The Architecture Underlying Syntactic Processing

Rutger Petrus Gerardus van Gompel

Department of Psychology
University of Glasgow

Submitted for the degree of PhD to the Higher Degrees Committee of the Faculty of Social Sciences, University of Glasgow.

September 1999

Abstract

In this thesis, I report five eyetracking experiments that tested current sentence processing theories. So far, most research has attempted to discriminate between various sentence processing theories by investigating whether non-syntactic sources of information can be employed immediately in syntactic ambiguity resolution. Two-stage theories such as the garden-path theory (e.g., Frazier, 1987a; Rayner, Carlson, & Frazier, 1983) claim that the use of non-syntactic information is delayed, whereas interactive or constraint-based theories (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Tyler & Marslen-Wilson, 1977) claim that all sources of information can be employed immediately. The experiments in this thesis focussed on a different aspect of current sentence processing theories, which has been largely ignored. They investigated whether the architecture of the sentence processor involves reanalysis or competition. Two-stage theories claim that processing difficulty occurs when an initially adopted syntactic analysis has to be revised, whereas most current constraint-based theories stipulate that competition between two or more syntactic analyses that are activated in parallel makes a sentence difficult to process.

The experiments in this thesis investigated reanalysis and competition by testing globally ambiguous syntactic structures and contrasting them with structures that are disambiguated (either to one analysis or the other). Constraint-based competition theories predict that competition occurs in globally ambiguous sentences which do not have a bias for one structure over another, because two syntactic analyses are about equally activated. No such competition should occur in disambiguated sentences, because only one analysis is supported by the disambiguating information. In contrast, traditional two-stage theories such as the garden-path theory predict that the processor initially adopts the structurally preferred analysis. When the disambiguation is inconsistent with this analysis, reanalysis should occur. Reanalysis should not occur when a sentence is disambiguated toward its preferred analysis, or when a sentence is globally ambiguous.

The eyetracking experiments in this thesis showed that disambiguated sentences are more difficult to read than globally ambiguous sentences. These results are incompatible with competition as a mechanism of syntactic ambiguity resolution, and therefore disconfirm the predictions of most current constraint-based theories. They are also problematic for some two-stage theories, because processing difficulty occurred in sentences where the disambiguation was toward the structurally preferred analysis.

In this thesis an alternative model, the *unrestricted race model*, is proposed, which explains the results in a straightforward manner. The unrestricted race model claims that the alternative analyses of a syntactically ambiguous sentence are engaged in a race. The analysis that is constructed fastest is adopted. The model stipulates that the analysis that receives most support from both syntactic and non-syntactic sources of information usually wins the race. When two analyses are about equally supported, as in *balanced ambiguities*, each analysis is adopted about half the time. Consequently, when the sentence is disambiguated (toward one analysis or the other), it is inconsistent with the analysis on half the trials, and therefore reanalysis should occur on those trials. Thus, the disambiguated sentences are more difficult than the ambiguous sentences, where reanalysis does not occur. Balanced ambiguities contrast with *biased ambiguities*, where there is a preference for one analysis. The unrestricted race model predicts that in such ambiguities, the processor adopts the preferred analysis on nearly all trials. Therefore, reanalysis should occur on very few trials when the disambiguation is consistent with this preference.

Acknowledgements

During my PhD, several people have been of great help and support. Firstly, Martin Pickering, my supervisor, was always available to discuss new ideas and help me with whatever problem I had. Without Martin's help and advice, my PhD would have been much more difficult. I am very grateful to him. I had stimulating discussions with many other people, and many gave me advice and help. I would especially like to thank Simon Liversedge and Christoph Scheepers for their discussions, ideas and great help. This thesis also benefited from the help and discussions of Simon Garrod (my second supervisor), Chuck Clifton, Steven Frisson, Asifa Majid, Andrew Stewart, Patrick Sturt and Matt Traxler. I would like to thank them as well. Finally, I would like to thank my examiners, Keith Rayner and Sara Sereno, for their helpful comments and for reading this thesis in such a short time.

The studies in this thesis were conducted in collaboration with Martin Pickering, Christoph Scheepers, Matt Traxler and Simon Liversedge. Martin assisted me in constructing the materials for the experiments that are reported in Chapter 4-6, Matt helped me with the materials in Chapter 4, and Simon with the materials in Chapter 5 and 6. Christoph Scheepers ran the eye-tracking experiment that is reported in Chapter 7 and helped me constructing the items.

Declaration

I declare that this thesis has been composed by myself and that the research reported here has been conducted by myself unless otherwise indicated.

Contents

List of Tables

List of Figures

Abbreviations

ADV	Adverb
ANOVA	Analysis of variance
C	Complementiser
CP	Complementiser phrase
DET	Determiner
I	Inflection
IP	Inflectional phrase
ms	milliseconds
N	Noun
NP	Noun phrase
PP	Prepositional phrase
S	Sentence
SPEC	Specifier
V	Verb
VP	Verb phrase
Z	Zero

1. Overview of the thesis

This thesis explores the architecture that underlies the human sentence processor. Since the 1970s, numerous theories about the architecture of the human sentence processor have been developed that attempt to explain how this processor deals with syntactic ambiguities in language. In Chapter 2, I outline the most influential of these theories, and describe the evidence that has been obtained for or against the various theories.

Sentence processing theories can roughly be divided into two classes. According to one class of models, which I will term two-stage models, the processor works in two stages. In the first stage, the processor makes an initial analysis on the basis of a restricted range of information. In the second stage, the processor checks this analysis against other sources of information. If the initial analysis is incorrect or unlikely to be correct, reanalysis occurs. This process of reanalysis causes the processor to slow down. The most influential of such two-stage theories has been the garden-path theory (e.g., Frazier, 1987; Frazier & Rayner, 1982), which is described in section 2.4. According to the garden-path theory, the initial analysis is based on linguistic tree structure information only. Other two-stage models (e.g., sections 2.7 and 2.8) assume that the initial analysis is based on other sources of information.

The second class of theories, usually called interactive or constraint-based theories, stipulates that there is only a single processing stage. If a sentence is syntactically ambiguous, the syntactic analyses are activated in parallel and the processor uses all possible sources of information simultaneously to foreground one of the analyses. According to most constraint-based theories (e.g., MacDonald et al., 1994; McRae et al., 1998), processing difficulty results from competition between the analyses. This occurs when two or more syntactic analyses are about equally activated by the various sources of information.

Most experimental research has attempted to discriminate between the two classes of models by investigating whether non-syntactic sources of information are initially ignored in sentence processing. In Chapter 2, I describe experimental results on the use of plausibility, frequency, and discourse information in syntactic ambiguity resolution. Some research has shown that these sources of information can be used very rapidly (confirming the predictions of constraint-based theories), whereas other studies

show that the use of this information is delayed. Hence, these studies do not satisfactorily distinguish between the theories.

In Chapter 3, I suggest that contrasting competition with reanalysis is an alternative way to test current sentence processing theories. Two-stage theories claim that processing difficulty is due to reanalysis, whereas most constraint-based theories claim that competition causes processing disruption. Constraint-based theories predict that processing disruption should occur in globally ambiguous sentences in which there is no preference for one analysis over the other. In such sentences two or more syntactic structures are about equally activated and therefore severe competition should occur. In contrast, disambiguated sentences should be easy to process, because one analysis is more strongly activated than its alternative(s). Two-stage theories make very different predictions. Reanalysis occurs when a sentence is disambiguated toward an analysis which is different from the initially adopted analysis, and this should cause processing difficulty. But reanalysis should never occur in globally ambiguous syntactic structures.

Section 3.5 describes an alternative to current sentence processing models, the unrestricted race model (e.g., Van Gompel, Pickering, & Traxler, 1999; in press). In this model, the various analyses of an ambiguous syntactic structure are engaged in a race to be constructed first. The analysis that is supported most by the various sources of information (both syntactic and non-syntactic) is most likely be constructed fastest and will usually be selected. The model claims that when the various sources of information strongly favour one analysis over its alternative (a biased ambiguity), this analysis will nearly always be adopted. In contrast, when two analyses are about equally preferred (a balanced ambiguity), each analysis will be adopted about half the time. The unrestricted race model is a reanalysis model: when the selected analysis is inconsistent with later information, reanalysis occurs.

The subsequent chapters (Chapter 4-7) describe eyetracking experiments that tested the unrestricted race model and contrasted it with other sentence processing theories. In order to distinguish between the theories, they focussed on a previously largely ignored aspect of the sentence processing architecture, namely whether reanalysis or competition occurs during syntactic ambiguity resolution. This was achieved by using both globally ambiguous and disambiguated sentences.

In Experiment 1 (Chapter 4), syntactic ambiguities such as (1) were investigated in an eyetracking experiment.

1a. The hunter killed only the poacher with the rifle not long after sunset. (ambiguous)

1b. The hunter killed only the leopard with the rifle not long after sunset. (VP attachment)

1c. The hunter killed only the leopard with the scars not long after sunset. (NP attachment)

In these ambiguities, the PP *with the rifle/scars* can modify either the verb *killed*, resulting in VP attachment, or the NP *the poacher/leopard*, resulting in NP attachment. Three conditions were tested. The condition in (1a) is globally ambiguous, because the PP *with the rifle* can either be VP or NP attached. This contrasts with (1b) and (1c), where only one analysis is plausible. Pre-tests established that there was no clear preference for either VP or NP attachment before the point of disambiguation: the ambiguity was *balanced.*

According to the garden-path model, the processor should initially adopt the VP attachment analysis, because this is the structurally simplest analysis. Because plausibility information is inconsistent with this analysis in (1c), reanalysis is predicted to occur here, resulting in processing difficulty. In contrast, no reanalysis should occur in (1a) and (1b), because plausibility information is consistent with the initial analysis.

Constraint-based competition models predict that competition should occur in the ambiguous condition. Because the ambiguity is balanced and because plausibility information does not favour either analysis, such models predict that severe competition should occur in the ambiguous condition (1a). No such competition should occur in the disambiguated conditions (1b) and (1c), because only one analysis is supported by plausibility.

The unrestricted race model makes exactly the opposite predictions. Because the ambiguity is balanced, each analysis should be selected about half the time. Consequently, in the disambiguated conditions the initial analysis would be implausible on about half the trials, and therefore reanalysis should occur on about half of the trials for both disambiguated conditions. Both conditions are predicted to be more difficult than the ambiguous condition, where the initial analysis is always plausible.

The results from the eyetracking experiment in Chapter 4 supported the unrestricted race model. First-pass regressions and regression-path times for the region

immediately following the point of disambiguation showed that sentences such as (1a) were easier to read than (1b) and (1c). These results are inconsistent with competition, and therefore provide evidence against most current constraint-based theories. In addition, they do not support the garden-path theory, which predicts that VP attachment is always initially adopted.

The subsequent eyetracking experiments elaborated on this finding. A very similar pattern of results was observed in Experiment 2 (Chapter 5), which investigated relative clause attachment ambiguities such as (2).

2a. The basement of the tenement that used to contain lots of junk was very gloomy. (ambiguous)

2b. The tenement of the landlord that used to contain lots of junk was very gloomy. (NP1 attachment)

2c. The landlord of the tenement that used to contain lots of junk was very gloomy. (NP2 attachment)

Again, semantically disambiguated sentences were more difficult to read than globally ambiguous sentences. However, it is noted in Chapter 5 that the results from ambiguities such as (1) and (2) might be consistent with competition models if one analysis is somehow selected in the ambiguous region that precedes the disambiguation (e.g., at *with the* in [1]). When the disambiguation is reached, only one analysis is available to the processor. If the disambiguation is inconsistent with it, the alternative has to be reactivated, which would cause difficulty and give rise to reanalysis-like effects.

In order to test this hypothesis, Experiment 3 (Chapter 5) used sentences such as (3), which do not contain an ambiguous region.

3a. The basement of the tenement containing lots of junk was very gloomy. (ambiguous)

3b. The tenement of the landlord containing lots of junk was very gloomy. (NP1 attachment)

3c. The landlord of the tenement containing lots of junk was very gloomy. (NP2 attachment)

In such sentences, the processor would not be able to select one analysis before the disambiguation. Although the results from Experiment 3 were somewhat equivocal, a subsequent replication with very similar materials in Experiment 4 (Chapter 6) showed the same pattern as for the ambiguities with delayed disambiguation: both disambiguated conditions were more difficult than the ambiguous condition. I conclude that the syntactic processor works in two stages. In the first stage, it selects one analysis, and in the second stage, it revises this analysis if necessary.

Experiment 4 also addressed another issue. Given that the processor works in two stages, what is the mechanism to select one analysis in the first stage? It is hypothesised that this initial selection process can involve either competition or a race. If it involves competition, any syntactically ambiguous sentence (regardless of whether one analysis is implausible, as in constraint-based theories) should be more difficult to read than an unambiguous sentence. In contrast, a race process predicts that globally ambiguous sentences are no more difficult than unambiguous sentences. In order to test these hypotheses, unambiguous sentences such as (4) were used.

4. Almost at once Colin noted that the tenement containing lots of junk was very gloomy.

No differences between the globally ambiguous and the unambiguous condition were observed in Experiment 4, although both conditions were more difficult to read than the disambiguated conditions. This supported a race process, such as the one incorporated into the unrestricted race model.

Finally, Experiment 5 (Chapter 7) investigated a completely different syntactic ambiguity, a subject-object word-order ambiguity in German. Pre-tests established that this ambiguity was biased, which contrasts with the balanced ambiguities in the other experiments. For biased ambiguities, the unrestricted race model predicts that readers nearly always adopt the preferred analysis. Hence, when the disambiguation is towards this analysis, little difficulty should occur, because reanalysis occurs on very few trials. When the disambiguation supports the alternative, non-preferred analysis, however, processing difficulty should be severe. This was borne out by the data: the preferred analysis condition was easier than the non-preferred analysis condition. However, there

was some indication that the preferred analysis condition was also somewhat more difficult than the ambiguous condition, suggesting that the processor adopted the non-preferred analysis on a small proportion of trials, in accordance with the unrestricted race model.

In sum, the experiments showed a clear pattern: globally ambiguous sentences were easier to read than disambiguated sentences. This pattern is inconsistent with constraint-based competition models which predict that competition should occur in the ambiguous conditions. It is also difficult to reconcile with some two-stage theories such as the garden-path theory, which predicts that readers should initially adopt the structurally preferred analysis. In the final chapter (Chapter 8), I argue that it is unlikely that the effects that were observed in the eyetracking experiments reflect late processes. In addition, it is argued that explanations that assume that readers did not process the ambiguity in the globally ambiguous conditions cannot satisfactorily account for the data. I conclude that the unrestricted race model provides the best explanation for the results and that reanalysis rather than competition underlies the architecture of the human sentence processor.

2. Syntactic processing: theories and experimental studies

2.1. A precursor of current sentence processing models: Bever (1970)

Many of the basic notions of current sentence processing models were introduced in a seminal paper by Bever (1970). He argued that the language processor employs a number of heuristic or perceptual strategies to uncover syntactic structures from the surface form of a sentence. If these strategies did not apply, the complexity was predicted to increase. This idea contrasted with previous sentence processing theories such as the derivational theory of complexity (e.g., Miller, 1962; see Fodor, Bever, & Garrett, 1974 for a discussion), which claimed that the complexity of a sentence could be explained by the number of transformational rules needed to derive its surface structure.

Although the specific perceptual strategies that Bever proposed are not incorporated into current sentence processing theories, Bever's framework can be taken as the starting point of current sentence processing theories. An important feature of Bever's strategies was that they focussed on how readers and listeners deal with (temporarily) ambiguous sentences, a feature shared with most current sentence processing theories. Bever introduced sentences such as (1) into the literature. This type of sentence causes extreme processing difficulty, often called a garden-path, because one has the impression that the sentence processor has followed a dead end and has to go back to find a new analysis.

1. The horse raced past the barn fell.

Sentence (1) is temporarily ambiguous: until *fell* the sentence can either be analysed as a simple main clause, as in (2), or as a sentence containing a so-called reduced relative clause with a structure as in (3).

2. The horse raced past the barn and fell.
3. The horse driven past the barn fell.

Bever proposed the simple main clause analysis is preferred over the reduced relative clause alternative as a result of two perceptual strategies. According to the first strategy, a noun + verb sequence is analysed as actor + action, unless this is implausible. According to the second strategy, the first noun + verb sequence is analysed as a main clause if the verb is not somehow marked as being part of a subordinate or relative

clause (e.g., *driven* is marked as part of a relative clause). In sentence (1) these two strategies conspire so that *the horse raced* is analysed as an actor + action sequence in a main clause. However, this preferred analysis is incorrect, as indicated by the verb *fell*, which accounts for the complexity of the sentence.

Not all of Bever's strategies were syntactic in nature. Indeed, Bever claimed that semantics guides the interpretation of ambiguous structures, and that, in fact, most normal processing of sentences is carried out with little regard to structural strategies. For example, Bever claimed that (4) is easy to process, because semantic plausibility indicates that the main clause reading is unlikely (horses are unlikely to send anything).

4. The horse sent past the barn fell.

Bever proposed about a dozen strategies to account for the complexity of various kinds of sentences, usually containing a syntactic ambiguity. Because they were usually very specific and applied to only a few sentence structures, it is unclear how they would account for all ambiguities in language. Presumably, one would need a large number of strategies to cover other syntactic ambiguities.

2.2. Kimball (1973): structural strategies

Kimball (1973) was one of the first researchers who addressed the issue of processing difficulty in more detail and attempted to explain the whole range of syntactic complexity effects by adhering to only a few general strategies. In contrast to Bever, he assumed that the sentence processor employed syntactic tree structure information, and claimed that processing difficulty in temporarily ambiguous structures could be explained in terms of preferences for certain syntactic tree structures that are used in generative grammar (e.g., Chomsky, 1957; 1965 for early versions). Kimball proposed seven principles of surface structure parsing that should be able to account for syntactic complexity effects:

1. Top down: Parsing in natural language proceeds according to a top-down mechanism.
2. Right association: Terminal symbols optimally associate to the lowest nonterminal node.
3. New nodes: The construction of a new node is signalled by the occurrence of a grammatical function word.

4. Two sentences: The constituents of no more than two sentences can be parsed at the same time.
5. Closure: A phrase is closed as soon as possible, i.e., unless the next node parsed is an immediate constituent of that phrase.
6. Fixed structure: When the last immediate constituent of a phrase has been formed, and the phrase is closed, it is costly in terms of perceptual complexity ever to go back to reorganise the constituents of that phrase.
7. Processing: When a phrase is closed, it is pushed down into a syntactic (possibly semantic) processing stage and cleared from short-term memory.

The first principle, Top down, predicts that the parser first builds the top S node. For example, when the first word of a sentence is *the*, the S node is build before the NP and DET nodes, as shown in Fig. 2.1.

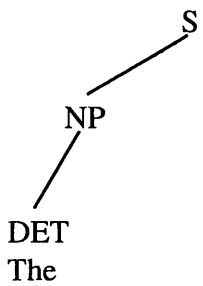

Fig. 2.1. Building of the top S node.

The second principle stipulates that attachment of a phrase low in the tree structure is preferred to attachment high. It is the first of the principles that aims to explain how certain ambiguous syntactic structures are processed. For example, it accounts for the preference to associate *yesterday* in (5) to the final subordinate clause rather than to the main clause or the first subordinate clause. As the tree structure in Fig. 2.2 illustrates, the subordinate clause is lower in the tree structure than the main clause.

5. Joe said Martha expected that it would rain yesterday.

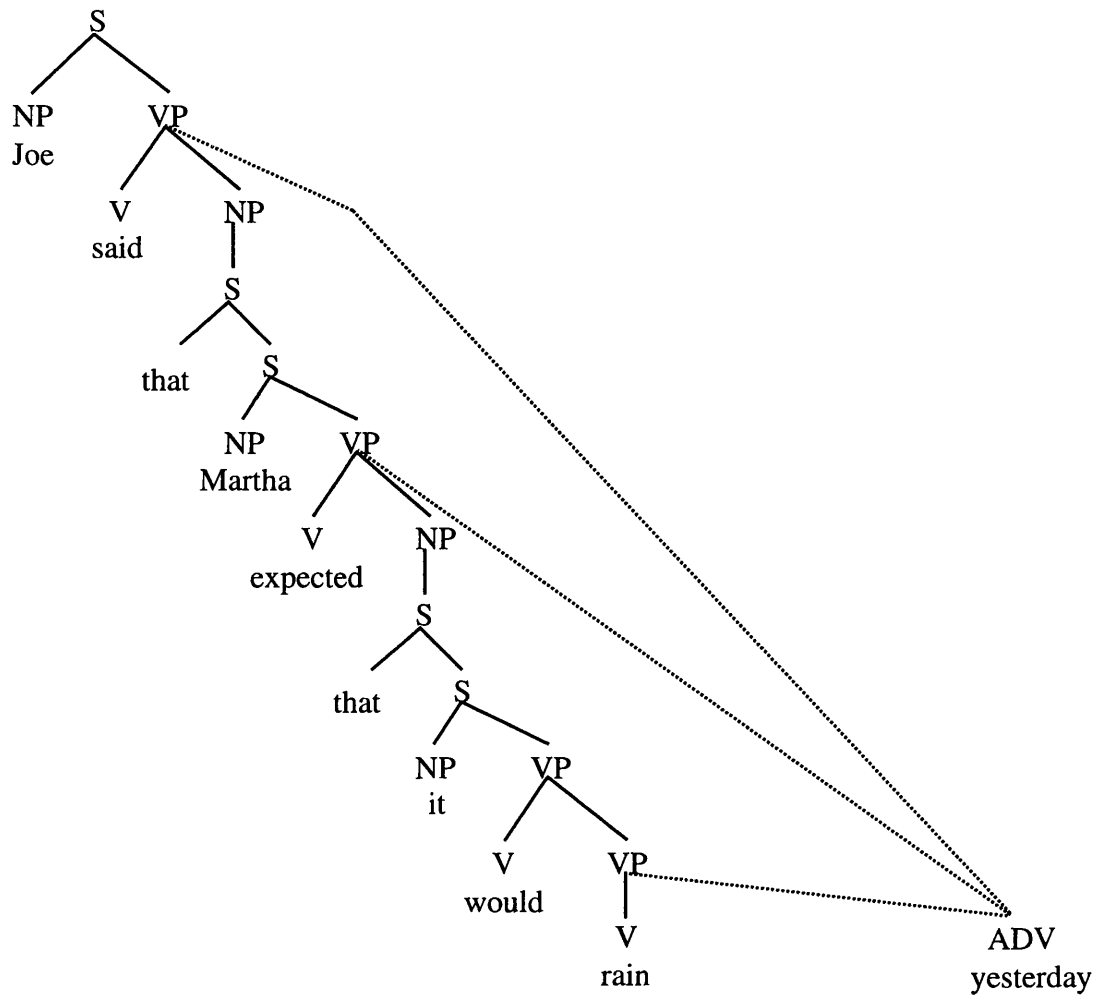

Fig. 2.2. Tree structure of (5).

