easy reanalysis).

On the other hand, semantic/thematic information has been demonstrated to influence parsing decisions very early, although to a different extent for different constructions. For PP-attachment sentences, Taraban and McClelland (1988) found in a self-paced reading study that subjects preferentially attached the PP to the VP or NP, depending on the over-all plausibility for either of these attachments in differing subject-verb-object combinations⁶⁸.

The general psycholinguistic evidence on thematic influences and subcategorization information in parsing looks fairly similar: early influences are clearly demonstrable, but distinctive experiments capable of distinguishing the predictions of different types of models are hard to come up with.

4.1.3.3 Conclusion

Summing up, there seems to be ample evidence that subcategorization information is used *extremely* early in the course of sentence processing (e. g. Clifton, Frazier & Connine, 1984). On the other hand, there is still no convincing evidence which decisively demonstrates that subcategorization information is only used during a second stage where the structure initially built without regarding subcategorization constraints is checked against this information and eventually rejected. "Shallow" thematic information, such as agents being typically animate, may also be used quite early in structure building (Trueswell et al., 1994) and gap filling (e. g. Tanenhaus et al., 1989). Information at even higher levels, such as *a priori plausibility* (world knowledge) and contextual pragmatics will be discussed in section 4.2.2.

In addition to the weak evidence on the filter approach, an explanation of many of the findings on early use of subcategorization information depends upon the followup assumption that the processor uses this information to guide the *reanalysis*. But why would a processor first ignore information when it is in principle capable of using it in a guiding fashion? Note that in an *Occam´s razor* type of argument, a model that initially ignores lexical information is *not* more parsimonious in its presumptions. Detailed lexical information *must* be used anyway and *is* used very early. The most parsimonious approach is therefore one which does *not* distinguish multiple stages of processing within the parser.

However, what the data provided by Mitchell and colleagues do suggest is that lexical preferences do not guide the parser in a completely unconstrained way. The best-fitting approach then, is one that uses detailed lexical information immediately, but only in circumstances which permit an interpretation of the input *as early as possible*. In such an approach, called here a *modulated* lexical guidance model, the different types of lexical information may be represented in a unified piece of information and used at a single stage in processing. *Parametrized Head Attachment* (Konieczny et al.,

^{68.}Note, however, that as far as I am aware, this result has never been replicated.

1994, see chapter 2.16) fits in quite well, and the SOUL mechanism, as described in chapter 5, even better.

Today Frazier and Clifton (1996) seem to be aware of the considerably weak empirical support for the unrestricted (i. e. valid for any type of subcategorization ambiguity) *filtering* approach. They claim:

> "... some advocates of structurally based parsing theories have hypothesized that subcategorization information is used only to guide reanalysis (...), not to guide initial analysis (...) (Adams, Clifton, and Mitchell 1994; Ferreira and Henderson 1990; Frazier 1987, Mitchell 1989), and have provided evidence in favor of their hypothesis⁶⁹. However, whether or not subcategorization information is used is not crucial to garden path theories; subcategorization is grammatical information that could be available to a modular, structural parser." (p. 17)

This assertion is a bit strange in the light of the extensive debate on this issue, lasting over fourteen (!) years until today, with the authors being among the major contributors. They continue:

> "The only thing that is crucial to such theories is that verb-disambiguated subcategorization information is not *necessary* for parsing a sentence ... A parser of the sort assumed in garden path theories need not wait until a verb carrying relevant subcategorization information is received ... Demonstrating that the verb subcategorization information can be used when it is available does not entail that it must be present before parsing of the input may begin." (p. 17)

That, of course, is something we can agree on, in particular in the light of the evidence from the experiment reported in section 4.1.2.

To conclude, it is reasonable to assume that lexical subcategorization information and possibly even shallow thematic information is used during initial structure assembly. Thus, the constraints on a valid model of human parsing have been narrowed down substantially: parsing decisions are dependent on the availability of lexical heads, but structure can nevertheless be built without them being available. If, however, a lexical head can project its detailed subcategorization and thematic information, it will do so.

Before a detailed parsing mechanism can be specified, two more questions will have to be answered. Firstly, *Parametrized Head Attachment* as it is described in section 2.16, is compatible with a locally parallel account, where all structural alternatives are built first and filtered by parsing principles only in a second stage. In the first part of the following section, I will weigh the empirical evidence including that from the preceding sections with respect to the question of how the parser handles ambiguities.

^{69.}Note that all of these approaches have been disputed in detail above.

Are structural alternatives built sequentially, are several alternatives considered in parallel, or is the construction of fully determinate representations postponed?

Secondly, the influence of higher level information on initial parsing decisions still remains to be discussed. Lexical and, to a certain extent, even thematic information can both be regarded as grammar internal information. Thus, immediate use of these kinds of information does not violate a modularity hypothesis (Fodor, 1983), where parsing cannot be guided by extralingual information. The question of whether higher level information can influence initial structure assembly will be discussed in the light of an eye-tracking experiment, where the potential influence of pragmatically biased contexts on PP-attachment preferences is investigated.

4.2 The architecture of the human sentence processing mechanism

4.2.1 Serial, parallel, and minimal commitment models

In chapter 2, sentence processing models were coarsely classified into three groups of models (i.e. serial, parallel, and minimal commitment models). I will draw somewhat finer distinctions in this section, following the classification discussed by Mitchell (1994). Depending on how structural ambiguities are handled, different predictions from each of the model classes arise for the processing load in ambiguous (e.g., *the answer* in 142) as well as disambiguating regions (e.g., *was wrong* in 142).

(142) The scientist believed | the answer | was wrong.

The predictions of the (classes of) models will be discussed in view of empirical results from psycholinguistic literature as well as from the experiments presented here.

4.2.1.1 Predictions

Serial models

In the group of serial models, following a depth-first strategy in structure building, I presented Wanner's (1980) ATNs (see section 2.5), Kimball's (1973, 1975) parsing principles (see section 2.4), Ford, Bresnan, and Kaplan's (1982) *Theory of Syntactic Closure* (see section 2.7), Frazier and Fodor's *Sausage Machine* (see section 2.4), as well as its successor, *Garden Path Theory* (Frazier & Rayner, 1982, Frazier, 1987; see section 2.6), Abney's (1987) *Licensing Structure Parser* (see section 2.8), and Mitchell's (1994) *Tuning Hypothesis* (see section 2.15). In some of these models (e.g., Wanner's ATNs, Ford, Bresnan, and Kaplan's model, the *Sausage Machine* and the *Garden Path Theory*), the positions where structural alternatives exist are annotated or "tagged" to allow for an easy reentry in cases of initial analysis failure. Furthermore, parsing principles may either fully determine the choice of a structural alternative or bias the choice in a more probabilistic way such that the preferred analysis is not always chosen but only most of the time.

Predictions in the ambiguous region. Ideally, serial models without annotations do not predict any difference in processing load between ambiguous structures and structures disambiguated in favor of the preferred reading. If local ambiguities are tagged for later recovery, however, the tagging operation may result in a somewhat increased processing load.

Predictions in the disambiguating region. In the disambiguating region, garden-path effects are predicted for disfavored structures in serial models, irrespective of whether the ambiguous region is tagged or not.

If parsing principles influence the choice of a structural alternative probabilistically, an ambiguity effect is to be expected: even for the preferred structures the "wrong" analysis has been selected initially in a proportion of the cases, so some reanalysis will always be necessary if an ambiguity is resolved later in the sentence. If, on the other hand, parsing principles determine the initial choice every time, "resolving" an ambiguity in favor of the preferred reading should ideally not be more costly than processing comparable structures in unambiguous sentential contexts. Empirically, these accounts are very difficult to distinguish. Principles are only assumed to determine structural choice, "other things being equal". If the less preferred structure is occasionally pursued, this can be attributed to "noise" due to other things not having been equal in the respective materials.

Parallel models

In the second class of models, several structural alternatives can be considered in parallel (see Gibson's - see section 2.10, MacDonald et al.'s - see section 2.11, and Stevenson's - section 2.12 - models, as well as the *unification space*, Kempen and Vosse 1989, and others). In almost all of these models, the human parser is assumed to be resource-limited to some extent. In MacDonald et al.'s and Stevenson's models, the processing load will generally increase with the number of structures that have to be considered. In Gibson's approach, on the other hand, structures exceeding some complexity metric are considered impossible to process. Keeping up several structures is not assumed to be costly per se.

In Just and Carpenter's (1992) READER system, the amount of parallel processing possible for the parser depends on the individual working memory capacity of the reader. If the memory capacity is exceeded, all analyses but one are discontinued. Otherwise, structural alternatives may be pursued in parallel. Thus, so-called high-(memory)-span⁷⁰ readers are assumed to be able to process more structures in parallel than so-called low span readers.

Predictions in the ambiguous region. In parallel multiple constraint models (e.g. Mac-Donald et al., 1994), the competition between structural alternatives should result in increased processing loads in ambiguous regions. In Gibson's parallel account, however, keeping up several structures does not necessarily increase processing load (see section 2.10).

Predictions in the disambiguating region. In parallel models, the predictions for the disambiguating region depend on the amount of parallelism permitted in the ambiguous region. Unconstrained parallelism without any weighing of the structural alternatives would not predict any effects in the disambiguating region because the ultimately correct structure can be selected from the full set.

If different strengths are attributed to alternative readings, as in MacDonald et al.'s (1994) and Stevenson's (1993, 1995) competition models, the weights of each alterna-

^{70.}In the reading span test (Daneman & Carpenter, 1980), subjects have to read sentences aloud and memorize the last word of the sentence. The more "last words" they can remember, the better their reading span. This test has been criticized extensively, since it is by no means certain that it measures anything like "syntactic working memory capacity" (see Frazier, in press; Strube, personal communication).

tive will have to be adjusted. The size of the effect should depend on the amount of re-weighing to be done. A "garden-path" effect is thus not (necessarily) attributed to reanalysis effort, but to the fact that the process of re-weighing is assumed to take longer for less preferred readings. Even for "preferred" structures some reweighing should be necessary in competition models leading to an ambiguity effect in the disambiguating region. If an alternative was abandoned in the ambiguous region because some complexity constant was exceeded, such as in Gibson's (1991) model, severe garden path effects are predicted.

Delayed attachment and minimal commitment

In the third class of parsing accounts, the structural representation of the sentence is not fully specified in cases of ambiguity. The integration of constituents may be postponed until disambiguating information shows up (e.g. Perfetti, 1990; Kennedy et al., 1989; Murray & Watts, 1995), or the structural representation may be partially underspecified (Frazier & Clifton, 1996, see section 2.14; Gorrell, 1995, see section 2.13; Marcus, Hindle, & Fleck, 1983; Sturt & Crocker, 1995; Weinberg, 1993).

Predictions in the ambiguous region. If structure building is postponed in cases of ambiguity, ambiguous regions may be processed faster, because less operations have to be carried out. This is not the case, however, if underspecified representations are constructed, which are structurally equivalent to the preferred reading, as in Gorrell's account. In this case, no ambiguity effect should be observed, as in untagged serial models.

Predictions in the disambiguating region. If structure building has been postponed in the ambiguous region (as in *wait-and-see* models like Perfetti's, 1990, or Kennedy et al.'s, 1989), taking up the operations in the disambiguating region may result in an increased processing load. The processing load associated to each alternative may vary, depending on how much additional structure building has to be done. In minimal commitment models like Gorrell's (1995) or Weinberg's (1993), where underspecified representations are built, an increased processing load is predicted when structural information has to be added, which is only necessary in non-minimal structures, of course. Severe garden-paths are assumed to result if a previously built structure has to be revised.

4.2.1.2 Evidence

Processing load in the disambiguating region

Throughout this thesis I have presented a broad variety of examples of parsing difficulties induced when a segment of a sentences forces an initially disfavored reading to be adopted finally. Such evidence has usually been taken to indicate that the parser employs a serial search strategy. While unrestricted parallel approaches can surely be ruled out on the basis of such observations, a closer look at the data reveals that other approaches apparently turn out to be compatible at least in some cases. In the following sections, I will briefly discuss some studies relevant to this question. Sentence (143) is one of the most extensively discussed cases in psycholinguistic literature on human parsing.

(143) John knew the answer to the physics problem | was correct.

Frazier and Rayner (1982) found that subjects were garden-pathed when the disambiguating region (was correct) forced the S-complement reading. Holmes, Kennedy and Murray (1987), however, argued that the increased reading times obtained were not due to a garden-path effect but to additional integration effort when a sentence structure has to be built compared to when only an NP is needed. Their interpretation was supported by the fact that unambiguous complement sentences (e.g. knew that ...) seemed to induce an equal amount of processing effort in the "disambiguation" region, as they established in a self-paced reading experiment. Consequently, the parser was claimed to avoid a commitment to one of the alternatives in an ambiguous sentence until sufficient information becomes available to rule out all but one alternative. Frazier and Rayner (1987) disputed this interpretation of the self-paced reading data and provided evidence from an eye-movement study suggesting that subjects did not show garden path effects in unambiguous S-complement continuations. In their reply, Kennedy et al. (1989) reported evidence from another eye-movement study confirming Holmes et al.'s (1987) interpretation and suggesting that Frazier and Rayner's (1987) findings were due to a line-break artifact in the presentation technique used. Ferreira and Henderson (1990) again provided evidence in support of Frazier and Rayner's (1982) initial interpretation. Trueswell et al. (1994), while confessing that subjects were indeed garden pathed, argued that Ferreira and Henderson's results were at least partly due to the fact that the majority of the verbs they used had a preference to expect an unreduced sentential complement, such that the lack of a complementizer (that) might have masked other lexical (preference) effects. Finally, Kennison (forthcoming) established that this can hardly be true and provided further evidence in support of "garden pathing" in NP versus S-complement ambiguities.

Interestingly, this type of ambiguity has been taken as a "supporting example" by the proponents of all kinds of models. Pritchett (1992) attributes the lack of a "conscious" garden path effect to the re-interpretability of the NP *the answer* ... within the *On-line Locality Constraint* (see section 2.9). Similarly, Gibson (1991) predicts no garden path since both alternatives can be pursued in parallel (see section 2.10). Proponents of minimal commitment models (e.g. Gorrell, 1995) also predict only a limited processing penalty when the non-minimal S-complement reading is forced, since the analysis can easily be accomplished by adding further structural predicates to the description (see section 2.13). Summing up, the evidence available on NP/S-complement ambiguities, although initially taken as substantiating serial models, does not seem to be as decisive as one would wish.

Clearer evidence against the complete delay of a commitment comes from an experiment conducted by Mitchell, Corley, and Garnham (1992). In Mitchell et al's. (1992) experiment, subjects had to read sentences like (144).

(144) a. Don shouted to the assistant that (he) / had been / ...

- b. Don forced the assistant that (he) / had been $/ \dots$
- c. Don and the assistant that (he) / had been / ...

If the pronoun *he* is included in sentence (144a), the part of the sentence starting with *that* can be either read as an (object) relative clause or a sentential complement of the verb *shouted*. If the pronoun *he* is not present, however, the only viable reading is the (subject) relative clause reading, due to the fact that in a complement clause, a subject would be missing before the verb.

Sentence (144b) only allows for the relative clause readings, since the verb *forced* does not take a sentential complement.

Similarly, sentence (144c) does not even provide a verb that could take a complement clause, such that including or excluding the pronoun *he* both results in unambiguous subject relative clause or object relative clause reading, respectively.

The authors found that if the first verb takes a complement, as in sentences like (144a), subjects were garden pathed in the *had been* segment, if the ambiguity was resolved there towards the (subject) relative clause reading (i. e., if the pronoun *he* had been excluded in the preceding segment). The authors claimed that when such verbs were used, there must have been an early commitment towards the preferred complement-clause reading (as predicted by most accounts, including PHA), resulting in a highly reliable effect, which they could demonstrate to persist even in strongly biased contexts (Mitchell et al., 1992, experiment 2).

This result alone, however, does not suffice to confirm the assumption of a depthfirst complement preference, as the authors point out. The effect might also be due to the fact that only when the pronoun *he* was not present in the preceding section was the ambiguity resolved in the *had been* segment, while in the other case the ambiguity remained. The same result would therefore also have been expected from a model that would not commit itself early in the case of an ambiguity.

In sentences (144bc), however, the reading times in the *had been* segment did *not* differ for the inclusion or exclusion of the pronoun *he*. What does this tell us? Both cases are unambiguous, but when the pronoun *he* is included, the ambiguity is already resolved in the segment *preceding* the *head been* segment. Thus, if *disambiguation per se* were the relevant factor for reading time differences, reading times on the *had been* segment for *subject* relative clauses (without *he*) should have been increased, contrary to what the authors actually found (cf. Mitchell et al. 1992).

The pattern of results as a whole is thus incompatible with a general underspecification assumption as well as with an unrestricted parallel approach, since mere disambiguation processes can be ruled out as being responsible for the obtained effects. Correspondingly, subjects must have committed themselves to the complement reading extremely early, namely, as soon as the potential complementizer *that* has been read, which is the only ambiguous word in these sentences. Note that although the data rule out the assumption that the parser waits for decisive information to make any decision, they are still compatible with a minimal commitment account like Gorrell's in which the parser will indeed commit itself to the simpler S-complement analysis.

The data presented for NP-attachment preference in section 4.1 also bear strongly on the issue. The evidence central to processes in the disambiguating region are illustrated in figure 149. The structures investigated in the experiment described in section 4.1 are repeated here for expository reasons as (145-148).

(145) Daß der Arzt der Sängerin ein Medikament gegeben hat, wußte niemand.

That the doctor [the singer_[fem, {gen/dat}]] a remedy *given* has, knew nobody. Nobody knew that the doctor has given a remedy to the singer.

- (146) Daß der Arzt der Sängerin ein Medikament *entdeckt* hat, wußte niemand. That the doctor [the singer_[fem, {gen/dat}]] a remedy *discovered* has, knew nobody. Nobody knew that the doctor of the singer has discovered a remedy.
- (147) Daß der Arzt dem Sänger ein Medikament gegeben hat, wußte niemand. That the doctor [the singer_[masc, dat]] a remedy given has, knew nobody. Nobody knew that the doctor has given a remedy to the singer.
- (148) Daß der Arzt des Sängers ein Medikament entdeckt hat, wußte niemand.

That the doctor [the singer_[masc, gen]] a remedy *discovered* has, knew nobody. Nobody knew that the doctor of the singer has discovered a remedy.

In (145) and (146), attachment of the case ambiguous NP *der Sängeringenitive or dative* is determined by the participle (*gegeben* or *entdeckt*) which is either di-transitive forcing the "dative" reading of the ambiguous NP or transitive, forcing the "genitive" reading. As illustrated in (149), only for the di-transitive verb, forcing VP-attachment, increased processing load was found in the disambiguating region. There was no ambiguity effect at all. Processing a transitive verb after a case ambiguous NP was

not more costly than processing any of the structures which were disambiguated on the structurally ambiguous NP (*dem Sänger* in 147, or *dem Sänger* in 148).



The complete lack of an ambiguity effect in the disambiguating region supports serial models where parsing principles determine the choice of structural alternatives in an all-or-none fashion. Weighed parallel competition models and approaches where parsing decisions are postponed are less compatible with the data.⁷¹

On the other hand, there appears to be evidence from a number of experiments which are claimed to support competition-based models. Unfortunately most of these experiments do not really bear on the issue, since they only established frequency and plausibility effects for structures that are disambiguated to what most serial accounts would predict to be the structurally less preferred readings, such as in (150), taken from (Trueswell et al., 1994).

(150) The defendant examined by the lawyer turned out to be unreliable.

Two shortcomings render these experiments less interesting (cf. Frazier, in press): first, increasing the number of less preferred structures increases the danger of artificial strategic effects, because subjects are very sensitive to experimental cues and become increasingly aware of the ambiguity throughout the experimental session. Second, showing evidence for plausibility effects on the ambiguous region or shortly afterwards in less preferred structures does not really differentiate between competition models and serial models with an efficient reanalysis component.

Apparently more informative is an experiment by Pearlmutter and MacDonald (in press): they investigated main verb/reduced relative ambiguities like (151a, b; Pearlmutter & MacDonald, in press) which were disambiguated in favor of the usually easier main verb reading.

^{71.}Note, however, that these models cannot be fully ruled out by a non-effect which might be due to the experimental techniques not being sensitive enough. On the other hand, as pointed out in chapter 2, eye-tracking is surely one of the most sensitive techniques available.

(151) a. The soup cooked in the pot but was not ready to eat.

b. The soup bubbled in the pot but was not ready to eat.

The verbs used were either equi-biased with respect to transitivity (i.e. equally frequent in their transitive or intransitive reading, 151a) or strongly biased towards the intransitive reading (151b). Only for the transitive reading of *cooked* in (151a) is a reduced relative clause reading temporarily viable. Thus, (151a) is temporarily ambiguous whereas (151b) is fairly unambiguous. In these experiments, so-called high span readers showed longer reading times in ambiguous (151a) than in unambiguous (151b) sentences within the ambiguous region (at the preposition, e.g., *in*), and at the beginning of the disambiguating region (the conjunction, e.g., *but*) whereas low span readers did not show any ambiguity effects, and they were slightly more accurate in answering questions related to the sentences.

To account for these results, Pearlmutter and MacDonald claim that high span readers, due to their increased working memory capacity, can and do consider competing readings if they are plausible. Low span readers, on the other hand, only consider the "easier" or lexically preferred reading. To substantiate this claim, the respective plausibility of the different readings was obtained in rating studies. Only for high span readers did Pearlmutter and MacDonald establish correlations of reading times with plausibility ratings at the beginning of the disambiguating region (the conjunction *but*), which were even stronger than correlations with the lexical frequency of the different readings of the respective verb.

This interpretation of the data can be criticized for various reasons, as Frazier (in press) points out. A major criticism is that one of the major predictions of multiple constraint competition models is disconfirmed by Pearlmutter et al.'s very data. Competition and thus processing load should be greatest when approximately equally strong constraints compete. A frequency based main verb preference competing with a more plausible reduced relative reading should therefore be particularly difficult. Nothing like that is reported in Pearlmutter and MacDonald.

Additionally, it is not very easy to explain why high span readers who, at least temporarily, consider an ultimately false alternative structure were more accurate in the comprehension task than low span readers who only consider the correct structure from the very beginning.

Finally, an ambiguity effect was only found for high-span readers (one third of the subjects). It is likely that they do not primarily differ from low span readers with respect to their working memory capacity, but with respect to motivation (cf. Frazier, in press) and sensitivity to experimental cues, such as ambiguity of the materials. If more motivated subjects read the experimental materials more carefully it is not surprising that they are more sensitive to plausibility.⁷²

^{72.}The increased attention to plausibility established for high span readers, by the way, does not necessarily reflect an increased parsing specific capacity, but (if any) more general working memory capacity related to the central executive (Strube, personal communication).

Processing load in the ambiguous region

As Mitchell (1994) points out, the "vast majority of studies reveal that the ambiguous materials are no more difficult and no less difficult to process than appropriately matched unambiguous materials". As we have seen above, the results from the Pearlmutter et al. study cannot be counted as an exception to this general picture.

The NP-attachment experiment presented in section 4.1.1.4 is an interesting instance of this rule. The summed first pass reading times of the "ambiguous regions", namely, the noun of the first NP (*Arzt*) and the determiner of the second NP (*der, dem,* or *des,* respectively), were *not* increased in the ambiguous conditions (*der*) compared to the unambiguous preferred NP-modifier (*des, genitive*) reading. Only disambiguation towards the less preferred VP-attachment (*dem, dative*) reading led to increased summed gaze durations, as illustrated in (152).

These data are not only incompatible with parallel competition models, but also with minimal commitment models where parsing operations are postponed in cases of ambiguity: only if there is a very early commitment in favor of the NP-modifier reading of the second NP can the increased processing load for the dative-marked NP be explained.



4.2.1.3 A preliminary conclusion

All in all, the vast majority of the data strongly suggest that the human sentence processor parses sentences in a depth-first manner, i.e. one structural alternative at a time, without any delays in commitment. From the previous sections, we know that it builds up structures incrementally taking the availability and the respective subcategorization properties of lexical heads into account.
4.2.2 Autonomous vs. interactive processing

The attachment preferences discussed so far have been attributed to syntactic and lexical processes performed by the human sentence processor. A question that still remains to be answered is that of whether or not the initial parsing decisions in the syntactic processor can be guided by higher level information, such as restrictions imposed by the context in which a sentence occurs. Some researchers even claim that parsing preferences are not at all determined by preferences in the syntactic processor, but by semantic and pragmatic influences alone. I will turn to this question in the following sections.

4.2.2.1 Attachment preferences in referentially biasing contexts

Models of the garden-path family where initial parsing decisions are only based on syntactic information have been criticized by, among others, Crain and Steedman (1985) as well as Altmann and Steedman (1988). They assume that some aspects of meaning and pragmatics can guide parsing in an – at least weakly – interactive manner. In their weakly interactive parallel model, all syntactically possible analyses at an ambiguous position in a sentence are carried out in parallel without restricting preferences occurring at the syntactic level. Pragmatic information is used immediately to choose the most plausible analysis. According to the principle of *Referential Support* (p1, Altmann & Steedman, 1988), potentially noun-modifying constituents following a definite NP should not be attached to the NP, if the simple NP already provides sufficient information to identify a unique referent in the discourse model.

(p1) The Principle of Referential Support

An NP analysis which is referentially supported will be favored over one that is not. (Altmann & Steedman, 1988, p. 201)

This hypothesis was substantiated by a series of experiments where structurally ambiguous sentences were presented in pragmatically biasing contexts. In an experiment reported in Altmann & Steedman (1988) sentences with a semantically disambiguated PP-attachment ambiguity (155a, b) were preceded by an NP-attachment biasing context (153) or a VP-attachment biasing context (154). In the NP-attachment biasing context, there were two potential referents for the object NP (*the safe*), so a modifier was needed to provide enough information to unambiguously identify a referent. In the VP-biasing context, on the other hand, only one potential referent was introduced, therefore the simple (non-modified) object NP sufficed to establish reference.

(153) NP-supporting context

A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was a safe with a new lock and a safe with an old lock.

(154) VP-supporting context

A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was a safe with a new lock and a strongbox with an old lock.

(155) a. NP-attached target

The burglar blew open the safe with the new lock and made off with the loot.

b. VP-attached target

The burglar blew open the safe with the dynamite and made off with the loot.

What do Crain and Steedman predict for isolated sentences, such as those investigated in the experiment presented in section 4.1.1? Since there was no explicit context in this experiment, no reference based preference can be derived automatically. But, according to Crain and Steedman (1985) there is no such thing as a null-context. Even isolated sentences carry their presuppositions as to pragmatic requirements which have to be fulfilled to render the respective sentence plausible. According to the principle of *Parsimony*, readers avoid interpretations with too many unsupported presuppositions.

(p2) The Principle of Parsimony

A reading which carries fewer unsupported presuppositions will be favored over one that carries more.

For modified NPs in isolated sentences, a set of entities has to be presupposed in the universe of discourse so that the modifier is needed to uniquely identify one of them. A simple NP does not carry any such presuppositions. With respect to PPattachment phenomena, the principle of *Parsimony* therefore predicts attachment to the VP irrespective of verb placement, when the sentences are read without a surrounding context. Although these predictions for sentences presented in the "null context" contradict the findings presented in section 4.1.1, it has yet to be shown how verb-second and verb-final constructions are processed when presented in referentially biasing contexts.

The question of whether or not contextual influences can override preferences seemingly determined by syntactic or lexical principles has been heavily disputed in the psycholinguistic literature. Crain and Steedman's (1985) evidence for the influence of pragmatically biasing contexts in sentential complement / relative clause ambiguities was soon argued to be irrelevant for initial parsing decisions because of the coarse experimental techniques used (sentence reading times and grammaticality judgements). In more sensitive self-paced reading and eye-tracking studies, Ferreira and Clifton (1986; Clifton & Ferreira, 1989) found a residual preference for the main verb ("*Minimal Attachment*") reading in main verb / reduced relative ambiguities even in contexts biasing the reduced relative reading.

Altmann & Steedman (1988) criticized Ferreira and Clifton's (1986) experiments for not including appropriate control conditions. In the experiment described above they claimed again that they had established an influence of context overriding any potential syntactic biases. Unfortunately, some inconsistencies in Altmann and Steedman's results as well as some confoundations with the ordering antecedents in the context (cf. Rayner et al., 1992; Britt, 1994) made these results look much less convincing. Furthermore, their findings could only be established in self-paced reading studies so far. Thus, the elimination of the residual effect might simply be due to decreased reading speed induced by this experimental technique. Altmann et al. (1992), however, argue that the residual effects for syntactic attachment preferences in eye-tracking experiments can only be found in a minority of cases in a regression contingent measure.

A delayed influence of pragmatic biases was established by Rayner et al. (1992) as well as Mitchell et al. (1992). Most convincing is Mitchell et al. 's (1992) experiment, where sentences were disambiguated very early, so that the experiment really tapped into immediate influences on parsing decisions. For sentences like (144a) described in section 4.2.1.2, Mitchell et al. found a strong contextual effect, but only in the segment following the *had been* segment. When the sentence was only disambiguated towards one of the readings in this segment (as it is the case when the pronoun *he* is included), subjects spent longer reading this segment if the disambiguation was supported by the context. However, the context was not able to effect the complement preference in the early region, indicating that contextual biases are only effective in a later stage of processing.

To account for the variation of results with respect to context effects, Britt, Gabrys, and Perfetti (1993; Britt, 1994) propose a somewhat different account on pragmatic effects on parsing decisions. In their *restrictive interactive model*, the influence of the *discourse model* is restricted to the attachment of optional verb-arguments. However, in the case that a PP fits a slot for an obligatory role of a verb, contextual information is ignored and the potential role-filler is initially attached to the VP by a subprocessor called *argument filler*. Low level phrasal units, such as PPs and NPs, are constructed autonomously by the so called *constituent builder*. This account is substantiated by self-paced reading studies, crossing referential biases with optionality of the ambiguous constituent.

Summing up, the picture arises that context cannot override strong lexical or syntactic attachment preferences as they are modeled by the *Principle of Preferred Role Attachment*. However, it still remains to be investigated whether in cases where a syntactic preference has been established in the absence of a strong lexical preference (i.e. in head final sentences) a principle like *Head Attachment* can be overridden by referential biases. In the following sections, I will discuss data from an eye-tracking experiment where PP-attachment ambiguities in verb-second and verb-final sentences were presented in pragmatically biasing contexts.

Materials and Design⁷³

The experimental sentences were manipulated according to a 2*2*2 within-subjects design (TABLE 13) with the factors *verb placement* (*verb-final* vs. *verb-second*), *semantic bias* (*VP-attachment biased* vs. *NP-attachment biased*) and *contextual bias* (*one possible referent* vs. *more than one possible referent*). The verbs used in the target sentences showed a preference to expect an instrumental PP, as established in pre-tests. For each experimental condition, one text was presented, resulting in eight texts per subject.⁷⁴ These eight texts were taken from a pool of 8 sets differing in content. The order of presentation was randomized. Every text was presented to an equal number of subjects. Additionally, 28 filler texts were included.

	context: R = 1		context: R > 1		
	semantic bias: VP-attachment	semantic bias: NP-attachment	semantic bias: VP-attachment	semantic bias: NP-attachment	
Verb-second	(156) (160a)	(158) (161a)	(156) (160b)	(158) (161b)	
Verb-final	(157) (160a)	(159) (161a)	(157) (160b)	(159) (161b)	

TABLE 13. Ex	perimental design.	The numbers in	parentheses refer	o the contexts
			r	

(156) Volker [bastelte [den Rahmen] [mit der Laubsäge]], ... Volker made the frame with the fretsaw, ...

- (157) Daß Volker [[den Rahmen] [mit der Laubsäge] bastelte], ...That Volker the frame with the fretsaw made, ..."That Volker made the frame with the fretsaw, ..."
- (158) Volker [verpackte [die Werkzeugkiste mit der Laubsäge]], ... Volker wrapped up the tool-box with the fretsaw.
- (159) Daß Volker [[die Werkzeugkiste mit der Laubsäge] verpackte], ...
 That Volker the tool-box with the fretsaw wrapped up, ...
 "That Volker wrapped up the tool-box with the fretsaw, ..."

In each target sentence, the PP *mit der Laubsäge* (*with the fretsaw*) can be attached syntactically to the direct object-NP as an attribute (NP-attachment), or to the VP specifying the instrument of the action (VP-attachment). However, sentences (156-159) are semantically biased (indicated by "[]" brackets) in such a way that world knowledge strongly biases the VP-attached interpretation in sentences (156, 157) and the noun-modifying interpretation in (158, 159). Similar to the experiment on PP-attachment in neutral contexts, the PP was not exchanged between the conditions in

^{73.}This experiment was run by Nicole Völker as part of her master's thesis supervised by Günter Kochendörfer, Gerhard Strube, and the author. The data are published in Konieczny, Hemforth, and Völker (1995).

^{74.}Admittedly, it would have been better to have more sentences per condition in particular for F2-analyses. Nevertheless, the data I will present cannot easily be attributed to single item artifacts because, for the most relevant conditions of pragmatic bias and verb-placement, the items differed only minimally between conditions.

this series in order to provide maximally parallel material, at least at the critical point of interest. Instead, the verb and the object were replaced so that the same PP can be interpreted as an argument of the verb, as in (156, 157), or as a NP-modifier, as in (158, 159).

Each target sentence was presented within a short text (160, 161), which introduced either a single referent (a: R=1), or two possible referents (b: R>1) of the direct object in the target sentence (e.g. one or two frames for targets for 156, 157 and one or two tool-boxes for 158, 159, respectively).

(160) Volker kümmerte sich um das Weihnachtspaket für seinen Neffen. Er sollte eine Werkzeugkiste bekommen, die eine Laubsäge enthielt. / Aber es mußten noch andere Weihnachtsvorbereitungen getroffen werden. /

(Volker took care of the Christmas package of his nephew. He was supposed to receive a tool-box which contained a fretsaw. / But other preparations for Christmas also had to be made.)

a. (R = 1): Volker fertigte das Weihnachtsgeschenk für seine Oma selbst. Sie sollte einen Bilderrahmen bekommen, der eine Einlegearbeit hatte./

(Volker made the Christmas present for his grandma himself. She was supposed to receive a picture-frame that had sections of inlaid work.)

b. (R > 1): Für seine beiden Omas fertigte Volker die Weihnachtsgeschenke selbst. Sie sollten beide jeweils einen Bilderrahmen bekommen, wovon der eine eine Einlegearbeit, der andere eine Blattgoldverzierung hatte. /

(For his two grandmas, Volker made the Christmas presents himself. Each of them was supposed to receive a picture-frame, one of which had sections of inlaid work, the other one was decorated with gold leaf.

Target:

Volker bastelte den Rahmen mit der Laubsäge, bevor er sie einpackte. (Volker made the frame with the fretsaw before he wrapped them up)/ (156) or Volker sorgte dafür, daß er den Rahmen mit der Laubsäge bastelte, bevor er sie einpackte. (Volker took care that he made the frame with the fretsaw before he wrapped it up /(157)

Er konnte seine eigene Laubsäge nicht finden.

(He could not find his own fretsaw.)

(161) Volker fertigte das Weihnachtsgeschenk f
ür seine Oma selbst. Sie sollte einen Bilderrahmen bekommen, der eine Einlegearbeit hatte./ Aber es mußten noch andere Weihnachtsvorbereitungen getroffen werden.

(Volker made the Christmas present for his grandma himself. She was supposed to receive a picture-frame that had sections of inlaid work./ But other preparations for Christmas also had to be made)

a. (R = 1): Volker kümmerte sich um das Weihnachtspaket für seinen Neffen. Er sollte eine Werkzeugkiste bekommen, die eine Laubsäge enthielt./

(Volker took care of the Christmas package of his nephew. He was supposed to receive a tool-box which contained a fretsaw.)

b. (R > 1): Volker kümmerte sich um die Weihnachtspakete für seine beiden Neffen. Jeder sollte eine Werkzeugkiste bekommen, die im einen Fall eine Laubsäge, im anderen Fall eine Kombizange enthielt./

(Volker took care of the Christmas package of his two nephews. Each of them was supposed to receive a tool-box, one of which contained a fretsaw, the other a pair of pliers.)

Target:

Volker verpackte die Werkzeugkiste mit der Laubsäge, bevor er zur Post ging. (Volker wrapped up the tool-box with the fretsaw before he went to the post-office) / (158) or: Volker sorgte dafür, daß er die Werkzeugkiste mit der Laubsäge verpackte, bevor er zur Post ging. (Volker made sure that he wrapped up the tool-box with the fretsaw, before he went to the post-office)/ (159)

Sie mußte unbedingt rechtzeitig ankommen./

(It absolutely had to arrive in time.)

In order to achieve minimal contrasts between the pragmatic bias conditions, the context was designed in such a way that both target sentences fit in easily. Since the targets contained different direct objects, both objects were introduced in both contexts (Thus, there was at least one frame and at least one tool-box in each of the contexts for the targets 156-159). The antecedent of the direct object was always introduced in the sentence immediately preceding the target sentence, in order to avoid a focus shift between antecedent and target (Rayner, Garrod, and Perfetti, 1992).

Whenever a sentence had to be presented in more than one line, the lines were separated by two empty lines. This was done in order to be able to count fixations that were only slightly below or above the material as fixations on the text.

Method

Procedure

The procedure was highly similar to that described for the experiments presented in the previous sections. The warming-up block consisted of two texts. Then the experiment, which was built up of six blocks, began. Each block was initiated by a brief calibration procedure and contained six randomly mixed texts. The text was presented in five segments, indicated by a slash ("/") in (160, 161). Each text was followed by a simple yes/no-question, which the subject was to answer by pressing one of two buttons (left-hand button: "yes", right-hand button: "no"). They answered with a high degree of accuracy, which did not vary across conditions.

Apparatus

The apparatus was identical to that used in the experiments presented in section 4.1.

Subjects

Thirty-two undergraduate students (native speakers of German) from the University of Freiburg were paid to participate in the study. All of them had normal, uncorrected vision and they were all naive concerning the purpose of the study. During an experimental session of 40 minutes, each of the subjects had to read 32 texts while their eye movements were monitored. Three subjects had to be excluded from the analyses because of inaccuracies or too many missing data.

Dependent Variables and Data Analyses

For the statistical analyses, the data were summarized for each word, except for the prepositional NP, which was treated as a single region, yielding two dependent variables, namely *first pass reading times*, and *regression path durations*, RPDs, (see chapter 3).

Exclusion of data

First pass reading times and regression path durations smaller than 100 ms were excluded from the analysis. If the sum of durations of erroneous fixations⁷⁵ in a sentence exceeded 200 ms, the data for this sentence were also excluded.

Accounting for word length effects

Although the PPs did not vary across the conditions, word length had to be accounted for for two reasons. Firstly, I was also interested in positions other than the PP, such as the verb, which do not contain the same word in each condition. Secondly, even at the PP, some data had to be occasionally excluded from the analysis (see para-

⁷⁵.i.e. a fixation not at a word in the target sentence but somewhere else on the screen

graph above). Word length was accounted for by *subtracting* from the *first pass reading times* 20 ms for each character, starting at the fourth character of each word.

Hypotheses

The locus of impact

Contextual bias. With respect to *referential support*, a preference for one of the alternative readings can be established immediately after the object-NP has been processed because a sufficient judgement can be made from the object-NP alone as to whether or not it successfully refers to a discourse referent. At this position, a preference for a modified NP must therefore have been established in multiple referent contexts, and a simple NP reading should be preferred in contexts introducing a unique antecedent.

Effects on reading times, however, may show up only when the referentially supported analysis fails for other reasons, such as its implausibility. It is therefore assumed that the earliest position at which the contextual bias may show its impact is a position where it conflicts with preferences imposed by other principles. In order to predict the position, I therefore have to identify the locus of impact of other constraining mechanisms, such as *Parametrized Head Attachment* and semantic bias⁷⁶.

Parametrized Head Attachment. The locus of impact of PHA differs between verbfinal and verb-second constructions. In verb-final constructions, *Head Attachment* forces the PP to be attached to the preceding NP, *as soon as a PP is started* (see Konieczny, Hemforth and Strube 1991). However, since short prepositions, such as "mit", are very often skipped, an effect might show up in the data only at the prepositional NP.

In verb-second constructions, however, the verb imposes its lexical expectations onto the parsing-process according to *Preferred Role Attachment*, as soon as *the verb* is read. Subcategorization requirements are thus directly projected into the structure.

Instrumental-PPs are regarded as optional complements. In the model presented here, they are predicted immediately when the corresponding subcategorizer, the verb, is processed, if it carries a lexically preferred expectation of the complement.

Semantic bias. It is obvious that the semantic bias cannot show its impact before the content of the PP is established. The earliest position possible is therefore the PP-noun, the last word of the PP in our sentences.

Interactions

Verb-second constructions. Since the verbs in this experiment have a preference to bind possible instruments, the verb-bias should interact with the contextual and the

^{76.}I will only discuss potential pragmatic influences on parsing decisions with respect to the predictions which can be derived from Parametrized Head Attachment in this section.

semantic bias, such that increased reading times can be expected when the context introduces two potential referents and thus supports the NP-modifying attachment of the PP, or the semantic bias renders the VP attachment implausible. This effect might already occur on the direct-object NP, if an optional instrumental PP has been predicted at the verb.

Verb-final constructions. If there is an initial analysis according to *Head Attachment*, reading times should be increased at the prepositional NP, when contextual and/or semantic bias force the simple NP-reading of the direct object.

An interaction of contextual and semantic bias can be expected, in such a way that an infelicitous combination, i.e. contextual and semantic bias forcing opposing interpretations, results in longer reading times than the felicitous combinations.

To sum up, it is most important that if the contextual bias alone guides the initial parsing preferences, no interaction with the placement of the verb should be expected, even more so because the contextual influence is established at the end of the direct object NP, while *Head Attachment* is supposed to be effective only at a later position in the sentence, namely, at the PP.

Results

All reading time measures were submitted to a full factorial 2*2*2 analysis of variance for repeated measures including the factors *verb placement*, *contextual bias*, and *semantic bias*.

Effects at the prepositional NP

TABLE 14 shows the mean adjusted RPDs and first pass reading times (in parentheses) at the prepositional NP (determiner and noun). I will go into details below.

Verb-placement. Not surprisingly, the PP was processed longer in verb-second sentences than in verb-final sentences (FPRT: $F1_{1, 28} = 13.36$, p < .01. $F2_{1, 7} = 15.22$, p <.01.; RPD: $F1_{1, 28} = 13.90$, p <.01. $F2_{1, 7} = 12.65$, p <.01.), which can be easily attributed to the clause-final wrap-up effect in verb-second sentences.

TABLE 14. Regression-path durations and first pass reading times (in parentheses) in ms at the prepositional NP by levels of *contextual bias* (R=1 vs. R>1), *semantic bias* (VP-attachment bias vs. NP-attachment bias), and verb placement (verb-second vs. verb-final)

	context: R = 1		context: R > 1	
	VP-attachment	NP-attachment	VP-attachment	NP-attachment
verb-second	332 (278)	333 (312)	365 (293)	528 (369)
verb-final	283 (236)	315 (223)	240 (205)	249 (212)

Of special interest, however, is the interaction of verb-placement with the other factors. With respect to these, the following picture arises:

Differing from previous studies on isolated sentences (see section 4.1.1), no reliable two way interaction of semantic bias and verb-placement was found at the prepositional NP (TABLE 15.).

TABLE 15. Regression-path durations and first pass reading times (in parentheses) in ms at the prepositional NP by levels of *semantic bias* (*VP-attachment biased* vs. *NP-attachment biased*), and *verb placement* (*verb-second* vs. *verb-final*).

	sem. bias: VP-attach	sem. bias: NP-attach
Verb-second	349 (285)	430 (341)
Verb-final	261 (221)	282 (217)

As can be seen in TABLE 16., however, there was a significant two-way interaction of *verb placement* and *contextual bias* with respect to RPDs.

TABLE 16. Regression-path durations and first pass reading times (in parentheses) in ms at the prepositional NP by levels of *contextual bias* (R=1 vs. R>1), and *verb placement* (*verb-second* vs. *verb-final*). The inferential statistics refer to the two-way interaction with respect to RPDs.

	context: R = 1	context: R > 1	subject analyses	item analyses
Verb-second	333 (295)	446 (331)	F1 _{1, 28} =8.88,p<.01	F2 _{1, 7} =21.35, p<.01
Verb-final	299 (229)	244 (209)		

Verb-second constructions. There was a reliable increase of the RPDs in the R>1 context, compared to the R=1 context within the verb-second condition (F1_{1, 28} = 6.99, p <.02; F2_{1, 7} = 37.17, p <.01). Although no reliable interaction of verb-placement and semantic bias could be established, first pass reading times were reliably increased (only for the F1 analysis, however: F1_{1, 28} = 4.42, p <.05), and RPDs marginally so (F1_{1, 28} = 3.46, p <.08), when a noun-modifying prepositional NP was read within a verb-second sentence (see TABLE 15.). These contrasts within the verb-second condition further confirm the hypothesis that processing difficulties result, when the R>1 context or the semantic bias supports the complex-NP reading, which had been ruled out by the role-preferences of the verb.

However, TABLE 14 shows that increased reading times in the NP-attachment biased condition are mainly due to the cases where context and semantic bias both contradict the initial analysis, namely VP-attachment. In these cases, processing seems to be severely hampered, leading to a significant reanalysis of large parts of the entire sentence, which show up in increased RPDs (R>1 vs. R=1 / NP-attachment biased: F1_{1, 28} =7.36, p<.02; F2_{1, 7} = 7.30, p<.04), resulting in a reliable two-way interaction between contextual bias and semantic bias for subject analysis (F1_{1, 28} =4.34, p<.05).

Verb-final constructions. There were no reliable effects with respect to either RPDs or *first pass reading times* in *verb-final* sentences. However, RPDs were increased in the R=1 condition (see TABLE 16; F1_{1, 28} = 2.79, p <.11; F2_{1, 7} = 3.49, p <.11). Though not reliable, this tendency is compatible with the assumption that NP-attachment was carried out first, according to *Head Attachment*, resulting in a conflict if the context supports the simple-NP reading (R=1).

Other positions

The words at positions other than the PP varied between the semantic conditions (see section 4.2.2.2). I will therefore only consider the effects of contextual bias and verb-placement at the noun of the direct object and only for verb-final constructions, at the verb.

Direct object NP. As can be seen in TABLE 14 there was a significant two-way interaction of *verb placement* and *contextual bias* with respect to RPDs, and a tendency for first pass reading times at the noun of the direct object.

TABLE 17. Mean adjusted *regression-path durations* and mean adjusted *first pass reading times* (in parentheses) in ms at the noun of the direct object by levels of *contextual bias* (R=1 vs. R>1), and *verb placement* (*verb-second* vs. *verb-final*). The inferential statistics refer to the two-way interaction with respect to RPDs (statistics for FPRTs are given in brackets).

	context: R = 1	context: R > 1	subject analyses	item analyses
Verb-second	204 (182)	280 (216)	F1 _{1, 28} =7.96, p<.01	F2 _{1, 7} =7.13, p<.04
Verb-final	231 (183)	233 (186)	(F1 _{1, 28} =2.68, p<.12)	$(F2_{1, 7} = 3.97, p < .09)$

It is obvious that this interaction results from the reliable contrasts within the verbsecond condition (RPDs: $F1_{1, 28}$ = 20.22, p <.01; $F2_{1, 7}$ = 8.24, p <.03; FPRTs: $F1_{1, 28}$ = 8.02, p <.01; $F2_{1, 7}$ = 9.44, p <.02). Thus, PHA and referential support show their impact as early as at the direct object in verb-second sentences. In verb-final sentences, however, the effect does not show up before the PP is processed.

The verb. It is hardly surprising that first pass reading times (F1_{1, 28} = 61.33, p <.01. F2_{1, 7} = 25.53, p <.01) and RPDs (F1_{1, 28} = 62.13, p <.001. F2_{1, 7} = 63.71, p <.001) were longer in verb-final sentences due to the clause wrap-up effect (see TABLE 18). Since the verb in verb-second sentences is of no further interest, I have only presented the results for verb-final sentences.

Although no reliable effects could be found at the verb in verb-final sentences, RPDS for R>1 condition were slightly increased (F1_{1, 28} = 2.87, p <.11), which results from the marginally reliable simple contrast (R=1 vs. R>1) within the verb-modifying condition.

TABLE 18. Mean adjusted *regression-path durations* and mean adjusted *first pass reading times* (in parentheses) in ms at the verb in verb final sentences by levels of *contextual bias* (R=1 vs. R>1), and *semantic bias* (*verbal instrument* vs. *noun modifier*). The inferential statistics refer to the simple contrast of contextual bias within the verb-instrument condition with respect to RPDs.

	context: R = 1	context: R > 1	subject analyses	item analyses
verb-instr.	247 (238)	337 (245)	F1 _{1, 28} =3.79, p<.07	F2 1, 7 = 2.13, p<.18
noun-mod.	272 (218)	278 (238)		

This result might look a little puzzling at first glance, since it appears to go against the findings at the PP-position. In the following section, however, I will argue that it fits perfectly to the processing model advanced in this thesis.

Discussion

It might seem surprising, at least at first glance, that the semantic bias of the sentences did not have an effect on processing times at the prepositional NP in verb-final sentences that were given in contexts, as it did in isolated sentences. However, several explanations can be considered: since the overall sentence reading times are smaller in contextually embedded sentences, time consuming high level semantic processes, such as evaluation of plausibility at the level of general world knowledge, might have occurred a few words later than in isolated sentences, especially since interfering contextual (referential) processes might have slowed down conceptual evaluation.

If we take another look at the examples, a further reason suggests itself: NPattachment violations, as in "*der Rahmen mit der Laubsäge*" (*the frame with the fretsaw*), might generally be "weaker" than VP-attachment violations, as in "*mit einer Laubsäge einpacken*" (*to wrap up with a fretsaw*). Nouns seem to impose much weaker constraints on their possible modifiers than verbs do. If so, semantic evaluation might have been delayed until the verb, at least in the case of verb-modifying semantic bias in the materials used in this experiment. Subjects might thus have behaved according to a "wait and see" strategy with respect to semantic evaluation in verb-final sentences.

Referential aspects, however, are inseparably tied to NPs, either simple or complex, and in cases of explicit anaphora, such as in our experiments, they do not require any inferences at a general conceptual level. I therefore assume that referential aspects are evaluated immediately and before other aspects of plausibility enter the game. Since these referential processes are supposed to be time consuming, to a small but definite extent, semantic effects might have been delayed even further, presumably beyond the subclause boundary. Unfortunately, the continuation of the sentence was not properly controlled, i.e, words after the comma differed across the conditions. Thus, I do not have access to the processing times beyond the subclause boundaries, where the semantics might have shown its effect.

Parametrized head attachment. If PHA, and in particular *Head Attachment* and *Preferred Role Attachment*, had not guided the first analysis, the interaction of contextual bias and verb-placement would be hard to explain. However, the impact of preferred role attachment in verb-second sentences was much stronger than that of head attachment in verb-final sentences, which was only marginally reliable. A quite similar interaction was also found in verb-second sentences even at the noun of the direct object. If we assume that the beginning of the PP ("*mit*") had been pre-processed within the perceptual span at the object noun, referential failure of the object in the R>1 context might have produced a conflict with the expectation of an instrumental PP imposed by the verb in verb-second sentences. In verb-final sentences, however, pre-processing is probably not sufficient as a trigger for *Head Attachment*, where no such expectation of a PP is effective⁷⁷. Interestingly then, contextual influences could not hinder the parser from adopting an analysis preferred by *Head Attachment*.

^{77.}As will be outlined in chapter 5.2, the preposition actually has to resort to *active attachment* in verb-final cases, as opposed to *predicted attachment* in verb-second sentences.

In verb-second sentences, increased reading times resulted at the object noun and during the PP when the contextual bias contradicted a preference imposed by *Pre-ferred Role Attachment*⁷⁸. At the PP, the highest penalty was observed when both contextual and semantic bias contradicted preferred role attachment. The highly increased regression path durations in these cases indicate an extensive reanalysis of the entire sentence.

In verb-final sentences, processing times were increased only at the PP if the contextual bias (R=1) contradicted the obvious head attachment of the PP to the preceding NP. Processing difficulties, however, were weaker than in verb-second sentences and only marginally reliable.

There are three possible reasons for this difference. Firstly, it might be due to the ease of recovery from an initial NP-attachment of the PP as a *modifier*, compared with recovering from the initial VP-attachment as a *complement*. Secondly, in verb-final constructions, a weak penalty might have resulted from a referentially *successful* object-NP (biasing the parser towards the non-preferred VP-attachment), whereas a referential *failure* in verb-second sentences (biasing the parser towards the non-preferred NP-attachment) might have caused a stronger penalty. Whereas the overall interpretation can sufficiently be established in the case of referential failure.

Note also that a NP-modifier, such as a PP, can very easily be re-interpreted as nonrestrictive while it remains attached to the NP. In the case of an R=1 context in verbfinal sentences it may have occurred that the PP was not reanalyzed as a complement of the verb at all, but re-interpreted as non-restrictive, still remaining a NP-modifier. This is likely to cause a much weaker penalty than in the case of a conflict in verb-second sentences, where the PP must be reanalyzed from a verb-complement to a restrictive modifier and thus be attached to a different host.

This interpretation is supported by the result that the highest processing times at the verb in verb-final sentences occurred, when the verb-modifying semantic bias forced the initially *head attached* PP, which was furthermore referentially supported by an R>1 context, to be reanalyzed as a complement of the verb, whereas reversed, the noun-modifying semantic bias does not seem able to initialize a reanalysis when it should have conflicted with the R=1 context. So, again, it seems to be the case that the PP has only been interpreted as non-restrictive in the R=1 context, still remaining NP-modifying.

Implications for parsing models

In general, the data are incompatible with all models that do not predict differential parsing preferences for verb-placement variations.

^{78.}The preposition was excluded from the data analysis, since it was skipped in about 80% of the cases, which further supports the assumption, that it had often been pre-processed at the object-noun.

The weakly interactive, parallel model of Altmann and Steedman (1988) does not assume initial *syntactic* preferences. In their model, the PP should have been attached according to referential support exclusively. Thus, the only type of interaction that should have occurred is one between contextual and semantic bias, in both verb-final sentences and verb-second sentences. In verb-second sentences, however, the most difficult condition is a "felicitous" one (Altmann, 1988), in which the context (R>1) and semantic bias (NP-modifying) both support the same attachment. Even more importantly, the interaction of contextual bias and verb-placement remains unexplained.

In Britt's model, it is not clear whether or not the discourse model is consulted to attach the PP in verb-final sentences. If it is, predictions are the same as those of the Altmann & Steedman model and thus, false. If, however, the discourse model requires a call from the argument filler it is not consulted, since no verb has been read up to the PP, which could have activated the argument filler. Thus, only the constituent builder could carry out an attachment, i.e. an attachment to the preceding NP. In this case, the encountered preference towards the [NP PP] attachment would be explained in much the same way as it is by any other head-licensing model, such as Abney's licensing structure parser. I have argued in section 4.1.2 that such models are inadequate, in that they require a licensing head in order to build structure at all. With respect to the verb-second sentences in our study, the model predicts an interaction of contextual and semantic bias at the PP, as the Altmann and Steedman model does, because the instrumental PPs were all optional arguments. Since the argument filler tries to match a preposition only against the obligatory slots of the verb, no attempt is made to attach the PP to the VP initially, so that it had to be cut off later. Therefore, the initial attachment only depends on the discourse model and can later conflict with the semantic bias. Again, this prediction contradicts the result that a "felicitous" condition (R>1/noun-modifying) was the hardest to process. This can only be explained by a strong initial attachment to the VP, before contextual and semantic bias enter the game.

Optional verb-complements, such as the "*mit*"-PPs in this experiment, are thus initially attached to the VP if the verb that bears the corresponding lexical preference has already been processed. It is therefore insufficient to distinguish only between optional complements and obligatory complements, as the restricted interactive model of Britt et al. (1993) does. It remains to be seen, however, whether the results of this research, especially the interaction of contextual bias and verb-placement, could be replicated if verbs without a preference for an instrumental PP were used. In any case, however, it could be shown that even optional complements are initially attached to their subcategorizer. The PHA model is therefore the only model under investigation here that fits the data presented and that meets most of the known empirical challenges.

4.2.2.2 Conclusion

Parametrized Head Attachment was demonstrated to be effective even when the target sentences were presented in contexts. The results presented in this section strongly suggest that the syntax processor operates independent of higher level influences. Nevertheless, contextual influences were demonstrated to be effective fairly early, and they seem to terminate before the conceptual evaluation of the proposed attachment. Therefore, a feedback channel between the syntactic processor and the higher level system must be assumed.

The architecture that emerges from this is illustrated in (162).



In a restricted sense, the syntax processor operates as a *module* (Fodor, 1983) in that its initial operations cannot be influenced by processes of modules following it. Nevertheless, there is at least some feedback from higher level systems that can stop the syntax module continuing an implausible analysis. Whether or not the higher level processes can *suggest* alternative analyses in case of a parsing failure, as proposed for the *thematic processor* (Frazier, 1987), has not yet been decided.

4.3 General discussion

From the evidence discussed in this chapter it must be concluded that the human sentence processing mechanism

- is highly *incremental*, i.e. it attaches each incoming item as it is encountered (Frazier, 1987a), *"linear parsing"* (Konieczny & Strube, 1995), (as opposed to *head-driven* or *head licensing* parsing), employing both bottom-up and top-down strategies,
- operates in a *serial* fashion, i. e. it pursues only one analysis at a time in the case of an attachment ambiguity (as opposed to *parallel* processing of multiple analyses)
- *commits* itself to a structural alternative *immediately* (as opposed to minimal commitment or to a "wait and see" strategy),
- uses *detailed lexical information*, including lexical subcategorization preferences, during the *initial stage* of structure assembly,
- resolves structural ambiguities according to *Parametrized Head Attachment* (Konieczny et al., 1994), i.e. depending on the presence or absence of the lexical head of the attachment sites, their lexical properties, and their respective distance to the item to be attached, and
- operates *autonomously* in its initial attachment decisions, i.e. higher level semantic and pragmatic processes are not capable of *directing* the initial attachment choice (as opposed to *strongly interactive* or *integrative* multiple constraint models).

5 The SOUL system⁷⁹

The main goal of this chapter is to present a model that satisfies the constraints posed by the psycholinguistic evidence elaborated in the preceding chapters. The model will be called the Semantics-Oriented Unification-based Language system (henceforth "SOUL"). As an implemented model, it is based on a modified Head-driven Phrase Structure Grammar (HPSG, Pollard & Sag, 1987, 1994), a modern, highly lexicalized grammar framework, currently predominant in computational linguistics. Before I can outline the core mechanism of the SOUL model, I will have to give a brief introduction to those aspects of HPSG that are relevant in the context of this thesis and present (and motivate) some necessary modifications and extensions. The overview will surely be incomplete, and the style of description as informal as possible. Since the purpose of this section will be to give the reader not familiar with HPSG at least an idea of the basic concepts, I will not provide the linguistic motivations for each of these concepts. Nevertheless, I expect the reader to be familiar at least with the basic concepts of Government and Binding Theory (Chomsky, 1981, 1986), as the most convenient way to introduce certain concepts of HPSG is to relate them to similar (or different) ones in GB.

Having described the linguistic knowledge then, I can then begin to introduce the SOUL-mechanism, which operates on the representations provided by HPSG. The crucial aspects of the SOUL mechanism will be illustrated in greater detail with some parsing examples. Finally, I will conclude with a discussion of the commonalities and differences of the SOUL mechanism with recent models of human sentence processing.

^{79.}The SOUL system, consisting of the feature-unification formalism SOUL-FR, an HPSGstyle grammar for a small subset of German, and the parsing component SOUL-P, was implemented in the Objective-C programming language and runs on computers with NEXTSTEP 3.2 or higher. The implementation details will not be presented here, since this thesis focusses on algorithmic and representational level aspects of the SOUL-mechanism and their implications and predictions of parsing phenomena. For more details see Konieczny (in prep).