Similarly, it accounts for the complexity of the unambiguous sentence (6): *to Berta* has to be attached to *gave*, which is very high up the tree structure due to the long direct object *the book that was about the skinning of cats in Alberta between 1898 and 1901*.

6. Joe gave the book that was about the skinning of cats in Alberta between 1898 and 1901 to Berta.

The other principle that deals with ambiguity preferences is principle 5, Closure. It predicts that the main clause in (7) is closed as soon as *the girl* has been encountered, so that it is analysed as the direct object of *knew*. Because this analysis is ultimately incorrect, Closure predicts that this sentence is difficult to process.

7. They knew the girl was in the closet.

Kimball claims that it also accounts for the complexity of (8). Assuming that *The horse raced past the barn* is parsed as a main clause (rather than a reduced relative)[1], Closure predicts that the main clause is closed on *barn* in (8), making the reduced relative reading inaccessible when *fell* is reached.

8. The horse raced past the barn fell.

Principle 3, New nodes, explains why (9) and (10) are not difficult: in both sentences *that* signals the construction of the subordinate clause, as *that* is a function word.

9. They knew that the girl was in the closet.
10. The horse that raced past the barn fell.

Principle 4, Two Sentences, is one of the principles that explains why processing difficulty occurs. Kimball that argues the processor is restricted by memory limitations: no more than two clauses[2] can be kept in short term memory. This explains the difference in complexity between (11) and (12)

11. The boy the girl kissed slept.
12. The boy the girl the man saw kissed slept.

When *the girl* in (11) is encountered, two clauses have been opened (a main and a relative clause) and have to be kept in short term memory. In (12), three clauses have been opened at *the man*. Because this exceeds the two-clause memory limitation of the processor, this sentence is extremely difficult to process. In fact, the Two Sentences principle follows from the final principle, Processing. This principle states that as soon a phrase is closed, it is shunted to a subsequent processing stage where short-term memory limitations do not apply.

Although the limitation on memory capacity is one of the explanations of processing difficulty in Kimball (1973) and in fact the main one in a subsequent paper (Kimball, 1975), it would not turn out to be very influential in most later sentence processing models (but see Gibson, 1991; 1998; Just & Carpenter, 1992; King & Just, 1991; Lewis, 1996; MacDonald, Just, & Carpenter, 1992, for more recent accounts of short term memory in sentence processing). The more influential principle turned out to be the principle of fixed structure (Principle 6), which introduced reanalysis difficulty to sentence processing models. Fixed structure stipulates that reanalysis of an already

[1] In fact, Frazier (1978) noted that this prediction does not follow from Kimball's principles, but that another strategy, minimal attachment, has to be stipulated.
[2] It would therefore have been more appropriate to call this principle Two clauses.

closed structure is costly. This also follows from the Principle 7, Processing. When a phrase has been cleared from short-term memory, it is not directly accessible for revision any more, and, therefore, reanalysis will be costly. This explains why (7) and (8) are hard to process: Kimball claims that in both sentences the main clause has been closed when the sentence is disambiguated (at *was* and *fell* respectively).

2.3. The sausage machine

Frazier (1978a), Frazier and Fodor (1978) and Fodor and Frazier (1980) proposed a parser which abandons the notion that sentence complexity is due to processing load, and claimed that complexity effects are entirely due to reanalysis difficulty. Their parser works in two stages. In the first stage, the Preliminary Phrase Packager (PPP) or *sausage machine* assigns lexical and phrasal nodes to groups of about six words. In the second stage, the Sentence Structure Supervisor (SSS) combines these packages of words together and constructs a complete interpretation of the sentence.

Frazier and Fodor argue that there is no need to postulate a principle of right association as in Kimball's parser. In their model, a similar principle, which attaches incoming phrases low into the tree structure simply follows from the limited six-window view of the PPP. This can be illustrated with sentence (13), in which the prepositional phrase (PP) *to Mary* can be attached either to the VP node of *read* or to the NP *the letter*. I will refer to this ambiguity as the VP/NP attachment ambiguity henceforth.

13. John read the note, the memo and the letter to Mary.

When the preposition *to* is reached, the words *note, the memo and the letter* are in the viewing window of the PPP. The preceding part of the sentence has been passed onto the SSS. Therefore, the verb *read* and the VP node are not available any more, and *to Mary* can only be attached low to the NP *the letter*. Frazier and Fodor noted that this preference depends on the number of words between the verb and the ambiguous prepositional phrase, because VP attachment is preferred in (14). This supports their model, as the verb is still in the viewing window when the PP is reached in this sentence, whereas it is inconsistent with the right association principle, because it predicts low attachment regardless of the length of the NP.

14. John read the letter to Mary.

In the same way, the model explains why there is a preference for low attachment in (5), and why structures such as (15) and (16), which are disambiguated toward high attachment, are difficult to process.

15. I met the boy whom Sam took to the park's friend.
16. The girl applied for the jobs that was attractive.

Frazier and Fodor argued that there is only one processing strategy that has to be motivated independently. This is the principle of *minimal attachment.* It stipulates that "each lexical item (or other node) is to be attached with the fewest possible number of non-terminal nodes linking it with the nodes already present" (Frazier & Fodor, 1978: p. 320). This explains the preference for VP attachment in (14). As can be seen in Fig. 2.3, attachment to the VP involves the postulation of fewer nodes than attachment to the NP. (The tree structures are identical to those in Frazier [1978a], see Abney, 1989; Frazier, 1990; Pritchett, 1992; Schütze & Gibson, 1999, for discussion and criticisms).

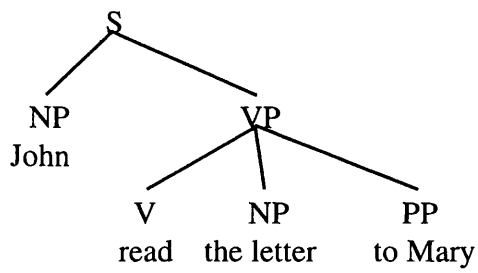

Fig. 2.3a. VP attachment tree structure for (14).

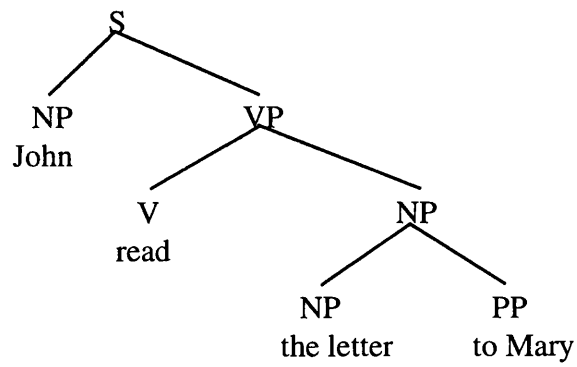

Fig. 2.3b. NP attachment structure for (14).

In a similar way, minimal attachment also accounts for the main clause preference in sentences such as (1).

Frazier and Fodor motivate the minimal attachment principle by assuming that the processor can build structures with the fewest number of nodes faster than structures that involve the postulation of extra nodes. Thus, minimally attached structures will be constructed faster than non-minimal ones.

The sausage machine is a reanalysis model. Difficulty occurs when an initially adopted structure has to be abandoned. Frazier (1978) sketches a way reanalysis might occur: when the processor is forced to reanalyse, it will first consider adding nodes, that is, it will first attempt non-minimal attachments rather than altering structures. This follows from the fact that it is argued that the parser continues to generate non-minimal

structures after a minimal attachment analysis has been adopted. Whenever the minimal attachment has to be abandoned, there will be non-minimal attachment analyses available to accept.

2.4. The garden-path model: parsing principles

Although the sausage machine would have a great impact on later developments in sentence processing theories, it was soon discovered that the notion of a viewing window could not be maintained. The sausage machine predicts that the PPP should make no attachment errors for structures that are within its window. However, Wanner (1980) noted that it was easy to come up with structures that are within the six-word viewing window, but still produce garden path effects. Examples of this are (17) and (18). Both sentences are completely within the viewing window, but both produce garden-path effects.

17. The horse raced yesterday fell.
18. John said Carol will come yesterday.

Frazier and Rayner (1982) abandoned the idea of a viewing window, and argued for the principle of *late closure*, introduced by Frazier (1978). The principle of late closure stipulates that incoming lexical items are preferably attached into the clause or phrase currently being processed (i.e., the lowest possible nonterminal node dominating the last item analysed[3]). This principle is very similar to the principle of right association (see Frazier, 1978 for a comparison).

Late closure accounts for the parsing preferences in (5) and (18), and the complexity of (6) and (15). Frazier and Rayner claimed that late closure operates in tandem with the principle of minimal attachment. In other papers (e.g., Frazier, 1978; 1987a) it was claimed that minimal attachment takes precedence over late closure. Whenever the two strategies conflict, minimal attachment wins. This model is usually referred to as the *garden-path model* of sentence processing. As argued in later papers (e.g., Ferreira & Clifton, 1986; Frazier, 1987a; Rayner et al., 1983) the garden-path model is a purely structural model: only information about syntactic tree structures determines the initial choice of the parser. Other potentially useful sources of information, such as semantic, contextual and frequency information, exert their influence only after the initial structural parse (see section 2.9 - 2.12).

[3] Informally, a node A dominates a node B if A is higher in the tree than B and if going through the tree from A to B does not require going upward. Nodes are also assumed to dominate themselves.

Frazier and Rayner (1982) investigated the principles of late closure and minimal attachment by measuring people's eye-movements while they were reading. This provides a very detailed record of on-line processing (see chapter 3). In one experiment, they tested temporarily ambiguous sentences such as (19b). In (19b), the NP *a mile* is temporarily ambiguous between the object NP of *jogs* and the subject of a following clause. Because the verb can either take an object argument or no argument, I will call these ambiguities NP/Z(ero) ambiguities henceforth (Sturt, Pickering, & Crocker, 1999). Late closure predicts that in NP/Z ambiguities, the ambiguous NP is analysed as the direct object NP of the preceding verb, because its VP is the current phrase.[4] However, this analysis is incorrect in (19b), because *a mile* is the subject of the following clause.

19a. Since Jay always jogs a mile this seems like a short distance to him. (late closure)
19b. Since Jay always jogs a mile seems like a short distance to him. (early closure)

Frazier and Rayner observed that reading times per character were longer for early closure sentences such as (19b) than for late closure sentences such as (19a). These findings supported the late closure principle. The difference between the two types of sentences occurred at the disambiguating region (*this seems* and *seems like*), that is, as soon as it became clear that the direct-object analysis was incorrect.

In their second experiment, they tested the minimal attachment strategy. They used sentences such as those in (20). In these so-called NP/S ambiguities, the initial verb can either take an object argument or a sentence complement. Minimal attachment predicts that the NP analysis is preferred: *the inheritance* in (20b) is attached to the VP of *claim* rather than to the sentence node of the following clause, because the latter involves the postulation of one extra node. However, because (20b) is disambiguated as non-minimally attached at *belongs*, processing difficulty should occur. Sentences such as (20a) were used as minimal attachment controls.

20a. The lawyers think his second wife will claim the inheritance.
(minimal attachment)
20b. The second wife will claim the inheritance belongs to her.
(non-minimal attachment)

[4] Minimal attachment does not discriminate here, because the main clause node is postulated in a top-down fashion as soon as the first word of the sentence is read. Hence, attachments to VP or S are equally minimal.

As predicted, reading times per character were longer for the non-minimal attachment sentences such as (20b) than minimal attachment sentences such as (20a)[5].

Frazier and Rayner argued that readers appeared to use three different strategies to recover from an initial misanalysis in (19b) and (20b). Processing disruption in these garden-path sentences was indicated by longer fixation durations (either followed by a regressive eye movement or not), regressions from the disambiguating region (*seems like* in [19b] and *belongs to* in [20b]) to an earlier part of the disambiguating region, or regressions from this region to the ambiguous region (*a mile* in [19b] and *the inheritance* in [20b]). The latter type of regressions occurred most frequently, and therefore Frazier and Rayner concluded that readers do not systematically backtrack through the sentence when they revise their initial analysis, nor do they start at the beginning. Rather, the results suggested that readers went back to where the misanalysis occurred.

Another structural strategy that has been incorporated in the garden-path model is the *active filler strategy* (Frazier, 1987b, Clifton & Frazier, 1989; Frazier & Clifton, 1989; Frazier & Flores d'Arcais, 1989). This strategy accounts for parsing preferences in *long distance dependencies* such as (21a). In this sentence, *who* is the direct object of *consider*, but it does not occur after it, as in most sentences. Hence, it is called a long-distance dependency.

21a. Who did the housekeeper from Germany urge the guests to consider? (temporarily ambiguous)

21b. The housekeeper from Germany urged the guests to consider the new chef. (unambiguous)

The active filler strategy assumes a transformational grammar, in which phrases are generated in a base position, and are subsequently moved by a transformation to another position in the sentence (e.g., Chomsky, 1981; 1986). When such as movement has taken place, it is assumed that a gap is left behind. For example, *who* in (21a) is moved from its position after *considered* to sentence initial position and leaves a gap behind. The active filler strategy stipulates that when there is an ambiguity as to where a gap has to be located, the earliest position is taken. Sentence (21a) contains such an ambiguity: *who* can be taken as the direct object of *urge* until it is disambiguated by *the guests*, which is *urge*'s true direct object. Because the incorrect gap position after *urge*

[5] Because (20b) is disambiguated by the full stop, Frazier and Rayner reported reading times for the entire sentences. This may be problematic, as the contents of the minimal and non-minimal attachment sentences were different. For example, (20a) also contains an NP/S ambiguity: *his second wife* can be minimally attached to the VP node of *think* or non-minimally to the subsequent clause (which is the correct reading).

is earlier than the correct gap position after *consider*, the active filler strategy predicts that the processor initially analyses *who* as the direct object of *urge*. Because this analysis turns out to be incorrect at *the guests*, reanalysis should occur, resulting in processing difficulty. This was confirmed by Frazier and Clifton (1989), who observed in a self-paced reading experiment that (21a) was more difficult to read in the regions following *urge* than in those following *urged* in the unambiguous condition (21b).

The active filler strategy also deals with certain word-order ambiguities in languages such as German and Dutch (see also Chapter 7). For example, the relative clause in the Dutch sentence in (22) is ambiguous between a subject and an object relative clause. In the subject relative clause, the relative pronoun is the subject of the relative clause and the NP *de boswachter* the object, whereas it is the other way around in the object relative clause.

22. Karl hielp de mijnwerker die de boswachter vond.
Karl helped the miner who the forester found.
"Karl helped the miner who found the forester."/
"Karl helped the miner who the forester found."

Because it is usually assumed that the basic word order in Dutch (and German) is subject-object-verb (Koster, 1975), the gap position for the relative pronoun *die* in subject and object relative clauses are as in (23a) and (23b) respectively (indicated by ___). Because the gap is earlier in subject relatives than in object relatives, the active filler strategy predicts a preference for (23a) over (23b). This has been confirmed by a number of studies in both German and Dutch (e.g., Frazier, 1987b; Frazier & Flores d'Arcais, 1989; Friederici, 1998; Hemforth, 1993; Schlesewsky, Fanselow, Kliegl, & Krems, in press; Schriefers, Friederici, & Kühn, 1995)

23a. Karl hielp de mijnwerker die ___ de boswachter vond. (subject relative)
23b. Karl hielp de mijnwerker die de boswachter ___ vond. (object relative)

2.5. Cross-linguistic evidence against late closure

A number of studies have investigated the application of the late closure principle in structures such as (24) in languages other than English. Late closure predicts that there should be a preference to attach the relative clause *who was in the school celebrating mass* to *the priest* (NP2) rather than to *the brother* (NP1), because the former is the

phrase being processed when the relative clause is encountered. Because the late closure strategy is a universal principle, this preference should apply to any language.

24. Martha cheered the brother of the priest who was in the school celebrating mass.

Cuetos and Mitchell (1988) tested structures similar to (24) in English and Spanish in two questionnaire studies. After each sentence, participants had to answer a question regarding the relative clause attachment. For example, the question in (24) was: *Who was in the school?* In the English materials, there was a preference to attach the relative clause to NP2 (e.g., *the priest*), consistent with late closure. However, in the Spanish materials, there was a preference for attachment to NP1 (e.g., *the brother*). Furthermore, three self-paced reading experiments showed that Spanish sentences that were disambiguated toward NP2 took longer to read than unambiguous control sentences in which there was only one NP to which the relative clause could be attached. Clearly, these results are problematic for the late closure principle, as it predicts a preference for NP2 attachment.[6]

The results were confirmed by Carreiras and Clifton (1993), who conducted a study using similar sentences in Spanish and found a NP1 preference. In contrast, when Carreiras and Clifton tested English translations of the same sentences, they did not find a significant preference for NP1 or NP2 attachment. It is unclear whether the difference between their study in English and Cuetos and Mitchell's study is due to differences in the materials or a difference in subject population, as Carreiras and Clifton used American English participants, whereas Cuetos and Mitchell used British English participants.

Further problems for the principle of late closure are posed by studies in other languages, which also show NP1 attachment preferences in structures like (24). NP1 attachment preferences have been observed in Dutch (Brysbaert & Mitchell, 1996), German (Hemforth, Konieczny & Scheepers, in press) and French (Pynte, 1998; Zagar, Pynte, & Rativeau, 1997), although there is some evidence that NP2 attachment is preferred in Italian (De Vincenzi & Job, 1993; 1995).

In a response to the findings of NP1 preferences, Frazier (1990) argued that although late closure is initially adopted, there are two reasons why NP1 attachment may be the (ultimately) preferred analysis in some languages. One strategy that may force readers to reanalyse their initial NP2 attachment is the *relativised relevance*

[6] It should be noted that Cuetos and Mitchell's reading experiments compared NP2 attachment sentences with unambiguous sentences rather than with NP1 attachment sentences. Hence, it may be premature to conclude that NP1 attachment is easier than NP2 attachment on the basis of their results. However, this argument cannot save the late closure principle, because it predicts that the processor always adopts NP2 attachment, and that, therefore, NP2 attachment sentences should be no more difficult than unambiguous sentences.

principle, which stipulates that a phrase is preferentially construed as being relevant to the main assertion of the sentence. This should result in a preference for NP1 attachment, because NP2 is not normally part of the main assertion of the sentence.

The relativised relevance principle, however, does not account for the difference between Spanish and English relative clause attachment. Frazier (1990) argued that this difference may be due to the fact that English has a genitive construction (as in *the priest's brother*), whereas Spanish does not. In English, ambiguous NP1 attachments can be avoided by using such a construction, as shown in (25).

25. Martha cheered the priest's brother who was in the school celebrating mass.

Frazier argued that because ambiguous NP1 attachment structures can be avoided in English, readers assume relative clauses are in fact NP2 attached. Therefore, NP2 attachment is preferred in English. In contrast, because genitive constructions do not exist in Spanish, NP1 attachment is preferred.

However, Frazier's proposals do not explain why NP1 attachment is also preferred in Dutch, which does have a possessive construction (although it is certainly less frequent than in English). More importantly, it is unclear why a reanalysis triggered by the relevance principle should make NP1 attachment *easier* to read than NP2 attachment. Assuming that processing difficulty occurs when the relativised relevance principle forces readers to reanalyse the initially adopted NP2 attachment, NP1 attachment sentences should be more difficult than NP2 attachment sentences, where no reanalysis should occur.