5.1 The competence base

Whereas *Parametrized Head Attachment* (see chapter 2.16), as a special purpose principle, has been formulated in a way that keeps it independent of any particular grammar framework, a *general* parser must operate on linguistic representations. If the focus is on *processing* issues, however, one could take the quite reasonable approach of assuming grammatical representations as general as possible and avoiding commitments to an elaborated linguistic grammar theory. On the other hand, relying on a *naive* grammar, such as a simplified set of phrase structure rules, bears the danger that it might turn out to be impossible to extend the grammar in a linguistically adequate way.

From the perspective of grammar theoreticians, performance data might provide valuable evidence about certain declarative assumptions, if competence and performance issues are closely linked. Only if a cognitive parser is based on an elaborate linguistic theory providing adequate representations of linguistic knowledge with a high degree of explanatory power, matches and mismatches of procedural implications of declarative knowledge and psycholinguistic constraints on natural language processing can be identified. In HPSG, for instance, the high degree of lexicalization of grammatical knowledge, i. e. the head-drivenness, will turn out as both advantageous and disadvantageous for modeling human sentence processing. Some problems will be shown to be only superficial, with solutions directly emerging from HPSG-internal facilities currently not available in other approaches.

5.1.1 Head-driven Phrase Structure Grammar (HPSG, Pollard & Sag, 1987, 1994)

The most important aspect of the HPSG way of representing linguistic knowledge is that *every* kind of linguistic object, such as the major category, subcategorization requirements, constituent structure (daughters), semantic content, etc., is modelled in virtually the same way, namely, with *feature structures* or *feature structure descriptions*⁸⁰. The uniform representation provides a multi-dimensional integrated representation of information, which in other contemporary linguistic frameworks requires multiple derivationally dependent layers (*strata*) of (constituent) structure. Government and Binding Theory (GB, Chomsky, 1981, 1986), for instance, distinguishes two levels (*strata*) of syntactic structure, s- and d-structure. Here, the logical and phonological form require s-structure representations from which they have to be derived. Standard GB is a *derivational* theory, in that certain derivational operations (move-alpha) are applied to structural representations to derive a different one. In doing so, the initial structure must often be modified or destroyed (Kiss, 1995).

⁸⁰ I will exclusively use the term *feature structure* throughout the thesis even in cases where others would talk about *descriptions* of feature structures. The term *description*, however, is not used in a consistent way in the literature: some (e. g. Kiss, 1995) use it to distinguish *complete* models of linguistic entities (feature structures) from only partial models or constraints on such entities (descriptions), others (e. g. Pollard and Sag, 1994) use it to distinguish the mathematical object (the graph or feature structure) from its descriptions, such as a set of first order predicate calculus terms, or *attribute-value matrixes*.

In HPSG, only one level (*stratum*) of syntactic structure is assumed. The uniform feature representation, however, is a powerful mechanism suited for expressing dependencies between distinct linguistic *levels* (dimensions), such as the syntactic, semantic or pragmatic features of a *linguistic sign* within one and the same feature structure. Ideally, no structure will have to be destroyed or modified during language processing. Instead, the feature structure representing the utterance will become increasingly (monotonically) specified at each level of representation simultaneously. The *integrated* theory of syntax and semantics distinguishes HPSG from derivational approaches such as GB.

This aspect is associated with the fact that HPSG is a strictly *declarative* grammar theory, meaning that the order in which the constraints are applied is completely irrelevant to the outcome. As a declarative grammar, the representations provided by the grammar theory are good candidates for the linguistic knowledge needed in both *language perception* and *language production*.

Both GB and HPSG are principle-based, and both are similar with respect to the high degree of lexicalization of grammatical information. Note, however, that both GB and HPSG grammars are not completely lexicalized, such as lexicalized TAGs (Joshi, 1985) or several variants of *Categorial Grammar* (e.g., Ades & Steedman, 1982), which leave no information (only certain *operations*) outside the lexicon. In HPSG, there is still a certain but small number of *constraints* independent of the lexicon, namely some principles and schemata which constrain the way in which phrase structures can be built.

With respect to the constituent structures assumed, no "extended projections" (Chomsky, 1991, Grimshaw, 1990), such as the *complementizer* (C)-phrase or the *Infl*-phrase, need to be proposed, again due to the richness of information inherent in the representation of linguistic objects themselves, i.e. due to the integrated multi-level (syntax, semantics, etc.) representation of signs in HPSG. The *constituent* structures can thus be kept extremely parsimonious, compared to those in GB. In summing up, these properties, and in particular its declarativeness, let HPSG appear fairly well suited for (cognitive) parsing.

Before I discuss the most important theoretical concepts in HPSG, I will have to sketch the basics of unification-based formalisms needed for the comprehension of theoretical concepts.

The description of the formalism is just intended to give the reader not familiar with HPSG an idea of what the basic formal aspects of HPSG are. The formalism will only be discussed very briefly. For a more detailed introduction, the reader is referred to Pollard and Sag (1987, chapter 2; 1994, chapter 1), Shieber (1986). For German introductions into formal aspects of HPSG, see Kiss (1995, chapter 1 and 2), and Müller (in prep., chapter 1).

5.1.1.1 The formalism

Mathematically, feature structures are graphs (more specifically: directed acyclic graphs, DAGs) consisting of nodes and arcs, both of which are labeled. Consider the following graph.

(163)



The graph in (163) models an object of the type *sign*, noticeable by the label of the root node. A *sign* has two *features* (arcs): PHON and SYNSEM⁸¹. PHON represents the phonological string of the utterance, abbreviated by the *list* of words henceforth, such as *<Mary lies>*.

Feature structures in HPSG are *sorted*: each label assigned to a node represents a particular *sort. Sorted feature structures* represent a partial description of a linguistically relevant generalized piece of information. The value of SYNSEM is restricted to the sort *synsem*. In (163), *synsem* is an atomic feature structure. In a complete model of the sign described, *synsem* represents a feature structure of its own; i. e. the node is the root node of another feature structure. Actually, the information in the *synsem* feature structure represents the "flesh" of a sign, namely, its syntactic features (category, subcategorization, etc.), as well as the semantic content and certain contextual features. This will be further elaborated after the basics of the formalism have been made clear.

Feature structures are usually described with *attribute-value matrices* (AVM), such as (164).

(164)

[sign	-
PHON	list
SYNSEM	synsem

The superscript in the upper left corner of the matrix indicates the *sort* of the feature structure. The attribute value pairs (PHON *list*) and (SYNSEM *synsem*) in (164) correspond to the respective arcs in feature structure (163). Next, consider matrix (165).

(165)

phrase	
PHON	non-empty-list
SYNSEM	synsem
DTRS	const-structure

⁸¹.I will ignore the QSTORE feature here, since it is irrelevant to this thesis.

Compared to (164), another attribute value pair has been added (DTRS *constituent-structure*), and the value of PHON has been specified to *non-empty-list*. (The value of the feature DTRS (daughters) will represent the constituent structure of the phrase.) The information in (165) is more specific than that in (164). Thus (164) *subsumes* (165). Note that the sorts of the respective feature structures also differ. *Phrase* in (165) is a *subsort* of *sign* in (164), as it is more specific than its supersort. In most HPSG grammar systems, such as SOUL, subsort relationships are explicitly coded in the grammar, so that a subsort can inherit the information from its supersort. Only more specific information has to be added in the subsort definition. Sorts can thus be hierarchically organized. A feature structure of the sort *sign*, for instance, can either be a *phrase* or a *word*.

(166) The *sign* hierarchy



Both a *word* and a *phrase* will thus inherit the information given in (166), but only a *phrase* will permit (and require) a constituent structure. For each sort, only certain attribute-value pairs are defined as *appropriate*. The appropriateness conditions on feature structures have two functions: on the one hand, they rule out the combination (unification) of feature structures that would entail inappropriate information and on the other hand, the appropriate definitions can be utilized in the search for a *complete* model, i.e. one that contains all the information that it can carry. Roughly speaking, language processing in this paradigm amounts to finding *complete* models (feature structures) of linguistic objects. I will return to this issue in section 5.1.1.7.

Sorts are therefore a powerful device that distinguishes HPSG from many other grammar frameworks. Valence-classes, for example, can be accounted for by a hierarchical system of sorts, which represent the respective feature structure of subcategorization requirements of *intransitive, transitive, strictly-transitive, ditransitive* verbs, etc.

Value sharing

Importantly, two features (arcs) can *share* the same value. They share the same value, if they point to the same node in the feature structure. Their value is thus *token*-identical, as opposed to *type*-identical, the latter meaning that although the two values appear to be the same, they might actually be different nodes in the feature structure. In attribute value matrices, value sharing is expressed by a numbered *tag*. In (167), for example, the tag [1] indicates that the (yet unspecified) value of the two HEAD-features in the phrase are identical.

(167)

phrase	
SYNSEM LOC CAT HEAD 1	
DTRS HEAD-DTR SYNSEM LOC CAT HEAD	1

Note that the "path" descriptions (e.g. SYNSEM | LOC | CAT | HEAD) in (167) is just an abbreviation for embedded feature structures each bearing only a single feature. Actually, (167) describes the feature structure given in (168)



Value sharing poses special constraints on the feature structure. Two feature structures can be unified, resulting in a feature structure containing all the information that has been specified separately in both. The resulting feature structure is thus subsumed by both initial structures, and it is the most general structure subsumed by these. Suppose structure (167) is *unified* with structure (169)

(169)

 $\begin{bmatrix} phrase \\ PHON < runs > \\ SYNSEM | LOC | CAT | HEAD \begin{bmatrix} verb \\ VFORM & finite \\ AUX & minus \end{bmatrix}$ $DTRS | HEAD-DTR | SYNSEM | LOC | CAT | HEAD ' \begin{bmatrix} verb \\ INV & plus \end{bmatrix}$

The resulting feature structure is given in (170).

(170)



Note that as the result of the value sharing in (167), the two initially separated HEAD values in (169) have been unified into a single feature structure in (170).

Value sharing is one of the most important features of the HPSG formalism. Many of the constraints expressed in the principles impose value sharing, mostly between certain features in distinct constituents of a phrase structure. The closest equivalent in standard GB is *move-alpha*, which covers comparable phenomena. However, *move-alpha* is usually thought of as an operation that only applies to constituents, whereas in HPSG *any* pair of features in the representation of a sign might share their values

(provided the value is appropriate for both), not just objects of the constituent structure.

Semantically, a feature structure *denotes* the set of entities that are described by the given features and their values. (170), for instance, denotes the set of objects of sort *sign*, whose SYNSEM features are restricted to *synsem* and whose PHON features are restricted to *list* (of course, this is a fairly underspecified sign). The less information a feature structure carries, the bigger the set of objects that it denotes. The unification of two feature structures denotes the intersection of the sets of objects denoted by the initial feature structures.

5.1.1.2 What do HPSG models represent?

HPSG is a system of linguistic *signs* (in roughly the Saussurean sense). As we have already seen, signs can be either phrases or words. Phrases have three root features, PHON, SYNSEM, and DTRS⁸². Metaphorically speaking, the feature DTRS (daughters) represents the phrase structure "skeleton", PHON the surface or the "skin" of a sign, and SYNSEM the "flesh". Especially for the latter feature, SYNSEM (SYNtax, SEMantics), there is no apparent equivalent in GB, since almost everything is expressed by virtue of the "skeleton", i. e. by (several *strata* of) constituent structure.

As they are retrieved from the lexicon, words carry the information that determines which kind of constituent structures can be built with the word as the *head* of the phrase. Consider (171), for example, representing the verb *eats*.



In the CATEGORY (CAT) feature, both the major category (*verb*) and the subcategory are specified. SUBCAT represents a list (< ... >) with two elements, a nominative (*nom*) NP, and an accusative (*acc*) NP. The SUBCAT list represents the *subject* (the first NP) and further *complements* of a head. The existence of an accusative NP indicates that *eats* is a *transitive* verb. The SUBCAT list is ordered such that it represents the *obliqueness* or *accessibility hierarchy* (Keenan and Comrie, 1977; cf. Pollard and Sag, 1987) of complements. In English, the order corresponds to the focus-neutral order in which the complements are realized.

^{82.}In standard HPSG, two further features are assumed: QSTORE (quantifier store) and RE-TRIEVED. I will ignore them, since they are completely irrelevant in the context of this thesis.

Lists (< ... >) can be understood as feature structures consisting of two features, e.g. FIRST and REST. While FIRST in the SUBCAT list of (171) takes a $sign^{83}$, REST takes just another *list* value, which can thus be extended recursively. As feature structures, *lists* are also *sorted*, meaning that linguistically relevant sorts of lists can be organized in a sort-hierarchy. The subcat-list hierarchy, for instance, would start with an unspecified sort that subsumes both *intransitive* (a one-element list containing only the subject) and *transitive* (a list with *at least* two elements, with its innermost REST feature unspecified). *Transitive* would subsume *strictly-transitive*, a list containing exactly two elements, *di-transitive*, and so on. Once defined, these sorts, like any other sort, can be referenced in the lexical entries. This particular sort-hierarchy will also play a special role in parsing, as described in section 5.2.

The CONTENT in (171) represents an entity that has been called *parametrized state* of affairs (psoa) or infon in Situation Semantics (Barwise & Perry, 1981, Devlin, 1991). It consists of a RELATION value (eat, in this case) and the arguments that this relation is supposed to take. In the approach taken here, the argument labels correspond to standard *thematic* roles (e.g. Grimshaw 1990). The thematic arguments in (171), AGENT and THEME, share their values with the elements on the SUBCAT list (note that they are not *identical* with the elements). The subscript of the tags in the subcat-list indicates that the tags represent the INDEX of the respective NPs, as described below. Note that the elements on the subcat list in (171) are abbreviations for the feature structures, whose LOCAL value is given in (172) and (173).



In standard HPSG, *person*, *numerus* and *gender* are part of the CONTENT, whereas case is assumed to be a syntactic (CAT) feature. At least for German, this is rather

^{83.}Actually, the elements on the SUBCAT list in standard HPSG (Pollard and Sag, 1994) represent only the SYNSEM value of signs. I will ignore this for reasons of simplicity.

unfortunate. Since *case* and the other *agreement* features co-vary in complex declination schemes, I assume the agreement features to be bundled in the AGR feature, replacing CASE as a head feature, as in (174).

(174)

cat	_		_
	agr CASE	case	
HEAD AGR	INDEX	index PER NUM GEN	per num gen

The INDEX value is then shared with the index value in the CONTENT. This approach allows a better organization of agreement information in a hierarchy of *agr*-sorts.

As illustrated in (172) and (173), the content of nouns bears the attribute RESTRIC-TIONS, whose value is a set of *psoas*. For referential NPs, each psoa in the RESTRIC-TION value places semantic conditions on the entities the indices can be anchored to in a given context (cf. Pollard and Sag, 1994, p. 26). For the noun *car*, for instance, the RESTRICTION value is a set containing the psoa given in (175), with the INSTANCE unified with the INDEX.

(175)

psoa	1
RELATION	car
INSTANCE	1

So far, I have provided some very basic insights into the organization of linguistic information in HPSG. Signs carry both syntactic and semantic features, and both interact by virtue of value sharing. The next thing to show is how lexical information interacts with extra-lexical information to constrain the way in which phrase structures are built.

5.1.1.3 Phrase structure: the "skeleton"

The DTRS (daughters) feature bears the immediate phrase structure daughters of phrases. HPSG distinguishes five kinds of daughters, each represented by a feature of its own: head-daughters (HEAD-DTR), complement-daughters (COMP-DTRS), adjunct-daughters (ADJ-DTR), filler-daughters (FILLER-DTR), and marker-daughters (MARKER-DTR). This is another aspect in which HPSG differs from GB, since GB does not distinguish different kinds of constituents independent of their position in the x-bar scheme. The qualitative labeling of daughters in HPSG, however, allows for the formulation of highly abstracted constraints over *single* level phrases, namely, the principles and schemata of HPSG.

Principles and schemata

Apart from lexical information, only a few very general constraints on the wellformedness of signs remain, namely, rule-schemata which are unspecified for *synsem* information and a number of principles of well-formedness further constraining the way in which signs can be combined (see Pollard & Sag, 1994, for a detailed introduction to HPSG). For the present purposes I do not need to present all principles in detail (i.e. the *head-feature principle*, the *subcategorization principle*, the *ID principle*, the *marking principle*, the *SPEC principle*, the *nonlocal feature principle*, the *trace principle*, the *subject condition*, the *weak coordination principle*, the *singleton REL constraint*, etc.). Instead, I will illustrate just a selection of the most important principles to give an idea of how principles can constrain the way structures may look.

The head-feature principle (HFP)

Principles take the form of an implication. The *head feature principle* (176) makes sure that phrases are projections of their heads: if a sign is a headed phrase, the HEAD values of the mother and head daughter have to be token-identical.

(176) Head feature principle

$\begin{bmatrix} phrase \\ DTRS & headed-structure \end{bmatrix} \implies$	
phraseSYNSEM LOC CAT HEADDTRS HEAD-DTR SYNSEM LOC CAT HEAD	1

In general, implications ("=>") are interpreted as follows: if a feature structure is subsumed by the condition (left hand side), it must also be subsumed by the conclusion (right hand side).

The subcategorization principle

Roughly speaking, the subcategorization principle (177) causes the complements of a head to be realized as constituents. Simultaneously, it constrains the way this can be accomplished. The subcategorization principle says that in phrases with a HEAD-DTR and COMP-DTRS, the SUBCAT list of the head-daughter is the concatenation⁸⁴ of the SUBCAT list of the mother with the list of complement daughters.

(177) Subcategorization principle

 $\begin{bmatrix} phrase \\ DTRS & headed-structure \end{bmatrix} \implies$ $\begin{bmatrix} phrase \\ SYNSEM | LOC | CAT | SUBCAT & 1 \\ \\ DTRS & \begin{bmatrix} head-comp-struc \\ HEAD-DTR | SYNSEM | LOC | CAT | SUBCAT & append(1, 2) \\ \\ COMP-DTRS & 2 \end{bmatrix}$

^{84.}List concatenation is expressed by the two-place relation *append*. In general, *relational dependencies* can be understood as a *dynamic* extension of the concept of value sharing (cf. Kiss, 1995): the value of the relation depends upon its arguments, which can change over time.

Note that the subcategorization principle does not constrain the number of the complements realized at any level of the phrase structure. This will be achieved by certain phrase structure schemata, which I will turn to now before continuing to describe further principles.

The phrase structure schemata

Immediate dominance schemata (ID-schemata) represent general constraints that phrases must satisfy in order to be well-formed. Basically, ID-schemata are under-specified descriptions of simple phrases bearing a single level of phrase structure. ID-schemata do not (generally) impose value sharing between constituents. Nevertheless, they constrain the form a phrase can take. Consider *schema1* (178).

(178) Schema1 (head-subject schema)

 phrase

 SYNSEM | LOC | CAT | SUBCAT <>

 DTRS
 head-comp-struc

 HEAD-DTR
 phrase

 COMP-DTRS
 < [] >

Schema1 describes a *saturated phrase* (i.e. a phrase with an empty subcat list "<>") whose head-daughter is a phrase and whose list of complement-daughters contains exactly one element. *Schema2* (179) describes an *almost* saturated phrase (its subcat list has exactly one element), whose DTRS value is an object of the sort *head-comp-struc* and whose head-daughter is a word.

(179) Schema2 (head-complement schema)

```
      phrase

      SYNSEM | LOC | CAT | SUBCAT < []>

      DTRS
      head-comp-struc

      HEAD-DTR
      word
```

Note that if *schema2* is unified with the *subcategorization principle* (177), all elements except the first have to be realized as complement daughters, since *schema2* restricts its subcat-list to contain exactly one element. Taken together then, *schema1* and *schema2* roughly correspond to the *X-bar* scheme in GB: the single complement daughter in *schema1* corresponds to the *specifier* position, while the complement-daughters in *schema2* correspond to the complement position(s), assuming a non-binary branching structure, though.

There is a third schema (180) for head-complement structures in HPSG, which describes a saturated phrase whose head daughter is a word. Unified with the subcategorization principle, the entire subcategorization list must be realized as the comple-

ment daughters (since the phrase is saturated). Schema3 thus constrains phrasestructures to be *flat*.

(180) Schema3 (head-subject-complement schema)

```
phraseSYNSEM | LOC | CAT | SUBCATDTRShead-comp-strucHEAD-DTRword
```

This is the scheme from which complements will usually be *topicalized*, which will be described later.

Besides the schemes for *head-comp-structures*, i. e. phrases with a head daughter and complement daughters, there are further schemes constraining phrases with different daughters. The *head-adjunct schema* (or short: *adjunct schema*) describes a phrase with a head-daughter and an adjunct-daughter. Adjuncts are assumed to modify their heads. For this purpose, they bear the HEAD-feature MOD (MODIFIED), which represents the expected head whose content can thus be accessed via value sharing. The adjunct scheme (181), then, establishes the connection between adjunct and head in that it unifies the MOD value of the adjunct with the SYNSEM value of the headdaughter.

(181) Adjunct Schema (simplified)

phrase		
DTRS	head-adj-struc HEAD-DTR SYNSEM] ADJ-DTR SYNSEM LOC CAT HEAD MOD	1

There are two more schemata, the *filler schema* and the *marker schema*, which need not be described in detail here. In summing up, there is a collection of six schemata, which constrain the form of phrases in that "every (headed) phrase must satisfy exactly one of the ID-schemata" (Pollard and Sag, 1994, p. 38), as expressed in the ID-principle. More formally, the ID-principle amounts to the implication given in (182), meaning that if an object is of the sort *phrase*, it must be unified with the *disjunction*⁸⁵ ("v" symbolizes the logical *or*) of ID-schemata on the right-hand side.

(182) ID-principle [phrase] => [schema1] ∨ [schema2] ∨ [schema3] ∨ [adjunct-schema] ∨ ...

5.1.1.4 Head projection

To illustrate the interaction of lexical information and the HPSG-principles including the ID-schemata, suppose the feature structure representing the verb *eats* (171) is unified with the head-daughter in *schema2* (179). As a headed-phrase, the resulting

^{85.}In the SOUL-system, principles like these are realized through a sort-hierarchy with (headed) *phrase* being the root sort and each schema being a subsort.

sign has to be unified with the *head-feature principle* (HFP) and the *subcategorization principle* (and all the others, of course). The HFP (176) percolates the HEAD features to the mother, thus causing the resulting scheme to be a VP. Note that *schema2* describes a phrase whose SUBCAT element is restricted to exactly one element. After unification with the *subcategorization principle*, only the first element on the SUBCAT list of the head daughter is thus passed to the mother, while the remaining complements are realized as complement-daughters.

As an *almost* saturated phrase (i.e. with a SUBCAT list containing exactly one element), *schema2* fits in *schema1* (178) as a head-daughter. The HFP causes the phrase to become a verbal-projection again. Since *schema1* is saturated, the *subcategorization principle* realizes the element on the subcat-list of the head-daughter as the subject daughter.

The resulting sign is a sentence-fragment, as illustrated in (183). Note that the signs in the tree are abbreviated to their category value.



Summing up, *schema1* and *schema2* compare to levels in the *X-bar* scheme (Jackendoff, 1977) known from GB. *Schema1* corresponds to the *X-max* level, and *schema2* to the *X'* level. Note, however, that the *X-bar* level is not represented explicitly, but implicitly through the number of elements on the SUBCAT-list. Note also that *schema2* in not necessarily binary-branching, as all complements of the lexical head are attached below it.

5.1.1.5 Content

The content of a sign is passed to the mother in a similar way as the HEAD-features are: the CONTENT value of the mother has to be the CONTENT value of the head-daughter. There is one exception, though. Adjuncts are assumed to be the functors that take their heads as arguments. In adjunct-schemata, the content of the mother is thus assumed to be passed from the adjunct-daughter, not from the headdaughter. The connection to the head-daughter is established in the adjunct itself: it bears (lexically) the HEAD-feature MOD, which represents a description of the head that it can combine with. Non-predicative adjectives, for example, carry an N'-fragment in their MOD-feature, through which the range of N's that they can combine with is narrowed down in several respects (e.g. *case* and *index*). Note that when an adjunct and a head are combined via the adjunct-schema (181), the SYNSEM-value of the head is unified with the MOD-value of the adjunct.

By having access (via MOD) to the head it combines with, the adjunct can establish the composition of its own and the head's contents. This is accomplished through building the union set of the RESTRICTIONS of the head and the RESTRICTIONS to be added by the adjunct.

Though motivated independently, MOD will be shown to help in building a structure incrementally.

5.1.1.6 A preliminary conclusion

This has been an extremely dense overview of the most important properties of HPSG in the current context. Many others have not been mentioned yet. The treatment of unbounded dependencies or "movement" of constituents in general, for example, is another important aspect of HPSG that makes it a powerful grammar framework. HPSG provides a mechanism borrowed from GPSG (Gazdar, 1982), that allows the analysis of movement phenomena such as unbounded dependencies without abandoning the weak generative capacity of context-free grammars. For that purpose, signs bear the NONLOCAL feature, which itself carries various features for a variety of non-local phenomena (i.e. "movements" of constituents beyond the scope of a head-projection), such as topicalization. One of these features is SLASH, through which a constituent can be "passed" from one phrasal level to the next, leaving a trace at the position where the constituent originated. The *trace* itself is a phonologically empty sign that establishes the connection between the LOCAL value of the sign that it substitutes and the SLASH value, by which the constituent is passed upwards until it reaches a *filler schema*, where the SLASH value is finally unified with the (LOCAL value of the) filler-daughter. Note that despite this rather procedural description of "movement", this mechanism is completely static in that it is accomplished by virtue of value sharing.

Moreover, the richness of sign representations allows the constituent structure needed to analyze sentences to be extremely parsimonious. For example, *raising* constructions, such as "*Chris believes Yuki to be happy*" are not analyzed by *movement* of constituents (which have to be assumed in GB due to the *theta criterion*). Instead of

moving the embedded subject into the object position of the matrix clause, (subject-toobject) raising is expressed through value sharing between the matrix object (on the SUBCAT list) and the subject specification on the complement's SUBCAT list. In HPSG, there is thus no need to posit an actual constituent (an NP-trace) for raising constructions, and hence the complement will be a VP, not an S.

In general, constituent structures in HPSG will be much more compact than their equivalents in GB, at the expense of rich feature representations, however. In the latest version of HPSG, (Pollard & Sag, 1994, chapter 9) traces have even been abandoned and replaced by a comparably powerful lexical mechanism that establishes the connection between SLASH and a LOCAL value in the subcategorization list of a lexical head. Consequently, HPSG assumes a minimum amount of vacuous structure and empty constituents, apparently rendering it well-suited for efficient parsing, since, for example, "empty productions" (i.e. postulations of gaps) can be avoided in many cases.

This takes us to the question of how a HPSG grammar is generally employed in the process of "parsing".

5.1.1.7 Type deduction versus parsing

Going into the details of the sorted feature formalism and the way it is processed is clearly beyond the scope and focus of this thesis. Nevertheless, some brief reflections appear necessary, in as much as the general properties of the unification based paradigm have to be discussed with respect to their general compatibility with a cognitive approach to parsing.

Roughly speaking, "parsing" in this paradigm amounts to satisfying the completeness criteria in an initially underspecified feature structure (a partial description of the model), a process that has been labeled *type deduction* or more generally, *symbolic constraint satisfaction*.

As an illustration of *type deduction*, imagine a feature structure, such as (184), given as a starting point. (184) partially describes a sign containing the entire sentence string in the PHON feature, plus the information that the sign is a (finite) verbal projection in the SYNSEM feature. Everything else is left unspecified.



The goal of sentence processing in general is to uncover the *meaning* of an utterance, which is (in parts) based on the value of the CONTENT feature. In this case, the problem (roughly) amounts to revealing a *complete* model of the linguistic object given by the sentence. To do so, the information available in the sortal annotations is utilized to reconstruct the missing portions of information in the partial description given in (184), including the semantic information in the CONTENT.

Sorted feature structures that ought to serve as *complete* models of linguistic entities are required to satisfy a variety of criteria. Among them, feature structures must be *totally well-typed* and *sort-resolved* (Carpenter, 1992). A feature structure is *totally well-typed*, if for each node in the feature structure all and only the features that are appropriate for the sort assigned to that node are actually present with the values that are assigned as appropriate for this sort.

A feature structure is *sort-resolved*, if each sort in a feature structure is a *maximal* (most specific) sort, i.e. one that is not further partitioned into subsorts. Especially the latter criterion is crucial for *"parsing"* in this paradigm.

In order to resolve the sortal constraints in the feature structure such that the semantic content of the sign can be returned, quite a lot of information has to be deduced, such as which constituents can be formed from which words, how they can be combined, how they map onto the arguments provided by the heads and so on.

Resolving the sortal constraints of a feature structure is conceptualized as a complex mathematical problem. It amounts to resolving a constraint equation system, i.e. finding the maximal sorts for every node (variable) in the feature graph that fit to the constraints imposed by the maximal sort of every other node. Resolving the sort at one node thus constrains the way in which the sort can be resolved at different nodes. Depending on the order by which the nodes in the feature graph (variables in the feature graph description) are accessed, finding a solution can be an extremely complex enterprise. Starting with the expansion of nodes which are only weakly specified can greatly expand the search space (the more general the sort, the higher the amount of potential maximal subsorts). The matter is further complicated by distributed disjunctions, i. e. disjunctions that contain unification tags to some node outside the scope of the sortal domain (the disjunction streched out by the sort), thus extending the scope of the actual disjunction. The wider the scope of a disjunction (the shorter the feature path at which it occurs), the more nodes are involved (interact with each other), multiplying the number of disjuncts that have to be matched against each other during unification.

Furthermore, in a pure HPSG system, the linguistic objects that would usually be considered the *lexical entries* are simply treated as the subsorts of the sort *word* (*lexical-sign*). During type-resolving, then, whenever the sort *word* had to be resolved, a disjunction consisting of the entire lexicon would have to be streched out and eventually narrowed down by the unification of constraints stemming from different nodes in the structure.

Not surprisingly, "pure" HPSG-based computational models have rarely surpassed the status of experimental systems (e. g. Franz, 1990).

Transparent parsing with HPSG?

Berwick and Weinberg (1985) have raised the issue of *type transparency*. A transparent parser employs the principles of the grammar directly, without compiling them into a more easily processable format, such as phrase structure rules. Transparency refers to concepts of the grammar theory, such as the *X-bar* scheme and the principles of GB, and *not* to the underlying formalism in which these concepts might be implemented. Implementing a transparent GB-parser is typically thought of as the coding the components of GB as serially ordered generators and filters applied to the respective output of the previous module (Berwick, 1991). In this sense, the theoretical constructs or *modules* might correspond to distinct stages of processing.

In a non-derivational approach such as HPSG, the principles and constraints of the grammar do not correspond to stages of processing. Processing issues are (ideally) completely eliminated at the level of theoretical concepts and moved to the level of the underlying formalism. Due to the declarativeness of the sorted feature formalism, there is in principle no procedural implication of linguistic constraints on how type deduction is to be carried out, even if there are more or less efficient ways to do so. From this rather *mathematical* point of view, processing issues are completely irrelevant to linguistic modelling. Nevertheless, great care has to be taken in how the constraints are computed in actual grammar implementations.

Clearly, the way in which type deduction and unification is carried out (see Uszkoreit, 1991) must by controlled in some way in order to keep the parsing problem tractable. In fact, there have been several successful attempts to build a parser working on HPSG-style grammars (e.g. DISCO, Uszkoreit et al., 1994). In many of these examples, however, the parser was implemented in a more traditional rule-based manner, e.g. a *chart parser*, operating on some pre-built schemata or rules (nevertheless transparently generated ones, such that declarativity is maintained).

The "control" approach taken here is similar to these latter examples in that the combination of signs is controlled by a parsing mechanism external to the HPSG formalism. Type resolution only operates on signs that have already been combined by external means, thus drastically reducing complexity. However, the principles and schemata of HPSG are *not* compiled into a set of context-free phrase structure rules which are then employed in a standard parsing algorithm. The way in which the grammar information is actually employed is closely linked to the information projected from the lexicon increment by increment, leaving extra-lexical constraints represented in much the same way as they are specified in the HPSG theory. From this point of view, the notion of *type transparency* might very well be applied, since the operations of the parsing mechanism in a *sign-based* approach, such as the one presented here, do indeed correspond to constructs at the level of grammar-theoretical concepts, but not at the level of the formalism.

5.1.2 HPSG as a cognitive model of linguistic knowledge?

The predominant attitude of linguists engaged in HPSG projects, on the other hand, is certainly a non-cognitive one (e. g. Kiss, personal communication). This

might be due to the "mathematical" perspective on language processing, which is clearly influenced by Gazdar´s *realistic* (Platonist) position. The founders of HPSG, Carl Pollard and Ivan Sag, however, seem to take a considerably different stance.

> "... On the other hand, in order to circumscribe our task, we do not charge our theory with providing a specific algorithm, though we would expect an adequate theory of language use to provide one. Thus we accept the conventional wisdom that linguistic theory must account for *linguistic knowledge* (...) but not necessarily for processes by which that *knowledge* is brought to bear in the case of individual linguistic tokens. Indeed, we take it to be the central goal of linguistic theory to characterize what it is that every linguistically mature human being knows by virtue of being a linguistic creature, viz. universal grammar. And a theory of a particular language - a grammar - characterizes what linguistic knowledge (beyond universal grammar) is shared by the community of speakers of that language. Indeed, from the linguist's point of view, that is what language is." (Pollard & Sag, 1994, p. 8, highlighting added by the author)

In speaking about linguistic *knowledge*, Pollard and Sag clearly take a mentalist point of view (although they argue with more caution in other examples). The mentalist position is even more explicitly expressed by Ivan Sag as he was interviewed by Anne-Marie Mineur and Gerrit Rentier (1993, TA!, studentenblad, computationele taalkunde, nr. 2.2):

"... And also the idea of taking seriously some division between competence and performance; having a theory of linguistic knowledge be embedded in a theory of language processing, a theory of performance. That is also an aspect of Chomskian linguistics that I would completely agree with. So with respect to at least two important criteria, I would regard myself a Chomskian. ..."

This, of course, is the ultimate coming out as a cognitivist, although the position expressed might still be compatible with a *weak* notion of competence. In the next quote, however, the position is further strengthened:

"... If the grammars offered by a linguistic theory are to be embedded into a *theory of human language processing*, then there are a variety of properties of language processing that might be expected to inform the design of grammar. For example, even the most superficial observation of actual language use makes plain the fact that language processing is typically highly incremental: speakers are able to assign partial interpretations to partial utterances (and quite rapidly, in fact). Thus, other things being equal, a theory of grammar that provides linguistic descriptions that can be shown to be incrementally processable should be regarded as superior to one that does not. ..." (Pollard & Sag, 1994, p. 11, emphasis added by the author) This is a fairly strong claim. In fact, it is certainly stronger than the one expressed in Chomsky's classical view on the relationship of competence and performance, since it posits that performance evidence *can* validate linguistic theorizing.

There are examples that Pollard and Sag take this commitment seriously. The complement extraction rule (Pollard & Sag, 1994, p. 378), for instance, is a replacement of the concept of *traces*, motivated by psycholinguistic findings on the direct association of "moved" constituents and the subcategorizer (Pickering & Barry, 1991).

To conclude, taking the representations provided by HPSG as the knowledge base employed by the human parser does not seem to be an unjustified enterprise.

5.1.3 Modifications and extensions in SOUL

5.1.3.1 SUBCAT

In the following feature structures, the SUBCAT feature is split into a SUBJECT (SUBJ) and a COMPS feature, with SUBJ taking the first elements on the SUBCAT list, and COMPS the rest. Although this split is linguistically motivated (Borsley, 1987, see Pollard and Sag, 1994, pp. 346), I am not going to discuss it in detail here. Additionally, a new type or daughter, namely, the SUBJ-DTR is introduced. Schema1, 2 and 3 must then be modified. Furthermore, I will assume that the *subcategorization principle* is collapsed with the schemata, resulting in the simplified versions of *schema1* (185) and *schema2* (186) (*schema3* should be obvious).



5.1.3.2 Complements and adjuncts

In Pollard and Sag (1987), adjuncts were assumed to be specified in the representation of the head-daughter. Nowadays (Pollard and Sag, 1994), adjuncts are understood as *semantic* heads (functors) that take their *syntactic* heads as arguments. Heads have thus lost their ability to pose constraints on the type of adjuncts they can combine with. This is unfortunate for various reasons. For example, if instrumental PPs were understood as modifiers, there would be no way of expressing the "expectation" of a verb for such a PP. Such an approach, however, seems to contradict the observations of Spivey-Knowlton and Sedivy (1994, cf. MacDonald et al., 1994) who found that *action* verbs tend to occur with *instrumental* or *manner*-PPs. Even some nouns, such as nouns related to communication (*mail, message*), seem to bear an affinity to *theme* PPs (*mail about s.th.*) and *goal* PPs (*message to Bill*). However, since such dependencies do seem to exist, "modifiers" like *instruments*, must be "known" to their heads somehow. A reasonable way to account for this would be to return to the original version of HPSG (Pollard and Sag, 1987).

The approach taken here, however, is more radical in that the notion of *adjuncts* is abandoned (at least) for verbs and certain argument types, such as *instruments*. Several linguistic criteria have been established to distinguish adjuncts from complements empirically. In the following sections, I will discuss a variety of them (cf. Pollard and Sag, 1987) in some detail, and I will argue that at least instrumental PPs cannot easily be classified as adjuncts and should rather be considered complements, contrary to the common treatment.

(187) Iterability

In general, two or more instances of the same adjunct type can combine with the same head, but this is impossible for complements (cf. Pollard and Sag, 1987, p. 136)

Pollard and Sag give an example that is supposed to demonstrate the adjunct status of instrumental PPs (188).

(188) Heather opened the rusty lock with a key, with a pair of pliers. (Pollard and Sag, 1987, p. 136)

The native speakers of English I consulted judge this sentence as grammatically borderline or worse. In German, an equivalent example would be almost impossible. According to the *iterability* criterion (187), instrumental PPs must therefore be complements, at least in German.

(189) Order-dependence of content

The contribution of adjuncts to semantic content can depend upon their relative order in a way which does not apply to optional complements. (Pollard and Sag, 1987, p. 135)

The relevance of (189) can be demonstrated by the fact that the two sentences in (190) clearly have different meanings due to the different ordering of the adjuncts.
(190) a. Kim jogs reluctantly twice a day.

b. Kim jogs twice a day reluctantly.

The interpretation of the sentence clearly seems to depend on which adverbial is within the scope of the other (although Pollard and Sag deny that quantificational scope is at stake here). Example (191), however, suggests that its interpretation does not seem to be strongly influenced by the ordering of the instrumental and the locative PP.

(191) a. Gerhard [eats [with chop-sticks] [in the restaurant]].

b. Gerhard [eats [in the restaurant] [with chop-sticks]].

Order dependency as a criterion for adjuncts does not appear then to be very decisive in the case of instrumental PPs. If anything, it renders them (optional) complements.

(192) Constancy of semantic contribution

Adjuncts can occur with a broad range of heads while seeming to make a uniform contribution to semantic content independent of the head (cf. Pollard and Sag, 1987, p. 136). The locative-PP *on the hill*, for example, bears the same meaning regardless of whether it occurs with *to camp, to jog, to meditate*, etc. (although this criterion appears rather vague). Nevertheless, instrumental PPs could again be rendered complements, since a PP[with], like *with the hammer*, for instance, appears to have fairly different meanings in the context of *hit, juggle*, or *decorate*.

To sum up, there is little reason to assume that instrumental PPs are not to be treated as optional complements. If this is so, an INSTRUMENT role must appear in the content of verbs. The content of *beobachten*, for example, can thus be partially described as in (193)⁸⁶.

(193)

psoa	-
RELATION	observe
EXPERIENCER	refo
THEME	refo
INSTRUMENT	refo

5.1.3.3 Optional complements and pruning

Pollard and Sag (1987) raised the notion of *optional* complements, i.e. complements that may, but need not, occur with their heads in a sentence. Some verbs, such as *donate*, bear complements that may be omitted, as illustrated in (194), while a similar constituent has to be lexically realized with other verbs, such as *put* (195).

^{86.}Note that thematic roles take *refos* (referential objects) in the SOUL approach, instead of *indexes. Refos* are, roughly, *restricted indexes* or *npros*, as they appear in the content of nouns (e.g. 172, 173). See also example (201) and footnote 88.

(194) a. Janet donated some money to the library.

b. Janet donated some money.

(195) a. Janet put the vase of flowers on the table.

b. *Janet put the vase of flowers.

Whether or not a complement is optional is specified in the SUBCAT (COMPS) list. Subcategorization lists with optional complements, as marked by the parenthesis in (196), amount to a logical disjunction of lists with and without the optional complement.

(196) *donate*: [COMPS: < ..., (PP)>] = [COMPS: < ... > v < ..., PP>]

While logically grounded, assuming disjunctions at the level of the COMPS-list is somewhat less neat from a procedural point of view, as the number of disjunct subcategorization frames to be maintained increases by factor 2^n (with *n* being the number of optional complements, leaving aside permutations of possible complement linearizations). In a depth-first parsing approach, the parser would have to backtrack and re-parse quite a lot, and even a chart-parsing account could only cope with the emerging complexity in a rather inelegant way.

In the SOUL approach, optional complements are distinguished from obligatory complements only in that they may be left *phonologically empty*. To represent this, no additional features or sorts have to be stipulated. Note that the feature PHON is assumed to take a value of the sort *phon-list* (*phl*). *Phon-list* is assumed to be partitioned into the subsorts *empty-phon-list*, and *non-empty-phon-list* (*nephl*), as illustrated in (197). Of course, *non-empty-phon-list* is a certain type of list (see section 5.1.1.2) taking phonological strings in the FIRST feature.

(197)



The only modification to standard HPSG needed is that the SUBCAT (COMPS) list contains complete *signs*, not just their SYNSEM values⁸⁷. If so, the complements' PHON values can be accessed through the SUBCAT list. If a complement is obligatory, as the *Locative*-PP for *put* (198), its PHON value can be restricted to *non-empty-phon-list* (*nephl*) meaning that the complement must bear a phonological string.

```
(198) put: [COMPS: < ..., PP[PHON: non-empty-phon-list]>]
```

On the other hand, optional complements, such as the *goal*-PP in the COMPS-list of *donate* (199), are proposed to be only restricted to *phon-list* (*phl*), thus allowing the PHON value to be resolved to *empty-phon-list* (*ephl*) or to *non-empty-phon-list* (*nephl*).

^{87.}In the overview of HPSG in the preceding sections, this assumption has already been made. Signs in the SUBCAT-list are also assumed in Pollard and Sag (1987).

(199) donate: [COMPS: < ..., PP[PHON: phon-list]>]

Optional complements can thus be treated like obligatory complements, i.e. they can simply be (head-) projected into the structure⁸⁸. Since there is only a single subcategorization list for all combinations of optional and obligatory complements, optional complements will therefore not generate disjuncts that have to be treated distinctly from each other during parsing.

If the surface string later turns out to not provide a corresponding substring, the PHON value can be set to *empty-phon-list*. I will call this operation *pruning* of projected daughters. Note, however, that a complement is never *deleted*; its PHON value is only set to empty.

5.1.3.4 Lexical preferences

Lexical preferences and HPSG. As demonstrated by Ford, Bresnan and Kaplan (1982), some verbs pose preferences for their possible arguments.

(200) a. Joe included the package for Susan.

b. Joe carried the package for Susan.

While the strongest lexical form for *carry* in (200b) includes a PP-complement, the strongest form of *include* does not (200a).

Since different multiple subcategorization constraints induced by optional complements are not treated as separate lexical forms with different strengths in standard HPSG, but as logical disjunctions of subcategorization lists, standard HPSG provides no obvious facility to represent *lexical preferences* at all. As a natural extension, however, the disjuncts could be ranked, such that each disjunct bears a strength value that determines which one is applied next during unification. However, as we have seen before, the approach to complement optionality taken here is rather different: elements are individually marked as optional or obligatory.

Thematic roles. Many researchers have suggested that lexical "expectations" (Ford et al., 1982) can be attributed to *thematic* properties of heads, rather than to functional or syntactic ones (MacDonald et a., 1994; Trueswell et al. 1993, Tanenhaus et al. 1989; Hanna et al., 1995).

The SOUL grammar proposes that thematic roles, or more specifically, role-takers, are marked as *highly* or *lowly salient*, such that some roles, e.g. *instrument*, are more

^{88.}Of course, not all kinds of lexical ambiguities, and even not all kinds of ambiguities within the same major category, can be unified into a single lexical representation. Lexical categories must sometimes have multiple lexical forms due to an entirely different mapping from complement to thematic structure. The verb *stop*, for instance, has an agentive (transitive) reading and an ergative (intransitive) reading. In the first case (*the policeman stopped the car*), the subject NP is the Agens of *stopping*, while the subject NP in the second form (*the car stopped*) is the Theme (Patiens). Therefore, the lexical forms can not be merged into a single form.

salient in the thematic frame of one kind of verb, e.g. *observe*, than in the frame of others, such as *to catch sight of sth*.

The feature structure which models a role filler in SOUL, namely a referential object (*refo*)⁸⁹, will thus provide the feature [SALIENCE sal], which can take either of the two values *high* or *low* (for the moment), as shown in (201).

(201)	i)
(201	IJ

[refo]
INDEX	index
RESTR	restr
SALIENCE	sal

For each role in the thematic frame of a head, e.g. the verb *beobachten ("observe*"), a salience value can be assigned, as shown in (202)

(202)

(~	υ	~	,	

psoa RELATION	observe]
INSTRUMENT	[<i>refo</i> SALIENCE	high]

Since the *refos* are token-identical with the content of NPs (or PPs) in the subcategorization list, the subcategorization list of *beobachten* can be abbreviated as in (203).

(203) beobachten: [COMPS: < ..., PP[PHON: phl]_[SALIENCE high]>]

The subscript in (203) indicates that SALIENCE is in the CONTENT of the particular complement. Other verbs, such as *erblicken* (*catch-sight-of*), although similar semantically, pose only a low salience to the instrument role. The content of *erblicken* ("*to catch sight of sth.*") is given in (204).

(204)

Ferritor (catchSightOf	
INSTRUMENT	[<i>refo</i> SALIENCE	low

The abbreviated subcategorization frame of *erblicken* is shown in (205).

(205) erblicken: [COMPS: < ..., PP[PHON: phl][SALIENCE low]>]

Salience and obliqueness are crucial to the parsing algorithm. In this section, however, the focus will remain on the representational issues for a while. It is necessary to give an account of SALIENCE that goes beyond stipulating certain values (although

^{89.}Note that a thematic role takes a *restricted index* as its value, different from standard HPSG, where a thematic argument of a relation takes only an *index*. The reason is that we can now model *selection restrictions* for certain roles in a more straightforward way, because the restrictions are now within the scope of the value of the role itself.

the values given here mirror empirical data from several questionnaire and closure studies). Of course, salience should be motivated independently. An account is needed that provides theory-internal conditions determining whether the SALIENCE value in certain thematic contexts is *high*, or in others *low*. Clearly, to give a complete account would mean a project of its own and absolutely beyond the scope of this thesis (which focuses on the *mechanisms* which put relevant information to use during language comprehension, but not primarily on the grammar (or semantic) theory. Nevertheless, some directions in research should be pointed out that appear fruitful.

5.2 Towards a theory of thematic salience

5.2.1 Frequency

Of course, *salience* might reflect the frequency in which a particular lexical head occurs in combination with its roles in every-day conversation. A *psychological* explanation like this has a considerable tradition in psycholinguistics: a number of lexicon-based accounts, such as Ford, et al.'s (1982) and MacDonald et al.'s (1994, see section 2. 11) attribute lexical preference effects to the *frequency* of the particular lexical form in every-day conversation. Frequency-based explanations, though not lexical ones, also play a role in Wanner's ATN and the *Tuning* approach (Mitchell et al., 1995). In the latter, parsing preferences are even exclusively attributed to the success history of the parser. Lexical preferences also occur in Gibson's model (1991, pp. 176). Although (or because) Gibson avoids a commitment to any "deeper" concept underlying lexical preferences, it seems likely that he assumes them to reflect lexical frequencies as well.

Clearly, if lexical preferences were shown to be solely determined by frequency, we would be set. However, this will probably not happen any time soon. Even if lexical preferences could be demonstrated to be related to frequencies, it is unclear whether this could be considered an explanation. It may very well be that frequencies only reflect preferences caused by principles in the production mechanism. And since people generally try to generate utterances as comprehensible as possible, the circle may soon be closed. If so, we will still remain in search of a theory of lexical preferences. In the following sections, I will briefly sketch some ideas which point in the direction of a more explanatory account.

5.2.2 Instruments and agents

A couple of *thematic properties* seem to be systematically related to salience.

It might be argued, for instance, that verbs like *erblicken* (*catch-sight-of*) do not permit an instrument at all, as the utilization of an instrument might presuppose an *intentional agent*. Such a claim, however, can easily be shown to be too strong. If not, sentence (206) would not have a sensible interpretation.

(206) The man was accidentally hurt with the knife.

Furthermore, there is evidence from a questionnaire study, conducted as a prestudy for experiment I, which shows the combination of instruments and non-intentional agents to be indeed feasible.

- (207) a. Den Mann beobachtete Peter gestern mit dem Fernglas. The man_{acc} observed Peter yesterday with the binoculars. Yesterday, Peter observed the man with the binoculars.
 - b. Den Mann erblickte Peter gestern mit dem Fernglas. The man_{acc} caught-sight-of Peter yesterday with the binoculars. Yesterday, Peter caught-sight-of the man with the binoculars.

Since the PP is not adjacent to the preceding NP, both sentences only permit the instrumental reading of the PP *mit dem Fernglas. Beobachten (to observe)*, in contrast to *erblicken (to catch sight of sth.)*, takes an intentional agent. Nevertheless, both sentences were rated equally plausible. This result further supports the assumption that the *instrument* role does not strictly presuppose an *intentional agent*. Therefore, *erblicken* as well as *beobachten* can be assumed to provide an *instrument* role.

What these examples tell us, however, is that *agentivity* or *intentionality* might be *related* to the *salience* of an instrument, since the fact that an action is carried out intentionally appears to increase the likelihood that an instrument is used, and vice versa. The relationship between agentivity (or intentionality) and the permission of an instrument role must be considered a *prototypical* one (at least as long as a more complete set of determinants is still not uncovered), rather than one of logical necessity.

In support of these assumptions, Spivey-Knowlton and Sedivy (1994, cf. Mac-Donald et al., 1994) observed that *action* verbs tend to occur with *instrumental* or *manner*-PPs. Some noun classes, such as nouns related to communication (*mail, message*) seem to bear an affinity to *theme* PPs (*mail about s.th.*) and *goal* PPs (*message to Bill*). In general, there seem to be various semantic/thematic interdependencies between the type of event, action or state expressed by the verb and its arguments, which are likely to determine the salience of a thematic role. I consider this line of research very fruitful.

5.2.3 Ontological necessity

Another lexical factor possibly influencing thematic salience is the *ontological necessity* of thematic roles. For instance, the verb *eat* requires both someone who eats and something to be eaten. Therefore, the *agent* and the *theme* are ontologically necessary. The verb *kick*, in contrast, does not necessarily require a theme, since "a kick" can very well be performed without something that is actually hit, e. g. as a dance figure or in a *tae kwon do* choreography. Ontological necessity might turn out as a crucial factor for *lexical preferences* in that it might determine the base-line that other influences modulate.

5.2.4 Conceptual properties

Another type of hypothesis related to thematic roles is that certain properties of the *concept* associated with a word might modulate the salience of its roles, if they only had enough time to do so. Specifically, some verbs conceptually entail the use of a specific instrument or are at least strongly associated with it. In order to *shoot*, for example, one needs a weapon that can be shot with; *cut* is strongly associated with a certain kind of edged instrument, whereas the association with an instrument is presumably much weaker for verbs like *read*. Verbs that express actions of building, creating, or modifying are likely to take an (artifactual) instrument, whereas, for example, *perceiving* verbs, such as *hear* or *see*, are not likely to take an instrument at all.

One might therefore raise the hypothesis that soon after a verb is read, its arguments become more or less salient depending on conceptual properties associated with it. Note however, that incorporating *conceptual* information in a model of lexical predictions is a fairly far-reaching hypothesis, as it appears to weaken the notion of *modularity* (Fodor, 1983). On the other hand, it must be kept in mind that the kind of interaction proposed is restricted to optional *complements*, which are syntactically present already. Moreover, allowing conceptual features to modulate the salience of arguments does *not* necessarily imply strong interaction. The latter would be assumed, for example, when all levels of information are competing with each other simultaneously, as in MacDonald et al.'s model.

In the approach proposed here, however, it is assumed that the parser hypothesizes attachments for an item irrespective of its lexical or conceptual features. These features will thus not influence the principled way in which an item tries to attach itself to the preceding structure. It is not assumed that the entire *conceptual* information associated with the current item is available at that very moment. On the other hand, by the time an item is to be attached, the preceding heads might have had enough time to modulate the *salience* of "forwardly" predicted arguments, and thus to determine which attachment sites become most visible to a later item to be attached. This difference is crucial. To illustrate the issue, consider sentence (208).

(208) Polizei erschoß 19jährigen mit Samuraischwert. (Headline in *Westfälische Rundschau*)

Police shot 19-year-old with samurai-sword.

Let us assume that the PP *mit Samuraischwert* is to be attached either to the preceding NP *19jährigen*, or to the verb *shot*. At this point in time, *shot* may have had enough time to render an instrument highly salient. If so, the PP will probably be attached to it, *even though* its own conceptual properties may have rendered this attachment inappropriate. This, however, seems to take place only at a later stage. The predictions thus meet the feeling of surprise most people have about this sentence.

Next, consider (209).

(209) The soup cooked ...

When the verb *cooked* is read, it can be attached as a past participle, introducing a reduced relative clause, or simply as a main verb. A strongly interactive model like MacDonald et al.'s will predict that a decision is based on the configuration of all types of information and their respective strengths. The model proposed here is considerably different: since only a verb is obligatorily predicted after the initial NP, there are no (predictions of) optional constituents whose salience could be modulated by some conceptual properties of the NP. If it is assumed that the conceptual information associated with the currently read item is integrated only later, the initial attachment decision for the verb will be determined by low-level properties, such as the lexical strengths of the respective forms. No interaction takes place in this case.

To sum up, conceptual properties might be integrated in the model presented here. Even if conceptual knowledge is assumed to be integrated at a later stage of processing, it could determine the *salience* or visibility of items to be expected later in the string. Whereas conceptual properties of currently read items cannot influence initial parsing decisions, those of preceding items may. This perspective may give a more differentiated view on *modularity* and *interaction*. The entire consequences of such an approach, however, remain to be investigated and will be the subject of future research.

5.2.5 Contextual influences on salience

At least intuitively, it is apparent that the salience of particular thematic roles can be strengthened by the supra-sentential context of the utterance. An instrument should, for example, be strongly expected after a question like "*With what did Peter hit Maria*?".

In a sentence completion task and an eye-tracking experiment, Hanna, Barker, and Tanenhaus (1995) looked at "*by*"-phrases in passive-constructions like (211a,b) which are locally ambiguous between an agentive "*by the director*" and a locative "*by the entrance*" reading.

(210) The artist decided to go to the gallery. Once he got there he wanted to

- a. know who had hung his prize painting.
- b. see where his prize painting had been hung.
- (211) Target
 - a. He was pleased to discover that his painting had been hung by the entrance.
 - b. He was pleased to discover that his painting had been hung by the director.

They found evidence for an interaction between a general bias for the agentive reading, verb-specific biases, and a bias introduced by the context (210a,b). Unfortunately, although the authors interpret the results of their experiment in the framework of a competition model, no fully straightforward conclusions can be drawn with respect to the time course of the various influencing constraints. More evidence of this kind will be helpful in providing a clearer picture of contextual factors constraining salience (see also Liversedge, Pickering, & Branigan, 1995).

5.2.6 Salience at run or compile time?

During processing, then, the parser could take semantic features like *agentivity* (and further factors contributing to salience) directly into account and base its predictions of forthcoming constituents on such factors only on demand. The approach taken in this thesis, however, is that the salience of thematic role fillers is pre-set (pre-compiled) into the lexical entries, such that the parser can make its decisions on the base of one value only. Note, however, that this does not imply that the salience value has to be fixed in the lexicon and cannot be modified throughout processing. If co-occurrence statistics turn out to be a viable basis for salience, long term changes to the salience features could be made each time the lexical entry is accessed in a particular thematic context.

Furthermore, the salience-features set for each lexical entry can be used to dynamically cumulate further influences on-line, yielding fine-grained interactions of lexical and contextual biases as reported in Hanna et al. (1995). The binary typed value of salience will therefore be modified to a continuous numerical value, ranging from 0 to 1. A threshold will then determine at which point the salience is to be considered *high* or *low*.

5.3 The processing mechanism

5.3.1 Modelling an adequate parser

There are generally two approaches to set up an empirically adequate parser. First, one could propose a general powerful parser, such as a parallel chart parser or an Earley parser, with an unrestricted (parallel) algorithm capable of pursuing all analyses at once, and build a more or less explicitly coded *heuristic component* suited for ambiguity resolution on top of it. The second approach is to set up a parser or mechanism whose capabilities are restricted in a principled way from the start, such that "heuristic" behavior emerges as an intrinsic property of the parsing mechanism. The former approach suffers for at least one reason: if one assumes a powerful base algorithm, there must be independent evidence that its power is needed in at least *some* circumstances. It is not sufficient, though necessary, to provide independent motivation for the heuristic component that constrains the parser. As long as there is no real reason to assume an architecture based on an unrestricted parser, I will take the second approach and propose a mechanism that is as simple as possible and parsimonious in its prerequisites.

The properties that such a mechanism must fulfill have been developed in the preceding chapters and are repeated here for convenience.

The human sentence processing mechanism

- is highly *incremental*, i.e. it attaches each incoming item as it is encountered (Frazier, 1987a), *"linear parsing"* (Konieczny & Strube, 1995), (as opposed to *head-driven* or *head licensing* parsing), employing both bottom-up and top-down strategies,
- operates in a *serial* fashion, i. e. it pursues only one analysis at a time in the case of an attachment ambiguity (as opposed to *parallel* processing of multiple analyses)
- *commits* itself to a structural alternative *immediately* (as opposed to minimal commitment or to a "wait and see" strategy),
- uses *detailed lexical information*, including lexical subcategorization preferences, during the *initial stage* of structure assembly,
- resolves structural ambiguities according to *Parametrized Head Attachment* (Konieczny et al., 1994), i.e. depending on the presence or absence of the lexical head of the attachment sites, their lexical properties, and their respective distance to the item to be attached, and
- operates *autonomously* in its initial attachment decisions, i.e. higher level semantic and pragmatic processes are not capable of *directing* the initial attachment choice (as opposed to *strongly interactive* or *integrative* multiple constraint models)

5.3.1.1 Visibility versus competition

As has been argued in the preceding chapter, the alternative analyses of a local ambiguity do not seem to *compete* with each other. As they stand, the facts indicate that an ambiguity is not even detected in most of the cases, suggesting that the

human sentence processor operates in a straightforward serial "depth-first" fashion. The question that then arises is how attachment preferences can be accounted for in a model that satisfies all the constraints worked out in the previous chapters. The approach that I take is reminiscent of the notion of *visibility*⁹⁰ (cf. Frazier, 1995). The *visibility* of an attachment site in the phrase marker representing the current sentence will determine whether or not an item will (preferably) attempt to attach itself to that site initially. Only one attachment site is visible at each state of processing, *overshadowing* the other sites. Whether or not an attachment site is visible is determined by partial representation of the sentence analyzed so far and the lexical items that participate in it, as will be described in the following sections. The *visibility* approach, then, motivates an *object-oriented* view on parsing in which the perspective of an item is taken which actively attempts to attach itself to a site that it can actually see.

5.3.2 Signs as objects

The HPSG framework focuses on linguistic *signs*. In the SOUL mechanism, consequently, *signs* play the central role in parsing (again). Signs are implemented in an object-oriented manner. As objects, they bear a declarative as well as a procedural component: besides the declarative component (the information represented in the feature structure), they come equipped with *methods* for combining themselves with other *sign*-objects.