2.6. Construal

Frazier and Clifton (1996) modified the garden-path model in order to explain parsing preferences in structures such as (24) and in other structures that they claimed have similar characteristics. They proposed that there are two types of syntactic relations, primary relations and non-primary relations. This distinction is roughly the same as the distinction between arguments and adjuncts that is often made in the linguistic literature (e.g., Pollard & Sag, 1987). According to Frazier and Clifton, primary relations include the subject and main predicate of a clause, and complements and obligatory constituents of primary phrases. Non-primary relations include all other syntactic relations, such as adjuncts and relative clauses. Frazier and Clifton assume that phrases that can (temporarily) be taken as primary phrases will be analysed as such. For example, the preferred reading of (26) is *the girl was told that Bill liked the story*, because *that Bill*

liked the story is a complement (primary relation) of *told*, whereas in the non-preferred reading, *the girl that Bill liked was told the story*, the relative clause *that Bill liked* is an adjunct (non-primary relation) of *the girl.*

26. John told the girl that Bill liked the story.

If a phrase is (temporarily) ambiguous between two possible primary relations, the parser employs only structural parsing principles such as late closure and minimal attachment. In such cases, the parser behaves exactly as in the garden-path model. Thus, minimal attachment still applies to structures such as (27) and (28), and late closure in structures such as (29). As in the original garden-path model, non-structural information can only influence processing after the application of these principles.

27. The horse raced past the barn fell.
28. The second wife will claim the inheritance belongs to her.
29. Since Jay always jogs a mile seems like a short distance to him.

If a phrase is ambiguous between two possible non-primary relations, the construal mechanism applies. The construal hypothesis stipulates that whenever a phrase cannot be analysed as instantiating a primary phrase, it is associated into the current thematic processing domain. The current thematic processing domain is formally defined as the extended maximal projection of the last theta assigner. For example, if the last theta assigner is a preposition, the thematic processing domain consists of the PP node in the tree structure and the nodes it dominates. The phrase is not attached in a determinate fashion as is the case for primary relations. When there is more than one possible attachment site within the domain, both structural and non-structural information are employed to select one analysis.

The application of construal can be illustrated with sentences (30) and (31).

30. I left the book that you were reading in the library.
31. The count ordered the steak with the sauce that the chef prepares especially well.

In neither sentence can the ambiguous phrase be attached as a primary phrase, so construal applies. In (30), the last theta assigner preceding the PP *in the library* is the verb *reading*. Thus, construal predicts that it will be associated into the subordinate clause and will not be not interpreted as part of the main clause. Similarly, in (31) the relative clause is associated to the PP *with the sauce*, because the preposition *with* is the

last theta assigner. This predicts a preference for NP2 attachment (to *the sauce*) over NP1 attachment (to *the steak*).

Although (32) is superficially similar to (31), construal makes a different prediction here. In contrast to *with*, the preposition *of* is not a theta-role assigner. Instead, in (32) *the relative* (NP1) assigns a theta role to *the boy* (NP2). Construal predicts that the relative clause is therefore associated to *the relative of the boy*. However, there are two possible attachment sites here: NP1 and NP2. Because the relative clause is associated to this domain rather than attached in a determinate fashion, both structural and non-structural strategies can come into play at this point. Whether NP1 or NP2 attachment is preferred depends on these strategies.

32. The teacher was talking to the relative of the boy who was in the hospital.

Evidence for the difference in parsing preference between (31) and (32) comes from Gilboy, Sopena, Clifton, & Frazier (1995; see also De Vincenzi & Job, 1993; Traxler, Pickering, & Clifton, 1998). They tested similar sentences in both English and Spanish in a questionnaire study of the kind used by Cuetos and Mitchell (1988). Structures containing the preposition *with* showed a preference for NP2 attachment, whereas generally no clear preference was observed for structures with the preposition *of*. An exception were structures such as (33), which showed a clear NP1 preference. Gilboy et al. argued that this preference is due to a non-structural parsing principle, the *referentiality principle*[7]. This stipulates that relative clauses are preferentially attached to NPs that are referential, that is, NPs that introduce or refer back to a discourse entity. Gilboy et al. argued that *water* in (33) is not referential in this sense, because it resists determiners (e.g., *the*), and that therefore, NP1 attachment is preferred in such structures.

33. John asked for the glass of water that was on the table.

Finally, Gilboy et al. also showed that the preference for NP1 attachment in structures with *of* was somewhat higher in Spanish than in English. They argued that in English, readers prefer NP2 attachment because NP1 attachment can be avoided by using a genitive construction, whereas this is not possible in Spanish (see section 2.5). Taken together, they argued that their results provided support for construal.

[7] Note that this principle is different from the relativised relevance principle that was discussed in section 2.5.

2.7. Arguments and adjuncts

As we have seen, one prediction of the construal model is that primary relations are preferred over non-primary relations. Because the distinction between primary and non-primary relations is closely related to the distinction between arguments and adjuncts, this corresponds to the preference for arguments over adjuncts. Precise definitions for argument and adjuncthood are difficult to give. A number of tests have been proposed to distinguish between arguments and adjuncts, but many of these tests have exceptions and they often yield different outcomes (Pollard & Sag, 1987; Schütze & Gibson, 1999). It is usually assumed that a phrase is an argument of a word (usually called the thematic role assigner) if the phrase is part of its subcategorisation frame. An argument phrase is lexically selected by its thematic role assigner, and therefore the meaning of an argument normally depends on the thematic role assigner. If a phrase is an adjunct of a word, it is not part of its subcategorisation frame and has a meaning that is independent of it. This is illustrated in (34) (Clifton, Speer, & Abney, 1991).

34a. The saleswoman tried to interest the man in a wallet during the storewide sale at Steigers. (VP attached argument)

34b. The man expressed his interest in a wallet during the storewide sale at Steigers. (NP attached argument)

34c. The man expressed his interest in a hurry during the storewide sale at Steigers. (VP attached adjunct)

34d. The salesman tried to interest the man in his fifties during the storewide sale at Steigers.
(NP attached adjunct)

In (34a) the PP *in a wallet* is an argument of the verb *to interest*, because its meaning depends on the verb. If the verb is replaced, the meaning of the PP changes (cf. *The saleswoman left the money in a wallet*). Similarly, the meaning of the PP in (34b) depends on the NP *his interest*. When the NP is replaced by another NP, the sentence becomes unacceptable (*The man expressed his concerns in a wallet*). In contrast, the PP *in a hurry* in (34c) is an adjunct modifying the verb *expressed*. When one replaces the verb (e.g., *The man left his money in a hurry*), the meaning of the PP remains the same. *In his fifties* in (34d) is also an adjunct, here modifying the NP *the man*. When one changes the NP, the meaning of the PP remains constant (e.g., *The salesman tried to interest the woman in her fifties*).

A number of sentence processing theories claim that argument attachment is preferred over adjunct attachment (e.g., Abney, 1989; Gibson, 1991; Pritchett, 1988;

1992). In Abney's (1989) model, the strategy to attach phrases as arguments rather than adjuncts has precedence over other strategies, such as the strategy to attach to verbs rather than nouns, and to attach low rather than high in the tree structure. In Gibson (1991) argument preferences arise because extra processing load occurs when a constituent does not receive a theta role and when a role assigned by a theta assigner is left unsatisfied.

In Pritchett's (1988, 1992) model, the preference for arguments follows from a more general property of the parser. Pritchett argues that the parser attempts to maximally satisfy the principles of syntax at every point during processing, which he calls *generalised theta attachment.* The syntactic principles that account for the argument preference in Pritchett's model are the case filter and the theta criterion. According to the case filter, every argument has to receive case from a case assigner (Chomsky, 1986), and according to the theta criterion, every argument has to receive exactly one theta role from a theta role assigner. Hence, when there is a case or theta role assigner, the parser will look for an argument that satisfies these principles. Because adjuncts do not receive case or theta roles, adjunct attachment is not preferred. Besides explaining parsing preferences for sentences such as (34), generalised theta attachment also predicts that (35a) is easier to process than (35b). Because the transitive verb *visited* in (35b) can assign a direct object case and a theme theta role, the parser will look for an argument to satisfy these principles of syntax. This will result in the incorrect analysis of *the doctor* as the direct object of the verb. This does not happen in (35a), because *arrived* does not assign a direct object case and can only assign a theta role to its subject. These predictions were supported by Adams, Clifton, & Mitchell (1998), who investigated sentences such as (35) in an eyetracking experiment and found that reading times for the disambiguating word *prepared* were longer in (35b) than in (35a) (see further section 2.11).

35a. After the young Londoner had arrived his parents prepared to celebrate their anniversary.
(intransitive verb)

35b. After the young Londoner had visited his parents prepared to celebrate their anniversary.
(transitive verb)

Although the results from Adams et al. confirmed one prediction of Pritchett's model, they do not provide evidence for the argumenthood preference that is predicted by Pritchett (1988; 1992) and Abney (1989). Using sentences such as (34) in an eyetracking and self-paced reading study, Clifton et al. (1991) explicitly addressed the

question of whether argument attachment is preferred to adjunct attachment by investigating the sentences in (34).

According to the garden-path model, minimal attachment applies to VP/NP attachment ambiguities such as (34). It claims that non-structural strategies such as a preference for argument attachment do not immediately have an effect on syntactic ambiguity resolution, although they may affect processing after the initial structural stage of processing. Hence, the garden-path model predicts that VP attachment (as in [34a] and [34c]) is at least initially preferred to NP attachment (as in [34b] and [34d]). This prediction was supported by Clifton et al.'s results: in both the self-paced reading experiment and in first-pass reading times[8] in the eyetracking experiment, reading times for VP attachment sentences were shorter than NP attachment sentences for the region consisting of the critical PP. No effect of argumenthood, nor an interaction between argumenthood and attachment site occurred at this point. At the next region (*during the storewide sale*) the conditions with arguments were read faster than those with adjuncts. The effect of argumenthood also occurred in first-pass regressions[9] out of the PP and in total times[10] for the PP and the region following it. Clifton et al. argued that these results provided evidence that the use of argumenthood information is delayed relative to the structural minimal attachment strategy, as predicted by the garden-path model.

However, the finding of an effect of argumenthood in first-pass regressions casts some doubts on this conclusion. First-pass regressions are arguably a very early measure of processing difficulty and it is unclear whether the processes they tap into are necessarily later than the processes measured by first-pass reading times. Rather, it is plausible to assume that readers have two strategies to recover from an initial misanalysis. One strategy is to spend longer in the region where the misanalysis is recovered, while another is to return to a preceding region to check previous input (see section 4.7). The latter strategy may even result in shorter reading times in the critical region when a misanalysis is detected (Rayner, Sereno, Morris, Schmauder, & Clifton, 1989). Furthermore, there are a number of aspects of Clifton et al.'s materials that can be criticised. Schütze and Gibson (1999) noted that they did not control the plausibility of the alternative readings and the number of words across conditions of their materials. Finally, some of Clifton et al.'s PPs were idioms, which may have resulted in fast reading times, and some PPs were not clear adjuncts.

Controlling for these factors, Schütze & Gibson (1999) conducted two experiments in which they contrasted NP attached arguments such as (36a) with VP

[8] The first-pass reading times measure is defined as the sum of all fixations starting with the first fixation in a region until the point of fixation is outside the region. If a reader passes a region before fixating in it, there is no first-pass time. (See section 4.7)

[9] The first-pass regressions measure is defined as the percentage of eye movements leaving a region to the left after entering the region for the first time. (See section 4.7)

[10] Total reading times are defined as the sum of all fixations in a region. (See section 4.7)

attached adjuncts such as (36b). Minimal attachment and the garden-path model predict a preference for VP attachment in such sentences, whereas an argument preference predicts a preference for NP attachment.

36a. The company lawyers considered employee demands for a raise but they didn't act until a strike seemed imminent.

36b. The company lawyers considered employee demands for a month but they didn't act until a strike seemed imminent.

They observed that argument attachment to the NP was easier than adjunct attachment to the VP in the region following the PP (*but they*). This is consistent with an argument attachment preference, but not with the principle of minimal attachment.[11]

2.8. Models of reanalysis

The models discussed so far assume a reanalysis mechanism in sentence processing. They claim that the parser initially makes an analysis on the basis of a restricted range of information (e.g., structural information or a preference for arguments), and when this analysis turns out to be inconsistent with other information, reanalysis of the initially adopted structure occurs. This reanalysis process increases processing time. Until recently, most models that assume such a reanalysis component have focussed almost entirely on initial processing and ignored the question of how the reanalysis process itself proceeds. Similarly, nearly all experimental studies have investigated initial analysis, usually focussing on what sources of information are immediately available. But clearly, if one assumes a reanalysis component, it is essential to develop principled accounts of reanalysis. A number of more recent sentence processing models have done this.

An attractive way of handling reanalysis is provided by a class of models which are called *deterministic parsers*. Parsing of a sentence is deterministic as long as no revisions have to be made to earlier assumptions about the syntax of the sentence. In Marcus (1980), deterministic parsing is achieved by giving the parser a limited look-ahead. Because of this look-ahead, the parser can wait until later disambiguating information is reached, obviating the need for reanalysis (for other proposals of delay models see Frazier & Fodor, 1978; Frazier & Rayner, 1987; Holmes, Kennedy, & Murray, 1987; Kennedy, Murray, Jennings, & Reid, 1989; Perfetti, 1990). An important

[11] It is not consistent at least in the way it was defined by Frazier (1990) and Clifton et al. (1991). Schütze and Gibson show that under certain tree structure assumptions, minimal attachment makes the same predictions as a preference for argument attachment.

hypothesis of Marcus' is that all sentences that are parsed without conscious difficulty can be parsed deterministically. In contrast, sentences that cannot be parsed without conscious difficulty need a different parsing mechanism. It is important to note that conscious difficulty is usually taken to be a severe form of processing difficulty. For example, Pritchett (1992, p. 7) describes conscious difficulty as "unrecoverable parsing errors which cannot be corrected without the conscious invocation of higher rational and nonautomatic cognitive processes".

In a later refinement of the Marcus (1980) model, Marcus, Hindle, and Fleck (1983) proposed what they called *description theory.* In description theory, the parser computes descriptions of tree structures, rather than tree structures themselves. Marcus et al. argued that these descriptions contained information about dominance relations between nodes. This results in an underspecification of the tree structure. For example, as can be seen in Fig. 2.4, when a node α dominates β, there is no specification as to how many nodes are in between α and β. This enables the parser to make some revisions in the tree structure without changing tree structure descriptions and maintaining determinism.

Fig. 2.4. α dominates β.

The look-ahead mechanism that was part of the Marcus (1980) and Marcus et al. (1983) models is difficult to reconcile with experimental evidence for incremental interpretation and the mere existence of garden-path effects within their proposed window size. If the parser delayed its analysis until disambiguating information is reached, it should not be garden-pathed. Hence, later deterministic models have abandoned the look-ahead mechanism (Berwick & Weinberg, 1984; Gorrell, 1995; Sturt & Crocker, 1996; Weinberg, 1993). In addition, the underspecification that is due to only specifying dominance relations in the Marcus et al. model results in few cases where reanalysis is necessary. Hence, later models have introduced further constraints on deterministic parsing (Gorrell, 1995; Sturt & Crocker, 1996).

Gorrell (1995) proposed that deterministic parsing involves not only preserving dominance relations, but also precedence relations. Gorrell makes a distinction between

primary and secondary relations.[12] The primary relations are dominance and precedence. These relations cannot be altered without causing conscious processing difficulty. This contrasts with secondary relations, such as theta role assignment, c-command and case assignment, which can be revised without difficulty. The contrast in processing (37) and (38) illustrates the distinction.

37. The Australian woman saw the famous doctor had been drinking quite a lot.
38. Before the woman visited the famous doctor had been drinking quite a lot.

Gorrell's model predicts that (37) is easier to process than (38). In both sentences, *the famous doctor* is initially analysed as the direct object of the preceding verb. This is due to the principle of simplicity, which states that the parser builds as few tree structure nodes as possible. This principle is very similar to the minimal attachment principle[13]. In both sentences, this initial tree structure needs to be changed after reanalysis. As shown in the simplified structure in Fig. 2.5, both dominance and precedence relations are preserved during this reanalysis in (37): the NP *the famous doctor* is dominated by the VP and topmost S node both before and after reanalysis. The fact that more dominance relations are added is of no importance, as long as the original relations are preserved. Furthermore, all nodes that precede the NP before reanalysis, also precede it after it (e.g., the V-node). As shown in Fig. 2.6, this contrasts with the reanalysis in (38). Here, the dominance relations for the NP *the famous doctor* are changed: for example, the VP node dominates the NP before, but not after reanalysis.

[12] Note that the use of primary and secondary relations here is completely different from the use of these terms in construal (Frazier & Clifton, 1996).

[13] Note, however, that according to the garden-path model, minimal attachment does not apply in (38), because the top S-node is projected as soon as the complementiser *before* is parsed. Instead, late closure applies.

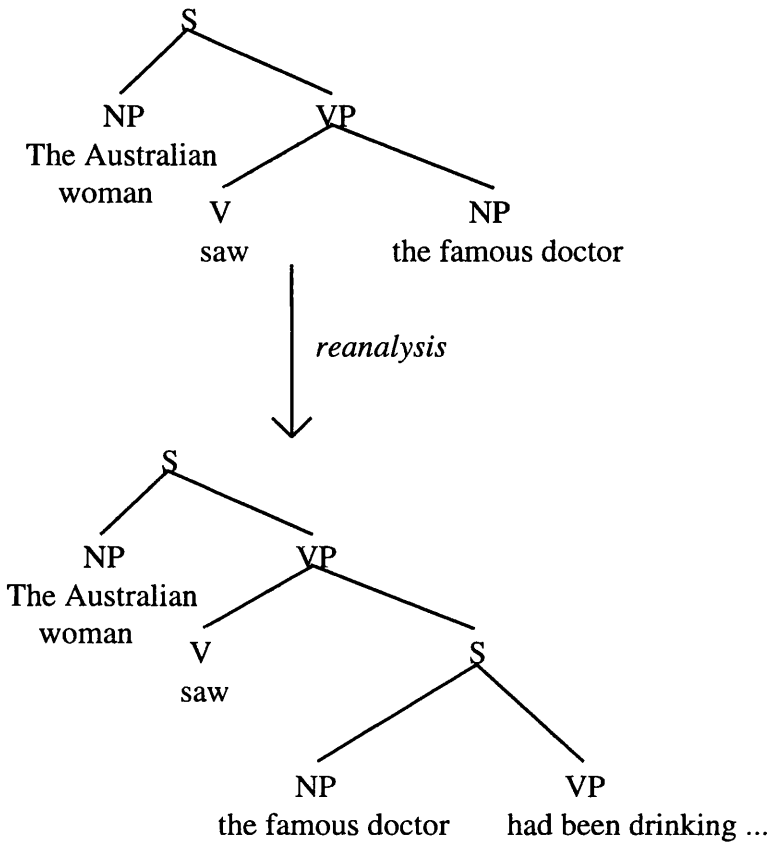

Fig. 2.5. Reanalysis of (37).

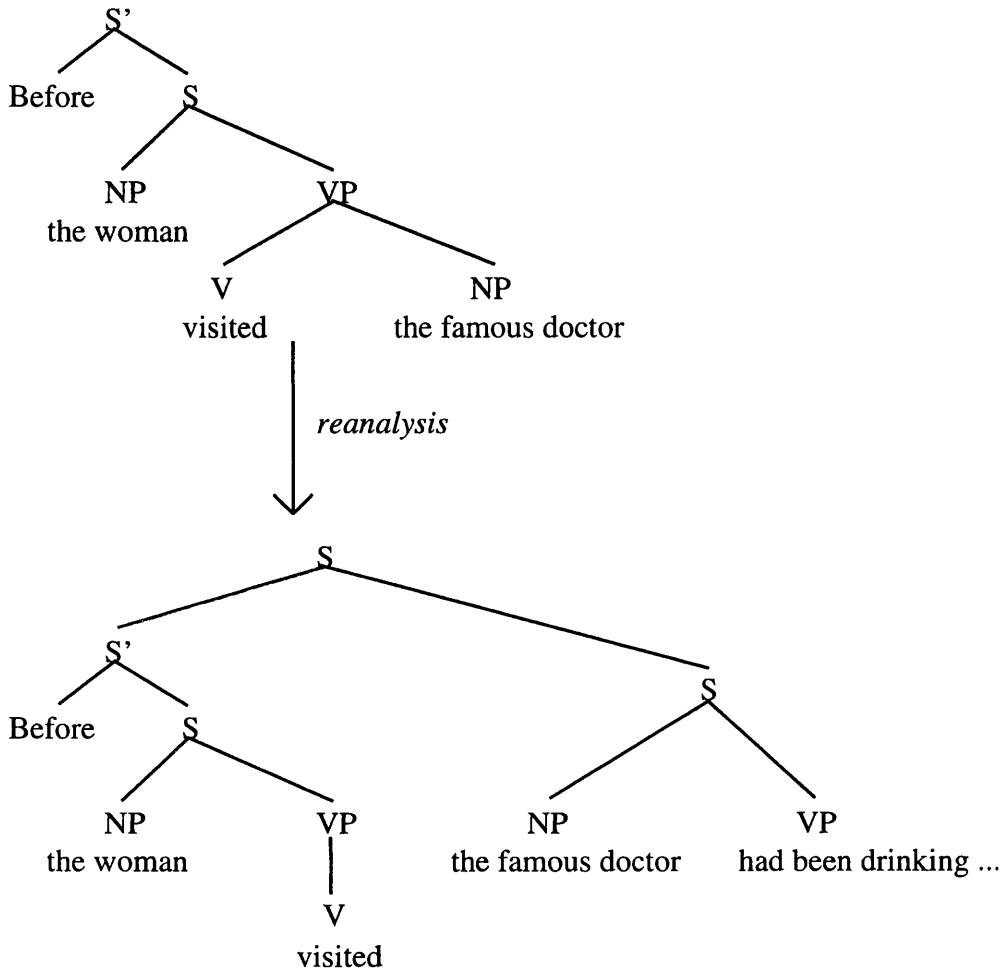

Fig. 2.6. Reanalysis of (38).