Parsing starts at the lexicon: as soon as a word is encountered, a lexical sign (or several signs) is (are) retrieved and *activated*. When activated, a lexical sign behaves in a principled way, as described below. As long as it has not attached itself to the sentence structure built so far (called the *current sign*), the sign retrieved from the lexicon is called *active sign*.

As an important property of the parsing mechanism inherent in signs, the current sign (object) carries a stack of *predicted signs*, which are "offered" to the active sign as possible attachment sites. The stack of predicted signs plays an important role in ambiguity resolution, as will become clear soon.

The *active sign* always attempts to perform one of the following methods in order of appearance:

1. *attach* self to the *most visible* node in the current sign (if there is one). In order to do so, consider sign-*predictions* first. Only if no such attachment is possible, search actively for attachment sites in the current sign (both types of search for an attachment site will be described in detail in one of the next sections). If it is completely impossible to attach self to the current sign in the current state,

^{90.}Note that the actual definition of visibility given here (see below) is different from Frazier's (1995), which recurs *i*. to the time course of node postulation, and *ii*. to *perceptually given packages* (e.g. phonological phrases). None of these issues play a role in the model presented here.

2. *project* self up to the next phrasal level, as a daughter in one of the basic HPSGrule-schemata, but never higher than sentence level. As a part of projection, unify the new phrasal unit with the principles of well-formedness in order to rule out illegal projections. Then start again with 1.

If either the attachment (1.) worked, or the projection (2.) reached the sentence level, the current sign

3. *predicts* the subsequent constituents in order to offer markedly visible attachment sites for subsequent items. This is accomplished by collecting the yet unfilled phrasal nodes in the current sign that are marked *obligatory* or *highly salient*. These nodes will be called *predicted signs* henceforth. Sibling daughters at each level of the phrase structure are put into a set. Since the collection algorithm starts at the top node and proceeds downwards, the lowest set of predicted signs is on top of the prediction stack. The elements of the top-most set on the *prediction stack* are called *predicted nodes*.

5.3.3 Predicted and active attachment

The two types of *attachment*-methods in 1. are called *predicted attachment* and *active attachment*. During the *predicted attachment* search, the active sign attempts to unify itself with one of the *predicted nodes*. If it fails to do so, the active sign enters the *active attachment* search mode for a phrasal node in the current sign which permits attachment to itself.

The *active attachment* search starts at the word to the left of the active sign and proceeds upwards along the right edge backbone of the current sign, always to the immediate dominator of the current constituent if the attachment is prohibited at the current level.

Furthermore, the *active attachment* search incorporates two distinct types of attachment mechanisms. Firstly, if the phrasal node currently under investigation provides a set of pointers to already projected but unpredicted – i.e. unsalient – complement daughters, the active sign attempts to unify itself with one of these first (*projected complement attachment*). If that fails, the active sign attempts one of several *lowering* operations (Sturt & Crocker 1995), e.g. *adjoining*, which are *not* supposed to be costly. These operations will be specified in detail in a later section.

The basic mechanism is summarized in (212).

(212) The SOUL mechanism

1.ATTACH

- a. predicted attachment
- b. active attachment
 - projected attachment
 - lowering (e.g. adjoining)

2.PROJECT

3.PREDICT

5.3.3.1 In terms of visibility

The visibility of the attachment sites is thus hierarchically ordered:

- *i.* Most *visible* are those complements of preceding heads that are marked as *obligatory* and/or *salient*. These have been predicted top-down, such that lower predictions overshadow higher ones. Only predictions at the lowest level (*predicted nodes*) are accessible.
- *ii.* If the attachment to one of the *predicted nodes* fails, the active sign can access the most recently read word and start a search for an attachment site by proceeding from the bottom to the top of the right edge of the current sentence (the lower, the more visible).

Before I describe both kinds of attachment operations in more detail, some necessary prerequisites must be introduced. For *predicted attachment* it is important to focus on the *prediction stack*, i. e. what kinds of elements are pushed onto it and deleted from it and under which circumstances.

5.3.4 States

The active and the current sign are linked to a data structure (an object) called *state*. Whenever the active sign performs one of its basic methods, a new state is generated and initiated with a pointer to its predecessor state. When a parsing operation fails at some point, the predecessor state can be recovered and an alternative method to the one that failed can be evaluated. The sequence of states amounts to something comparable to an *agenda* in a standard parsing approach (see Allen, 1987), except that there is no control structure (the parser) that is driven by the options on the agenda only. Instead, the active sign proceeds through a fixed program, memorizing earlier states only to be able to continue with an alternative operation if something goes wrong. In order to be able to backup to a previous state properly, the active sign or the current sign (or both) are copied before they are modified by the operation in the new state.

5.3.5 Bottom up projection

Projection, i.e. the bottom-up postulation of a mother node, plays an important role in the parser. In the SOUL system, projection is accomplished by virtue of unifying the active sign with one of the daughters of a HPSG ID-schema, which is immediately unified with the principles of the grammar. New sister nodes will be created through projection, depending, for example, on the subcategorization requirements of the active sign. No explicit grammar rules are required, since a "left corner" alike prediction effect is achieved by instantiating one of the general ID-schemata with the constraints posed by the lexical sign and the principles of well-formedness.

Projection is prohibited if the sign was already attached to an obligatory node in a previous state.

5.3.6 The prediction stack

The prediction stack collects the top-down predictions of heads and non-heads. The collection mechanism can be described as a depth-first scanning procedure proceeding through the constituent structure of a sign starting at the root node. Sign-nodes, i. e. nodes representing a feature-structure of the sort *sign*, carry two boolean variables, called *isLexicallyFilled* and *isEmpty*. The first flag, *isLexicallyFilled*, indicates whether or not all sub-constituents have already been found in the input string. During the scanning procedure, the routine stops at a node which is marked as lexically filled and backtracks to another path.

The second flag, *isEmpty*, marks the opposite: it is set to YES if none of its sub-constituents have already been identified in the input string, and to NO if at least one has been found. If a node is marked as *not empty* and *not lexically filled*, a "*construction flag*" (c-flag henceforth) is built, indicating that a sign is only partially filled, in that some sub-constituents are still expected to be *empty*. A c-flag is a specially marked data object that consists of a pointer (indicated by the asterisk character *) to the current node (cf. Allen, 1987). C-flags will be indicated by "[]" in the text, such that "[*NP]" is a c-flag pointing to a node representing a noun phrase. (Note that the construction flag notation "[]" must not be confused with HPSG feature structure abbreviations, marked in the same fashion. The context should always indicate the appropriate reading, however.)

Whenever a c-flag is built from a node, it is pushed onto the sign's prediction stack followed by the set of the *empty* immediate daughters of the node. Sets of nodes will be indicated by "{}". When there is only one element in the set, the parentheses will be omitted.

If a phrasal node has an unfilled but non-empty immediate daughter, the scanning procedure continues to proceed to that node. Note that at each level of constituent structure, there is never more than one unfilled but non-empty immediate daughter, due to the fact that each word in the input string is attached immediately.

Consider the structure representing the fragmentary sign (213) after the first word *the* has been read (ignore for a moment the question of precisely how this structure is set up).



The scanning procedure builds the prediction stack given on the right of (213). The left-most element (N') is the top-most element, followed by an NP c-flag. You can read the stack as "when the N' has been found in the input string, the NP is (could be) complete". Consequently, the next predicted sign would be a VP, which has to be found to complete the S. (Note that the predicted signs on the stack are just pointers to the corresponding node in the structure.)

Now let us have a look at structure (214) built as soon as the transitive verb *gibt* (*give*s) has been integrated into the structure.



In German, the order of the complements is not fixed, although there is a preference to expect the dative case NP before the accusative NP (Pechmann et al., 1994). The set indicates that both NPs are predicted at the same time⁹¹ and thus visible to the active sign.

Whenever an item is attached to a predicted node, the node's *isEmpty* flag is set to NO and the predicted sign is turned into a new c-flag for that node, while all empty nodes below it are pushed above it in the prediction stack. Thus, if the word (*dem*) is integrated into the structure, the prediction stack is altered to (215).



⁹¹.Note that the set is ordered to express canonical ordering.

5.3.7 The treatment of optional complements

Empty nodes are only put onto the prediction stack if the feature structure it represents is marked as *obligatory* (*non-empty-phon-list*) or as *salient* (+SAL). Optional complement nodes with an unsalient content will *not* be predicted. Instead, non-predicted nodes are collected in a set, which is attached to (associated with) the mother's node for use in the *active attachment* procedure.

As an example, consider structure (216).



The verb *erblickte* (*caught sight of*) was shown to subcategorize for an optional PP with a lowly salient content. Even though the PP is projected into the structure as soon as the verb is integrated, it will be ignored by the scanning procedure and it will thus not appear on the prediction stack.

5.3.8 Overshadowing predictions

The visibility of a predicted node is determined by whether or not the node is within the top-most set of predicted nodes on the prediction stack. All nodes but the top-most are *overshadowed* by the top-most predictions. Note that each set of predicted nodes is followed by a *c-flag*, representing the mother node of the predicted nodes in the set. After a complete phrasal node was found in the input string and thus removed from the set of predicted nodes, the recovery of the c-flag causes this phrasal layer to be put in the well-formed substring table to make it reusable for possible reanalysis operations. The construction flag can then be deleted from the prediction stack making the next set of nodes become visible again.

5.3.8.1 When is a construction flag deleted?

Construction flags representing *clauses* have to be treated differently, however. Clause boundaries are known to behave as "barriers" for a variety of linguistic and psychological phenomena. Whenever a clause boundary is reached, certain *compacting* operations seem to be initiated that reduce the irrelevant structural or phonological waste, such that only the content of that clause has to be maintained. However, as the content has been integrated into the discourse model, it is then further elaborated and integrated. There are a number of studies that provide evidence in support of these assumptions. Caplan (1972) found that when subjects had to decide whether or not a certain word had occurred in a sentence, such as *night* in (217), they responded significantly quicker after they listened to sentence like (217a) than after a sentence like (217b), even though the distance of the word to the end of the sentence was exactly the same.

(217) a. Make your call after six, because *night* rates are lower.

b. Whenever one telephones at *night*, rates are lower.

The surface representation of the final clause was obviously held in a more readily accessible state than that of the preceding clause.

Furthermore, Jarvella (1979) demonstrated that people have better verbatim memory for clauses that they are currently processing than for previous clauses. Subjects made reliably more errors in recalling the surface structure of a fraction of material when they were asked to do so after a clause boundary succeeding the material, than when they were asked before the end of the clause had been reached. Reproducing the sentence *meaning*, however, was not affected by clause boundaries.

Finally, sentence boundaries seem to induce ultimate integration and re-checking processes, as they show up in sentence wrap up effects (Just and Carpenter, 1987) in many reading studies.

To sum up, there is plenty of evidence showing that clause boundaries are special. The most straightforward way to account for clause boundary phenomena in the framework of SOUL is to apply the integration and compact operations (which have yet to be specified) whenever a clausal c-flag has to be purged from the prediction stack. Note that integration operations may proceed even when a clausal c-flag is reached, indicating a *potential* clause boundary. Compacting operations, however, must be restricted to circumstances with strong evidence that the clause boundary is actually reached, since it would be difficult to integrate further sentence adjuncts if the constituent structure has already been deleted. Such "strong" evidence could stem from punctuation in written language or certain intonation markers in spoken language, or from the fact that there is no other possible option for integrating an item into the clause. Accordingly, a clausal c-flag can only be released when the existence of a clause boundary has been verified. Predicted constituents outside the current clause will thus not be visible for the active sign until strong evidence for a clause boundary becomes available⁹².

For non-clausal c-flags, it is unclear whether they should be released early or late (i. e. when it is reached for the first time or only when it has to be passed for external reasons, respectively). Unless there is independent evidence for a delay in releasing, I will assume early release for non-clausal c-flags⁹³.

^{92.}The treatment of clause boundaries proposed here is reminiscent of Milsark's (1983) suggestion to limit the size of the window of the PPP in the *sausage machine* (Frazier and Fodor, 1978, see section 2. 4) to a *clause*, instead of six words, in order to avoid the need for stipulating *Right Association* as a principle operating within the limits of the PPP.

5.3.9 Ambiguity resolution

To illustrate how the parsing mechanism operates and how ambiguities are resolved, I will proceed through some examples. Consider (218) first.

(218) Peter beobachtet den Polizisten mit dem Fernglas.

We will ignore the beginning of the sentence for a moment (see section 5.2.10) and step in after the verb *beobachtet* (*watches*) and the direct object NP *den Polizisten* (*the policeman*) have been integrated into the structure. When the preposition *mit* (*with*) is read, it will attempt to attach itself to the current sign. (For expository reasons, I assume the preposition already to be projected to a PP here.) *Predicted attachment* is always attempted first, and thus the active sign looks for a visible node on the prediction stack of the current sign to unify with.



Fortunately, an instrumental PP has been predicted at that position (i. e. it is on top), which the active sign can easily unify with, resulting in structure (220).



^{93.} Note, however, neither of the predictions under consideration in this thesis will depend upon the outcome of this question.

We skip the integration of the next determiner and step in when a noun is to be integrated. (Again, the lexical sign *Fernglas* (*binoculars*) is assumed to be projected one level upwards into a N'-phrase already.)



As before, the attachment succeeds during its first attempt, since an N' had already been predicted, such that structure (222) is built.



Finally, only the c-flags are left on the prediction stack, which may now be released one-by-one from left to right (top to bottom), except the clausal c-flag [*S]). For each c-flag, the corresponding phrase is added to the well-formed substring table.

Consider sentence (223) next.

(223) Peter erblickt den Polizisten mit dem Fernglas.

The only difference is that the verb has now been replaced by *erblickt* (*caught sight of*) whose instrument role is only lowly salient. The PP will thus not be predicted, as we can see in (224). We step into the process when a PP fragment attempts to attach itself into the structure. The direct object NP was the last predicted complement so that all but the clausal c-flag had already been released.



Since no PP is visible, predicted attachment fails and the PP has to resort to *active attachment*.

5.3.9.1 Active attachment

Within *active attachment*, the active sign makes use of its pointer to the left neighbor word. Neighbor pointers are established whenever a new active sign is instantiated. The *neighbor* pointer defines the starting point for the active search for an attachment site. If the node currently pointed to does not provide an attachment site, the process is passed on to the mother node in the syntactic tree structure, while the active sign's *neighbor* pointer is set to the new node. However, proceeding to the mother node is only permitted if the c-flag (on the prediction stack) corresponding to the current node can be released. Thus, if there are obligatory requirements left below the current node, proceeding to the mother node is prohibited.

The active attachment search generally provides two ways for finding an attachment site (a third one will be introduced later) at each node. During the first, called *projected attachment*, the current *neighbor* node is requested to pass the set of yet unfilled unpredicted daughters. (Remember that each node has a pointer variable set to the set of its *unpredicted* daughters, which has been set during the scanning procedure in a straightforward way.) The request will usually succeed at nodes at the X' level (but also at *schema3*, of course), since phrases at this level can have complement daughters.

If a non-empty set of unpredicted daughters is returned, the set is scanned for a matching (*unifiable*) sign node. Nodes that do not match are deleted from the set. If a consistent node is found, however, the active sign unifies itself with it. If the attach-

ment has to be revised during a later stage of integration, the node is deleted from the set as well, such that the active attachment search can step in at the right point.

Secondly, if *projected attachment* fails the current *neighbor* node is asked whether it is permitted to *adjoin* (cf. Joshi, 1985) the active sign *below* it. Permission is generally only given either at nodes representing maximally projected (saturated) phrases, or nodes representing an instance of the adjunct scheme, such that later adjuncts will appear within the scope of former adjuncts⁹⁴. Nodes that allow *adjoining* are called *entry-nodes*.

Adjoining

There are two alternative mechanisms for adjoining, depending on the underlying grammar. If one assumes *flat* phrases, i. e. phrases that may branch for multiple complements, these phrases must be allowed to take (multiple) adjuncts as well, since adjuncts may be mixed with complements in the German Mittelfeld (cf. Kasper, 1993). Therefore, each phrase does not only have a list of complement daughters, but also a list of adjunct daughters. Syntactically, adjoining simply amounts to adding another constituent to the list of adjunct daughters, such that no re-structuring is required. Semantically, however, the situation is more complicated. Since adjuncts are assumed to be functors that take their heads as arguments, adjunct daughters have been proposed to be the *semantic heads* of phrases. This, however, renders the assumption of multiple adjunct daughters of a single phrase somewhat difficult to handle (see Kasper, 1993, for suggested solutions).

The adjunct phrase in standard HPSG, however, is binary branching, i. e. no further daughters are permitted besides one head daughter, the *syntactic head*, and one adjunct daughter, the *semantic head*⁹⁵. With such an adjunct scheme, adjoining a constituent to a phrase that has already been constructed requires the structure to be split at some point. Since this is the process I adopted in my approach (at least for NPs), I will briefly sketch the procedure in the next sections.

The second version of *adjoining* proceeds as follows: if an *entry-node* is found during the active search, a copy of the binary branching *adjunct-scheme* is instantiated with its HEAD-DTR set to the HEAD-DTR of the entry-node, and its ADJ-DTR (adjunct daughter) set to the *active sign*'s root node⁹⁶.

^{94.}Note that this second sub-procedure could be canceled if we also canceled the notion of adjuncts in general. In this case, what is now called *adjoining* would be subsumed by the first procedure, the search for unpredicted complements.

^{95.}Standard HPSG is therefore a somewhat inconsequent mix of flat phrases with binary branching phrases, which, especially for German, produces a variety of problems yet to be solved.

^{96.}The adjunct scheme passes its entire subcategorization requirement from the head daughter to the mother. In terms of the *Tree-Adjoining Grammar* (Joshi, 1985), the adjunct scheme thus behaves like an *auxiliary tree*, in that the root node (the mother) has the same syntactic category and bar-level as one of its foot nodes (the head daughter). Note, however, that the adjunct daughter is supposed to be the *semantic head* of an adjunct phrase in HPSG.

Finally, the HEAD-DTR of the entry-node is then set (modified) to the root node of the instantiated *adjunct scheme*, as illustrated in (225).



Once the active sign has been adjoined, the *isLexicallyFilled* instance variable of the entry-node is set to the value of the variable of the former active sign, which is NO in this case. Each node recursively notifies its mother node to set its *isLexicallyFilled* to this value as well.

Note that since we operate with complex feature graphs and not with *descriptions* of mere syntactic structures, as in D-theory (Marcus, Hindle & Fleck, 1983), *adjoining* is a *non-monotonic* operation, although non-monotonicity is limited to the constituent structure portion of information in the sign's feature structure.

Let us continue proceeding through example (223) to further illustrate the mechanism. The first *phrasal* node in the right edge of the current sign, the N' projection of the last preceding word, does not provide any unpredicted nodes (since *Polizist* does not take arguments), nor does it permit adjoining. The search thus proceeds to the mother node (NP[acc]), as illustrated in (226). This node is a maximal projection and thus permits *adjoining*.



The adjoining procedure applied to the active sign at the entry node reveals structure (227).



Note that the entire right edge of the structure was put onto the prediction stack after scanning, such that the modified phrases can eventually be added to the wellformed substring table again.

When the words *dem Fernglas* are read, they are integrated into the predicted NP, of course. Since all predictions have been satisfied then, all c-flags are released again except the last element, the clausal c-flag. Suppose a *full-stop* (".") is read next. A full

stop is certainly strong evidence that the sentence boundary has been reached. When the clausal c-flag has to be deleted, any empty constituent that is marked as optional is pruned from the tree⁹⁷, as illustrated in (228).



To sum up, we have seen how the prediction mechanism determines the visibility of attachment sites for the active sign. The SOUL mechanism makes use of the *salience* of thematic roles to offer predictions about future attachments. Whenever the active sign attempts to attach itself to the current sign, the predicted nodes are the first choice. Only when no valid prediction is offered, does the active sign start to scan the right edge to the current sign from the bottom to the top for a potential attachment site. In the current examples, in which a PP could be ambiguously attached to either the VP or the preceding NP, the initial attachment is determined by the lexical properties of the verbs, namely, the salience of the instrument role. If a PP is salient, as in the case of *beobachtete* (*watched*), it will be predicted and then found in the input string, resulting in the structure given in (229).

(229) Peter [beobachtete [den Polizisten] [mit dem Fernglas]].

In the case of *erblickte* (*caught sight of*), however, the instrument role is only weakly salient and thus no *with*-PP is predicted, such that the PP can only attach itself actively, succeeding in adjoining itself to the direct object NP.

5.3.9.2 Is the attachment preference a function of the verb's predictions only?

There are a variety of aspects in the mechanism outlined so far which have not been discussed and exemplified yet. First of all, not only verbs can pose predictions. Some nouns take arguments as well, in particular derivatives from verbs. The noun *discoverer*, for example, is supposed to be derived from the verb *to discover*, and since the verb takes an argument, the noun can be considered to take an argument as well.

^{97.}Of course, if we assume that the entire constituent structure is deleted at sentence boundaries, pruning of single constituents would be obsolete. For expository reasons, however, I take the more conservative view of pruning here.

(230) Carlos entfernte den Entdecker von Amerika gestern von der Liste der eingeladenen Gäste.

Carlos deleted the discoverer (of/from) America yesterday from the list of the invited speakers.

What are the predictions of the SOUL mechanism for the initial attachment preference of the ambiguous PP *von Amerika* in sentence (230)? By the time the preposition *von*, which can be taken by both the verb *entfernte* and the noun *Entdecker* in German, is activated, the structure and prediction stack given in (231) have already been assembled.



This time, there are two predicted PPs on the prediction stack (the arcs indicate to which node in the structure they point). Clearly, the noun's predictions *overshadow* those of the verb, which are thus invisible to the active sign. The preferred attachment is then the one to the most recent head, despite the verb's expectation. It should have become clear by this that the SOUL mechanism predicts a preference for *low attachment* in all constructions in which the lower head itself poses (matching) predictions, like the noun *interest* in sentence (232)⁹⁸ or the verb *eat* in (233).

- (232) Steve discussed his interest in the Volvo. (Abney, 1987)
- (233) Yuki gave Don who was eating {the food, the cider}. (Kamide and Mitchell, 1995)

^{98.}Note that the initial argument-attachment prediction has been questioned by Clifton, Speer and Abney (1991). This study was discussed in detail in section 4.1.3.1, where it was argued that the data the authors provide allow for fairly different interpretations.

5.3.9.3 Is the attachment preference determined by the lexical predictions of the heads that serve as potential attachment sites?

Consider the sentences (234) and (235).

(234) John put the book Bill was reading in the study. (Fodor & Frazier, 1980)

(235) Joe carried the package that I included for Susan. (Ford et al., 1982)

In (234), the verb *put* obligatorily subcategorizes a Locative-PP. The second verb, however, does not, so the PP *in the study* should be attached to *reading* as an adjunct. Quite similarly, in sentence (235) the "high" verb (*carried*) seems to predict a *for*-PP (Beneficiary), while the low verb (*include*) does not (Ford et al., 1982). If the predictions of the heads alone determined the preference to attach the PP, it should clearly be attached to the high VP in both cases. This does not seem to be the case. Most native speakers and the majority of researchers (e.g. Fodor and Frazier, 1980; Ford et al., 1982; Abney, 1987, 1989, Mitchell, personal communication) share the intuition that the PP is preferentially attached *low*, i. e. to *reading* in (234)⁹⁹ and to *included* in (235), respectively.

^{99.}Unfortunately, there has, to the best of my knowledge, never been an on-line reading study run to investigate this issue, which is quite bizarre in the light of the fact that these structures have been argued with for more than fifteen years. Some researchers have claimed to have different intuitions about such sentences (e. g. Gibson, 1991; Stevenson, 1995). Gibson felt that a sentence like *"Janet put the book Lyn was reading in the study in the rack"* causes some difficulties at the second locative PP and concludes that the preference for the first PP must have been *high* attachment. This conclusion, however, is by no means compelling, since the difficulty might just as well have been caused by the fact that the second PP might have tried to attach itself low, namely to the noun *study. Study in the rack*, then, might have caused plausibility problems such that the second PP as well as the first PP had attempted to attach themselves to the low VP, which doesn't work because the locative role had already been occupied by the first PP (assuming that it is not possible to specify two completely different locations for the same event).

What does the SOUL mechanism predict? Let us have a look at the structure built from sentence (236) when the preposition *in* wants to attach itself.



Recall that clausal c-flags cannot be released from the prediction stack before there is "strong" evidence that the clause boundary has been reached. Therefore, all elements on the prediction stack underneath the clausal c-flag including the predicted locative PP are overshadowed, i. e. not visible to the active sign. The most visible attachment site for the PP in the sentences (234) and (235) is thus the low VP, meeting the intuitions about attachment preferences in these sentences.

So far, the SOUL mechanism has been demonstrated to account for a variety of attachment preferences. The preferences were shown to be determined by the lexical prediction of the heads involved, their order in the input string (their position in the structure) and the major category of the head, as in the recent example of low attachment to a verb, whose maximal projection is a sentence which overshadows predictions from higher heads.

The attachment ambiguities considered so far share the property that the lexical heads of the potential attachment sites precede the ambiguous phrase to be attached. In the next sections, it will be demonstrated how the SOUL mechanism deals with constructions, in which at least one potential attachment site has its lexical head yet to come when a constituent is about to be attached. In order to treat such constructions properly, we will have to take a short excursion to the general problem of parsing such constructions incrementally with a lexicalized head-driven grammar, such as HPSG.

5.3.10 Incremental parsing with a head-driven grammar

Modern grammars like HPSG differ from traditional rule-based grammars in one important respect, among others. In order to avoid redundancy on the one hand and to achieve a higher level of generalization on the other, grammatical information is no longer coded in a set of *constructive* grammar rules (e.g. context free phrase structure rules), which could be used to generate structure. Instead, grammatical information is organized in a hierarchical system of lexical sorts of, for example, subcategorization information, leaving only few general *constraints* on the well-formedness of structures outside the lexical system.

Although both types of grammars (*constructive* vs. *constraint-based*, see Crocker & Lewin, 1992) can be shown to be at least weakly equivalent in generative capacity, both approaches differ strongly with respect to the parsing processes they imply.

In languages like German, verbs often succeed their complements and adjuncts at the surface. Whereas the assembly of structure in *constructive* grammars is independent of the notion of heads and complements and their particular order, parsing such structures with a lexicalized, constraint-based grammar in a transparent fashion (Berwick and Weinberg, 1985) is not at all straightforward. Without the head-information of the particular piece of structure to be built, it is at least extremely inefficient, if not impossible, to generate (project) structure, since nothing prevents the parser from massive overgeneration in the absence of lexical constraints.

Parsing strategies have been proposed which reflect the properties of lexicalized head-driven grammars in parsing, both in computational linguistics (e. g. *head-driven parsing*, Kay, 1989) but also in psycholinguistics (e. g. *licensing structure parsing*, Abney, 1987, 1989). Though considerably different, all such approaches share the view that structure building is restricted to circumstances in which it is *licensed* by the head. An incremental parser proceeding through the sentence from left to right would thus have to delay the attachment of constituents to the structure.

In chapter 4.1, I provided strong evidence contradicting these predictions of headdriven parsing approaches (see also Hemforth, 1993; Hemforth, Konieczny & Strube, 1993; Bader and Lasser, 1994). For a *head-driven grammar* framework, a problem then emerges: how can structure be assembled (efficiently) in the time course of sentence processing in a *linear* fashion (i. e. word-by-word, with each word being integrated into the sentence structure immediately, Konieczny & Strube, 1995) while the lexical head is still absent?

In the next paragraphs I will present some approaches to this problem (see Konieczny & Hemforth, 1994), some of which arise from within the linguistic system of the HPSG itself, whereas in other cases additional assumptions have to be made in order to let the proposed process of incremental interpretation meet to the experimental findings.

5.3.10.1 SPEC and MOD in noun phrases

There are some features internal to HPSG and motivated by independent linguistic reasons which allow for incremental parsing in some restricted circumstances. I will briefly discuss how such facilities can be employed in parsing noun-phrases, such as (237).

In "standard" HPSG (Pollard & Sag, 1994), and contrary to many other approaches, nouns are considered to be the heads of determiners and not vice versa, i.e. noun phrases are assumed, as opposed to determiner phrases (Abney, 1987; but see Netter, 1993, for a DP analysis in a HPSG-style grammar)¹⁰⁰. Consequently, NPs are head-final phrases with determiners (specifiers) being arguments of the noun. For reasons of semantics that have to do with quantifier scope (see Pollard and Sag, 1994), determiners are equipped with the head feature SPEC (*SPECIFIED*). When the determiner is projected into a schema as a specifier (details following), the SPEC value is unified with the SYNSEM-value of the head-daughter (via the SPEC-principle, Pollard and Sag, 1994). Since a noun subcategorizes for the determiner in turn, if projected into the schema as a head daughter, SPEC information can thus be regarded as *inverted subcategorization information* (ISI, see Konieczny & Hemforth, 1994). In the following example I will briefly illustrate how this information can be employed to build structures incrementally.

(237) den guten Schauspieler

the good actor (accusative case)

First, the determiner *den* is lexically constrained to combine only with a noun (N') that bears the disjunctive agreement restrictions *singular, masculine, accusative,* or *plural, dative.* Since *den* is a lexical sign without any subcategorization demands, the determiner can be maximally projected to DetP by virtue of ID-*schema3.* As a maximal projection the DetP will then succeed in projecting itself as the subject-daughter in *schema1.* After applying (unifying) the principles of well-formedness to the phrasal fragment, in particular the *SPEC principle,* the SYNSEM-value of the head daughter is set token-identical to the SPEC-value of the determiner (tag [2] in (238)). The *Head Feature Principle* (HFP) results in the unification of the head value of the head-daughter with that of the mother (tag [1] in (238)). Consequently, the SPEC-information in determiners causes a fragmentary NP to be built, although its lexical head, the noun,

^{100.}We will not discuss the validity of this assumption in this paper. Although DP-alike approaches might turn out being more adequate, the standard approach provides a welcome occasion to demonstrate transparent incremental parsing of head-final structures. Note, however, that in a DP-alike analysis this problem wouldn't even occur.

has not yet been processed. The predicted N' bears the agreement restrictions specified at the determiner.