Gorrell's model also correctly predicts that reanalysis in (39) is costly.

39. Ian gave the man the report criticised a demotion.

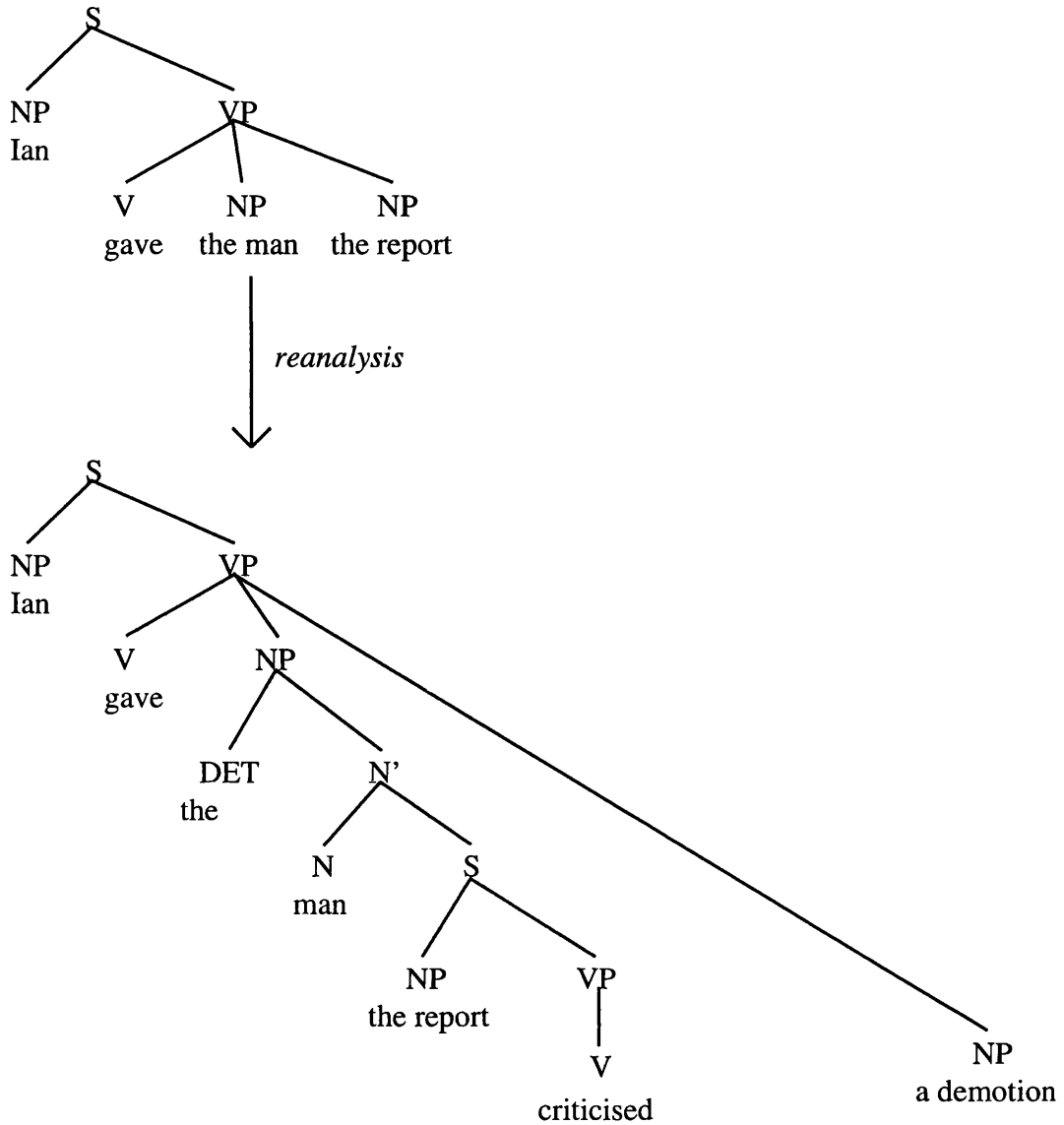

Fig. 2.7. Reanalysis of (39).

In Gorrell's model *the report* is initially analysed as the direct object of *give*. Again, this is due to the principle of simplicity. As shown in Fig. 2.7, NP1 (*the man*) precedes NP2 (*the report*) in such a structure. After reanalysis, however, NP1 dominates NP2. Because a node cannot be preceded by another node when they are involved in a dominance relation, the precedence relation has to be revised. This causes a conscious garden-path effect. Crucially, these predictions are different from what is predicted by Marcus et al.'s (1983) model. According to this model, reanalysis in (39) should be easy, because dominance relations are preserved: the NPs *the man* and *the report* remain dominated by the S and VP node after reanalysis. However, Gorrell claims that this prediction is disconfirmed by intuitions.

Sturt and Crocker (1996) proposed a model of reanalysis that is similar to Gorrell's. As in Gorrell's model, both dominance and precedence relations need to be preserved for a parse to be deterministic. This follows from their hypothesis that only two tree structure additions are possible: simple attachment and tree lowering. This is claimed to make the model somewhat more predictive than the Gorrell model. Sturt and Crocker further argue that initial attachment preferences are the same as attachment preferences during reanalysis. This is shown by (40). Assuming that *the man who believes the countess* is initially analysed as the direct object of *know* and *the countess* as the direct object of *believe*, rather than as the sentence complements of respectively *know* and *believe*, reanalysis should occur when *killed* is encountered. At this point, the structure can be reanalysed in two ways: either *know* can take a sentence complement, or *believe* can. Because low attachment is preferred in initial parsing, Sturt and Crocker predict that this should also be the case in reanalysis, and therefore *killed* is predicted to be part of the sentence complement of *believe* in the preferred interpretation. If this prediction is correct, disambiguation toward low attachment as in (40a) is easier than disambiguation toward high attachment as in (40b) (see Sturt, 1997 for off-line evidence of this).

40a. I know the man who believes the countess killed herself.
40b. I know the man who believes the countess killed himself.

A different type of reanalysis model is proposed by Pritchett (1988). Although Pritchett's model explains why certain initial parsing preferences occur, as described in the previous section, its main importance rests in the way it deals with reanalysis. Pritchett's model does not share the determinism assumption Instead, Pritchett proposed the *theta reanalysis constraint*, which stipulates that reanalysis is costly when a theta marked constituent has to be reanalysed as outside its original theta domain.[14] This explains why reanalysis is easy in (37), but not in (38). *The famous doctor* initially receives a theme theta role from *saw*. After reanalysis, the whole sentence complement clause receives this role. Because *the famous doctor* is dominated by the S node of the complement clause, this reanalysis is predicted to be easy. In contrast, the reanalysis in (38) is difficult. *The famous doctor* initially receives a theta role from *visited*, but has to be removed from this theta domain upon reanalysis, because on the correct analysis it receives a theta role from *drinking*.

In Pritchett (1992), the theta reanalysis constraint was revised. According to both Pritchett (1988) and Pritchett (1992), *her books* in (41) is initially analysed as an

[14] In Pritchett's model, a constituent α is in the theta domain of β if it receives a theta role from β or is dominated by a constituent that receives a theta role from β.

indirect object (*her*) plus direct object (*books*), because this maximally satisfies the theta roles that *gave* assigns (but see Clifton, Kennison, & Albrecht, 1997 for evidence against this). This analysis subsequently has to be revised, because the whole constituent *her books* has to be a direct object. According to the theta reanalysis constraint, this reanalysis should be costly, because *books* has to be removed from the original theme role and is not dominated by it after reanalysis. However, Pritchett claims that this does not correspond with intuitions about the processing of (41).

41. They gave her books to Ron.

Pritchett (1992) proposed the *on-line locality constraint* to account for sentences such as (41). This constraint states that the position of the constituent in the tree structure after reanalysis must be governed[15] or dominated by its position before reanalysis, otherwise reanalysis is impossible. It bears some similarity to the dominance constraint on reanalysis in Marcus et al. (1983) and the precedence and dominance constraints in Gorrell (1995). As can be seen in Fig. 2.8, the original position of the NP node for *books* governs the position after reanalysis, because it governs the NP node of *her books*.

[15] Government is defined as: α governs β iff α m-commands β and every γ dominating β dominates α, γ a maximal projection. M-command is defined as: α m-commands β iff α does not dominate β and every γ that dominates α dominates β, γ a maximal projection. In Fig. 2.8 this means the NP *books* governs all nodes under the VP node.

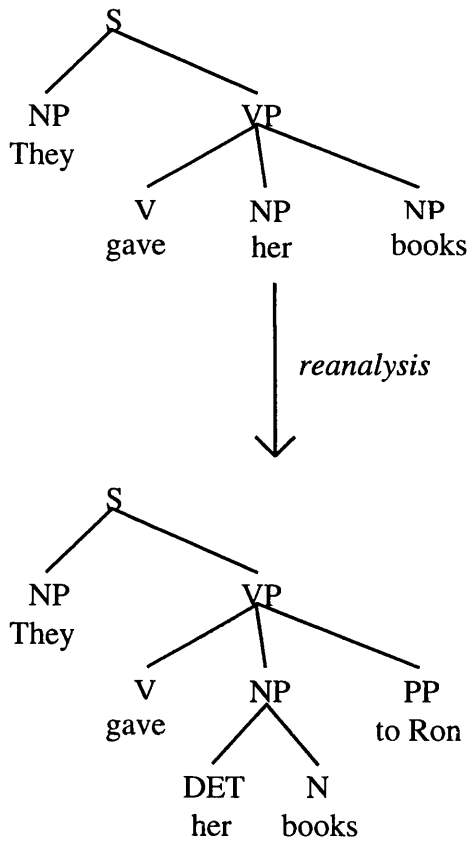

Fig. 2.8. Reanalysis of (41).

Another approach to reanalysis is found in Gibson (1991) and Jurafsky (1996). Both propose a parallel model of syntactic ambiguity resolution. As long as there is no great preference for one syntactic analysis over another one, both are activated, as in other parallel models (e.g., MacDonald et al., 1994; McRae et al., 1998). Because the non-preferred analysis remains activated, little difficulty should occur when the structure is disambiguated towards this analysis. However, both Gibson (1991) and Jurafsky (1996) propose that if one analysis is much preferred over another one, the non-preferred analysis is dropped. This results in a costly reanalysis process when the sentence is subsequently disambiguated towards the dropped analysis, because the analysis is not available any more. The two models differ in the way they account for syntactic preferences. In the Gibson (1991) model, the preferences arise from differences in processing load between alternative structures. The processing load for a particular structure at a certain point in the sentence depends on a number of factors, such as whether there are unassigned theta roles or whether there are more recent attachment sites available. In Jurafsky (1996) syntactic preferences are determined by the relative frequency of rewrite rules to form a particular structure.

Although the discussion above shows that quite a few models of reanalysis have recently been developed (for other models see e.g., Bader, 1998; Fodor & Inoue, 1994;

1998; Inoue and Fodor, 1995; Frazier & Clifton, 1998), experimental data on reanalysis processes are very scarce. Most models of reanalysis rely on intuitive data. This can be dangerous, because judgements of acceptability may be affected by individual differences between researchers and priming between structures (Branigan, Pickering, Liversedge, Stewart, & Urbach, 1995). Furthermore, because such judgements cannot be made on-line, it is unclear whether they reflect initial difficulty, reanalysis difficulty, or both. Finally and most importantly, the acceptability judgements are usually done on single sentences, so that they may reflect particular properties of one sentence. Usually non-syntactic factors such as plausibility and subcategorisation preferences which have been shown to have effects on sentence processing difficulty are not taken into account, and if any direct comparisons between structures are made, they usually differ on multiple dimensions which may affect acceptability. The problem is particularly important, as the distinctions of processing difficulty that the models make become more and more fine grained. For example, the relative distinction between conscious and unconscious difficulty may not always be very obvious and there may be gradations in between.

However, there have been some experimental studies of reanalysis. Sturt et al. (1999) contrasted reanalysis difficulty for NP/S and NP/Z ambiguities such as (37) and (38) in a self-paced reading experiment, controlling for plausibility and the frequency of the alternative syntactic structures. As described before, most models of reanalysis predict that reanalysis in NP/Z ambiguities is more difficult than in NP/S ambiguities (e.g., Gorrell, 1995, Inoue & Fodor, 1995; Pritchett, 1988; 1992; Sturt & Crocker, 1996. They found that both (37) and (38) produced difficulty compared to unambiguous base-line conditions, but crucially, that this difference was much more pronounced for (38) than (37). This suggests that reanalysis occurred for both types of constructions, but that reanalysis was much more difficult for (38), confirming the predictions of most models of reanalysis.

A number of other experimental studies have investigated whether the length of an ambiguous region affects difficulty of reanalysis. Frazier and Rayner (1982) investigated NP/S and NP/Z ambiguities similar to (37) and (38) in an eyetracking experiment and varied the length of the ambiguous NP. They observed that reading times at the point of disambiguation were longer for sentences with long ambiguous regions than for sentences with short ones. Frazier and Rayner hypothesised that reanalysis may be more costly after a long ambiguous region because semantic interpretation lags behind syntactic analysis. When a region is sufficiently long, readers may have committed to a semantic interpretation of the ambiguous phrase, so that reanalysis is difficult. In contrast, when a region is short, semantic analysis may not have occurred yet, making reanalysis relatively easy.

Similar findings for NP/Z structures were reported in Warner and Glass (1987) and Ferreira and Henderson (1991), using a grammaticality judgement task. Ferreira and Henderson tested two alternative hypotheses to Frazier and Rayner's (1982) explanation of the length effect. Following Warner and Glass, they hypothesised that the length effect may be due to a greater complexity of long regions. However, this was refuted by an experiment in which they varied the syntactic complexity of the ambiguous NP in terms of number of nodes, while keeping its length constant: no difference was observed between the conditions which differed in syntactic complexity. Finally, Ferreira and Henderson tested whether the length effect may be due to the position of the head noun of the ambiguous NP. They compared sentences such as (42a), in which the head noun (*dog*) is at the end of the NP, with sentences such as (42b), in which it is at the beginning.

42a. While the boy scratched the big and hairy dog yawned loudly.
42b. While the boy scratched the dog that is hairy yawned loudly.

It was found that when the head noun occurred at the end of the ambiguous region, reanalysis effects were smaller than when it occurred at the beginning. On the basis of these results, Ferreira and Henderson argued that the ambiguous NP is analysed as a direct object and assigned a theta role as soon as the head noun is reached. When disambiguation toward the alternative Z analysis immediately follows the head noun, reanalysis is relatively easy, because the thematic structure of the alternative analysis has not decayed very much. In contrast, when the head noun is followed by other words, this structure has decayed, and reanalysis is difficult.

There are two problems with this explanation. Firstly, Frazier and Rayner (1982) observed a length effect even though in most of their materials the length difference was due to the words *preceding* the head noun. Secondly, the head noun position effect seems to be induced by the grammaticality judgement task Ferreira and Henderson used, since two subsequent self-paced reading studies have failed to replicate the effect of head noun position (Ferreira & Henderson, 1993; Sturt et al., 1999; but cf. Konieczny, Hemforth & Scheepers, in press)

Pickering and Traxler (1998; for similar studies see also Clifton, 1993; Holmes, Stowe, & Cupples; 1989; Schmauder & Egan, 1998) investigated the effects of plausibility on reanalysis. Using eyetracking, they tested NP/Z ambiguities such as (43) and NP/S ambiguities such as (44). In the (a) versions, the NP analysis is plausible until the sentence is syntactically disambiguated by the verb *amused* or *harmed.* In contrast, the NP analysis is implausible in the (b) versions.

43a. As the woman edited the magazine about fishing amused all the reporters. (plausible NP analysis)

43b. As the woman sailed the magazine about fishing amused all the reporters. (implausible NP analysis)

44a. The criminal confessed his sins which upset kids harmed too many people. (plausible NP analysis)

44b. The criminal confessed his gang which upset kids harmed too many people. (implausible NP analysis)

Pickering and Traxler found that the (b) versions produced more difficulty at or just after the post-verb noun (*magazine/sins/gang*) than the (a) versions. The earliest effects were obtained in first-pass regressions from the noun (for [43]) and post-noun region (*about fishing/which upset kids*, for both [43] and [44]). Pickering and Traxler argued that this suggested that readers initially attempted to form a direct object analysis, and were garden-pathed in the (b) versions because this analysis was implausible. The reading time pattern was reversed following the disambiguation: more first-pass regressions from the post-verb region (the words following the disambiguating verb) occurred in (43a) than (43b), and more from the verb region in (44a) than (44b). This result is consistent with two explanations. When the direct object reading was implausible, readers may have reanalysed this analysis before the disambiguation, so that no reanalysis was necessary at the disambiguating verb. Alternatively, reanalysis at the disambiguation may be easier when plausibility supports the reanalysis than when it does not. In either case, the study stresses the importance of semantic plausibility for reanalysis, as it shows that plausibility affects the ease of reanalysis.

2.9. Modularity and the garden-path model: plausibility information

Most of the models discussed so far share the assumption that some sources of information potentially useful to the processor are initially ignored.[16] This assumption has been most forcefully argued for by proponents of the garden-path theory (e.g., Ferreira & Clifton, 1986; Frazier, 1987a; Rayner et al., 1983). It follows from their hypothesis that the language processor is modular (Fodor, 1983; Forster, 1979). Fodor (1983) lists a number of characteristics of modular systems. One of them is

[16] Exceptions are construal and Gorrell's (1995) model. In construal, multiple sources of information can be employed initially in non-primary attachments (but not in primary attachments). Gorrell claims that in his model, multiple sources of information can be employed as well, although he mainly focuses on structural principles.

informational encapsulation, that is, some information that is potentially useful for a system is ignored, because communication between distinct systems is limited. The garden-path model shares this assumption. It assumes that structural parsing principles are informationally encapsulated: structural parsing principles such as minimal attachment and late closure apply even in the face of other potentially useful but non-structural information such as plausibility and lexical and discourse information. The garden-path model stipulates that non-structural information is consulted only after an initial structural analysis has been made. If this information is inconsistent with it, the initial analysis is revised, which causes processing disruption.

Rayner et al. (1983) tested the modularity hypothesis by investigating whether the parser employs semantic plausibility information to process syntactically ambiguous structures. They investigated sentences such as (45) in an eyetracking experiment. These sentences contain a VP/NP attachment ambiguity (cf. [14]). The PP following *cop* can either be attached to the VP, in which case it modifies the verb (as in "see with the binoculars"), or to the direct object NP (as in "the cop with the revolver"). Minimal attachment predicts that modifying the verb is preferred (VP attachment, see Fig. 2.3). This analysis is plausible in (45a), but not in (45b).

45a. The spy saw the cop with binoculars but the cop didn't see him.
(minimal attachment plausible)

45b. The spy saw the cop with the revolver but the cop didn't see him.
(minimal attachment implausible)

The experiment showed that first-pass reading times for the region *binoculars/revolver but the cop didn't see him* were longer for sentences such as (45a) than for sentences such as (45b). Rayner et al. took this as evidence that the processor initially adopts the structurally simplest reading and ignores plausibility information. They argued that there are two independent processors that work in parallel in sentence processing. The first processor is a structural processor. It follows the principles of minimal attachment and late closure and ignores non-structural information such as plausibility. The second processor is the *thematic processor*, which examines the alternative structures of a phrase and selects the most plausible one. Although the two processors work in parallel, the initial analysis of an ambiguous structure is determined by the structural processor. After one analysis has been selected on the basis of structural information, this analysis is compared with the output of the thematic processor. If this output structure is the same as the structurally simplest analysis, processing of the sentence is easy. However, if it is different, the processor will have to reanalyse the initially adopted structure, which produces processing difficulty. This

explains the difference in complexity between (45a) and (45b). In (45a) the minimal attachment reading is plausible, so that no reanalysis occurs. But in (45b), it is implausible, and therefore reanalysis is predicted to occur.

In another eyetracking experiment, Rayner et al. (1983) investigated the use of plausibility in ambiguities such as (46).

46a. The florist sent the flowers was very pleased.
(reduced relative implausible)

46b. The performer sent the flowers was very pleased.
(reduced relative plausible)

46c. The performer who was sent the flowers was very pleased.
(unambiguous relative clause)

46d. The performer sent the flowers and was very pleased with herself.
(unambiguous main clause)

Sentences (46a) and (46b) contain the same temporary ambiguity as (1): until *was* the sentence can either be analysed as a simple main clause or as a sentence containing a reduced relative clause. Both structures are disambiguated as reduced relative clauses. Sentence (46c) is an unambiguous relative clause control, and (46d) is a (temporarily ambiguous) main clause. According to the garden-path model, the reduced relative clause sentences should be difficult to read compared to the controls, because minimal attachment predicts that in the temporarily ambiguous sentences, *sent the flowers* should initially be analysed as a main clause rather than a reduced relative (e.g., Frazier, 1978). In (46a) the ultimately incorrect minimal attachment analysis is plausible. However, in (46b) it is not: performers do not normally send flowers. Here plausibility information supports the ultimately correct reduced relative analysis.

Rayner et al. hypothesised that if the sentence processor uses plausibility information for its initial analysis, (46b) should be easier to read than (46a). But if the initial analysis is only based on structural information such as the principle of minimal attachment and plausibility information is initially ignored, readers should experience equal difficulty in both sentences in early measures of processing difficulty. Both (46a) and (46b) should cause processing difficulty as compared to (46c) and (46d).

Indeed, this is what Rayner et al. found. First-pass reading times per character for the disambiguating region (*was very* or *and was*) were longer for both reduced relative conditions (46a) and (46b) than for the unambiguous relative clause (46c) and the simple active sentence (46d). There were also more regressions from the disambiguating region and the words following it in the relative clause conditions than

in the unambiguous conditions.[17] Hence, Rayner et al. concluded that readers did not immediately adopt the most plausible reading of a sentence, but based their initial analysis on structural principles independently of plausibility information. This provided support for the informational encapsulation of the syntactic processor.

The plausibility manipulation in (46) is relatively weak and may in fact have been too weak to facilitate the reduced relative analysis. For example, it is quite possible for a florist to be sent flowers, or for a performer to send flowers. Ferreira and Clifton (1986) used a more rigorous manipulation of plausibility in a very similar experiment. They manipulated the animacy of the first NP using sentences such as (47). In (47a) the animate NP *defendant* is a plausible subject of *examined*, whereas the inanimate *evidence* in (47b) is not. Clearly, this manipulation is stronger than the one Rayner et al. used: the inanimate *the evidence* in (47b) is impossible as the subject of *examined*, whereas *the performer* is merely unlikely as the subject of *sent*. Sentences (47c) and (47d) were control sentences containing an unreduced relative clause.

47a. The defendant examined by the lawyer turned out to be unreliable.
(main clause plausible)

47b. The evidence examined by the lawyer turned out to be unreliable.
(main clause implausible)

47c. The defendant that was examined by the lawyer turned out to be unreliable.
(unambiguous relative clause control 1)

47d. The evidence that was examined by the lawyer turned out to be unreliable.
(unambiguous relative clause control 2)

Although the animacy manipulation was more rigorous than the plausibility manipulation used by Rayner et al., the results were very similar. The eyetracking experiment showed that first-pass reading times in the disambiguating region (*by the lawyer*) were longer for the reduced than the unreduced relative clauses. The reduced relative clauses (47a) and (47b) did not differ. Importantly, first-pass reading times on the verb (*examined*) were longer when the first NP was inanimate than when it was animate. Therefore, Ferreira and Clifton concluded that animacy information was available to the reader at the point of ambiguity, but that it did not guide initial syntactic processing. They took this as strong support for a modular approach to syntactic processing.

[17] Note, however, that Rayner et al. did not provide statistics for the comparisons between conditions.

2.10. Interactive and constraint-based models: use of plausibility information

The assumption of modularity that is part of the garden-path model contrasts strongly with proposals of *interactive* and later *constraint-based theories.* In these theories, which were developed as an alternative to modular models, all sources of information that are potentially useful for the processor can be employed immediately. In such models, there are no restrictions on the communication between the various sources of information that can be used and no separable stages.

The first interactive models of language processing were proposed by Marslen-Wilson and Tyler (1975; 1980; Marslen-Wilson, 1975), who showed in a number of on-line word-recognition experiments that various sources of information affected word processing extremely rapidly. In an early study testing the predictions of interactive models for syntactic processing, Tyler and Marslen-Wilson (1977) investigated the use of plausibility information in sentences such as (48). In these sentences, *flying* is ambiguous between an adjective reading (as in *As they glide gracefully over the city, flying kites are beautiful to watch*) or a verb reading (as in *If you know how to handle sudden gusts of wind, flying kites is very good fun*). When the preceding clause is as in (48a), the adjective reading is most plausible, whereas the verb reading is most plausible when it is as in (48b). This plausibility manipulation was confirmed by completions of the same fragments. After each sentence, participants had to name either a verb consistent with the adjective reading (*are*) or verb reading (*is*).

48a. As they glide gracefully over the city, flying kites ... (adjective bias)
48b. If you know how to handle sudden gusts of wind, flying kites ... (verb bias)

No overall differences were obtained between naming *is* or *are*, indicating that there was no preference for either reading. But crucially, naming times for *is* in (48a) were longer than for *are*, whereas the opposite pattern was observed for (48b). Thus, naming times were longer when the verb was consistent with the plausibility bias of the sentence than when they were not. This showed that plausibility affected syntactic processing extremely rapidly, at the point where the disambiguation occurred. Tyler and Marslen-Wilson argued that this result is most consistent with interactive theories, and does not support theories which claim that there is a considerable delay in the use of plausibility information. Although some researchers have pointed out a number of confounds in Tyler and Marslen-Wilson's materials (Cowart & Cairns, 1987; Townsend & Bever, 1982), this conclusion was supported by Farrar and Kawamoto (1993), who controlled for these factors in two self-paced reading experiments using similar structures.

As mentioned in the preceding paragraph, findings of early use of plausibility information are not necessarily incompatible with modular models. As Rayner et al. (1983) argued, reanalysis may occur when an initially adopted analysis is inconsistent with plausibility information. This may occur very rapidly, as suggested by Pickering and Traxler's (1998) experiments. Such a reanalysis process toward the more plausible reading may have occurred in (48) before the disambiguation (*is* or *are*) was reached. Because neither Tyler and Marslen-Wilson nor Farrar and Kawamoto investigated processing times before the disambiguating verb, this explanation cannot be ruled out.

Taraban and McClelland (1988) investigated the claims of the garden-path model more directly. They tested the set of VP/NP attachment materials that were used by Rayner et al. (1983), and compared them with another set of materials such as (49). The latter set was specifically constructed to yield an NP attachment expectancy, whereas the Rayner et al. set was found to have a VP attachment expectancy. Participants' expectations were assessed by asking them whether the noun (e.g., *article/government*) was an expected continuation of preceding sentence fragment.

49a. The reporter exposed corruption in the article. (VP attachment)
49b. The reporter exposed corruption in the government. (NP attachment)

Taraban and McClelland tested both sets of materials in two self-paced reading experiments. They observed that reading times for the Rayner et al. materials were longer for the NP than the VP conditions, replicating Rayner et al.'s eyetracking study. However, for the other set of materials, NP attachment was faster than VP attachment.

The NP attachment preference for some materials is clearly inconsistent with the principle of minimal attachment and the garden-path model, which predict a VP attachment preference. Instead, Taraban and McClelland (1988; 1990; McClelland, St John, & Taraban, 1989) favoured interactive models, and argued that attachment preferences are determined by semantic expectancies. In order to investigate what these expectancies are based on, they conducted another experiment. They presented their participants with the same type of sentences as in the other experiments, but had four conditions, as shown in (50).

50a. The janitor cleaned the storage area with the broom. (expected noun)
50b. The janitor cleaned the storage area with the solvent. (less expected noun)
50c. The janitor cleaned the storage area with the manager. (less expected role)
50d. The janitor cleaned the storage area with the odor. (less expected attachment)

The noun *broom* in (50a) is fully consistent with participants' expectations. The noun *solvent* in (50b) was rated as less expected and less plausible than *broom*, but has the same thematic role (an instrument for cleaning) and is also VP attached. *Manager* in (50c) has a less expected thematic role (accompaniment role), but is similar in plausibility and has the same attachment site. Finally, *odor* in (50d) is neither the expected role nor the expected attachment site, although it is has a similar plausibility to (50b) and (50c).

The self-paced reading experiment revealed that reading times were faster in (50a) than (50b) on the word following the noun, which shows that expectancy for a particular meaning (independent of thematic roles) and/or plausibility affect processing. Thematic role expectations produced the most robust effect: (50c) was more difficult to read than (50b) on the three words following the critical noun. Finally, (50c) and (50d) did not differ from each other[18], which suggests that the expectation for a particular attachment site did not impact on processing. Taken together, these results are consistent with Taraban and McClelland's proposal that readers create expectancies on the basis of the semantic content of the sentence and that processing difficulty occurs when the continuation of the sentence is inconsistent with it. Although semantic expectations and plausibility had some effect, thematic role expectations appeared to be strongest. Because such expectations are semantic, these results are inconsistent with modular, structural sentence processing models (see Carlson & Tanenhaus, 1988; Tanenhaus & Carlson, 1989, for an alternative model of thematic role expectations).

More recently, Trueswell, Tanenhaus, and Garnsey (1994) claimed to find further support for interactive models. They investigated reduced relative/main clause ambiguities that were very similar to those used by Ferreira and Clifton (1986) (see [47]). They noted that there were a number of problems with this study, which may account for Ferreira and Clifton's findings. Firstly, they observed that in about half of the materials that Ferreira and Clifton used, the animacy manipulation in sentences such as (47b) was very weak: the inanimate noun could be a plausible subject of the verb. Secondly, they criticised the display mode that Ferreira and Clifton used. Because the line break occurred after the same word in the ambiguous and unambiguous conditions and because the unambiguous conditions contained two extra words (*that was*), the first line was longer in the unambiguous than the ambiguous conditions. Trueswell et al. claimed that this difference might have resulted in generally longer reading times for the ambiguous conditions. Finally, they questioned the appropriateness of unreduced relative clauses as control conditions. Given that the unambiguous conditions contained extra words, eye-movement landing sites may be different from the ambiguous

[18] Condition (50d) was slightly easier to read three words after the critical word, but this effect was not significant by items.

conditions. Furthermore, the structure for unreduced relatives is different from reduced relatives, which may affect the reading process.

Because of the potential confounds in Ferreira and Clifton (1986), Trueswell et al. repeated the experiment, using a stronger plausibility manipulation and a one-line display. Furthermore, they compared reduced relative clauses employing unambiguous verbs such as *drawn* in (51a) with unreduced versions such as in (51b).

51a. The poster drawn by the illustrator was used.
51b. The poster that was drawn by the illustrator was used.

First-pass reading times for (51a) and (51b) did not differ, suggesting that unreduced relative clauses that were used by Ferreira and Clifton were an appropriate control condition for reduced relatives. However, the pattern of results that Trueswell et al. obtained for the experimental sentences such as (47) was different from that obtained by Ferreira and Clifton. First-pass reading times for the conditions with ambiguous verbs showed in interaction between animacy and ambiguity (reduced vs. unreduced relative clauses) in the disambiguating region (*by the lawyer*). Reduced relative clauses beginning with an animate noun took longer to read than their unreduced controls, but no difference was observed between relative clauses beginning with an inanimate noun and their controls, suggesting that the animacy manipulation prevented readers from garden-pathing. This result was replicated in a second experiment with new but similar materials.

Trueswell et al. also conducted a correlational analysis between first-pass reading times and the plausibility of the inanimate noun as the patient/theme of the verb in the relative clause. Plausibility was measured in an off-line rating task. On the verb (*examined*) and the disambiguation (*by the lawyer*), they obtained a negative correlation indicating that reading times were faster the more plausible the noun was as a patient/theme. Importantly, there was no evidence for early reanalysis before the point of disambiguation which might have explained the results. Although some processing difficulty obtained at the verb for materials with a weak animacy manipulation, no such effect was observed for those materials where the animacy manipulation was strongest.

Trueswell et al. claimed that their results were consistent with interactive models of syntactic ambiguity resolution. They proposed that syntactic structures are activated in parallel, with the degree of activation depending on the support the structures receive from the various sources of information or constraints. They claimed that various non-syntactic constraints, such as lexical frequency information (see section 2.11), discourse information (see section 2.12) and plausibility are employed to activate syntactic analyses. In their study, the plausibility of a noun-phrase as an agent or a patient/theme

of a verb influenced the activation of the reduced relative and main clause analysis. The stronger plausibility activated the reduced relative clause, the easier such sentences were. This explains the results of the correlational analysis, which showed that garden-path effects range on a continuum.

In McRae et al. (1998) and Spivey and Tanenhaus (1998), this kind of constraint-based model was worked out in more detail. This model, which they call the *competition-integration model*, assumes that various constraints activate multiple syntactic structures. An important assumption of the model is that this results in a process of competition between alternative structures of a syntactic ambiguity. When the various constraints activate one structure much more strongly than the others, little competition occurs and little processing difficulty is predicted. But as two or more structures become more equally activated, competition increases, which should result in more processing difficulty.

The McRae et al. and Spivey and Tanenhaus models differ only in small details, most of which do not concern us here. However, the McRae et al. model is somewhat more predictive, in that it assumes that constraints that apply to sentence comprehension also apply to sentence production. Although this is also implicit in the Spivey and Tanenhaus model, it had not been worked out in that paper. McRae et al. assessed sentence production preferences using completion studies, in which participants had to complete sentence fragments. The model predicts that if a sentence fragment is predominantly completed as one particular structure, this structure should also be preferred in comprehension tasks. This structure should be easy to read, because there is little competition with alternative structures. But the weaker the preference for a single structure in the completions, the more competition should occur in comprehension, and the more difficult the sentence is predicted to be.

McRae et al. demonstrated their model with sentences such as (52). The noun *crook* in (52a) in combination with the verb *arrested* makes the reduced relative plausible, whereas the main clause alternative is not. Conversely, the noun-verb combination makes the main clause analysis plausible in (52b), while the reduced relative is not. McRae et al. first conducted a completion study. Participants had to complete sentences after *arrested*, *arrested by*, *arrested by the*, or *arrested by the detective*. Except for the final completion, where no differences were observed, more reduced relative clause completions occurred for (52a) than (52b), which showed that in offline production, the plausibility manipulation had an effect. At the first completion (after *arrested*), the main clause analysis was preferred for both (52a) and (52b), but this preference disappeared for the completions after *by* and the words following it.

52a. The crook arrested by the detective was guilty of taking bribes.
(main clause implausible)

52b. The cop arrested by the detective was guilty of taking bribes.
(main clause plausible)

They first modelled these completion data. At the verb (*arrested*) three constraints were included in the model: a general bias for the main clause analysis, the thematic fit between the first NP and the verb's thematic roles, and the frequency of the verb appearing as a past participle versus a past tense. At the following words, further constraints were included, namely the support for the reduced relative analysis received from *by* and *the*, and the thematic fit of the noun in the *by*-phrase, which determines whether this noun should be taken as an agent or not. As shown in Fig. 2.9, the model assumes that there are two interpretation nodes, one for the main clause analysis, and one for the reduced relative analysis. These interpretation nodes can be considered as reflecting the activation that the main clause and reduced relative analysis receive. The interpretation nodes are fed by the constraint-nodes. Each constraint (thematic fit, main clause bias and past participle frequency) has two *constraint nodes*. One of these nodes feeds into the main clause interpretation node, and the other into the reduced relative clause interpretation node. The activation they feed depends on the degree to which each constraint favours one analysis over the other and the weight each constraint has relative to other constraints. The weights were determined by calculating the best fit with the completion data. After the constraint nodes have fed the interpretation nodes, the interpretation nodes feed the same activation back to the constraint nodes. This backward activation is added to the already existing activation of the constraint nodes and subsequently normalised so that the total activation of a pair of constraint nodes is 1. The model recursively carries out this cycle of forward activation, backward activation, and normalising. In this way, one of the interpretation nodes gradually receives more activation, whereas the other one gradually receives less activation. The process continues until one of the interpretation nodes reaches a criterion level of activation given by the formula:

$$\text{criterion level} = 1 - \Delta\text{crit} * \text{cycle}$$

where Δcrit is a constant between 0 and 1. Hence, according to this formula, the degree of activation that needs to be reached for stopping the recursion becomes lower the more cycles the processor has gone through. McRae et al. claim that the more processing cycles it takes to reach the criterion level, the more severe the competition between the analyses is, and the more severe the processing difficulty.

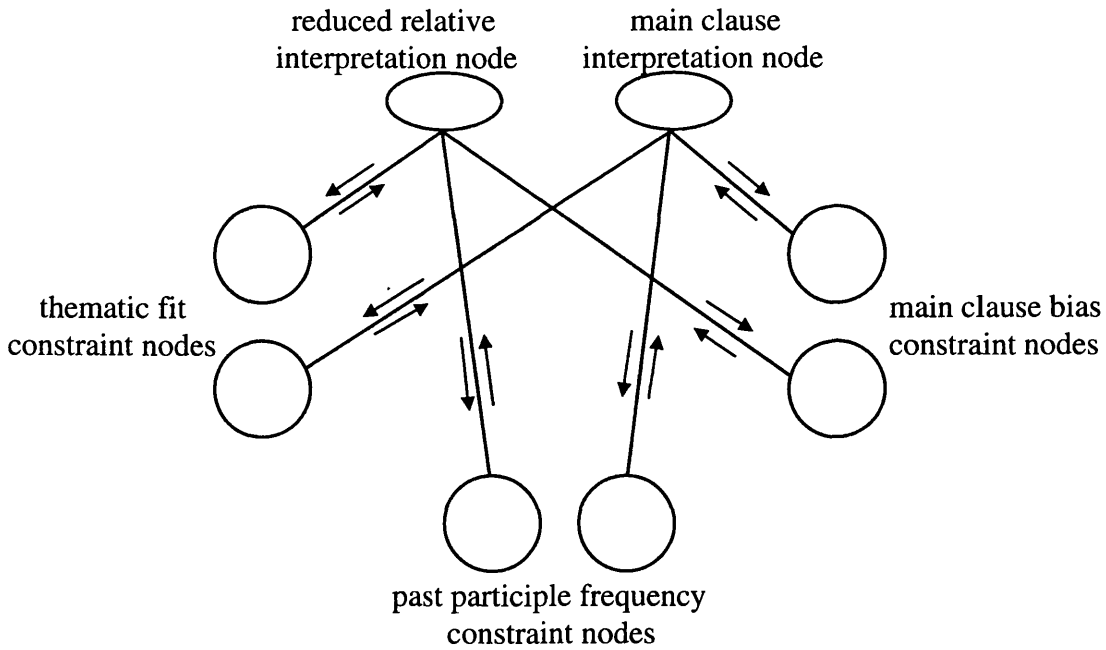

Fig. 2.9. Interpretation and constraint nodes in the competition-integration model.

The competition-integration model works incrementally. At each word it goes through the process of competition until it reaches the criterion level. When this has been reached, the final activation level carries over to the processing of the next word. It is assumed that this level contributes half of the total activation for the next word. This results in reanalysis-like effects. For example, if the activation of the main clause is .90 and the activation for the reduced relative is .10 at word n, and new constraints entering at word n+1 result in .10 and .90 activation respectively, both analyses are equally supported. As a result, it will take many processing cycles to reach the criterion level, and therefore, severe competition occurs.

McRae et al. accurately modelled their completion data this way, although this is not particularly surprising, given that the weights of the constraints were based on exactly the same data. Using the same constraints and weights, they predicted reading times for sentences such as (52a) and (52b) with their model. According to the model, competition in the region consisting of the verb and *by* (*arrested by*) should be relatively high in (52a), because the structural bias for the main clause analysis conflicts with plausibility of the noun as the subject of the sentence and the reduced relative support from *by*[19]. Relatively little competition should occur in (52b), because the model predicts that the main clause preference in combination with its plausibility overrules the information provided by *by*. The pattern is reversed at the next word. Because in (52b) the main clause activation at *arrested by* is higher than the reduced relative

[19]The preference for the past participle over the past tense interpretation of the verb was also predicted to have an effect, although it was not particularly strong for the McRae et al. materials. For ease of exposition, this constraint is ignored here.

activation, but the constraint provided by *the* favours the reduced relative, competition should occur as soon as *the* is reached. Little competition should occur at *the* in (52a), because the reduced relative activation was already fairly high at *arrested by*. Indeed, this pattern was confirmed by a self-paced reading experiment with the same materials as the completion study, which McRae et al. interpreted as support for their model.

2.11. Frequency information

Another battleground for modular and interactive models has been the use of frequency information in sentence processing. One source of frequency information, and the one that has received most attention in sentence processing, is lexical frequency information. Lexical information provides information about the number of arguments of a word. For example, the transitive verb *visited* subcategorises for two arguments, a subject and a direct object, as shown in (53). In contrast, the subcategorisation frame of the intransitive verb *arrived* has only one argument role, as shown by the ungrammaticality of (54b).

53a. The man visited.
53b. The man visited his parents.

54a. The man arrived.
54b. *The man arrived his parents.

Lexical preferences are more subtle. For example, although *sang* is grammatical when it occurs with two arguments, as shown in (55b), it occurs more frequently with only one (55a). Similarly, *visited* occurs more often with two arguments than with one.

55a. The man sang.
55b. The man sang the hymn.