(238) Subject-projection of "den" (abbreviated)



Pre-nominal modifiers operate in a similar fashion. Inflected attributive adjectives like *guten* are unambiguously adjuncts. Despite their syntactical optionality, adjuncts in HPSG are supposed to be the *semantic heads* of a phrase. It is their content-value that is passed to the mother, not that of the head-daughter. Since adjuncts semantically modify the head-daughter, they come with the head feature MOD (MODIFIED), which is quite similar to the SPEC-feature of determiners and markers. However, in contrast to these, adjuncts are *per definitionem* not subcategorized for. An AP (a saturated adjective phrase) succeeds in projecting itself as an adjunct-daughter in the *adjunct-schema* (schema5). In this schema, the head of the phrase is constrained by the unification of the MOD-value of the adjective with the SYNSEM-value of the head-

daughter (tag [2] in (239)). As a *phrasal* sign the fragment has to be unified with all principles of well-formedness that can be applied to it.

(239) Adjunct-projection of the adjective "guten" (abbreviated)



The application of the *Head-Feature Principle* (HFP) will eventually pass the projected head features to the mother (tag [1] in (239)), resulting in a N' phrase.

Since the resulting phrase is a non-saturated N' looking for a subject, it can be unified with the predicted N' head daughters of the detP-projection (238), as illustrated in (240).

(240)



Finally, when the noun *Schauspieler* (*actor*) is read, it can be head-projected and directly integrated into the partial phrase resulting in structure (241).



After the integration the NP is constrained to be unambiguously accusative cased (singular), since the dative-plural variant is ruled out by the noun *Schauspieler* (242).

(242)



5.3.10.2 Inverted subcategorization information

Hence, a noun phrase can be constructed incrementally even before its head has been processed. The head-selection features SPEC and MOD are HPSG-specific devices which undermine head-licensing restrictions to a certain degree. From a processing point of view, especially SPEC represents the interesting idea of inverted subcategorization information (ISI). If we only had a comparable device for "usual" complements, we could propose, for example, fragmentary VPs or sentences from potential verb complements, such as the subject-NP or any of the objects.

Let us follow this idea for a little while. Suppose we let the lexical rules that generate morphology and case assignment equip nouns with a feature, call it IP-SPEC, for purposes of *incremental processing*. Since nominative case nouns, for example, cannot have any other function than to be the subject of a sentence, it can be assumed that they IP-specify an unsaturated V', just as determiners specify a N'. Thus, unifying a nominative NP with the principles (including an IP-equivalent to the SPECprinciple) after projection would automatically create a sentence fragment with its subject instantiated, analogous to the construction of an NP in the section above. In these cases, the parser makes use of the principles and schemata of the competence grammar directly, i.e. the grammar system does not need to be compiled into a more "parsable" format like context-free phrase structure rules¹⁰¹.

On the other hand, the lexicalized approach to head-prediction is associated with a variety of drawbacks. Firstly, lexicalized ISI information is clearly redundant, since the constraints it imposes are completely covered by subcategorization information. The second problem is associated with the fact that morphological information is often highly ambiguous. The German article "die", for example, is not restricted to the nominative case, but can combine with several casus, depending on the constellation of the other agreement features. Nouns are often morphologically even less informative. Since the ISI information of a noun is tied to the case it can take, the lexical representation would have to be vastly enlarged, since each CASE alternative would have to be doubled by an IP-SPEC disjunct. It might therefore be advisable to consider further approaches.

5.3.10.3 Generalized sentence templates

We have seen that HPSG, as an instance of a lexicalized head-driven grammar, cannot account for phenomena of linear parsing such as the subject-initial expectation in a completely *transparent* (Berwick & Weinberg, 1985) way. Since at least some information has to be added in order to allow adequate processing, one might argue that the processing mechanism works on pre-compiled structures instead of the representations HPSG provides. A number of proposals have been made about what representations pre-compiled from a HPSG grammar might look like (e. g. Neumann, 1993; Netter et al. 1994). The approach taken here, however, is to keep to the minimum amount of pre-compiled structures needed to account for the data, i. e. to remain *maximally* transparent (cf. Konieczny & Hemforth, 1994).

What seems to be missing is an equivalent to a phrase structure rule like (243) that could pose a "top-down" prediction of a nominative NP into the process even before a verb which eventually assigns a case to its complements is read.

(243) S \rightarrow NP[nom] VP

What seems to come closest is the instantiation of the basic HPSG rule schemata with what could be considered a *generic verb*, a verb the only thing known about being that it takes a (grammatical) subject and, of course, that it is a finite verb.

Example (244) shows (a simplified version of) such a sentence-template for English main clauses, which I have adopted for German main clauses as well¹⁰². Contrary to

¹⁰¹. This, of course, can be regarded as a slight commitment to *type transparency* (Berwick & Weinberg, 1985).

¹⁰² Note that there is a debate still going on about how the German VP should be analyzed (see Uszkoreit, 1986; Kiss & Wesche, 1991; Netter, 1992, Kasper, 1993, Frank, 1995). Since this discussion is far beyond the scope of this thesis, I will restrict myself to the approach given above.

approaches (e.g. Uszkoreit, 1986, adopted by Pollard and Sag, 1994) that assume any constituent in the *Vorfeld* to be fronted by virtue of a topicalization mechanism (e.g. *movement*), I take the SVO structure to be the canonical sentence structure in German that does not entail certain *focus*-related restrictions involved in topicalization (see also Hemforth, 1993; Travis, 1991, for a structural approach in GB). It is thus assumed that the subject-initial construction is plainly base-generated as a maximal projection of a finite verb.

(244) Canonical main-clause template



Complement topicalization, however, is analyzed as suggested by Uszkoreit (1986), Pollard & Sag (1994) and Kasper (1993), i. e. as a topicalization from a flat scheme (*schema3*), as illustrated in (246).

Note that the slashed constituent in the verb (tag [4]) is not yet linked to one of its complements. The template thus works with any complement to be topicalized, since the connection is only established when the feature structure of the verb is unified with the lexical head-daughter, by virtue of the complement extraction rule (see Pollard & Sag, 1994, pp. 378).

As reported in Konieczny and Hemforth (1994), I assume an ordered set of generalized templates, each one representing one of the general sentence types, such as subclauses, wh-questions, and others. It has not yet been decided whether the order function is based on the *frequency* (or *probability*) of the structures, on the *simplicity* of the structures built by means of whether or not a schema involves *extraction* of a com-



(246) Template for main clause structure with a topicalized complement.

plement, or even on certain pragmatic constraints, for instance, that *new* information should not be provided in the subject position, and *given* information should precede *new* information, which, taken together, results in a subject initial preference (cf. Niv, 1993).

5.3.10.4 Parsing with templates

Suppose the parser starts reading a sentence beginning with an NP. As soon as the determiner is read, it will be projected to an NP, as outlined above. This NP then *triggers* the instantiation of one of the sentence templates. In order to find a matching template, the ordered set of sentence templates is serially searched in such a way that the more preferred templates are found earlier than the less preferred ones. Note, however, that I will not claim any specific order here, except that the subject-initial template is the top-most element.

5.3.10.5 Head-final constructions

Note that the VP has not been compiled into a complete set of possible extensions, in order to reduce complexity. If each possible complement structure is compiled into a discrete template, the number of templates currently assumed would have to be multiplied by the number of possible VP-extensions.

All clausal templates are still unspecified for COMP-DTRS (note that the complement daughters are collected in a list structure, which is the value of the (single) feature COMP-DTRS. Thus, the tree node indicated by the branch ending in "-o" in the template structures does not represent a constituent, but the value of COMP-DTRS, i.e. a node in the *feature* structure.

How can complements be attached in verb-final sentences, such as (248)?

(248) ..., daß Peter das Buch las.

..., that Peter the book read.

"that Peter read the book"

Suppose a complementized sentence (a *marked* sentence in the terminology of HPSG) has been predicted before the complementizer *daB* is read. Complementizers (markers) *specify* a sentence (a saturated verb projection) by virtue of the SPEC feature, so that an S-node is projected after the marker has been integrated into the marker-scheme, as illustrated in (249).

(249) projection of a complementizer



Note that the head is marked as -INV (not inverted), meaning that the finite verb has to occur in the final position.
When the marker is attached to the predicted S'-node, an S-node is predicted, as shown in (250).



Schema3 is the only ID-schema that allows the lexical head to occur in the final position. When the next word, the proper noun *Peter*, is read, it triggers the instantiation of a sentence template that should be compatible with *schema3*. This template is given in (251).



The NP *Peter* is attached as the subject NP, resulting in the structure given in (252). The next word, the determiner *das*, has to resort to the *active attachment* search in order to attach itself to the structure, because the prediction stack does not provide a matching attachment site. When the search arrives at the VP, there is no set of projected (but unpredicted) daughters.



In this case, the process will access the value of COMP-DTRS, which is restricted to sort *verb-comps* via tag [4] in (251), as illustrated in (252).

Verb-comps is the most general sort for complement structures. In the lexical sorthierarchy, the sort *verb-comps* is partitioned into several subsort-levels that correspond to valence-classes of verbs such as *intransitive, transitive, ditransitive* and so on, as in (253), which constrain how the complement structure can be extended¹⁰³.

(253) Sort-hierarchy of verb-complements



¹⁰³·For reasons of simplicity I assume that complete signs are subcategorized instead of *syn-sem*-structures. The *verb-comps* hierarchy in this case is a part of the *verb-subcat* hierarchy.

The sort name is used as a pointer to a node in the hierarchical system of sorts. Here, a subsort might be found that allows for the active sign to be integrated. Starting at that node, the sortal hierarchy is searched through until either a sort is found that requires (at least) this item as a subcategorized-for element. Such a sort is immediately found, and thus the sort *verb-comps* is further narrowed down to *transitive*.



Since *transitive* is unspecified for the REST-feature (i.e. the list is open-ended), a further object NP could further restrict the sort to *ditransitive*, and so on. The sortal restrictions thus constrain the type and number of constituents that can be attached to a head-final phrase, and therefore allow a structure to be built in a tractable manner. In this regard, they serve the same purpose as phrase structure rules in a standard parsing approach. Note, however, that the sort-hierarchy is a system of *lexical* sorts, i.e. its information is directly linked to the lexical entries and motivated by the requirement of expressing linguistic generalizations to be used in the *lexicon*.

The sortal information enables us to build complement structures without a head. Additionally, when the complement-structure becomes more and more restricted to subsorts of *verb-comps*, these restrictions are immediately unified into the feature-structure of the verb-head itself via tag [4] in (251). In this way the predicted verb is successively modified to be more specific each time a new complement has been attached. When the verb is eventually read, it only has to be unified with the predicted node without any need for an additional check of subcategorization requirements.

5.3.11 Attachment preferences in verb-final constructions

So far, I have demonstrated that incremental processing and immediate attachment of items can be accomplished even within a head-driven grammar approach, if the grammar system provides facilities, such as the hierarchy of sorts, that can be accessed to overcome the head-licensing approach.

We are now in a position to focus on attachment ambiguities in verb-final constructions, such as (255).

(255) ..., daß Peter den Mann mit der Krawatte fesselte.

..., that Peter the man with the tie fettered.

"... that Peter fettered the man with the tie."

As in the verb-second version of the sentence, the PP *mit der Krawatte* can be attached to the preceding NP or to the VP. We have seen that a rule-based top-down approach employing the principle of *minimal attachment* incorrectly predicts the PP to be attached initially to the VP. The SOUL mechanism, on the other hand, proceeds quite differently. When the preposition *mit* is to be integrated into the structure, there is no visible attachment site on the prediction stack due, of course, to the absence of a subcategorizer in that position. The active sign must then resort to *active attachment* starting at the preceding noun and succeeding at the NP-node, below which it can adjoin itself.



Clearly, the next two words *der Krawatte* (*the tie*) are seamlessly integrated into the predicted NP within the PP, resulting in the initial analysis given in (257).

(257) ..., daß Peter [den Mann mit der Krawatte] ...

The empirical finding of an initial [NP PP] preference is accounted for perfectly. Similarly, in verb-final sentences like (258), the ambiguous NP *der Professorin* bears either *dative* or *genitive* case and must therefore be attached to the VP or to the preceding NP, respectively.

(258) ..., daß der Doktorand der Professorin zusätzlichen Urlaub abgetrotzt hat.

Again, in the absence of a verb that could have imposed another NP onto the stack, the active detP *der* has to resort to *active attachment* and succeeds in adjoining itself to the preceding NP-node. Obviously, the analysis given in (259) will be the first one found, once again corresponding to the empirical findings (see chapter 4.1).

(259) ..., daß [[der Doktorand der Professorin] zusätzlichen Urlaub abgetrotzt hat].

In general, the *visibility* approach implemented by the prediction stack and the active attachment search will result in attachment preferences predicted also by *head attachment*. Whenever a potential attachment site still lacks its lexical head, it will generally be less visible than any other site that lies lower on the right edge backbone. Lower sites, however, will be those whose lexical head have been read previously. In summing up. The SOUL-mechanism entails a *serial* version of *head attachment*. Note, however, that the SOUL-mechanism is not just a special purpose *ambiguity resolution* mechanism (or principle), such as *parametrized head attachment*, but a complete parsing approach operating in a very general manner regardless of whether or not an ambiguity is actually present.

5.3.12 Modulated lexical guidance

Lexical expectations have been demonstrated to influence initial parsing decisions. As argued in chapter 4.1.3.1, lexical information will not prevent the parser from ungrammatical attachment *hypotheses* in every case. Recall sentence (260), for instance.

(260) After the child sneezed the doctor prescribed a course of injections.

The empirical findings on such sentences suggest that the NP following the first verb (*the doctor*, in this example) will *not* automatically be hypothesized to be the subject of the embedded clause, even though the strict intransitiveness of the verb *sneezed* could have guided the parser to do so. As illustrated in (261), the NP fragment pro-

jected from the determiner *the* is forced to search for an attachment site actively, as no NP is visible on the predictions stack.



Note that even though a sentence node is predicted somewhere on the stack, it is overshadowed by the clausal c-flag of the embedded sentence. (The NP will therefore not *trigger* the instantiation of a sentence template that could have provided an attachment site for the subject.) During the attachment search, then, the NP will reach the VP node and request unpredicted complements of the verb, most likely without success in the case of *sneezed* (note, however, that *transitivity* information does not seem to be as restrictive as it is supposed to be in all cases, e. g. in "*the child sneezed the napkin off the table*.").

If the active sign succeeds in discovering a complement to unify with, even if that would require an extremely rare lexical form of the verb to be active, it will unify itself with this complement. In this case, the attachment might only be annulled when the following noun provides further information that allows for plausibility effects to arise. If no complements are found, however, the active sign will attempt to *adjoin* itself to the structure, with a certain probability of success. Again, plausibility might only rule out such an attachment at the end of the NP then. Only if the active attachment search eventually passes the S-node can the clausal c-flag be released from the stack and the predicted matrix-clausal S-node become visible, finally permitting the attachment as the subject NP of the embedded sentence.

To sum up, the subcategorization restrictions of the verb cannot prevent the active sign from *attempting* to attach itself to the VP. It can nevertheless prohibit the actual attachment, if it strictly rules out further complements. It must be stressed that the SOUL mechanism will *at no stage of processing* permit ungrammatical attachments. It is thus considerably different from Mitchell's two stage approach (Mitchell, 1989), which allows a structure to be built at the first stage, even if subcategorization information could have blocked the attachment. Only in the second stage is this information employed to check the attachment and to revise it if necessary. Such a *filtering* account must be distinguished from a single-stage lexical approach like the SOUL mechanism outlined here. SOUL can be considered a lexical guidance model, because

i. lexical properties determine the visibility of attachment sites which further (partially) determine the initial attachment (note that in the case of "... *erblickte den Polizisten mit dem Fernglas*" the VP-attachment is not even attempted in the first run), and *ii.* all kinds of lexical information are utilized at once to block inviable attachments.

However, lexical information may only guide the parsing process within the limits defined by the SOUL mechanism; it cannot modify the basic way in which the SOUL mechanism operates, i. e. the order in which the basic operations are applied, so that a sign will never, for example, *project* itself further upwards before it attempts to attach itself to the current structure. Lexical information can only influence parsing decisions in circumstances in which the SOUL mechanism gives it the opportunity to do so. In that sense, SOUL takes a *modulated lexical guidance* approach.

5.3.13 Revision of initial attachments

So far we have seen how the SOUL mechanism processes sentences until a first parse is found. The basic mechanism accounts for initial attachment preferences obtained in a variety of ambiguous sentence types. But how does the parser recover from a false analysis?

(262) Hans erblickte den Leoparden mit dem Fernglas.

First, consider sentence (262). The preposition had been attached to the preceding NP during the active attachment search. The noun *Fernglas* (binoculars) however, will probably cause the *interpreter* to reject this attachment due to its limited plausibility. The rejection causes the active sign to recover another option to proceed at the latest decision point. The latest decision point was the state before the attachment to the predicted N' node was established. Which further options does the active sign have at that point? In this particular example, not many. As a first attempt, it can initiate the active attachment search, but since the predicted N' node is marked *obligatory*, the NP node may not be passed. Since no attachment site for the noun could be found, it attempted to project itself further upwards. However, this is prohibited because the noun has already been attached to an obligatory, predicted node. In general, it can be shown that if such an attachment has already taken place, no further projection is required¹⁰⁴.

Since projection is blocked as well, there is no option left for the noun other than to recover its predecessor state, i.e. the state in which the NP projected from the determiner had been attached to the NP node predicted by the preposition. Again, the active attachment search fails immediately since the PP lacking an obligatory NP node must not be surpassed. As before, then, projection at any level of this item is also prohibited, since it had already been attached to an obligatory, predicted node.

^{104.}Note that since adjoining can be accomplished lazily, i.e. initiated on demand by a subsequent item that wants to be adjoined, there is no need to allow recursive head-projection in an adjunct scheme to *predict* an adjunct in advance.

Next, a state is recovered in which the PP projected from the preposition is about to be adjoined to the NP. At that point, the active attachment search *may* surpass the NP since there are no obligatory requirements blocking the process to proceed further upwards. The active search then reaches the VP and discovers the non-empty set of projected, but unpredicted complements of the verb, including the instrumental PP, which turns out to be unifiable with the active sign. When the control passes to the current sign to predict the NP and initiate a new item to be read, it will firstly look into the well-formed substring table for an NP starting at the current position. Since the NP has already been processed, it can be integrated seamlessly.

To sum up, the revision of the initial attachment turned out to be easy, firstly because the amount of backtracking was limited to the small number of options to be proceeded through, secondly, because an alternative attachment site for the ambiguous PP could be found by simple continuing the active attachment search at that point were it failed, and thirdly, due to the fact that a constituent (the NP) could be recovered from the well-formed substring table. Similarly, recovery from the wrong analysis is equally easy in sentence (263).

(263) Hans beobachtete den Leoparden mit der Wunde.

As we have seen above, *beobachtete* predicts an instrumental PP. The *interpreter* will probably reject the noun *Wunde* as a plausible instrument of *beobachten*. Backtracking is thus initiated, proceeding just as sketched for the previous example. When the preposition is reached, the PP initiates the active attachment search, finally succeeding in adjoining itself to the NP *den Leoparden*. Again, the NP can be recovered from the well-formed substring table.

Lowering

The active attachment search proceeds along the right edge of the current sentence structure. As we have seen, signs may adjoin themselves to appropriate hosts. There are some cases which appear to be easily processed even if they contain an item that fails to attach itself actively to the structure. Consider sentence (264), which is an instance of the type of sentence discussed in various places in chapter 4.

(264) John heard the noise was due to a design flaw (Gorrell, 1996)

Although some processing difficulty is observable at the second verb *was*, the sentence is by no means hard to process. The NP *the noise* is first attached as the complement of *heard*, such that *was* lacks a subject. Several proposals have been made in psycholinguistic literature to account for the ease of the reanalysis in these constructions. Abney (1989) has proposed the *steal action*, which allows a constituent accessible on the right edge of the structure to be re-attached to a subsequent item. Similarly, in Pritchett's model, the NP can be reanalyzed within the On-Line Locality Constraint (see chapter 2.9). D-theory based models (Marcus, Hindle and Fleck, 1983; Weinberg, 1993, Gorrell, 1995, Sturt and Crocker, 1995) take (264) to be an instance of *lowering* where the NP can be "reanalyzed" by adding further *dominance* predicates to the structure description.

The approach taken here adopts the *lowering* approach (Sturt and Crocker, 1995), however without adopting structural determinism and d-theory, for reasons discussed in section 2.13 and 4.2.1. While *right-edge availability* (cf. Abney, 1989) is a crucial factor, another one is whether the item to be attached can easily substitute or replace the lowered constituent in the structure. In (264), thus, the NP-complement can simply be replaced by a sentential complement. Now consider (265).

(265) While Peter was mending the sock fall off his lap.

The sentence is more difficult to process than (264), although the NP *the sock* is also available through the right edge. However, *the sock* has to be reanalyzed as the subject of the matrix clause, which cannot substitute the previous NP.

Lowering, in the more specific sense, is thus incorporated into active attachment: if the active sign is a *head* which requires a complement to the left, it will search for a node on the right edge that *i*. satisfies the requirement and *ii*. can be replaced by the active sign (see Konieczny et al., in prep, for a more detailed discussion).

5.4 Discussion

5.4.1 Coverage of ambiguity resolution phenomena

In this section, I will quickly proceed through a variety of examples of attachment ambiguities to demonstrate the wide coverage of SOUL. The examples have been presented as evidence for the Garden Path model in chapter 2. 5 (cf. Frazier and Clifton, 1996).

For (266b), SOUL predicts a *strong* garden path effect resulting from the prediction of a main verb.

(266) Main clause / reduced relative

- a. [The horse raced past the barn] ?fell.
- b. [The horse [S' raced past the barn] fell].

Recovery would need to re-activate an abandoned lexical form of the verb, which cannot easily be done within the SOUL mechanism.

A garden path is also predicted for (267b), at least for NP-complement biased verbs, as it has been found in the studies of Rayner & Frazier, 1982; Ferreira & Henderson, 1990).

(267) NP versus S complement

- a. John [knew the answer to the physics problem] ?was wrong.
- b. John knew [_S the answer to the physics problem was wrong].

In SOUL, the preference is due to *attach* before *project*, such that the NP *the answer* ... attaches itself before it considers to project itself to an S. Recovering is comparably easy, since *lowering* (Sturt and Crocker, 1995) allows the NP to be reanalyzed as the subject of the sentential complement (Konieczny et al., in prep).

In (268), the difficulty arises from the conjunction being initially attached low due to the *active attachment* search proceeding upwards along the right edge of *John kissed Miriam*.

(268) NP conjunction versus S conjunction

- a. Jacob [kissed [Miriam and her sister]] ?laughed.
- b. [Jacob kissed Miriam] and [her sister laughed].

Sentence (269) is of the same type as the German "*deleted the discoverer of/from* …" example given in (230).

(269) PP-attachment to VP/NP

- a. Sandra [wrote [a letter] to Mary].
- b. Sandra wrote [a letter to Mary].

The complement prediction of *letter* overshadows that of *wrote* resulting in *most recent head attachment*.

Complement sentences, such as *that Bill liked the story* in (270a) are certainly predicted, while relative clauses, such as *that Bill liked* in (270b), are not.

(270) Complement / relative clause

a. John [told the girl [that Bill liked the story]].

b. John told [the girl that Bill liked] the story.

Predicted attachment precedes *active attachment*, resulting the obtained preference for (270a). Recovery is comparably difficult, as in (266), since the attachment of the ambiguous *that* to the (top-down) predicted complement sentence as a complementizer effects the lexical alternative (relative pronoun) to be abandoned.

In (271), clearly, the NP *the dog* will favor attaching itself as a predicted complement over projecting itself to a (reduced relative) clause.

(271) Attachment of NP as second object / relative on first object

- a. Fred [gave the man the dog] ?bit the package.
- b. Fred gave [the man the dog bit] the package.

In (272), the predicted attachment of the NP as a complement of the verb *mending* is favored over projection.

(272) Direct object versus subject of S2

- a. While [Mary was mending the sock] ?fell off her lap.
- b. While Mary was mending [the sock fell off her lap].

Sentence (273) has already been dealt with in example (236). A clausal construction flag overshadows the prediction of *put*, leading to a preference to attach *in the library* low.

(273) Attachment of PP to lower clause / higher clause

a. I put [the book that you were reading in the library]?.

b. I put [the book that you were reading] in the library.

As an adjunct, the *when*-clause in (274) will take the lowest attachment site, leading to the tempus mismatch at the end.

(274) Attachment of S to lower clause / higher clause

a. Fred will realize [that Mary left when the party ?starts].

b. Fred [will realize [that Mary left] when the party starts].

Similarly, the adverbial *yesterday* in (275) is incompatible with the *future* tempus of the most visible attachment site, the lower clause.

(275) Attachment of adverb to lower / higher clause

- a. We remembered [that the assignment will be due ?yesterday].
- b. We [remembered [that the assignment will be due] yesterday].

5.4.2 SOUL and the Garden Path model

The GP model and the SOUL mechanism, and various others, share the view that only one analysis is pursued in the case of a local ambiguity. Both the GP-model and SOUL operate in a serial depth-first fashion. There is also a certain resemblance between *Predicted Attachment/Active Attachment* and *Minimal Attachment/Late Closure*, respectively. Predicted Attachment will usually cause *high* attachments while *Active Attachment* and *Late Closure* share the focus on *recency*. The similarity is only superficial, however. First, there is a recency component in *Predicted Attachment* as well, since only the lowest predictions are visible at all. Second, *Active Attachment* will not necessarily result in the lowest attachment, since all nodes along the right edge are visible as attachment sights.

At first glance, it might also appear that *attach* before *project* emulates *minimal attachment*. Note, however, that although *attach* before *project* and *Minimal Attachment* appear to produce the same predictions for examples like (276a-c), the predictions are different in sentences like (277).

(276) a. John knew the answer was correct.

- b. Jill knew at 7:30 Bill would arrive.
- c. The horse raced past the barn fell.

(277) Peter wußte, daß der Hund der Bibliothekarin eine Wurst gestohlen hat.

Peter knew, that the dog the librarian_[gen, dat] a sausage stolen has.

- a. "Peter knew that the dog has stolen a sausage from the librarian." Or:
- b. "Peter knew that the dog of the librarian has stolen a sausage."

Whereas the GP model predicts the "minimal" interpretation (277a) to be preferred, SOUL predicts (277b) to be preferred due to *Active Attachment* to the lower host.

Similarly, the VP attachment of the PP *mit dem Fernglas* in (278) is preferred in the GP model, whereas SOUL predicts NP-attachment, as outlined in section 5.2.11.

(278) Mir wurde erzählt, daß Martin das Pferd mit dem Fernglas erblickte.

I was told, that Martin the horse with binoculars caught sight of.

"I was told that Martin caught sight of the horse with binoculars."

The major reason for the difference in predictions is that *adjoining* a phrase, e.g. a PP to an NP, is considered an especially time-consuming revision process in the GP-model, but not so in SOUL. *Adjoining* the NP *der Bibliothekarin* to the NP *der Hund* is performed within the mechanism outlined by SOUL, but only during reanalysis in the GP-model. *Minimal Attachment* and *attach* before *project* have an intersecting set of

predictions (of garden-pathing), namely those cases in which only an additional step of bottom-up projection of the ambiguous item succeeds in providing the analysis that turns out to be the correct one in the end (e.g. 276a-c). However, since the retrospective introduction of non-minimal nodes is not necessarily costly in SOUL, nonminimal structures can very well be preferred here, which is never the case in the GPmodel.

Moreover, the attachment predictions for PP-attachment sentences like (279a) are not at all subject to *attach* before *project* in SOUL, but to *Minimal Attachment* in the GPmodel. In SOUL, the initial attachment is determined in these cases by the *visibility* of the potential attachment sites, namely, whether or not the hosts are *predicted* (VP attachment of the PP in 279a, also for the PP *von der Liste* in 279d, NP attachment of the PP in 279b) and, if not differing in this regard, whether one site is more accessible than another, either by being on top of the prediction stack (NP-attachment in 279c, also for the PP *von Amerikas Inseln* in 279d), or being lower in the right edge backbone of the current sign (NP-attachment in 279e).

(279) a. Peter observed the cop with binoculars.

- b. John expressed his interest in the Volvo. (Abney, 1987; Clifton, Speer & Abney, 1991)
- c. El fisico dedujo las conclusiones del experimento. (Igoa, 1995)
- d. Chuck entfernte den Entdecker von Amerikas Inseln gestern von der Liste der eingeladenen Gäste.

Chuck deleted the discoverer (of/from) America's islands yesterday from the list of the invited guests.

e. Marilyn erblickte den Einbrecher mit dem Fernglas.

Marilyn caught sight of the burglar with the binoculars.

The GP-model predicts an initial minimal (VP) attachment preference for all these sentences (thus predicting a garden path for sentence 279d). SOUL predicts a VP attachment preference only for sentence (279a). In summing up, the SOUL model accounts for the data better than the GP-model¹⁰⁵.

5.4.3 SOUL and the Licensing Structure Parser (Abney, 1987, 1989)

SOUL shares a more differentiated view on complements and adjuncts with the *Licensing Structure Parser* (LSP). This similarity, however, is only limited. In SOUL, the treatment of complements depends upon the salience of the referential objects they represent. Thus, complement attachment is not necessarily preferred over adjunct attachment, as can be seen in sentence (279e): even though *with the binoculars* can be a complement of the verb, the adjunct attachment to the preceding NP is preferred ini-

^{105.}Note that the general VP-attachment preference found in many studies is likely to be due to the choice of verbs used in the study. In Rayner, Garrod & Perfetti (1992), for instance, the authors report an initial "minimal attachment" preference in PP attachment sentences, but later admit that they excluded those sentences which showed a contrary preference in pre-studies.

tially, while the licensing structure parser would have predicted VP-attachment. However, complement before adjunct attachment does hold in SOUL in the sense that only complements bear the potential to be predicted at all¹⁰⁶. As discussed in chapter 4.1.3.3, the empirical results available confirm an argument over adjunct advantage (see also: Schütze and Gibson, 1996).

5.4.3.1 Verb attachment and low attachment

Furthermore, the second principle in the LSP, *Verb Attachment*, has no direct equivalent in SOUL. The principle is motivated by intuitions on sentences such as (280).

(280) John put the book he was reading in the library.