It is not entirely clear whether this kind of lexical information should be considered as part of syntax. It clearly needs to be represented at the word level, but in the cases such as (54b), it has syntactic effects, as shown by the ungrammaticality of this sentence. Thus, strict subcategorisation information could in principle be incorporated in syntax-first models of sentence processing. For example, Frazier (1987a, p. 562) phrases the principle of late closure as: "if grammatically permissible, attach new items into the clause or phrase currently being processed." Given that subcategorisation

information can determine whether a sentence is grammatical or not, late closure should not apply if it violates subcategorisation information. However, in a number of papers, proponents of the garden-path model (Frazier, 1987a; 1989; 1995; Frazier & Clifton, 1989; Clifton and Frazier, 1989) have claimed only structural information is employed during the initial stage of processing, and that other sources of syntactic information such as subcategorisation information are initially ignored.

This view contrasts with approaches which do claim that lexical information is used immediately. For example, Fodor (1978) proposed that readers use lexical expectations in long-distance dependencies such as (56a).

56a. Which student did the teacher walk to the cafeteria with?
56b. Which student did the teacher walk to the cafeteria?

In this sentence, *which student* has to be associated with the preposition *with*, but it does not occur next to it. According to transformational grammar, it has been removed from this position, leaving a gap behind (see section 2.4). However, before *with*, another possible gap position exists after *walk*, because *which* student can be its direct object, as illustrated by (56b). This results in a temporary ambiguity.

Fodor argued that every verb ranks its subcategorisation frames. Whether a gap is posited in temporarily ambiguous long distance dependencies such as (56) depends on the ranking of the direct object option relative to the gap option. The ranking is determined by the frequency with which a verb occurs as a transitive or intransitive verb. Verbs that are usually used transitively rank the direct object option above a gap, whereas verbs that are usually intransitive rank the gap highest. Because *walk* is more frequently used as an intransitive than a transitive verb, this predicts that no gap is posited after it, and that (56a) should be easier to process than (56b). This prediction differs from the predictions of the garden-path model, which assumes that the active filler strategy (section 2.4.) applies to long-distance dependencies. This strategy predicts that a gap will be posited as soon as possible, regardless of lexical frequency information.

Ford, Bresnan, and Kaplan (1982) extended the ranking of subcategorisation frames to structures that do not involve long-distance dependencies. They claimed that there are two parsing principles that deal with lexical preferences: the *principle of lexical preference* and the *principle of syntactic preference*.[20] The principle of syntactic preference stipulates that syntactic categories are ranked according to their strength, which is presumably determined by the frequency of occurrence of the categories.

[20] Ford et al. propose two other principles, *final arguments* and *invoked attachment*. As they do not directly deal with lexical preferences, they are not explained here.

Assuming that an NP category is ranked higher than an S category, this accounts for the preference to analyse *that* in (57) as a determiner within an NP (as in [58]) rather than the complementiser of an S (as in [59]).

57. That silly old-fashioned ...
58. That silly old-fashioned joke is told too often.
59. That silly old-fashioned jokes are told too often is well-known.

According to the principle of lexical preference, phrase structure rules that are coherent with the strongest (presumably the most frequent) lexical form of the predicate are preferred. This can be illustrated with (60).

60. The woman positioned the dress on that rack.

The verb *positioned* can occur either with or without a PP complement. Because the PP complement option is assumed to be the strongest lexical form of this verb, (60) is preferentially be interpreted as *the dress was positioned on that rack* rather than as *the dress on that rack was positioned.* However, if the PP complement option is ranked lower than the analysis without PP complement, the latter analysis is preferred. This is the case for a verb like *want* in (61).

61. The woman wanted the dress on that rack.

Ford et al.'s model is a serial model. The processor adopts one analysis, and if this analysis becomes untenable, reanalysis to a lower ranked alternative occurs. This assumption is fundamentally different from current constraint-based models, which assume a parallel processor in which multiple analyses are activated according to their frequency of occurrence (and other constraints). Examples of such models are Trueswell et al. (1994), McRae et al. (1998) and Spivey and Tanenhaus (1998), which were described in section 2.10. Other parallel models that assign an important role to frequency information are Tabor, Juliano, and Tanenhaus (1997), which focuses mainly on syntactic category ambiguities such as in (57), and MacDonald et al. (1994).

MacDonald et al. (1994) explicitly argue that syntactic ambiguity resolution is a form of lexical ambiguity resolution. They claim that whenever a word is read, its meaning and lexical category (e.g., whether it is a noun or verb) are activated along with its possible subcategorisation frames which are linked to tree-structures. The activation of a word's meaning, category and subcategorisation frames depends on both the frequency of the alternatives and contextual support. Studies on lexical ambiguity

resolution have shown that processing of a word is slow when contextual information supports the subordinate meaning of an ambiguous word (e.g., Duffy, Rayner, & Morris, 1988; Rayner and Duffy, 1986). Hence, it follows that syntactic processing is slow when previous context supports a low-frequent subcategorisation frame of a syntactic ambiguity. In addition, because lexical ambiguity studies have also shown that ambiguous words are difficult to read when two meanings are about equally activated (e.g., Rayner & Duffy, 1986), McDonald's model also predicts that syntactic ambiguity resolution is difficult when multiple subcategorisation frames are activated to an equal extent. Hence, the model is very similar in spirit to other current constraint-based theories, such as McRae et al. (1998) and Spivey and Tanenhaus (1998).

A great deal of studies has investigated the use of lexical subcategorisation information. A few studies support the garden-path claim that such information is delayed. Ferreira and Henderson (1990) investigated NP/S ambiguities such as (62), using both eyetracking and self-paced reading methodology. In these sentences, the noun following the verb is syntactically ambiguous between a direct object NP (as in *He forgot Pam*) or the subject of a sentence complement (as in the ultimately correct reading in [62]). In (62a) and (62b), the verb is biased toward the sentence complement reading, whereas it is biased toward the direct object reading in (62c) and (62d). Sentences (62a) and (62c) are temporarily ambiguous until *needed*, whereas the two other versions are disambiguated toward the sentence complement reading by *that*.

62a. He wished Pam needed a ride with him.
(sentence complement bias, temporarily ambiguous)

62b. He wished that Pam needed a ride with him.
(sentence complement bias, unambiguous)

62c. He forgot Pam needed a ride with him.
(direct object bias, temporarily ambiguous)

62d. He forgot that Pam needed a ride with him.
(direct object bias, unambiguous)

Ferreira and Henderson's eyetracking experiment showed that the first fixation duration in the disambiguating region (*needed*) was longer in the temporarily ambiguous sentences than in the unambiguous sentences, and crucially, that the two temporarily ambiguous sentences did not differ significantly. This indicated that the very early garden-path effect in the ambiguous sentences was independent of subcategorisation preferences. A similar effect was found in total reading times in this region. The results were further supported by a self-paced reading experiment which showed prolonged reading times on the disambiguating region for both ambiguous conditions compared to

their unambiguous controls. The self-paced reading experiment showed that the effect of subcategorisation preferences was delayed relative to the structural preference: an interaction between bias and sentence type (ambiguous vs. unambiguous) occurred at the region following the disambiguation (*a ride*). In this region, ambiguous sentences with an NP bias verb were more difficult to read than their unambiguous controls, whereas ambiguous sentences with an S biased verb were not. Ferreira and Henderson took this as evidence for the delayed use of subcategorisation preferences, and as support for the garden-path model.

A study that supported the same conclusions is Mitchell (1987), who used NP/Z ambiguities such as (63) in a self-paced reading study. In sentence (63b) *his parents* can either be the direct object of *visited* (the NP analysis) or the subject of the following clause (the ultimately correct Z analysis). Late closure predicts a preference for the NP analysis. In (63a), subcategorisation information associated with *arrived* precludes this analysis, whereas it is consistent with subcategorisation information of the verb *visited.*

63a. After the young Londoner had arrived his parents prepared to celebrate their anniversary. (intransitive verb)

63b. After the young Londoner had visited his parents prepared to celebrate their anniversary. (transitive verb)

Mitchell argued that if the processor employs subcategorisation information immediately, readers should be garden-pathed at *prepared* in (63b), but not in (63a). Assuming that late closure correctly predicts an NP preference in (63b), the NP analysis should be adopted initially. However, this analysis turns out to be incorrect at *prepared*, and therefore, processing difficulty should occur at this point in (63b). No such difficulty should occur in (63a), because the use of subcategorisation information forces the processor to adopt the Z analysis immediately. In contrast, if the processor initially ignores subcategorisation information, it should adopt the NP analysis in both sentences. This should lead to processing difficulty at *his parents* in (63a), because the NP analysis is impossible and the processor has to reanalyse. The results from Mitchell's self-paced reading experiment were consistent with the latter prediction: the viewing display containing the sentence up to and including the NP *his parents* took longer to read in sentences such as (63a) than (63b). This pattern was reversed in the next region containing the rest of the sentence: the intransitive verb condition took less time to read than the transitive verb condition. This suggested that readers had reanalysed the initially adopted NP analysis into an S analysis before the verb in the second clause (*prepared*) was reached in (63a), so that no reanalysis was required at the verb any

more. In contrast, in (63b) reanalysis occurred at or after the verb, resulting in processing difficulty in the final region.

Mitchell's results were highly dependent on the presentation mode he used: (63a) was more difficult than (63b) only when the first viewing window included the NP *his parents*. When the first viewing window ended at the main verb (*arrived/visited*), the two conditions did not differ significantly in the region starting with *his parents*. Hence, it is unclear from Mitchell's experiment how readers process sentences such as (63) in normal reading. It is possible that the NP preference in (63a) was artificially induced when participants saw *arrived his parents* in one window, because they assumed that the end of the window was the end of the clause.

Adams et al. (1998) investigated the same sentences in an eyetracking experiment, in which readers were presented the whole sentence at once, as in natural reading. They found no reading time differences between (63a) and (63b) on *his parents*, but participants experienced more difficulty on the verb in the second clause (*prepared*) in the transitive than in the intransitive condition. This suggests that readers adopted the Z analysis very early on in the intransitive condition, so that no reanalysis occurred on the NP following the intransitive verb. In the transitive condition, the NP analysis is initially adopted, which results in reanalysis at the verb in the second clause because it is only consistent with the Z analysis. Thus, this experiment indicated that the results from Mitchell's (1987) experiment were due to the particular presentation mode.

Adams et al. argued that one explanation that is consistent with their data is the lexical filtering hypothesis. This hypothesis has been formulated to account for findings of early use of lexical information in the garden-path model. Frazier (1987) claimed that the processor initially adopts the structurally simplest analysis, independently of subcategorisation information. When this initial analysis subsequently turns out to be incorrect (at *prepared*), the lexical filter comes into play. If the initial analysis is inconsistent with subcategorisation information, but an alternative analysis is consistent, lexical information facilitates reanalysis. But if the initial analysis is consistent with subcategorisation information, but the alternative analysis is not, reanalysis is difficult. This accounts for Adams et al.'s data: reanalysis at *prepared* is easy in the intransitive condition, but difficult in the transitive condition. Note that according to Frazier's (1987a) proposal, subcategorisation information is employed in a different way than plausibility and discourse information: the latter sources of information trigger reanalysis (Rayner et al., 1983; Clifton & Ferreira, 1989), whereas subcategorisation information only has an effect after reanalysis has been triggered by another source (e.g., the ungrammaticality of *prepared* if the preceding NP has been analysed as a direct object. In later papers (e.g., Clifton and Frazier, 1989; Frazier & Clifton, 1989) the

lexical filter hypothesis has been modified, so that lexical frequency information can also trigger reanalysis. However, as it is unclear when subcategorisation information triggers reanalysis, and when it merely confirms or disconfirms a reanalysis the predictions of the lexical filter hypothesis remain somewhat underspecified.

Clearly, the data from Adams et al. can straightforwardly be explained by constraint-based models (e.g., McRae et al, 1998; Spivey & Tanenhaus, 1998; McDonald et al., 1994) and other models in which subcategorisation information determines the parser's initial analysis. The data fit well with a number of other studies that show that lexical frequency information is used very early on in the parsing process.

For example, Trueswell, Tanenhaus, and Kello (1993) investigated NP/S ambiguities very similar to those used by Ferreira and Henderson (1990). They raised a number of criticisms to Ferreira and Henderson's experiments. Firstly, they noted that the verb-bias in the Ferreira and Henderson materials was relatively weak. Secondly, they argued that the unambiguous conditions which contained the complementiser *that* may not have been good base line conditions, because there may be a general preference for a sentence complement to be introduced by *that*. If the complementiser is absent, the processor may not expect a sentence complement. Finally, they observed that many of the ambiguous NPs following the verb were actually implausible direct objects of the verb and that about half of the sentences in the experiments contained a sentence complement structure. This may have resulted in the disappearance of any difficulty for the sentence complement structures, even in sentences with a direct object biased verb. Trueswell et al. argued that this, together with the possible complementiser effect, may have caused slower reading times for the ambiguous conditions compared to the unambiguous *that*-complement conditions.

Trueswell et al. used materials that were more strongly biased (as measured by a completion study) and a smaller proportion of sentences with sentence complements.[21] They conducted a naming, a self-paced reading and eyetracking experiment. In the naming study, participants listened to sentence fragments such as (64) and had to name one of the pronouns in italics as quickly as possible. The pronoun *he* has nominative case and is consistent with the S reading, whereas *him* has accusative case, which is consistent with the NP reading.

[21] Although Trueswell et al. noted that some of Ferreira and Henderson's (1990) may have contained ambiguous NPs that were in fact implausible objects, they did not control for this factor.

64a. The old man observed *he/him* (direct object bias, complementiser absent)

64b. The old man observed that *he/him* (direct object bias, complementiser present)

64c. The old man insisted *he/him* (sentence complement bias, complementiser absent)

64d. The old man insisted that *he/him* (sentence complement bias, complementiser present)

For the unambiguous fragments containing *that*, naming times were longer for *him* than *he* irrespective of the bias of the verb. For the fragments without complementiser, there was an interaction between verb bias and case of the pronoun. Naming times to *him* were faster than *he* in (64a), but the opposite pattern was observed in (64c). This indicated that lexical information was used very early. Trueswell et al. also correlated the difference in *he* naming latencies between (64a) and (64b) with the frequency of *that* continuations in their completion study. A significant positive correlation was obtained, indicating that naming times were affected by *that* expectations.

The eyetracking experiment which tested sentences such as (65) showed an interaction between verb bias and presence of complementiser at the disambiguating verb phrase (*was in the*). First-pass reading times at the disambiguating verb phrase were longer for (65a) than (65b), but no difference was observed between (65c) and (65d), indicating that lexical preferences affected processing. Trueswell et al. also observed that the landing position of the eye differed depending on whether a complementiser was present or not. This resulted in shorter first fixation times at the ambiguous noun when the complementiser was present compared to when it was absent. It may explain why Ferreira and Henderson (1990) observed a similar first fixation effect. Finally, there was a significant correlation between the frequency of *that* completions and first-pass reading times at the ambiguous NP in the disambiguating region, consistent with the findings in the naming study. The self-paced reading study showed similar results to the eyetracking study.

65a. The student forgot the solution was in the book.
(direct object bias, complementiser absent)

65b. The student forgot that the solution was in the book.
(direct object bias, complementiser present)

65c. The student hoped the solution was in the book.
(sentence complement bias, complementiser absent)

65d. The student hoped that the solution was in the book.
(sentence complement bias, complementiser present)

Similar effects of lexical frequency have recently been found in an eyetracking experiment by Garnsey, Pearlmutter, Myers, and Lotocky (1997). They investigated NP/S ambiguities using verbs that were biased toward the NP reading (e.g., *accepted* in [66a,b]), verbs biased toward the S reading (e.g., *admitted* in [66c,d]), and verbs that were balanced (e.g., *acknowledged* in [66e,f]). Estimations of the biases were obtained from a completion study. Garnsey et al. also manipulated plausibility: the NP following the verb was either plausible as its direct object (in [66a,c,e]) or not (in [66b,d,f]). They reasoned that if syntactic ambiguity resolution is similar to lexical ambiguity resolution, plausibility should have a strong effect for balanced sentences, but little or no influence on NP biased sentences. They compared the sentences in (66) with unambiguous control conditions that included the complementiser *that* after the verb in the main clause.

66a. The talented photographer accepted the money could not be spent yet.
(NP bias, plausible direct object)

66b. The talented photographer accepted the fire could not have been prevented.
(NP bias, implausible direct object)

66c. The ticket agent admitted the mistake had been careless and stupid.
(S bias, plausible direct object)

66d. The ticket agent admitted the airplane had been late taking off.
(S bias, implausible direct object)

66e. The sales clerk acknowledged the error should have been detected earlier.
(balanced, plausible direct object)

66f. The sales clerk acknowledged the shirt should have been marked down.
(balanced, implausible direct object)

First-pass reading times for sentences with NP biased verbs were slower at the disambiguation (e.g., *could not* in [66a]) than for unambiguous controls containing *that*, whereas no such effect was found for the S biased verbs. Neither pattern interacted with

plausibility. In contrast, first-pass reading times for balanced verbs at the disambiguating region did interact with plausibility: no difficulty relative to the unambiguous control was observed when the NP was implausible, whereas plausible NPs were marginally slower than the *that* condition. This suggests that plausibility has an effect only when verb bias is absent.

Garnsey et al. did not observe a correlation between *that* preference and reading times, as in Trueswell et al. (1993). They did observe a correlation between the strength of the verb bias and the difference between the ambiguous and unambiguous conditions in the disambiguating region. This effect was only obtained when plausibility supported the less frequent analysis, suggesting that small frequency differences only had an effect when the two analyses were about equally supported. These findings pose difficulties for models such as proposed by Ford et al. (1982), which assume that subcategorisation frames are ranked. In such models, either the NP or S subcategorisation frame is ranked highest, resulting in either an NP or an S preference, but there should be no gradations in between.

Trueswell (1996) investigated a slightly different form of lexical frequency. Instead of manipulating subcategorisation frame frequencies, he manipulated the frequency of the verb as a past participle or a past tense: *recorded* in (67a) is most frequently used as a past participle, whereas *searched* in (67b) is most frequently a past tense. The past participle form is compatible with the reduced relative reading in (67), while the past tense form is compatible with the main clause reading. In both sentences, the main clause interpretation is implausible because the sentence-initial inanimate NP cannot be the subject of the subsequent verb, as in Trueswell et al. (1994) (See section 2.10).

67a. The message recorded by the secretary could not be understood.
(past participle bias)

67b. The room searched by the police contained the missing weapon.
(past tense bias)

In a self-paced reading experiment, Trueswell found that sentences such as (67b) produced difficulty at the disambiguation (*by the police*) relative to an unambiguous control, but this difficulty was eliminated for sentences such as (67a). The initial NP in this experiment was always inanimate, which has been shown to facilitate the reduced relative reading (Trueswell et al., 1994). In another experiment, Trueswell investigated similar sentences starting with an animate NP which was plausible as the subject of the subsequent verb. For such sentences, processing difficulty was observed regardless of the frequency of the past participle form, although facilitation occurred later in the

sentence for sentences containing a verb with a high frequent past participle form. Trueswell concluded that both frequency and animacy information affected syntactic ambiguity resolution.

A great deal of other studies have shown results that are consistent with this conclusion (e.g., Boland, Tanenhaus, Garnsey, & Carlson, 1995; Clifton, Frazier, & Connine, 1984; Holmes et al., 1989; Mitchell & Holmes, 1985; Stowe, Tanenhaus, & Carlson, 1991). However, Pickering, Traxler, and Crocker (1999) have recently reported a number of studies that might seem inconsistent with it. Pickering et al. investigated both NP/S and NP/Z ambiguities. As both structures led to the same conclusions, only the NP/S ambiguity will be discussed here. They used sentences such as (68), similar to those of Garnsey et al. (1997).

68a. The athlete realised his goals somehow would be out of reach.
(plausible direct object)

68b. The athlete realised his shoes somehow would be out of reach.
(implausible direct object)

In (68a) *his goals* is plausible as the direct object NP of *realised*, whereas *his shoes* in (68b) is not. As shown by completion studies, the main verb (e.g., *realised*) occurred more frequently with a sentence complement than with a direct object NP. However, despite this frequency bias, reading times at the region following the ambiguous NP (*somehow*) were longer if the NP was plausible as a direct object than if it was not, suggesting that lexical preferences were initially ignored. Indeed, these results are incompatible with accounts that assume that lexical frequency information is the only factor that determines initial parsing preferences in ambiguity such as (68). If they were, the S analysis should be the preferred interpretation, because it is most frequent for the verbs that were used.

However, some current constraint-based theories claim that lexical frequency information is not the only type of frequency information that the parser uses. For example, Tanenhaus and Trueswell (1995), Trueswell (1996) and McRae et al. (1998) argued that the processor uses both lexical frequency information and more coarse grained frequency information such as an overall preference for the NP analysis over the S analysis. Although this may seem a somewhat *ad hoc* assumption, Mitchell, Cuetos, Corley, and Brysbaert (1995) pointed out that, in fact, this falls out quite naturally from some interactive simulations (e.g., Juliano & Tanenhaus, 1994). Because in many such theories word nodes are connected to each other in the network, the activation of a word's subcategorisation frames is not only determined by how often the word itself occurred with the subcategorisation frames, but also by how often other words occurred

with these frames. Therefore, the activation of subcategorisation frames is influenced both by specific and more coarse grained frequencies. This is consistent with Pickering et al.'s results. This study shows that there must be a general bias toward the NP analysis in (67) (either structural or due to coarse-grained frequency), but leaves open the possibility that lexical frequency has some effect.

Clearly, the Pickering et al. study is also consistent with purely structural models (although many other studies on frequency information are not) and with models that assume that the processor employs coarse-grained frequency information about particular structures, but ignores frequency information associated with individual words. Mitchell et al. (1995; Cuetos, Mitchell, & Corley, 1996) explicitly proposed the latter type of parser. They mainly focussed on relative clause structures such as in (24), repeated here as (69).

69. Martha cheered the brother of the priest who was in the school celebrating mass.

They claimed that frequency information about any NP1+P+NP2+relative clause structure is represented under the same entry. Thus, parsing preferences should be independent of the particular NP, preposition or relative clause in the ambiguity.

However, there are clear problems with this proposal. As we have seen, Gilboy et al. (1995) showed that the preference for NP1 or NP2 attachment greatly depends on both the preposition (*of* vs. *with*) and the NPs (referential vs. non-referential NP). Even more problematic for coarse-grained frequency accounts is that in a language such as Dutch, relative clause attachment to NP2 is more frequent than to NP1 (Mitchell & Brysbaert, 1998), whereas Brysbaert and Mitchell (1996) showed that NP2 attachment was in fact more difficult to read than NP1 attachment. Similarly, Hindle and Rooth (1993) showed that NP attachment is more frequent than VP attachment in VP+NP+PP structures such as (45), whereas most reading studies show a preference for VP attachment (e.g., Rayner et al., 1983). Spivey-Knowlton and Sedivy (1995) explained such preferences by adhering to a more fine-grained lexical frequency account. They observed that for action verbs (e.g., *hit*), VP attachment occurred more frequently than NP attachment, whereas they did not find a frequency bias for psych (e.g., *expect*) and perception verbs (e.g., *see*). Consistent with such an account, they found that NP attachment was more difficult to read than VP attachment for action verbs, but no difference for psych and perception verbs.[22]

[22] This was true only when they used an indefinite direct object NP. A VP attachment preference was observed when the NP was definite. However, Spivey-Knowlton and Sedivy argued that an indefinite direct object NP represents the neutral context condition, whereas a definite NP results in a contextual bias toward VP attachment.

Recent evidence against frequency accounts comes from Gibson, Schütze, and Salomon (1996) and Gibson and Schütze (1999). These studies tested three-site conjoined NP attachments such as (70) (see also Gibson, Pearlmutter, Canseco-Gonzalez, & Hickok, 1996). In these sentences, the NP *one about a woman with a dog/ one with a dog* can be conjoined with either *a joke* (NP1), *a man* (NP2) or *an umbrella* (NP3). In (70a) plausibility disambiguates it toward NP1 conjunction, and in (70b) toward NP2 conjunction.

70a. The talkshow host told a joke about a man with an umbrella and one about a woman with a dog but hardly anybody laughed. (NP1 conjunction)

70b. The talkshow host told a joke about a man with an umbrella and one with a dog but hardly anybody laughed. (NP2 conjunction)

The results from Gibson and Schütze's self-paced reading experiment showed that NP2 attachment was more difficult than NP1 attachment in the disambiguating region (*about a woman/ with a dog*), although NP2 attachment was more frequent according to Gibson et al.'s (1996) corpus counts. This frequency bias toward NP2 appeared when collapsed over all NP1+P+NP2+P+NP3+conjunct sentences but also in more fine grained-analyses. Clearly, these results are problematic for coarse-grained frequency models.

2.12. Discourse information

The same controversy surrounding the early use of frequency and plausibility information exists for discourse information. As is the case for other types of non-structural information, the modular garden-path theory predicts that discourse information is initially ignored, whereas interactive and constraint-based theories claim that it is employed immediately. Crain and Steedman (1985) proposed the *referential theory* that was specifically designed to deal with the use of discourse information in syntactic ambiguity resolution. In this theory, a syntactic processor proposes syntactic alternatives, while discourse and semantic information select among these alternatives. They termed it a *weak interactive processor*, as opposed to strong interactive models, in which discourse and semantic information propose syntactic structures.

Crain and Steedman proposed the *principle of parsimony* to explain parsing preferences arising from properties of discourse. This states that if there is a reading that carries fewer unsatisfied presuppositions or entailments than any other, that reading will be adopted. The principle can be illustrated with (71).

71a. The psychologist told the wife that he was having trouble with her husband.
(complement clause)

71b. The psychologist told the wife that he was having trouble with to leave her husband.
(relative clause)

Sentences (71) are temporarily ambiguous: *that he was having trouble with* can be the start of a complement clause of *told*, as in (71a), or a relative clause modifying *the wife*, as in (71b). Crain and Steedman argued that in the absence of a context (the so-called null-context), the relative clause reading is disfavoured, because it contains an unsatisfied presupposition, namely that there is more than one wife in the context and that the relative clause selects one of them. This presupposition is not required in the complement clause reading, which is therefore preferred. Thus, the null context is predicted to result in the same preferences as a context mentioning one wife.

Crain and Steedman claimed that if the extra presupposition of the relative clause reading is satisfied, this reading becomes the preferred one. This can be achieved by mentioning more than one wife in the preceding context. The complement clause reading then becomes less preferred, because it is unclear which wife is referred to.

Crain and Steedman tested the predictions of the principle of parsimony with a grammaticality judgement task. Sentences such as (71) were preceded by either a complement clause supporting context (one referent) or a relative clause supporting context (two referents). As predicted, they found that complement clauses were judged grammatical more often than relative clauses in one referent contexts, while the pattern was reversed for two-referent contexts. This finding is inconsistent with the garden-path theory, which predicts a minimal attachment preference for (71a) over (71b).

However, there are a number of disadvantages to the grammaticality judgement task that Crain and Steedman used. In the first place, the task may tap degree of grammaticality of a sentence rather than processing difficulty, which is not necessarily the same. Secondly, because this task is relatively unnatural, it may be subject to strategic influences. Finally, the task does not provide an on-line record of processing, so it cannot distinguish between early and late use of context.

Hence, Ferreira and Clifton (1986) tested Crain and Steedman's principle of parsimony using self-paced reading and eyetracking, which provide an on-line record of sentence processing. They used reduced relative clause ambiguities in four conditions, shown in (72).[23]

[23] The context is shortened here for ease of exposition.

72a. John ran a tape for one of his editors, and he showed some photos to the other.
The editor played the tape agreed the story was a big one.
(reduced relative, supporting context)

72b. John gave a tape to his editor and asked him to listen to it.
The editor played the tape and agreed the story was a big one.
(main clause, supporting context)

72c. John brought out a tape for one of his editors, and told him to listen carefully to it.
The editor played the tape agreed the story was a big one.
(reduced relative, neutral context)

72d. John brought out a tape for one of his editors, and told him to listen carefully to it.
The editor played the tape and agreed the story was a big one.
(main clause, neutral context)

Crain and Steedman argued that reduced relative clauses presuppose the existence of a set of entities (e.g., a set of editors in [72]). Because this presupposition is unsatisfied in a null-context, a reduced relative clause is more difficult to process than a main clause. Instead of using a null-context, Ferreira and Clifton (1986) used what they called neutral contexts such as in (72c) and (72d), which permit both the reduced relative and main clause analysis, but do not strongly bias toward one or the other. They compared (72c) and (72d) with (72a) and (72b), in which the context supported the ultimately correct interpretation. In (72a) two editors are mentioned, which satisfies the presupposition of the reduced relative analysis, and therefore referential theory predicts that it should not cause processing difficulty. It should be no more difficult than (72b), in which the context supports the main clause interpretation. These predictions contrast with those of the garden-path theory, which predicts that the main clause analysis is always preferred (as a result of minimal attachment), regardless of context.

Ferreira and Clifton's self-paced reading experiment showed that the reduced relative conditions (72a) and (72c) both produced difficulty in the disambiguating region (*agreed*) compared to the main clause conditions (72b) and (72d). In fact, the difference between the reduced relative and main clause condition was larger in the supporting context conditions than in the neutral conditions, contrary to the predictions of the principle of parsimony. Although the effects failed to reach significance in Ferreira and Clifton's eyetracking experiment, the pattern of results was the same. Furthermore, in another eyetracking experiment, context had no effect on VP/NP attachment ambiguities such as (73). Ferreira and Clifton showed that VP attachment structures such as (73a)

were preferred to NP attachment structures such as (73b) regardless of whether context supported NP attachment or not.

73a. Sam loaded the boxes on the cart before his coffee break. (VP attachment)
73b. Sam loaded the boxes on the cart onto the van. (NP attachment)

Thus, the data supported a modular approach to syntactic processing. These findings are consistent with a number of other studies that failed to find evidence for early use of discourse information (Mitchell, Corley, & Garnham, 1992; Murray & Liversedge, 1994; Rayner, Garrod, & Perfetti, 1992).

However, Ferreira and Clifton's experiments were criticised by Altmann and Steedman (1988). Specifically, they argued that no context can be truly neutral with respect to discourse presuppositions and that one should compare contexts supporting one reading with contexts supporting another reading to maximise the chance of obtaining an effect. They did this in two experiments using VP/NP attachment ambiguities as in (74). In (74a) and (74b) the ambiguity is resolved in a manner consistent with the context: in (74a), the VP attachment analysis is supported by the context, which mentions one lock, whereas, in (74b) the NP attachment analysis is supported, because two locks are mentioned. In a similar way, the contexts in (74c) and (74d) are inconsistent with the way the ambiguity is resolved.

74a. 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 which had a new lock and a strongbox which had an old lock.
The burglar blew open the safe with the dynamite and made off with the loot.
(VP attachment, consistent context)

74b. 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 which had a new lock and a safe which had an old lock.
The burglar blew open the safe with the new lock and made off with the loot.
(NP attachment, consistent context)

74c. 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 which had a new lock and a safe which had an old lock.
The burglar blew open the safe with the dynamite and made off with the loot.
(VP attachment, inconsistent context)

74d. 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 which had a new lock and a strongbox which had an old lock.
The burglar blew open the safe with the new lock and made off with the loot.
(NP attachment, inconsistent context)

Altmann and Steedman's self-paced reading experiments showed that both the VP attachment and NP attachment sentences were easier when preceded by a supporting context than when the context did not support this reading. In fact, the NP attachment conditions were easier to read than the VP attachment sentences, opposite to the claim of minimal attachment, which predicts a preference for VP attachment.

Although this finding seems to present difficulties for the garden-path theory, Clifton and Ferreira (1989) noted that the NP attachment preference was most likely an artefact, because the lexical items in the PP were introduced further away in the VP attachment conditions (*the dynamite*) than in the NP attachment condition (*the new lock*). Indeed, in a self-paced reading experiment Clifton and Ferreira showed that when the lexical items in the VP attachment conditions occurred after those in the NP attachment conditions, this resulted in a VP attachment preference.

Clifton and Ferreira argued that Altmann and Steedman's results were in fact consistent with the garden-path model. According to the garden-path model, (74c) should cause processing difficulty, because the minimal attachment reading results in a definite NP (*the safe*) which does not refer to a single antecedent. Because of this infelicitousness, processing difficulty should occur. Furthermore, condition (74d) should be relatively difficult compared to (74b), because context should facilitate reanalysis of the initially adopted VP attachment in (74b) but not in (74d). For these reasons, Clifton and Ferreira (1989) argued that one should use neutral contexts, as in Clifton and Ferreira (1986).

Furthermore, Ferreira and Clifton criticised Altmann and Steedman's self-paced reading task. They argued that this task may be sensitive to strategic influences and may not always give the same results as eyetracking studies (Ferreira & Henderson, 1990; Kennedy & Murray, 1984), whose results arguably reflect more natural reading processes. In self-paced reading it is also difficult to distinguish between initial processing and later reanalysis effects, because reading times in self-paced reading are usually unnaturally long so that initial and later processing may be conflated.

Despite these criticisms, a number of eyetracking studies have also found support for early use of discourse information (e.g., Altmann, van Nice, Garnham, & Henstra, 1998; Altmann, Garnham, and Dennis, 1992; and Altmann, Garnham, and Henstra, 1994; Britt, 1994; Britt, Perfetti, Garrod, & Rayner, 1992; but see Mitchell et

al., 1992; Mitchell and Corley, 1994 for criticisms of some of these studies). For example, Britt et al. (1992) found early effects of discourse using eyetracking to investigate VP/NP attachment ambiguities similar to (74) and observed that VP attachment was preferred in neutral contexts, but that this preference was neutralised after NP attachment supporting contexts. However, they did not find a NP attachment preference after such contexts, as in Altmann and Steedman (1988) and as predicted by the referential theory. Furthermore, they did not find such a neutralisation for reduced relative/main clause ambiguities. This suggests that the effects of context are more subtle than referential theory claims. This is further supported by a study by Britt (1994), who observed context effects for VP/NP attachment ambiguities with verbs that optionally subcategorised for a PP, but not for verbs that obligatorily subcategorised for it. This view is consistent with current constraint-based theories, which claim that context has initial effects on sentence processing, but does not necessarily overrule preferences that exist when sentences are presented in isolation (e.g., MacDonald et al., 1994; Spivey & Tanenhaus, 1998; Spivey-Knowlton & Sedivy, 1995; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993).

Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995) argued that context may not always be immediately accessible during reading, because it must be maintained in memory, which may explain why context effects are often relatively weak. They reasoned that if the relevant context remains in the visual field of the reader, so that it does not have to be remembered, context should exert an immediate effect on syntactic processing. In their experiment, participants heard VP/NP attachment ambiguities such as (75), in which the PP *on the envelope* is temporarily ambiguous between VP and NP attachment.

75. Put the saltshaker on the envelope in the bowl.

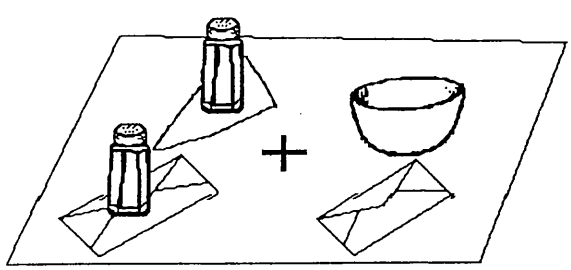

Fig. 2.10. NP attachment supporting context (from Eberhard et al., 1995).

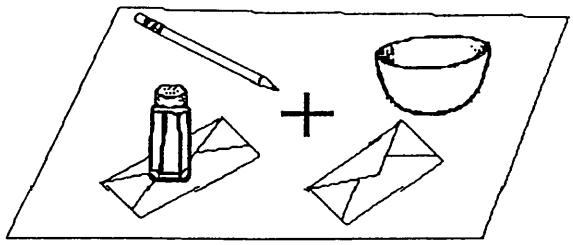

Fig. 2.11. VP attachment supporting context (from Eberhard et al., 1995).

At the same time, the participants were presented with one of the two visual contexts in Fig. 2.10 and 2.11 and were asked to perform the action described by the sentence while their eye-movements were monitored (cf. Allopenna, Magnuson, & Tanenhaus, 1998; Sedivy, Tanenhaus, Chambers, & Carlson, 1999 for a similar methodology). The context in Fig. 2.10 supports the NP attachment analysis, because *on the envelope* selects one of the two saltshakers. The context in Fig. 2.11 supports the VP attachment analysis, because only one saltshaker is present. Tanenhaus et al. showed that when participants were presented with contexts such as Fig. 2.11, they looked at the incorrect destination (the envelope without anything on it) about half of the time after hearing *envelope*. This indicated that they initially analysed the structure as involving VP attachment. Importantly, such eye-movements occurred relatively rapidly (about 500 ms) after participants heard the word *envelope*, which suggests that this kind of methodology taps into early processing. In contrast, when they were presented with the context in Fig. 2.10, they rarely looked first at the envelope without anything on it, but nearly always looked at the correct destination, the bowl. In fact, there was no difference with an unambiguous NP attachment condition such as (76), indicating that the NP attachment was always initially adopted, and the offset of the eye movement was equally fast in both conditions.

76. Put the saltshaker that's on the envelope in the bowl.

Hence, this suggests that context can, if it is sufficiently easy to access, overrule syntactic preferences that exist in the absence of a context.

3. Reanalysis vs. competition

3.1. Two-stage and constraint-based theories

As we have seen in the previous chapter, current sentence processing theories can roughly be divided into two fundamentally different classes which make very different claims about the architecture of the language processor. The two classes differ with respect to two main points: they differ with respect to the sources of information that can be used initially in sentence processing, and with respect to how processing difficulty occurs.

According to one class of theories, often called two-stage theories, the sentence processor bases its initial analysis a limited range of information. In the second stage of processing, this analysis is checked against other sources of information. Processing difficulty is due to reanalysis in such models: if information that was initially ignored is inconsistent with the initial analysis, reanalysis occurs, and this results in processing disruption. The most influential of such models has been the garden-path theory (e.g., Frazier, 1987a; Frazier & Rayner, 1982; Rayner et al., 1983), which assumes that only structural syntactic information is used in the first stage. Other two-stage theories assume that information such as argumenthood (e.g., Pritchett, 1988; 1992; Abney, 1989) is used initially.

The second class of theories are the constraint-based theories (e.g., MacDonald et al., 1994; McRae et al., 1998; Spivey & Tanenhaus, 1998; Tabor et al., 1997). Since the early 1990s, constraint-based theories have replaced the garden-path theory as the dominant account of sentence processing. All constraint-based theories stipulate that the processor works in a single stage and that it can employ all sources of information immediately. Constraint-based theories assume a parallel processor: different possible analyses of an ambiguous syntactic structure are activated simultaneously by the various sources of information. Through a process of competition and inhibition, one analysis will ultimately reach a certain threshold level (e.g., McRae et al., 1998). Processing difficulty arises as a result of such a competition-inhibition process. When one analysis receives more support than the others, little competition occurs between the analyses, resulting in little processing difficulty. But when one or more alternative analyses are

about equally activated, there is a strong competition between the analyses, which should lead to processing difficulty.

Nearly all experimental studies have attempted to discriminate between two-stage and constraint-based theories by investigating whether various non-syntactic sources of information are employed immediately in sentence processing, or whether there is a delay relative to syntactic strategies. As we have seen, this has generated a great deal of important research in sentence processing. However, this research has also reached somewhat of a stalemate. Although it seems clear that many sources of information can be employed extremely rapidly, proponents of two-stage theories claim that this does not necessarily disconfirm such theories. Possibly, reanalysis due to non-syntactic sources of information may occur very rapidly, so that current methodologies such as self-paced reading and even eyetracking may not be sensitive enough to detect structural garden-path effects (e.g., Clifton & Ferreira, 1989). Similarly, proponents of constraint-based theories claim that findings of delayed use of non-syntactic information do not necessarily disconfirm their models, as the manipulations of non-syntactic information may simply be too weak to show up in early measures of processing.

In an attempt to resolve this stalemate, this thesis focuses on an aspect of sentence processing that has often been neglected in sentence processing research. This is the question of whether reanalysis or competition causes processing disruption. This question is a very important one, because it discriminates between two fundamentally different architectures of sentence processing and distinguishes between the two main approaches in syntactic processing. Two-stage theories claim that processing difficulty arises when the processor has to reanalyse, whereas constraint-based theories claim that processing difficulty results from competition. If we can determine whether processing difficulty is due to reanalysis or competition, we can narrow down the range of possible processing models. Furthermore, it can give us insight into whether syntactic and lexical are related processes, or whether they are fundamentally different.

3.2. Competition in lexical ambiguity resolution

According to most current constraint-based theories, syntactic ambiguity resolution is a form of lexical ambiguity resolution (e.g., MacDonald et al., 1994;

Trueswell et al., 1994). As a consequence, syntactic ambiguity resolution should show the same characteristics as lexical ambiguity resolution. Therefore, on the basis of results that have been obtained in lexical ambiguity resolution, we should be able to derive predictions for syntactic ambiguity resolution.

One observation that has been made in lexical ambiguity resolution is that in a neutral context, ambiguous words with two meanings that are about equally frequent (balanced words) are read more slowly than unambiguous controls that are matched for frequency. In contrast, ambiguous words for which one meaning is more frequent than the other (biased words) are read equally fast as unambiguous control words. For example, in an eyetracking study Rayner and Duffy (1986; Duffy et al., 1988) found that balanced words such as *pitcher* produced longer first-pass reading times than unambiguous words with the same frequency, whereas first-pass reading times for biased words such as *port* were equally long as for frequency-matched controls.

These results can be interpreted as a competition between the alternative meanings of ambiguous words. When an ambiguous word occurs in a neutral context and its two meanings are about equally frequent, both meanings are about equally activated. Therefore, competition occurs and reading times for such a word are slow. In a similar way, competition also occurs when context supports one meaning, but there is a frequency bias towards the other meaning. In contrast, little competition occurs when frequency supports only one meaning and the ambiguous word occurs in a neutral context, or when context

Other studies on lexical ambiguity resolution have examined the processing of ambiguous words in a biasing context. Using eyetracking methodology, Duffy et al. (1988; see also Dopkins, Morris & Rayner, 1992; Rayner & Frazier, 1989; Rayner, Pacht & Duffy, 1994) investigated the processing of both balanced and biased words when they were preceded or followed by a context biasing towards one of the meanings. For the balanced words, the contextual bias was always toward the most frequent meaning, whereas it was toward the less frequent meaning for biased words. Examples of the conditions that Duffy et al. tested are presented in (1).