Even though *put* provides an obligatory locative, *in the library* appears to be attached to *reading* preferentially, as predicted by *verb attachment*, thus producing a garden-path when no further locative is provided at the end of the sentence.

In SOUL, however, such an effect results from the combination of two architectural properties: firstly, predictions will never pass the border of an embedded sentence, or, in other words, an embedded sentence overshadows predictions outside their scope, and secondly, a subclause (or embedded sentence) is only then considered complete if the active sign is unable to integrate itself into it (or if the end is triggered by punctuation, of course). Therefore, the prediction of a locative complement is blocked by *he was reading*, such that *in the library* is attached actively, succeeding in attaching itself to *reading*. Note that *verb-attachment* appears to be a one-purpose stipulation, whereas the sentence blocking mechanism in SOUL is motivated independently (see section 5.2.8.1).

Finally, *low attachment*, as pointed out by Frazier (1990), would lead to a false prediction in sentence (276b), repeated here for convenience:

(281) Jill knew at 7:30 Bill would arrive.

The failure results from the emphasis on *low*, because if *at 7:30* was attached to the VP of the embedded sentence, which is below the VP of the matrix clause, it would definitely be attached lower than if it was attached to the matrix clause VP. The preference is of course quite the opposite, namely, to attach *at 7:30* to *knew*, rather than to *arrive*¹⁰⁷.

^{106.} Moreover, if the decision is to be made about the attachment as either a complement or an adjunct to the *same* attachment site, complement attachment is also preferred in SOUL. However, since adjunct attachment is attempted immediately after complement attachment within the (automatic) part of the SOUL mechanism, only negligible cost is predicted for the revision of complement attachment in favor of adjunct attachment.

^{107.}This argument could be countered (I am not aware of whether or not Steven Abney ever did so), however, by pointing out that the parser prefers to *attach* an item over *shifting* it, which would be required for producing the intermediate S-node above the PP *at 7:30*.

5.4.3.2 Head placement effects

Both SOUL and the LSP come to the same prediction in sentences such as (277 and 278, repeated here for convenience), but for very different reasons.

(282) Peter wußte, daß der Hund der Bibliothekarin eine Wurst gestohlen hat. Peter knew, that the dog the librarian_[gen. dat] a sausage stolen has.

a. "Peter knew that the dog has stolen a sausage from the librarian." Or:

b. "Peter knew that the dog of the librarian has stolen a sausage."

(283) Mir wurde erzählt, daß Martin das Pferd mit dem Fernglas erblickte.

I was told, that Martin the horse with binoculars caught sight of.

"I was told that Martin caught sight of the horse with binoculars."

Since attachment only takes place in the LSP when it is licensed by the head (*the licenser*), VP attachment of the ambiguous constituent in (277) and (278) is ruled out in general. Of course, if it were not, VP attachment would probably be predicted for both cases. But as it stands, the LSP predicts *attach before shift*, which amounts to NP attachment in both sentences.

As was outlined in section 5.2.10, SOUL, on the other hand, is not restricted to head-driven (head-licensing) parsing, thus being compatible with empirical data on incremental processing (see section 4.1.2). In SOUL, the NP attachment preference in (277) and (278) results from the combination of the facts that *i*. heads in a phrase-final position can never (don't have an opportunity to) predict their complements, so that complements can only attach themselves *actively*, and *ii*. the NP is lower and is as such more accessible than the VP in the right edge backbone during the active search.

To sum up, though the LSP seems to fit most data better than the GP-model, the fit is far from being perfect, and the principles themselves amount to mere stipulations describing a rather narrow range of data.

5.4.4 Syntactic Preferences (Ford, Bresnan & Kaplan, 1982)

SOUL shares a more subtle view on lexical preferences with the syntactic preference model of Ford et al. (1982), a view which goes beyond the mere complement adjunct distinction. Whereas in LFG complement frames are considered the relevant location to represent lexical preferences, it is the *thematic* frame which is supposed to model such preferences in SOUL. The latter type of representation allows interpretational influences to modulate the salience of role-takers in the content quite easily, whereas the relationship between interpretational issues and the syntactic (functional) frame is less direct.

Further differences between SOUL and the syntactic closure model arise from the fact that in the latter, lexical preferences can only influence parsing in combination with three further principles, *invoked attachment, final arguments* and *syntactic prefer*-

ence, which modulate hypothesizing and attaching. The closest equivalent of hypothesizing is "attempt to attach" in SOUL, which fails or succeeds immediately, however. Note that in the Syntactic Closure theory, several expansions of grammar rules are hypothesized in parallel at a given state in processing, though only one option is then executed while the others remain on the agenda for later back-ups. In SOUL, topdown prediction is limited to sisterhood prediction and sentence-schema instantiation, but alternatives at that point of processing are not hypothesized in parallel, but one-by-one and only on demand. For being able to do so, each sign memorizes which hypothesis has to be generated next if the current attempt to proceed fails. The implementation of signs in SOUL thus bears resemblance to the "stream"-concept of the Scheme programming paradigm (Abelson & Sussman, 1985). On the other hand, hypothesizing all options at a give point can amount to quite a lot of options being generated which will probably never be considered when processing is halted after the first successful parse has been found. In SOUL, each option is evaluated as it is generated. The final evaluation is attachment. In the Syntactic Closure model, the attachment of a phrase cannot take place before the entire phrase has been analyzed, such that issues of interpretation (F-structure) have to be delayed until the end of the phrase. In SOUL, however, attachment is word-by-word, such that every new word can be evaluated immediately.

Due to the interdependencies of *lexical preference, invoked attachment,* and *final argument,* modifier attachment to the most recent head (complement) can only take place for the very last complement in the functional frame of the verb. SOUL and the lexical preference model may therefore come to a different prediction in sentences such as (284).

(284) Peter observed the dealer in the entrance hall with binoculars.

In SOUL, the attachment of the PP *in the entrance hall* to the NP or to the VP depends on whether or not the locative is treated as a salient complement of *observed* (presumably it is not, resulting in NP attachment), whereas the *Syntactic Closure* model would have to attach it to the VP in any case because there is still another lexically preferred complement to come, such that *final argument* cannot apply for the locative.

In quite a different interpretation of the model, it could also happen that the preferred 3-place verb-frame including the (instrumental) PP is rejected after the locative turns out unable to work as a PP-comp. If so, the locative would indeed be treated as following the final argument, the direct object. But then the PP *with binoculars* would also have to be treated as a post-final argument phrase and thus be attached to *the dealer* (or to *the entrance hall*?), contradicting the intuitions about the preferences in this sentence.

In general, the lexical preference model appears to have severe problems when it is applied to sentences differing from rather simple PP-attachment examples. Although the notion of *lexical preferences* is still an important one in modern models, in order to work correctly it depends on rather weakly motivated principles.

5.4.4.1 Modulated vs. unrestricted lexical guidance

More generally, the SOUL mechanism, as a *restricted* lexical guidance model, differs from unrestricted lexical models, such as the *Theory of Syntactic Closure*, in another important aspect: SOUL can account for the "Mitchell"-effect in sentences with "intransitive" verbs like (285) in a fairly straightforward way.

(285) After the child sneezed the doctor prescribed a course of injections.

Mitchell (1987), as well as Adams et al. (in prep.), have found that the NP following the first verb (*the doctor*) appears to be attached to the preceding verb (*sneezed*) for a short period of time, although its subcategorization restrictions should have blocked the attachment (see chapter 4.1.3.2). Note that the *Syntactic Closure* model is *prediction-driven*, i. e. if the (preferred) lexical form of a verb does not subcategorize an element that matches with the current input, the attachment to the VP is not even hypothesized¹⁰⁸ (leaving aside the possibility for attachment as an adjunct). Consequently, the NP the doctor is initially (and successfully) hypothesized as the subject of the matrix clause, such that no effect on the NP should be expected, contrary to the results.

SOUL, on the other hand, is *visibility-driven*: if no matching attachment site is predicted, the NP (or even its determiner) performs an active search to attach itself to the waiting sign's right edge. On this search, it might find an unpredicted complement (as in the case of some verbs in Adams et al.'s study) to unify with (to attach to), thus allowing for plausibility effects to arise.

As a not strictly prediction driven model, SOUL might have problems to account for the data reported by Clifton, Frazier, and Connine (1984, see chapter 4.1.3.1), repeated here for convenience. Subjects read sentences such as (286) and (287).

(286) The baby-sitter {a. read, b. sang} the @ story to the sick child.

(287) The baby-sitter {a. read, b. sang} to @ the sick child.

The sentences were presented externally paced (300 ms + 50 ms break) in a stationary window. After the first word following the verb (indicated by "@"), the presentation was discontinued until the subjects performed a lexical decision task on an unrelated word, which was presented at a separate location on the screen.

The lexical decision times obtained indicated that preferences had a very early impact: subjects performed faster when the word following the verb matched its preferred argument frame (908 ms and 877 ms mean secondary task reaction time for

^{108.}*Hypothesizing*, in this context, means the generation of a parsing option that is put onto the agenda of the LFG-parser. In a different sense of *hypothesizing*, the attempt to generate such an option is in fact pursued, since all elements on the right hand side of the VP-rule (see chapter 2.7) are checked for compatibility before the parser resumes to different options already on the agenda. Thus, even though the VP rule generally provides an NP on the right hand side, the principle of final argument blocks the parser from *hypothesizing* such an attachment, in the sense given by Ford et al. (1982).

286a and 287b, respectively) than when it mismatched the frame (1000 ms and 1008 ms mean secondary task reaction time for 286b and 287a, respectively).

It has been argued (Frazier, personal communication) that SOUL fails to predict the converse result effected by preferred *intransitive* verbs like *sang*: if a low salience for the theme (direct object) is assumed for *sing*, the parser would have to resort to *active attachment* for both the PP and the NP continuation, resulting in about the same processing load for both continuations.

The argument is based on the assumption that the verbs used in this experiment differ only with respect to their preference to predict a direct object, i.e it rests on the assumption that the PP could not be taken as a preferred argument as well. Both the NP- and PP-attachment would then have to be carried out via the *active-attachment* mechanism in the case of a preferentially *intransitive* verb, whereas with preferentially *transitive* verbs, the NP, but not the PP can be attached via *predicted* attachment.

Note, however, that the prepositions given in the sentences could always possibly have started a prepositional complement by the time the secondary task was performed¹⁰⁹, namely, *before* the PP was completed. It seems legitimate to assume that many verbs (of both experimental types) in the materials provide a salient PComp argument started by that particular preposition. The difference in the transitivity preference, then, might have had an important side effect: in the absence of a predicted direct object on the prediction stack (i.e. in the case of a preferred *intransitive* verb), the predicted PComp is immediately accessible for the active sign. In the case of a preferentially transitive verb, however, it fails to attach itself via *predicted* attachment, since a direct object NP is predicted in the first place of the prediction stack. The transitivity preference (to propagate a direct object NP) has the side effect of *overshadowing* the PComp prediction. Thus, this fact alone would suffice to explain the data.

An even closer look to the materials, however, discovers a related, but even stronger pitfall.

- (288) Mary often {a. called, b. worried} her niece but the two got along extremely well.
- (289) Mary often {a. called, b. worried} about her niece but the two got along extremely well.

In some examples, such as (288) and (289), the preferred intransitive verb *worried* strongly predicts the prepositional complement, here starting with *about* ..., whereas the PP can only be attached as an adjunct for the preferentially transitive verb *called*. In these cases, the transitivity preferences are confounded with a *reverse* PComp preference. The "intransitivity effect" might therefore also be due to a simple *PComp preference* effect.

^{109.} Note that although some of the sentences in the materials were continued with an adjunct PP, such as *with his entire family*, this is not yet clear when the preposition *with* is read, which can also introduced an instrument.

5.4.5 Visibility and competition revisited

In chapter 4, I have outlined that the psycholinguistic evidence available strongly suggests a serial model of ambiguity resolution. The model presented here stresses the notion of *visibility* of attachment sites, as opposed to having competing analyses or attachment sites (as in CAPERS; Stevenson, 1993; see section 2.12). Crucially, these two approaches do not only differ in their predictions for ambiguous sentences, but also for unambiguous ones. For unambiguous sentences, competition-based models generally do not predict parsing difficulties whatsoever, whereas visibility-based models predict at least some latencies in cases where the valid analysis is not visible in the first place (cf. Frazier, 1995).

Clifton, Dickey and Frazier (forthcoming) ran an eye movement study on sentences like (290ab) which set up an interesting instance of this case.

- (290) a. Unfortunately [the space capsule everyone objected to] had been built by NASA *and cost* the taxpayers a fortune.
 - b. Unfortunately [the space capsule everyone objected to after it was built by NASA] *had cost* the taxpayers a fortune.

In the first case (290a), the critical region *(and cost)* introduces a VP coordination, while the critical region *(had cost)* in the second (290b) sentence follows the complex subject NP, thus constituting the matrix verb complex. In both cases, the subject NP was modified by a reduced relative clause, but only in the latter case (290b) does the relative clause cover all words up to the critical region.

Reading times on *had cost* in (290b) were higher than reading times on *and cost* in (290a), indicating that coming out of an embedded sentence is in fact harder than continuing the analysis within the clause currently being processed. Note that both sentences are unambiguous, such that the processing difficulty in (290b) cannot easily be explained by a competition mechanism. On the other hand, this result strongly suggests that clause boundaries in general act as processing barriers, as predicted by the lazy release mechanism for clausal c-flags. Note that the clausal c-flag seems to overshadow even the prediction of the obligatory matrix verb, certainly being one of the more relevant constituents of the sentence.

Thus, visibility-based models like SOUL turn out to be appropriate in their processing predictions for ambiguous structures as well as for unambiguous structures, while competition-based models fail in both cases.

5.5 Avoiding memory load: semantics-oriented parsing

We have seen that the parser can be misguided in a way that makes it almost impossible to recover from the garden-path. Processing breakdown can also occur in unambiguous structures, such as center-embedded sentences. As outlined in chapter 2.1.1, sentences with too many *center embedded subclauses*, such as (291), are known to be almost impossible to understand, even though they are syntactically well-formed.

(291) The woman that the boy that the friend visited liked laughed.

In the approach taken here, such phenomena are considered to be only indirectly induced by *syntactic* properties of the sentence. To clarify the issue at hand, some basic assumptions about the *semantic* processor (see chapter 4.2.2.7) will be outlined at least briefly here.

As illustrated in chapter 4.2.2.7, I assume that the mental model of the discourse is a substantial part of the semantic processor where it is subject to working memory capacity limitations. The discourse model is assumed to be structured, roughly as outlined in DRT (Kamp and Reyle, 1993) or in (Habel, 1986). So-called *referential objects* (*refos*, see section 5.1.3.4) introduced by noun phrases are either added (in the case of indefinite NPs) to the discourse model or mapped onto already existing *refos* (mostly definite NPs) as soon as they are encountered. When a sentence is processed, *refos* become *propositionally connected* (also: *integrated* or *chunked*). In SOUL, refos are incrementally connected *during parsing*, by virtue of HPSG *psoas* in heads. Thus, whereas syntactic structure assembly is highly incremental (see section 4.1.2), the *integration* of refos depends on the availability of the head.

Crucially, the introduction of new refos is assumed to consume short-term memory resources unless they can be integrated into the discourse structure (the referential net) When a sentence is read, a refo should be identified in the discourse model, or, if it must be added to the model, it should be connected to the discourse net as soon as possible, to keep the memory load as small as possible. Where too many refos cannot be connected, as in (291), and must be kept separate until the verbs come along, the working memory capacity is obviously exceeded.

We have seen earlier, that the central parsing principle of this work, PHA, serves the purpose of guiding the syntactic analysis in a way that allows the immediate integration of new refos into the current mental model ("*semantics-oriented*" processing). PHA could thus be regarded as being subsumed by the more general *immediate semantic integration* (292) principle (Konieczny et al., 1991, see also Crocker, 1993).

(292) immediate semantic integration principle (ISIP)

Unintegrated referential objects (refos) induce memory load.

Although one might be tempted to replace PHA by the more general ISIP principle, I have presented data indicating that attachment preferences are established even before a NP was completed (see chapter 4.1.1.4, Experiment III), i.e. even before a *refo* could be established that then had to be integrated. Such results go against ISIP as a guiding principle in sentence processing. In general, the data strongly suggest that the syntax processor proceeds autonomously in a serial fashion.

Nevertheless, the attachment predictions of PHA and SOUL are *compatible* with ISIP. The reason for this is that the parser is suited to helping the comprehension system in a most efficient way, namely, in providing analyses which minimize memory load.

It is this kind processing load that the SOUL mechanism is suited to avoid. In always attempting to attach items immediately, the parser can minimize the number of unconnected referential objects. In predicting highly salient complements, SOUL achieves the best possible integration of an item that can then be integrated into the maximally elaborated net of entities. It is economy in this sense that qualifies SOUL as a *semantics oriented* parser.

5.6 Summary

In this chapter, I introduced the SOUL system as a sentence processing mechanism operating with rich grammatical representations as provided by *Head-driven Phrase Structure Grammar* (Pollard and Sag, 1987, 1994).

The high degree of lexicalization of grammatical knowledge, i. e. the head-drivenness of HPSG, turned out as both advantageous and disadvantageous for modeling human sentence processing. On the one hand, focusing on heads fits the emphasis of heads in PHA, the head *prediction* mechanism in SOUL, and the importance of bottom-up parsing with active lexical heads in SOUL.

On the other hand, the predictions of the "natural" parsing strategy of head driven grammars, the head-licensing strategy was shown to be incompatible with the experimental findings. As a solution, either process-specific information can be supplied in the competence base, or certain special-purpose knowledge structures, namely generalized sentence templates, have to be proposed. The constructive use of inverted subcategorization information (ISI) in HPSG provides an exceptional mechanism for building structures from non-heads. On the other hand, since ISI for verb complements is not linguistically motivated, it requires an undesirable and potentially illegitimate modification of the competence ontology for performance purposes only. Sentence templates, however, can be compiled directly from the grammar preserving nothing but information from the grammar-base. On the other hand, this method can be regarded as the most intransparent one, because linguistic knowledge has to be pre-compiled into "more usable" representations. However, the deviation from transparency is limited to parsing of non-complements, since for complements, the already existing sort-hierarchy for complement sorts can be employed.

The processing of verb-final and verb-initial structures is carried out differently: whereas heads in the initial position can project their subcategorization requirements into the structure, the successive attachment of complements in head-final structures results in a sortal inference on the head's subcategory. However, since in head-final constructions constraining lexical information is available to permit the attachment of the complements of subsequent heads, processing these structures is not significantly disadvantageous compared to head-initial structures (cf. Mitchell, 1989, Frazier, 1989)

The parsing process is driven by signs, which are implemented as objects with methods for combinatorial purposes. The resulting system behavior can be described as *highly incremental*, in that each incoming item gives highest priority to attach itself to the sentence structure built so far. The concrete attachment-mechanisms implemented in signs were shown to account for a broad variety of parsing phenomena,

such as attachment preferences in structurally ambiguous sentences as well as processing difficulties in unambiguous structures. The SOUL mechanism will prefer analyses that allow the earliest semantic integration possible such that the interpretation will be in general minimally memory consuming.

6 Summary and concluding remarks

This thesis has presented a model of human sentence processing that accounts for a wide variety of processing phenomena, including, most importantly, attachment preferences in structurally ambiguous inputs.

The model satisfies important criteria established by empirical results from psycholinguistic experiments, partly provided in this thesis and partly known from literature. The criteria are as follows:

- 1. The human sentence processor operates highly *incrementally*, i.e. it attaches each incoming item as it is encountered, employing both bottom-up and top-down strategies (Frazier, 1987a; following a *"linear parsing"* strategy, Konieczny & Strube, 1995, as opposed to *head-driven* or *head licensing* parsing). A linear parsing strategy has been established in an eye-movement study on German subject-object asymmetries. The data strongly suggest that the first few words of the initial NP in these sentences were attached to the sentence structure even before the NP was complete and, of course, long before the verb could have *licensed* any attachment. Attachment processes were identified in the absence of licensing heads, rendering a *head-driven* parsing strategy inadequate, but suggesting a left-corner ("arc eager") parsing strategy.
- 2. Furthermore, the parser operates in a *serial* fashion, i. e. it pursues only one analysis at a time in the case of an attachment ambiguity, as opposed to *parallel* processing of multiple analyses. Evidence stems from an eye-tracking experiment on German NP-attachment ambiguities in verb final sentences. Although no ambiguity effect could be established, a processing penalty was observed when the sentence was disambiguated towards the non-preferred reading.

- 3. Interestingly, the same pattern of results was found even at the very first word of the ambiguous NP, the determiner, indicating an extremely early commitment to one alternative. The parser thus *commits* itself to a structural alternative *immediately* (as opposed to minimal commitment or to a "wait and see" strategy).
- 4. In chapter 4.1.3, I discussed the psycholinguistic literature on the question whether or not the parser can be guided by detailed lexical information in its initial parsing decisions. It was shown that none of the findings that were claimed to provide evidence for a lexical filter approach were ultimately convincing, and the results of many studies, among them experiment I in this thesis, show the guiding influence of lexical preferences. Thus, the parser uses *detailed lexical information*, including lexical subcategorization preferences, during the *initial stage* of structure assembly.
- 5. As established with PP-attachment ambiguities in German verb-final and verb second sentences (Experiment 1 and 2) and with NP-attachment ambiguities in German verb-final sentences (Experiment 3), the parser resolves structural ambiguities according to *Parametrized Head Attachment* (Konieczny et al., 1994), i.e. depending on the presence or absence of the lexical head of the attachment sites, their lexical properties, and their respective distance to the item to be attached. In verb-final sentences, for instance, the attachment to the preceding constituent was preferred, whereas in verb-second sentences, attachment to the VP was quite possible, but determined by the lexical properties of the verb.
- 6. These preferences prevailed even in the biasing contexts of the fifth experiment presented in section 4.2.2, such that the parser can be assumed to operate *autonomously* in its initial attachment decisions, i.e. higher level semantic and pragmatic processes are not capable of *directing* the initial attachment choice (as opposed to *strongly interactive* or *integrative* multiple constraint models).

In chapter 2, I discussed a wide variety of psycholinguistic models (*Perceptual Strategies*, Bever, 1970; Kimball's seven principles, the Sausage Machine, ATNs, the Garden Path Theory, Syntactic Closure, the Licensing Structure Parser, Pritchett's Generalized Theta Attachment and On-Line Locality Constraint, Gibson's weighted parallel model, the Multiple Constraint Model, MacDonald et al., 1994; the connectionist competition model CAPERS, Stevenson, 1993, 1995; Gorrell's theory of syntax and parsing, Construal Theory, Frazier and Clifton, 1996; and the Tuning Hypothesis, Mitchell, 1994). I was able to demonstrate that none of these models satisfies all the criteria established in chapter 4.

While *Parametrized Head Attachment* (Konieczny et al., 1994), as a parsing *principle* to be integrated into a more complete parsing account, does not *violate* any of the criteria, it turns out to be vague in several respects. The SOUL (Semantics-Oriented Unification-based Language) mechanism was proposed as a general parsing mechanism to overcome these shortcomings.

SOUL is an implemented, strictly incremental serial parser based on *Head-driven Phrase Structure Grammar* (*HPSG*, Pollard & Sag, 1994). SOUL immediately utilizes the rich representation of HPSG models in a semi-transparent way. Incremental parsing is established by virtue of underspecified sentence schemata. Except for these, no *compilation* of grammatical constraints into a more parsable format is necessary, since

the parser can utilize *subcategorization sorts* already provided by the HPSG system in the absence of a licensing head to rule out invalid attachments and predict the forth-coming head.

Nevertheless, lexical subcategorization preferences are projected into the structure immediately and determine the *visibility* of the potential attachment sites for an ambiguous item. The mechanism can be summarized as follows: HPSG sign-tokens, i.e. words or phrases, are represented as objects and come equipped with methods for combining themselves with other sign-objects. Parsing starts out from the lexicon: as soon as a word is encountered, a lexical sign is "activated". The active sign attempts to apply one of the following methods:

- 1. attach to the most "visible" (criterion given below) site in the current sentence. If it is impossible to find an attachment site,
- 2. project up to the next phrasal level, as a "daughter" of one of the basic HPSG-ruleschemes. Retry 1.

"Visibility" is hierarchically ordered:

a. Most visible are those complements of preceding heads marked as lexically preferred. These are predicted top-down, such that lower predictions overshadow higher ones. Only predictions at the lowest level are accessible. Attachment to one of these nodes is called "predicted attachment".

b. If predicted attachment fails, the active sign can access the most recently read word and start a search for an attachment site by proceeding from the bottom to the top of the right edge of the current sentence (the lower, the more visible).

"Active attachment" (b.) includes certain "lowering" (Sturt and Crocker, 1995) options, such as simple adjoining, which are not supposed to be costly. If an initially successful attachment fails at a later stage, the revision process proceeds as if the attachment had failed originally.

It was demonstrated with numerous examples that the SOUL mechanism covers an almost complete range of psycholinguistic phenomena on initial parsing preferences and easy revision. The initial parsing preferences in verb-initial (verb-second) and verb-final structures match those of Parametrized Head Attachment" (Konieczny et. al, 1994, 1995).

To illustrate the mechanism for a last time, consider the German verb-final and verb-second examples given in (293) and (294).

(293) Peter fesselte den Mann mit der Krawatte.

Peter fettered the man with the tie.

(294) ..., daß Peter den Mann mit der Krawatte fesselte.

..., that Peter the man with the tie fettered.

"... that Peter fettered the man with the tie."

When the PP *mit* ... attempts to attach itself to the current sentence structure, an instrumental PP is predicted as a salient complement of *fesselte* (*fettered*) in (293), while no such prediction could have been posed in (294). Thus, the most visible attachment site in (293) is the VP (complement), whereas active attachment renders the direct object NP to be the most readily accessible node in the structure.

Thus, the SOUL mechanism, as well as PHA, initially build structures which can be semantically interpreted as soon as possible. Note that the PP would have been left uninterpreted in the verb-final sentence (294) if it had been initially proposed to be a complement or adjunct of the verb. The strategy emerging can therefore be considered "semantics-oriented", minimizing on-line memory load. This is why language processing is easy, at least in general.

References

- Abelson, H. & Sussman, G. J. (1985). *Structure and Interpretation of Computer Programs*. Cambride, Massachusetts: MIT Press.
- Abney, S. (1987). Licensing and parsing. *Proceedings of NELS*. Amherst: University of Massachusetts, 17.
- Abney, S. (1987). The English noun phrase in its sentential aspect. MIT, Cambridge: PhD Thesis.
- Abney, S. (1989). A computational model of human parsing. Journal of Psycholinguistic Research, 18.1, 129-144.
- Abney, S., & Johnson, M. (1991). Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research*, 20(3), 233-250.
- Adams, B. C., Clifton, C., Jr., & Mitchell, D. C. (1991). *Lexical guidance in sentence parsing.* Poster presented at the meeting of the Psychonomics Society. San Francisco.
- Adams, B., Clifton, C., & Mtchell, D. (1994). *Lexical guidance in sentence processing: Further evidence for a filtering account.* University of Massachusetts, Amherst: Manuscript.
- Ades, A., & Steedman, M. (1982). On the order of words. *Linguistics and Philosophy*, *4*, 517-558.
- Aho, A. V., Hopcroft, J. E., & Ullman, J. D. . (1974). *The Design and Analysis of Computer Algorithms.* Reading, MA: Addison-Wesley Publishing Company.
- Allen, J. (1987). *Natural language understanding*. Menlo Park, CA: The Benjamin/Cummings Publishing Company.
- Altmann, G. (1988). Ambiguity, parsing strategies, and computational models. *Language and Cognitive Processes, 3,* 73-97.
- Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191-238.
- Altmann, G. T., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language.*, *31*, 685-712.
- Baddeley, A.D. (1986). Working memory. Oxford: Oxford University Press.
- Bader, M. (1990). Syntaktische Prozesse beim Sprachverstehen: Theoretische berlegungen und experimentelle Untersuchungen. *Magisterarbeit an der Albert-Ludwigs-Universität Freiburg*.
- Bader. M. & Lasser, I. (1994). German Verb-Final Clauses and Sentence Processing: Evidence for Immediate Attachment. In C. Clifton, L. Frazier & K. Rayner (Eds.), *Perspectives in Sentence Processing* (pp. 225-242). Hillsdale, NJ: Lawrence Erlbaum.
- Barwise, J. Perry, J. (1981). Situations and Attitudes. The Journal of Philosophy, 81, 668-691.
- Berwick, R. C. (1991). Principles of principle-based parsing. In R. C. Berwick, S. P. Abney, & C. Tenny (Eds.), *Principle-based parsing: Computation and psycholinguistics.* Dordrecht: Kluwer.
- Berwick, R. C. & Weinberg, A. S. (1985). Deterministic parsing and linguistic explanation. *Language and Cognitive Processes, 1,* 109-134.
- Bever, T. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and development of language*. New York: John Wiley.
- Bever, T. G. (1970). The influence of speech performance on linguistic structure. In G. B.

Flores d'Arcais & W. J. M. Levelt (Eds.), *Advances in psycholinguistics*. Amsterdam: Elsevier (North-Holland).

- Bloomfield, L. (1933). Language. New York: Holt, Rinehart & Winston.
- Borsley, R. D. (1987). *Subjects and Complements in HPSG.* CSLI-Report. Menlo Park, CA, Stanford, CA: Center for the Study of Language and Information, CSLI-87-107.
- Bresnan, J. (1982). The mental representation of grammatical relations. Cambridge, MA: MIT Press.
- Bresnan, J., & Kaplan, R. M. (1982). Introduction: Grammars as Mental Representations of Language. In J. Bresnan (Ed.), *The Mental Representation of Grammatical Relations*. Cambridge, MA.: MIT Press.
- Bresnan, J., & Kaplan, R. M. (1982). Lexical Functional Grammar: A formal system for grammatical representation. In J. Bresnan (Ed.), *The mental representation of grammatical relations.* Cambridge, MA: MIT Press.
- Bresnan, J. W. (1978). A realistic transformational grammar. In M. Halle, J. W. Bresnan, & G. A. Miller (Eds.), *Linguistic Theory and Psychological Reality*. Cambridge, MA: MIT Press.
- Britt, M.A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language, 33,* 251-283.
- Britt, M.A., Gabrys, G., & Perfetti, C.A. (1993). A restrictive interactive model of parsing. Proceedings of the 15th Annual Conference of the Cognitive Science Society, Boulder, CO. Hillsdale, NJ: Erlbaum, pp. 260-265.
- Caplan, D. (1972). Clause boundaries and recognition latencies for words in sentences. *Perception & Psychophysics, 12,* 73-76.
- Carpenter, R. (1992). *The logic of typed feature structures.* Cambridge: Cambridge University Press.
- Carreiras, M., & Clifton, Jr. C. (1993). Relative Clause Interpretation Preferences in Spanish and English. *Language and Speech.*
- Chapin, P. G., Timothy, S. & Abrahamson, A.A. (1972). Two factors in perceptual segmentation of speech. *Journal of Verbal Learning and Verbal Behavior*, *11*, 164 - 173.
- Chomsky. (1957). Syntactic Structures. The Hague: Mouton.
- Chomsky. (1965). Aspects of the theory of syntax. Cambridge, MA: MIT Press.
- Chomsky N. (1959). A review of B.F. Skinner's 'Verbal Behavior'. Language, 35, 26-58.
- Chomsky, N. (1981). Lectures on government and binding. Dordrecht: Foris.
- Chomsky, N. (1986). Barriers. Cambridge, Mass.: MIT-Press.
- Chomsky, N. (1991). Some notes on economy of derivation and representation. In R. Freidin (Ed.), *Principles and parameters in comparative grammar* (pp. 417 454). Cambridge, MA: MIT Press.
- Chomsky, N. (1995). The minimalist program. Cambridge, MA: MIT Press.
- Chomsky, N., & Miller, G. (1963). Finitary models of language users. In R. D. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (Vol. 2). New York: John Wiley.
- Clifton, C., & Fereirra, F. (1989). Ambiguity in context. *Language and Cognitive Processes, 4,* SI 77-104.
- Clifton, C., Frazier, L., & Connine, C. (1984). Lexical expectations in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 23,* 696-708.
- Clifton, C., Jr., Speer, S., & Abney, S. P. (1991). Parsing arguments: phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language*, 30, 251-271.

- Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. R. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives* (pp. 320-358). Cambridge: Cambridge University Press.
- Crocker, M. (1993). Properties of the principle-based sentence processor. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society.* Hillsdale, N.J.: Erlbaum.
- Crocker, M.W. and Lewin,I. (1992). Parsing as Deduction: Rules versus Principles. In Neumann, B. (Ed.), 10th European Conference on Artificial Intelligence. (Vol. ECAI 92, pp. 508-512) John Wiley & Sons, Ltd.
- Cuetos, F., & Mitchell, D. (1988). Cross linguistic differences in parsing: Restrictions on the issue of the Late Closure strategy in Spanish. *Cognition, 30,* 73-105.
- Daneman, M. & Carpenter, P. A. (1980). Individuell differences in working memory and reading. *Journal of Verbal Learning and Behavior*, 19, 450-466.
- De Vincenzi, M. & Job, R. (in press). An investigation of late closure: the role of syntax, thematic structure and pragmatics in initial and final interpretation. *Journal of Experimental Psychology: Learning, Memory, & Cognition.*
- Devlin, K. (1991). Logic and information. Cambridge: Cambridge University Press.
- Dowty, D. R. (1989). On the semantic content of the notion of 'thematic role'. In Chierchia, G. et al. (Ed.), *Properties, Types and Meaning* (Vol. 2). Dodrecht: Kluwer Academic Press.
- Earley, J. (1970). An efficient context free parsing algorithm. *Communications of the Association for Computing Machinery, 13,* 94-102.
- Elman, J. L. & McClelland, J. L. (1984). Speech perception as a cognitive process: the interactive activation model. In N. Lass (Ed.), *Speech and language* (Vol. 10, pp. 360 380). San Diego, CA: Academic Press.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25,* 348-368.
- Ferreira, F. & Henderson, J. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language, 30,* 725-745.
- Ferreira, F. & Henderson, J. M. (1990). Use of Verb Information in Syntactic Parsing. *Journal of Experimental Psychology: Learning Memory and Cognition.*, *16(4)*, 555-568.
- Flores d'Arcais, G. B. (1982). Automatic syntactic computation and use of semantic information during sentence comprehension. *Psychological Research, 44,* 231-242.
- Flores d'Arcais, G. B. (1987). Syntactic processing during reading for comprehension. In M. Coltheart (Ed.), Attention and Performance XII: The Psychology of Reading (pp. 619-633). Hove: Erlbaum.
- Fodor, J. A. (1983). The modularity of the mind. Cambridge, Mass.: MIT Press.
- Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar.* New York: McGraw-Hill.
- Fodor, J. A., Garrett, M. F., & Bever, T. G. (1968). Some syntactic determinants of sentential complexity, II:Verb structure. *Perception and Psychophysics*, *3*, 453-461.
- Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, *9*, 427-473.
- Fodor, J. D. & Frazier, L. (1980). Is the HSPM an ATN? Cognition, 8, 417-459.
- Ford, M., Bresnan, J. & Kaplan, R.M. (1982). A comptence-based theory of syntactic closure. In Bresnan, J. (Ed.), *The mental representation of grammatical relations*. Cambridge, Mass.: MIT Press.

- Ford, M., & Dalrymple, M. (1988). A note on some psychological evidence and alternative grammars. *Cognition, 29,* 63-71.
- Franz, A. (1990). A parser for HPSG (Vols. CMU-LCL-90-3). Carnegie Mellon University: Laboratory for Computational Linguistics.
- Frazier, L. (1979). On comprehending sentences: syntactic parsing strategies. Bloomington, Ind.: IULC.
- Frazier, L. (1986). The mapping between grammar and processor. In Gopnik, I. & Gopnik, M. (Eds.), *From models to modules* (pp. 117 134). Norwood, NJ: Ablex.
- Frazier, L. (1987a). Sentence processing: A tutorial review. In M. Coltheart (Ed.), Attention and performance XII: The psychology of reading (pp. 559-586). Hove/London/Hillsdale: Lawrence Erlbaum.
- Frazier, L. (1987b). Syntactic processing: evidence from Dutch. *Natural Language & Linguistic Theory*, *5*, 519-559.
- Frazier, L. (1987c). Theories of sentence processing. In J. Garfield (Ed.), Modularity in knowledge representation and natural-language understanding (pp. 291 - 308). Cambridge, MA: MIT Press.
- Frazier, L. (1989). Against lexical generation of syntax. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 506-528). Cambridge, MA: MIT Press (Bradford).
- Frazier, L. (1990). Exploring the architecture of the language-processing system. In G. Altmann (Ed.), *Cognitive Models of Speech Processing* (pp. 409-433). Cambridge, Mass.: MIT Press.
- Frazier, L. (1990). Parsing modifiers: Special purpose routines in the human sentence processing mechanism. In Balota, D. A., D'Arcais, G. B. F. & Rayner, K. (Eds.), *Comprehension process in reading* (pp. 303-331). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Frazier, L. (1995). Syntactic processing complexity. Paper presented at the Einstein-Kolloquium, Postdam, July 1995.
- Frazier, L. & Clifton, C. . (1996). Construal. Cambridge, MA: MIT Press.
- Frazier, L., Clifton, C., & Randall, J. (1983). Filling gaps: Decision principles and structure in sentence comprehension. *Cognition, 13,* 187-222.
- Frazier, L. & Fodor, J. D. (1978). The sausage machine: a two stage parsing model. *Cognition*, *6*, 291-325.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psycholo*gy, 14, 178-210.
- Frazier, L., & Rayner, K. (1987). Resolution of syntactic category ambiguities: eye movements in parsing lexically ambiguous sentences. Journal of Memory and Language, 26, 505-526.
- Garnham, A. (1985). Psycholinguistics: Central topics. London: Methuen.
- Garrett, M. F., Bever, T. G., & Fodor, J. A. (1966). The active use of grammar in sentence perception. *Perception and Psychophysics*, *1*, 30-32.
- Gazdar, G. J. M. (1982). Phrase structure grammar. In G. K. Pullum & P. Jacobson (Eds.), *The Nature of Syntactic Representation.* Dordrecht: Reidel.
- Gibson, E. (1991). A computational theory of human linguistic processing: memory limitations and processing breakdown. Unpublished Ph.D. dissertation. Carnegie Mellon University:.
- Gibson, E., Pearlmutter, N., Canseco-Gonzalez, E,nriqueta, & Hickock, G. (in press). Recency preference in the human sentence processing mechanism. *Cognition.*

- Gilboy, E., Sopena, J., Frazier, L., & Clifton, C. (1995). Argument structure and association preferences in Spanish and English complex NPs. *Cognition, 54,* 131-167.
- Gorrell, P. (1995). Syntax and parsing. Cambridge: Cambridge University Press.
- Gorrell, P. (1996). Diagnosis and Determinism. New York: *Paper presented at the 9th annual CUNY conference on human sentence processing.*
- Grimshaw, J. (1990). Argument structure. Cambridge, Mass.: MIT Press.
- Günther, U. (1989). Lesen im Experiment. Linguistische Berichte, 122, 282-320.
- Habel, C. (1986). *Prinzipien der Referentialität: Untersuchungen zur propositionalen Repräsentation von Wissen.* Berlin: Springer.
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift as an ERP-measure of syntactic processing. *Language and Cognitive Processes*, *8*, 439-483.
- Hanna, J., Barker, C., & Tanenhaus, M. (1995). Integrating local discourse constraints in resolving lexical thematic ambiguities. Tucson, AZ: *Poster presented at the 8th annual CUNY conference on human sentence processing*.
- Hemforth, B. (1993). *Kognitives Parsing: Repräsentation und Verarbeitung sprachlichen Wissens.* Sankt Augustin: Infix.
- Hemforth, B., Konieczny, L., Scheepers, C. & Strube, G. (1992). SOUL-Processing: Semantikorientierte Prinzipien menschlicher Sprachverarbeitung. In G. Grz (Ed.), *KONVENS-92* (pp. 198-208). Berlin: Springer.
- Hemforth, B., Konieczny, L. & Scheepers, C. (1993). On Reanalysis. In B. Hemforth & G. Strube (Eds.), *Reanalysis and Repair*. IIG, Univ. Freiburg: Institutsbericht.
- Hemforth, B., Konieczny, L., Scheepers, C. & Strube, G. (1994). First Analysis, Reanalysis and Repair. *IIG-Berichte*. Freiburg:, 8/94.
- Hemforth, B., Konieczny, L., & Scheepers, C. (1994). On reanalysis: Head position and syntactic complexity. In B. Hemforth, L. Konieczny, C. Scheepers & G. Strube (Eds.), *First Analysis, reanalysis, and repair* (Vol. IIG-Berichte 8/94, pp. 23-50). Freiburg: Institut fr Informatik und Gesellschaft.
- Hemforth, B., Konieczny, L. & Scheepers, L. (1994). Probabilistic or universal approaches to sentence processing: How universal is the human language processor? In H. Trost (Ed.), *KONVENS94.* Berlin: Springer.
- Hemforth, B., Konieczny, L., & Strube, G. (1993). Incremental syntax processing and parsing strategies. *Proceedings of the 15th annual conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum, 539-545.
- Henderson, J. M. & Ferreira, F. (1990). The effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning Memory and Cognition, 16,* 417-429.
- Herrmann, T. (1985). *Allgemeine Sprachpsychologie*. Grundlagen und Probleme. München, Wien, Baltimore: Urban & Schwarzenberg.
- Hörmann, H. (1976). *Meinen und Verstehen. Grundzüge einer psychologischen Semantik.* Frankfurt am Main: Suhrkamp.
- Holmes, V. M. . (1987). Some Hypotheses About Syntactic Processin in Sentence Comprehension. In Coltheart, M. (Ed.), Attention and Performance XII: The psychology of reading (pp. 227-247). London: Lawrence Erlbaum Associated Ltd.
- Holmes, V. M., Kennedy, A., & Murray, W. (1987). Syntactic structure and the garden path. *Quarterly Journal of Experimental Psychology.*
- Igoa, J. M. (1995). Parsing decisions and the construal hypothesis: Attachment preferences in

primary phrases in Spanish. *Paper presented at the Second Symposium on Psycholinguistics.* Tarragona, Spain: 19-22 April 1995.

- Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning and Verbal Behavior, 23,* 612-624.
- Inoue, A. (1991). A comparative study of parsing in English and Japanese. University of Conneticut: PhD Thesis.
- Jackendoff, R. (1977). X-bar Syntax: A Study of Phrase Structure. Cambridge, MA: MIT Press.
- Jarvella, R. J. (1979). Immediate memory and discourse processing. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 13). New York: Academic Press.
- Johnson-Laird. (1983). Chapter 13: Parsing and Performance. .
- Johnson-Laird, P. N. (1983). Mental models. Cambridge, MA: Harvard University Press.
- Joshi, A. K. (1985). How much context-sensitivity is required to provide reasonable structural descriptions: tree adjoining grammars. In D. Dowty & L. Kartunnen (Eds.), Natural Language Processing: Computational and Theoretical Perspectives. Cambridge: Cambridge University Press.
- Juliano, C. & Tanenhaus, M. (1993). Contingent frequency effects in syntactic ambiguity resolution. Proceedings of the 15th annual conference of the Cognitive Science Society. Hillsdale, NJ, Erlbaum, 593-598.
- Juliano, C. & Tanenhaus, M. K. (1994). A constraint-based lexicalist account of the subject/object attachment preference. *Journal of Psycholinguistic Research, 23*, 459-471.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329-354.
- Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Boston: Allyn & Bacon.
- Just, M. A., Carpenter, P.A. & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111, 228-238.*
- Kamp, H. & Reyle, Y. (1993). From Discoures to Logic; Bd I und II. Dordrecht: Reidel.
- Kasper, R. (1993). Adjuncts in the Mittelfeld. In J. Nerbonne, K. Netter, & C. Pollard (Eds.), *German in Head-driven Phrase Structure Grammar* (pp. 39-70). Menlo Park: CSLI.
- Kay, M. (1989). Head-driven parsing. *Proceedings of the international workshop on parsing technol*ogy. Carnegie-Mellon University, Pittsburgh, PA:.
- Keenan, E. & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry, 8,* 63-99.
- Kempen, G., & Vosse, T. (1989). Incremental syntactic tree formation in human sentence processing: a cognitive architecture based on activation decay and simulated annealing. *Connection Science* 1, 3, 273-289
- Kennedy, A., Murray, W. S., Jennings, F., & Reid, C. (1989). Parsing complements: Comments on the generality of the principles of minimal attachment. *Language and Cognitive Process*es, 4, 51-76.
- Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*, 15-47.
- Kimball, J. (1975). Predictive analysis and over the top parsing. In J. Kimball (Ed.), *Syntax and semantics* (Vol. 4). New York: Academic Press.
- Kiss, T. (1995). Merkmale und Repräsentationen. Opladen: Westdeutscher Verlag.
- Konieczny, L. (1989). Zum Verstehen von strukturellen Ambiguitäten: Der Einfluß von Kontext, Weltwissen und lexikalischen Präferenzen. *Diplomarbeit an der Ruhr-Universität Bo*-

chum.

- Konieczny, L. & Hemforth, B. (1994). Incremental parsing with lexicalized grammars. In G. Strube (Ed.), *Current research in Cognitive Science at the Center for Cognitive Science* (Vol. IIG-Berichte 1/94, pp. 33-54). Freiburg: Albert-Ludwigs Universität, Institut für Informatik und Gesellschaft.
- Konieczny, L., Hemforth, B. & Scheepers, C. (1994). Reanalysis vs. internal repairs: Nonmonotonic processes in sentence perception. In B. Hemforth, L. Konieczny, C. Scheepers & G. Strube (Eds.), *First Analysis, reanalysis, and repair* (Vol. IIG-Berichte 8/94, pp. 1-23). Freiburg: Institut für Informatik und Gesellschaft.
- Konieczny, L., Hemforth, B., Scheepers, C., & Strube, G. (1995). PP-attachment in German: results from eye movement studies. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), *Eye movement research. Mechanisms, processes and applications* (Vol. 6, pp. 405-420). Amsterdam: North Holland.
- Konieczny, L., Hemforth, B., & Strube, G. (1991). Psychologisch fundierte Prinzipien der Satzverarbeitung jenseits von Minimal Attachment. *Kognitionswissenschaft, 2.*
- Konieczny, L., Hemforth, B., & Völker, N. (1995). Der Einfluß von Verbstellung und Kontext auf die Anbindung von Konstituenten. Kognitionswissenschaft, 5, 104 - 114.
- Konieczny, L., Scheepers, C., Hemforth, B. & Strube, G. (1994). Semantikorientierte Syntaxverarbeitung. In S. Felix, C. Habel, & G. Rickheit (Eds.), *Kognitive Linguistik: Repräsentationen* und Prozesse. Opladen: Westdeutscher Verlag.
- Konieczny, L. & Strube, G. (1995). SOUL: A Cognitive Parser. Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society (pp. 631-636). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Konieczny, L. (in prep.). The SOUL system: an object-oriented implementation of a feature unification formalism and a sign-oriented parser. *IIG-Berichte*. Universität Freiburg.
- Kutas, M. & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*, *11*, 99-116.
- Kutas, M. & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition*, *11*, 539-550.
- Kutas, M. & Hillyard, S. A. (1989). An electrophysiological probe of incidental semantic association. *Journal of Cognitive Neuroscience*, *1*, 38-49.
- Liversedge, S. (1994). *Referential contexts, relative clauses, and syntactic parsing.* University of Dundee: Unpublished doctoral thesis.
- Liversedge, S., Pickering, M., & Branigan, H. (1995). The comprehension of sentences that are ambiguous between agentive and locative interpretations. Tucson, AZ: *Poster presented at the 8th annual CUNY conference on human sentence processing*.
- MacDonald, M. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 4,* 35-56.
- MacDonald, M., Pearlmutter, N. & Seidenberg, M.S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101,4, 676-703.
- Marcus, M. (1980). A theory of syntactic recognition. Cambridge, MA: MIT Press.
- Marcus, M., Hindle, D. & Fleck, M. (1983). D-Theory: Talking about talking about trees. Association for Computational Linguistics, 21, 129-136.
- Marr, D. (1982). Vision. San Francisco: Freedman.
- Marslen-Wilson, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature, 244*, 522-523.

- Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, *189*, 226-228.
- Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8,* 1-71.
- McCloskey, J. (1988). Syntactic theory. In F. J. Newmeyer (Ed.), *Linguistics: the Cambridge survey* (Vol. 1, pp. 18 59). Cambridge: Cambridge University Press.
- McConkie, G. W. & Rayner, K. (1975). The span of the effective stimulus during an eye fixation in reading. *Perception and Psychophysics, 17,* 578-586.
- Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. . (1995). Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory and Cognition, 23,* 477-494.
- Miller, G. A. (1951). Language and Communication. New York: McGraw-Hill.
- Miller, G. A. (1952). Finite Markov processes in psychology. Psychometrika, 17, 149-167.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *53*, 81-97.
- Miller, G. A. (1962). Some psychological studies of grammar. *American Psychologist*, 17, 748-762.
- Miller, G. A. & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology*, *63*, 176 185.
- Mitchell, D. (1994). Sentence Parsing. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 375-410). San Diego: Academic Press.
- Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods for investigating iemmediate processes in reading. In D. E. Kieras & M. A. Just (Eds.), New Methods in Reading Comprehension Research (pp. 69 90). Hillsdale, NJ: Lawrence Erlbaum.
- Mitchell, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), Attention and Performance XII: The psychology of reading. London: Lawrence Erlbaum Associates Ltd.
- Mitchell, D. C. (1989). Verb-guidance and other lexical effects in parsing. *Language and Cognitive Processes, 4,* 123-155.
- Mitchell, D. C., Corley, M. M.B., & Garnham, A. (1992). Effects of Context in Human Sentence Parsing: Evidence Against a Discourse-Based Proposal Mechanism. *Journal of Experimental Psychology: Learning Memory and Cognition.*, 18, 69-88.
- Mitchell, D. C. & Cuetos, F. (1991). The origins of parsing strategies. In C. Smith (Ed.), *Current Issues in Natural Language Processing*. Austin: Center for Cognitive Science, University of Texas.
- Mitchell, D., Cuetos, F., Corley, M., & Brysbaert, M. (1995). The linguistic tuning hypothesis: further corpus and experimental evidence. Tucson, AZ: *Paper presented at the 8th annual CUNY conference on human sentence processing*.
- Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perceptin and Performance, 10,* 667-682.
- Müller, S. (in prep.). The Babel-System An HPSG fragment for German, a Parser and a dialog component.
- Murray, W. S. & Rowan, M. (1996). Tasks and timing: Early, mandatory, pragmatic processing. New York: Paper presented at the 9th annual CUNY conference on human sentence pro-

cessing.

- Murray, W. & Watts, S. (1995). Prosodic form and parsing commitments. Tucson, AZ: Paper presented at the 8th annual CUNY conference on human sentence processing.
- Netter, K. (1993). Towards a theory of functional heads: German nominal phrases. In J. Nerbonne, K. Netter, & C. Pollard (Eds.), *German in Head-driven Phrase Structure Grammar* (pp. 297-340). Menlo Park: CSLI.
- Netter, K., Kapser, R., Kiefer, B., & Vijay-Shanker, K. (1994). HPSG and TAG. *Paper presented at the 3e colloque international sur les grammaires d'arbres adjoints*. Rapport technique TA-LANA-RT-94-01, pp. 77-82, Paris 1994.
- Neumann, G. (1993). Application of Explanation-based Learning for Efficient Processing of Constraint-based Grammars. Saarbrücken: DFKI (submitted to CAIA-94, San Antonio, Texas).
- Niv, M. (1993). Resolution of syntactic ambiguity: the case of new subjects. *Proceedings of the 15th annual conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.
- O'Regan, J. K. & Levy-Schoen, A. (1987). Eye movement strategy and tactics in word recognition and reading. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 363-383). London: Erlbaum.
- Osterhout, L. & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language.*, 31, 785-806.
- Partee, B., Ter Meulen, A., & Wall, R. (1990). *Mathematical methods in linguistics*. Dordrecht: Kluwer.
- Pechmann, T., Uszkoreit, H., Engelkamp, J., & Zerbst, D. (1993). Word order in German verbfinal sentences: A linguistic and psycholinguistic problem. *Manuscript of the University of the Saarland.*
- Perfetti, C. A. (1990). The cooperative language processor: Semantic influences in an autonomous syntax. In Balota, D. A., d'Aarcais, G. B. F. & Rayner, K. (Eds.), *Processes in reading* (pp. 205-230). Hillsdale, New York: Lawrence Erlbaum Associates.
- Perfetti, C. A. & Lesgold, A. M. (1977). Discourse comprehension and sources of individual differences. In M.A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, NJ: Erlbaum.
- Pickering, M. & Barry, G. (1991). Sentence Processing without Empty Categories. *Journal of Language and Cognitive Processes*, 230-259.
- Pinker, S. (1994). The language instinct. New York: William Morrow & Co.
- Pollard, C. & Sag, I. A. (1987). *Information-based Syntax and Semantics*. CSLI Lecture Notes (No. 13). Stanford: Center for the Study of Language and Information, 1.
- Pollard, C., & Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar.* Chicago, London: University of Chicago Press and CSLI Publications.
- Pritchett, B. L. (1988). Garden path phenomena and the grammatical basis of language processing. *Language*, *64*, 539-576.
- Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.
- Rayner, K., Garrod, S., & Perfetti, C. A. (1992). Discourse influences during parsing are delayed. *Cognition*, 45, 109-139.
- Rayner, K. & Pollatsek, A. (1987). Eye movements in reading: A tutorial review. In M. Coltheart (Ed.), Attention and performance XII: The psychology of reading (pp. 327-362). London: Erlbaum.
- Rayner, K., & Pollatsek, A. (1989). The psychology of reading. Englewood Cliffs, NJ: Prentice

Hall.

- Rayner, K. & Sereno, S. (1994). Eye movements in reading: Psycholinguistic studies. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 57-82). San Diego: Academic Press.
- Rayner, K., Sereno S. C., Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes, 4,* SI 21-50.
- Rayner, K., Well, A. D., Pollatsek, A., & Bertera, J. H. (1982). The availability of useful information to the right of the fixation in reading. *Perception & Psychophysics, 31,* 537-550.
- Schank, R. C. (1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology, 3,* 552-631.
- Scheepers, C., Hemforth, B. & Konieczny, L. (1994). Resolving NP-attachment ambiguities in German verb-final constructions. In B. Hemforth, L. Konieczny, C. Scheepers & G. Strube (Eds.), *First Analysis, reanalysis, and repair* (Vol. IIG-Berichte 8/94, pp. 51-76). Freiburg: Institut für Informatik und Gesellschaft.
- Scheepers, C., Konieczny, L. & Hemforth, B. (1994). On the semantic representation of verbs in a role-restricted sort hierarchy. In G. Strube (Ed.), *Current research in Cognitive Science at the Center for Cognitive Science* (Vol. IIG-Berichte 1/94, pp. 55-66). Freiburg: Albert-Ludwigs Universitt, Institut fr Informatik und Gesellschaft.
- Schütze, C. T. & Gibson, E. (1996). The role of argument preference in English PP attachment. New York: *Poster presented at the 9th annual CUNY conference on human sentence processing.*
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.
- Shieber, S. M. (1986). An introduction to unification-based approaches to grammar. CSLI lecture notes, Number 4. Menlo Park: Center for the study of language.
- Skinner, B. F. (1957). Verbal Behavior. New York: Appleton-Century-Crofts.
- Steinhauer, K. & Friederici, A. (1995). *Subject-object asymmetries in German: ERP studies*. Freiburg: Paper presented at the IIG.
- Stevenson, S. (1993). A constrained active attacment model for resolving syntactic ambiguities in natural language parsing. University of Maryland: Doctoral dissertation.
- Stevenson, S. (1995). Arguments and Adjuncts: a computational explanation of asymmetries in attachment preferences. *Proceedings of the 17th annual conference of the Cognitive Science Society*, July 22-25, 1995. Mahwah, NJ: Lawrence Erlbaum.
- Stowe, L. A. (1989). Thematic structures and sentence comprehension. In G. E. Carlson & M. K. Tanenhaus (Eds.), *Linguistic Structure in Language Processing*. Dordrecht, etc.: Kluwer.
- Stowell, T. (1981). Origins of phrase structure. MIT: Doctoral dissertation.
- Strube, G., Hemforth, B., & Wrobel, H. (1990a). Auf dem Weg zu psychologisch fundierten Modellen der Sprachverarbeitung: Echtzeitanalysen des Satzverstehens. In S. Felix, S. Kanngieer & G. Rickheit (Eds.), Sprache und Wissen. Beiträge des DFG-Forschungsschwerpunktes Kognitive Linguistik. Opladen: Westdeutscher Verlag.
- Strube, G., Hemforth, B., & Wrobel, H. (1990b). Resolution of Structural Ambiguities in Sentence Comprehension: On-line Analysis of Syntactic, Lexical, and Semantic Effects. *The 12th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum, pp. 558-565.
- Sturt, P. & Crocker, M. (1995). Incrementality and monotonicity in syntactic parsing. *Edinburgh Working Papers in Cognitive Science, 10.*
- Sturt, P. & Crocker, M. (1995). Monotonic parsing and reanalysis. Tucson, Arizona: Paper pre-
sented at the 8th annual CUNY conference on human sentence processing.

- Sturt, P. & Crocker, M. (1996). Flexible monotonicity: command and construal in incremental sentence processing. New York: Paper presented at the 9th annual CUNY conference on human sentence processing.
- Tanenhaus, M., Carlson, G., & Trueswell, J. C. (1989). The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes, 4,* 211-234.
- Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influence of content-based expectations. *Journal of Memory and Language, 27,* 597-632.
- Travis, L. (1991). Parameters of phrase structure and verb-second phenomena. In R. Freidin (Ed.), *Principles and parameters in comparative grammar* (pp. 339-364). Cambridge, MA: MIT Press.
- Trueswell, J.C., Tanenhaus, M.K., & Garnsey, S.M. (1994). Semantic influences on parsing: use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33,* 285-318.
- Trueswell, J.C. Tanenhaus, W.K. Kello, C. (1993). Verb specific constraints in sentence processing: Separating effects of lexical perference from garden paths. *Journal of Experimental Psychology: Learning Memory and Cognition.*, 19, 528-553.
- Trueswell, John C. tanenhaus, Michael K. Kello, Christopher. (1993). Verb-Specific Constraints in Sentence Processing: Separating Effects of Lexical Preference From Garden Paths. *Journal of Experimental Psychology: Learning , Memory and Cognition.*, 19, 528-553.
- Uszkoreit, H. (1986). Constraints on order. Linguistics, 24, 884-906.
- Uszkoreit, H., Backofen, R., Busemann, S., Diagne, A.K., Hinkelmann, E., Kasper, W., Kiefer, B., Krieger, H., Netter, K., Neumann, G., Oepen, S., & Spackman, P. (1994). DISCO-An HPSG-based NLP system and its applications for appointment scheduling . *Proceedings* of COLING-94.
- Vitu, F. (1993). The influence of reading rythm on the optimal landing position effect. In G. d'Ydewalle & J. van Rensbergen (Eds.), *Perception and Cognition. Advances in eye movement research* (Vol. Studies in visual Information processing 4, pp. 181-192). Amsterdam: Elsevier.
- Wanner, E. (1980). The ATN and the sausage machine: Which one is baloney? *Cognition, 8,* 209-225.
- Wanner, E., & Maratsos, M. P. (1978). An ATN approach to comprehension. In M. Halle, J. W. Bresnan and G. A. Miller (Ed.), *Linguistic Theory and Psychological Reality*. Cambridge, MA: MIT Press.
- Weinberg, A. (1993). Parameters in the theory of sentence processing: minimal commitment theory goes east. *Journal of Psycholinguistic Research, 22,* 339-364.
- Woods, W. (1970). Transition network grammars for natural language analysis. *Communications of the Association for Computing Machinery, 13,* 591-602.
- Zagar, D. & Pynte, J. (1992). The role of semantic information and attention in processing syntactic ambiguity: An eye movement study. Paris: *Paper presented at the 5th conference of the European Society for Cognitive Psychology*, Paris, September 12 - 16.