1a. Because it was kept on the back of a high shelf, the pitcher was often forgotten. (balanced word, preceding context)

1b. Of course the pitcher was often forgotten because it was kept on the back of a high shelf. (balanced word, following context)

1c. When she finally served it to her guests, the port was a great success. (biased word, preceding context)

1d. Last night the port was a great success when she finally served it to her guests. (biased word, following context)

When the context followed the ambiguous word, reading times for balanced words were slow relative to their controls, whereas biased words did not differ from their controls. This replicated the results from Rayner and Duffy (1986). But when the context preceded the word, the pattern of reading times at the ambiguous word was reversed. Reading times for biased words were slower than their unambiguous controls, whereas reading times for balanced words did not differ from the control words.

Duffy et al. proposed the *reordered access model* to account for their data (see Dopkins et al., 1992 for further evidence for this model). In this model, both frequency and preceding context influence the speed of lexical access. In balanced words in a neutral context, this leads to a competition between both meanings, whereas no such competition occurs for biased words. When a preceding context biases to one meaning of a balanced word, context speeds access of this meaning and processing is easy. But when the preceding context biases toward the infrequent meaning of a biased word, both meanings of the word are equally supported, and competition occurs. Duffy et al. argued that biasing contexts do not select the appropriate meaning: the dominant meaning of a biased word remains activated even if prior context supports the alternative meaning.

MacDonald et al. (1994) proposed that syntactic ambiguity resolution proceeds in the same way as lexical ambiguity resolution. They argue that both frequency and context influence syntactic ambiguity resolution. As in lexical ambiguity resolution, a biasing context does not select one meaning of an ambiguous word. Although it boosts the activation of the less frequent structure of a biased syntactic ambiguity, the more frequent structure is also activated. However, when context biases toward one of the structures of a balanced syntactic ambiguity, only this structure will be activated.

From MacDonald et al.'s discussion, it follows that competition should arise when two syntactic structures are about equally activated, as in lexical ambiguity resolution. Although MacDonald et al. are most explicit in their claim that syntactic ambiguity resolution is a form of lexical ambiguity resolution, their predictions are similar to those of other current constraint-based theories which also claim that syntactic and lexical ambiguity resolution processes are closely related (e.g., Trueswell, 1996; Trueswell et al., 1993; 1994). Furthermore, as in other interactive and constraint-based theories (Bates & MacWhinney, 1989; McRae et al. 1998; Spivey & Tanenhaus, 1998; and Tabor et al. 1997), MacDonald et al. assume that processing difficulty results from competition between analyses that are about equally activated (but cf. Jurafsky, 1996; Pearlmutter & Mendelsohn, 1999).

3.3. Competition or reanalysis in syntactic ambiguity resolution?

Although there is general agreement that lexical ambiguity resolution involves a process of competition, there is much less certainty whether competition also occurs in syntactic ambiguity resolution. Although most current constraint-based theories assume that lexical and syntactic ambiguity resolution are similar, other researchers have claimed that syntactic and lexical ambiguity resolution are fundamentally different (e.g., Frazier, 1989; Rayner & Morris, 1991; Traxler, Pickering, & Clifton, 1998; Van Gompel et al., 1999; in press). Proponents of such a distinction usually assume that words are accessed in parallel, which can give rise to competition effects, whereas syntactic structures are constructed in a serial fashion, so that no competition effects can arise. Instead, they argue that syntactic ambiguity resolution incorporates a reanalysis mechanism.

Given the importance of the issue, it is somewhat surprising that very few studies have attempted to investigate competition and reanalysis. Furthermore, those studies which did claim to provide evidence for one or the other are in fact equivocal on a closer inspection.

Some of the first researchers who claimed to have found evidence for reanalysis were Frazier and Rayner (1982). One structure they tested were syntactically ambiguous NP/Z structures such as (2).

2. Since Jay always jogs a mile seems like a short distance to him.

In their eyetracking experiment, they found that reading times were longer in (2) than in an unambiguous control sentence. They argued that this difficulty occurred because the initially adopted NP analysis was ungrammatical. As a result, reanalysis ensued. However, recent constraint-based competition theories would claim that processing difficulty in (2) is due to competition. In such models, (2) produces processing disruption at *seems* because there is competition between the more frequent and/or structurally simpler NP analysis and the disambiguating information of *seems*, which indicates that this analysis is incorrect.

A similar argument can be made for Rayner et al. (1983). They showed that VP/NP attachment ambiguities such as (3b) took longer to read than (3a) and argued that this was due to reanalysis of the initially adopted VP attachment in (3b).

3a. The spy saw the cop with binoculars but the cop didn't see him.
(minimal attachment plausible)

3b. The spy saw the cop with the revolver but the cop didn't see him.
(minimal attachment implausible)

But again, current constraint-based theories would explain these results in terms of competition between the analyses (e.g., Spivey-Knowlton & Sedivy, 1995). Because the VP attachment analysis is preferred (presumably due to frequency information) but plausibility supports the alternative NP attachment analysis (as in *the cop with a revolver*), a process of competition arises in (3b). No competition arises in (3a), where both sources of information are favour the same analysis.

More recently, proponents of constraint-based theories have claimed to have found evidence for competition. For example, MacDonald (1994) tested sentences such as (4) which are temporally ambiguous between a main clause and a reduced relative clause analysis (as in the ultimately correct analysis in [4]). In (4a) the verb *captured* is obligatorily transitive, and in (4b) *fought* is optionally transitive.

4a. The ruthless dictator captured in the coup was hated throughout the country. (obligatorily transitive verb)

4b. The ruthless dictator fought in the coup was hated throughout the country. (optionally transitive verb)

5a. *The ruthless dictator captured just after dawn.

5b. ?The ruthless dictator captured just after dawn the army.

6. The ruthless dictator fought just after dawn.

When the verb is obligatory transitive and is followed by a PP, a main clause interpretation is very unlikely, as shown in (5). However, this interpretation is still possible for optionally transitive verbs, because they can be used intransitively, as illustrated in (6). Therefore, MacDonald argued that the reduced relative in (4a) should be easier to read than in (4b) following the point of disambiguation (*was*).

These predictions were confirmed by MacDonald's self-paced reading experiment. Interestingly, the pattern was reversed in the ambiguous region (the verb *captured/fought in the coup*). Sentences with optionally transitive verbs were read faster than a control condition with unambiguous reduced relatives, but sentences with obligatorily transitive verbs were not. MacDonald termed the processing difficulty arising in sentences with obligatorily transitive verbs the *reversed ambiguity effect.* She interpreted this as a competition between a preference for the main clause analysis (due to frequency or structural simplicity) and the constraints supporting the reduced relative analysis (due to the obligatorily transitive verb in combination with the following PP) in (4a). No such competition occurred in (4b), because the verb and the PP are consistent with the preference for the main clause analysis.

Although MacDonald's data are consistent with a competition mechanism, they can also be explained by reanalysis models. Such models predict that readers initially adopt the main clause analysis because this is the preferred analysis, but subsequently have to revise this analysis in (4a), because the obligatory transitive verb in combination with the PP makes this analysis highly unlikely. In (4b), nothing indicates that the initial analysis is unlikely until *was hated.*

The same argument applies for an experiment by McRae et al. (1998), in which sentences such as (7) and their unambiguous controls (disambiguated by *that was*) were tested.

7a. The crook arrested by the detective was guilty of taking bribes.
(implausible main clause, plausible reduced relative)

7b. The cop arrested by the detective was guilty of taking bribes.
(plausible main clause, implausible reduced relative)

In (7a) the reduced relative is plausible and the main clause analysis is not, whereas it is the other way around in (7b). In McRae et al.'s self-paced reading study, *arrested by* was presented in one window, because they claimed that in normal reading, readers do not normally fixate on short functions words such as *by*, and that such words are processed parafoveally when the verb is fixated.

McRae et al. observed an interaction between sentence type (7a) and (7b) and ambiguity (ambiguous vs. unambiguous sentences) in the region *arrested by*. In (7a) this region was much more difficult to read than in its unambiguous control, whereas this difference was less pronounced in (7b). This is consistent with competition: in (7a), the main clause, which is activated because it is structurally simple or highly frequent, competes with the plausibility of the reduced relative and the information provided by *by*, which support the reduced relative clause analysis; whereas in (7b) both the preference for a main clause and plausibility support the main clause analysis and only *by* supports the alternative. Thus, competition should be less here, and their reading-time data was consistent with this prediction.

Indeed, the reading times mirrored findings from a sentence completion task. Sentence fragments up to and including *by* were nearly always completed as a reduced relative clause in (7a), but in (7b) completions as a main and reduced relative were about equally frequent. McRae et al. took the reading time results in combination with the completion data as evidence for competition and as support for their constraint-based simulation. They went on to claim that the results are inconsistent with the garden-path model and with two-stage models in general. They supported this claim by building simulations of the garden-path model using the same competition model but with the modification that all constraints except the main clause bias were delayed.

These garden-path simulations made less accurate predictions than the constraint-based simulation in which all constraints were used simultaneously.

However, the assumption that two-stage theories incorporate competition is incorrect; instead, they employ reanalysis. Hence, the simulations do not show that constraint-based models predict processing performance better than two-stage models. They merely show that in a competition architecture, immediate use of all information gives a more accurate prediction of processing than when some information is delayed. In fact, reanalysis models can straightforwardly account for the data. On such accounts, the main clause analysis is preferred and initially adopted in both (7a) and (7b). In (7a), the analysis becomes unlikely when the region *arrested by* is encountered, because the main clause analysis is implausible and because the *by*-phrase usually indicates a reduced relative analysis. Hence, reanalysis occurs. In (7b), *arrested* in combination with the initial NP is plausible as a main clause and therefore no reanalysis should take place at this point. Although *by* supports the reduced relative interpretation, the completion data suggested that this constraint alone is not strong enough to trigger reanalysis on all trials, as the completion data showed that (7b) was completed as a reduced relative after *arrested by* ... on only about half of the trials. In contrast, reduced relative completions occurred on about 90% of trials for (7a).

This explains the pattern that McRae et al. observed. Reanalysis on *arrested by* occurred both in (7a) and (7b), explaining why both were more difficult to read than their unambiguous controls. However, reanalysis at this point occurred more often for (7a) than for (7b), resulting in increased reading times for (7a) relative to (7b). Finally, at the definitive disambiguation *was guilty*, the pattern was reversed: (7a) was easier than (7b). At this point, reanalysis occurred more often in (7b) than in (7a), as readers had often reanalysed the main clause analysis in the previous region in (7a), but had hardly ever done so in (7b). Thus, both reanalysis and competition models can account for McRae et al.'s results.

Finally, MacDonald, Just, and Carpenter (1992) claimed that their study showed that ambiguous structures are more difficult to read than unambiguous structures. This can be interpreted as evidence for competition. MacDonald et al. investigated sentences such as (8) in a self-paced reading experiment. Sentence (8a) and (8b) are both simple main clauses. However, (8a) is temporarily ambiguous: the sentence can be interpreted as a reduced relative interpretation until the full stop is encountered (as in *The*

experienced soldiers warned about the dangers before the midnight raid conducted the midnight raid.) In contrast, the unambiguous past tense form of *spoke* makes (8b) unambiguous. Sentence (8c) and (8d) contain relative clauses: (8c) contains a reduced relative clause that is ambiguous until *conducted*, whereas (1d) is unambiguous due to the relative pronoun *who.*

8a. The experienced soldiers warned about the dangers before the midnight raid. (main clause, temporarily ambiguous)

8b. The experienced soldiers spoke about the dangers before the midnight raid. (main clause, unambiguous)

8c. The experienced soldiers warned about the dangers conducted the midnight raid. (reduced relative clause, temporarily ambiguous)

8d. The experienced soldiers who were told about the dangers conducted the midnight raid. (reduced relative clause, unambiguous)

MacDonald et al. also measured their participants' verbal working memory as measured by a reading span task (e.g., Daneman & Carpenter, 1980; Just & Carpenter, 1992; King & Just, 1991). In such a task, participants read sets of sentences and have to remember the final words. The larger the set of sentences for which they can remember the final words, the larger their *reading span.* This has been argued to reflect verbal working memory capacity of readers.

The self-paced reading experiment showed that both readers with a low and a high reading span experienced more difficulty reading sentences such as (8c) than sentences such as (8d). This result was obtained in the region consisting of *conducted the midnight* and in the final region (*raid*). This result is not too surprising and is consistent with reanalysis models: it can be interpreted as evidence that readers initially adopted the main clause reading, but had to revise this analysis as soon as they reached the disambiguating verb *conducted.*

The results from the main clause conditions (8a) and (8b) were more interesting. MacDonald et al. observed that high span readers had more difficulty in the disambiguating region *raid* in (8a) than in (8b), whereas no such effect was observed for low span readers. They argued that high span readers kept both analyses in memory and

therefore had more difficulty resolving the ambiguity than low span readers, who only maintained the main clause analysis.

Even more importantly, reading times were longer in the region *before the midnight* in the temporarily ambiguous condition (8a) than the unambiguous condition (8b) for both high and low span readers. MacDonald et al. took this effect as evidence that readers retained both analyses in memory in the temporarily ambiguous condition, whereas only one analysis was available in the unambiguous condition. This can be interpreted as competition between the reduced relative and the main clause analysis in the ambiguous condition.

These results were replicated in two further experiments that MacDonald et al. conducted. They clearly pose difficulty for reanalysis models, which predict that ambiguous and unambiguous sentences should be equally difficult to read, as in either case only one analysis is pursued. However, the MacDonald et al. study was subsequently criticised by Pearlmutter and MacDonald (1995). They conducted a plausibility study on MacDonald et al.'s materials, and showed that the ambiguous conditions were rated as less plausible than the unambiguous conditions. They argued that this might explain the slower reading times for the ambiguous conditions in MacDonald et al. Furthermore, they argued that high span readers may be more sensitive to subtle plausibility differences than low span readers, and that this might have caused high span readers to experience more difficulty with the ambiguous conditions than low span readers. In a correlational study using materials similar to MacDonald et al., they provided evidence for this: ratings for the relative plausibility of the alternative interpretations correlated with reading times for high span readers, but not for low span readers. Therefore, the MacDonald et al. study does not provide conclusive evidence for competition in syntactic ambiguity resolution.

3.4. Traxler et al. (1998): Evidence for reanalysis

Thus, the aforementioned studies do not allow us to distinguish between reanalysis and competition. In order to resolve this debate, we need a case where reanalysis and competition frameworks make different predictions. Such a case occurs in globally ambiguous structures in which two analyses are roughly equally preferred (a

balanced syntactic ambiguity). Competition frameworks predict that such balanced globally ambiguous sentences should be more difficult to process than sentences that are disambiguated. In a globally ambiguous sentence, more than one syntactic structure remains activated throughout the sentence and therefore competition should occur between the analyses, which ought to produce difficulty. But when a sentence is disambiguated, only one analysis remains activated and therefore no such competition should arise. In contrast, according to reanalysis frameworks, globally ambiguous sentences in which both analyses are plausible should never be more difficult than disambiguated sentences, because reanalysis should never occur for globally ambiguous sentences.

Traxler et al. (1998) contrasted reanalysis and competition using both globally ambiguous and disambiguated sentences. They investigated relative clause attachment ambiguities such as (9-10)[1] (e.g., Brysbaert & Mitchell, 1996; Carreiras & Clifton, 1993; Cuetos & Mitchell, 1988; De Vincenzi & Job, 1995; Gibson et al., 1996; Gilboy et al., 1995).

9a. The son of the driver that had the moustache was pretty cool. (ambiguous)
9b. The car of the driver that had the moustache was pretty cool. (NP2 attachment)
9c. The driver of the car that had the moustache was pretty cool. (NP1 attachment)

10a. The steak with the sauce that was tasty didn't win a prize. (ambiguous)
10b. The steak with the sauce that was runny didn't win a prize. (NP2 attachment)
10c. The steak with the sauce that was tough didn't win a prize. (NP1 attachment)

All these sentences are syntactically ambiguous, in that the relative clause can be attached to the first noun phrase (NP1) or the second noun phrase (NP2) of the sentence. Each item had three conditions: a globally ambiguous condition (version a), a condition that was semantically disambiguated toward NP2 attachment (version b) and a condition semantically disambiguated toward NP1 attachment (version c). This plausibility manipulation was checked by off-line plausibility pre-tests.

[1] Three other conditions, which tested PP attachment ambiguities, were included in the experiment that tested sentences like (7). However, these conditions did not reveal any significant differences in reading times, and therefore no firm conclusions can be derived from them.

Further off-line pre-tests were conducted to determine whether there was a preference for either NP1 or NP2 attachment. As we shall see, this is important to derive predictions from some of the models. Participants had to indicate which NP they thought that the relative clause modified. There was a clear bias toward NP2 attachment in sentences such as (10), with the preposition *with*. Participants indicated that the relative clause attached to NP2 81% of the time. The bias was less pronounced in sentences such as (9), with the preposition *of*. Attachment to NP2 was chosen only 68% of the time. Because the plausibility norms indicated that the plausibility of the attachment sites did not differ, these percentages reflected the bias in the materials in the absence of plausibility information.

Competition models such as current constraint-based theories predict that competition should emerge in (9a). In this condition, the semantic plausibility constraint does not favour one analysis over the other, and neither is there a strong non-semantic preference that biases one analysis over the other. In contrast, no such competition should occur in (9b) and (9c), because plausibility favours only one analysis. In constraint-based theories, the degree of processing difficulty depends on the bias of the materials, as measured by the off-line pre-tests (e.g., McRae et al., 1998; Garnsey et al., 1997). This predicts fairly similar reading times for (9b) and (9c), as the difference in preference (68% vs. 32%) is relatively small. If there is any difference, NP2 attachment should be easier than NP1 attachment, because the off-line pre-test indicated that NP2 attachment is somewhat preferred.

Pre-tests also indicated that the bias for NP2 was stronger in (10) than in (9). Thus, the predictions for (10) are different. Little competition is predicted in the ambiguous condition (10a), because the non-semantic NP2 preference information activates one analysis much more than the other. If the bias is strong enough, the ambiguous condition should be no more difficult than the NP2 attachment condition (10b), in which both plausibility information and the bias for NP2 attachment favour the same analysis. Condition (10c) should be most difficult: there should be a competition between the preference for NP2 and plausibility information which disambiguates toward NP1.

The predictions of reanalysis models are different. Most models predict a preference for NP2 attachment (e.g., De Vincenzi & Job, 1993; Frazier, 1987a; Gilboy et al., 1995; Hemforth, Konieczny, & Scheepers, in press). For example, according to

the garden-path model (Frazier, 1978; 1987a; Rayner et al., 1983) the principle of late closure causes the processor to make an initial attachment to NP2 in both (9) and (10). It should not matter whether there is an off-line preference for NP2 attachment or not. When NP2 attachment is implausible, reanalysis should occur. This predicts reanalysis in (9c) and (10c), but no reanalysis should occur in the other conditions, because the initially adopted NP2 attachment is plausible.

According to construal theory (Frazier & Clifton, 1996) relative clause attachment constitutes a non-primary relationship. Therefore, the relative clause is associated in the currently active theta domain. In (10), the active theta domain for the relative clause attachment is *with* + NP2 (i.e., *with the sauce*), because *with* assigns a theta role to NP2. Hence, only NP2 is available as attachment site and therefore NP2 attachment should always be initially adopted in (10), just as in garden-path theory. The predictions are different for (9). Here, the active theta domain is the NP1 + *of* + NP2 complex (i.e., *the son of the driver*), because the preposition *of* does not assign a theta role to NP2. Construal stipulates that the relative clause is associated to this complex. At this point, both syntactic and non-syntactic information can be used to attach the relative clause to one of the NPs. Because plausibility information can be employed at this stage, the processor should somehow select the most plausible analysis. Construal does not specify how this happens: either reanalysis or competition may occur.

Traxler et al. (1998) conducted two eyetracking studies (their Experiment 1 and 3) to test the predictions of the theories for (9) and (10). For the biased structure in (10), readers experienced more difficulty in the NP1 attachment condition than in the ambiguous and NP2 attachment condition, which did not differ from each other. The earliest point at which this pattern occurred was in first-pass regressions out of a short region immediately following the disambiguating noun (in this case, *didn't win*).[2] This is consistent with other studies employing semantic disambiguation (Pickering & Traxler, 1998). The same pattern also obtained in total time on the critical noun (e.g., *tasty, runny, tough*).[3]

The results from the experiment testing sentences such as (9) were very different. Total times on the critical word (*moustache*) were shorter in the ambiguous

[2] First-pass regressions is defined as the percentage of times leaving a region to the left before going past it (see chapter 3).
[3] Total time is defined as the sum of all fixations in a region (chapter 3).

conditions than both disambiguated conditions, which did not differ from each other. This result provides strong evidence against competition-based frameworks. A competition mechanism predicts that the ambiguous condition should be more difficult to read than the disambiguated conditions, but the opposite was the case. In order to provide a more rigorous test of the competition mechanism, the items that were most balanced (ranging between 35% and 60% NP2 preference) were analysed separately. Competition frameworks predict that the competition between the two analyses in the ambiguous condition should be strongest for these items. However, the pattern of results for these items was exactly the same as for the entire set of items.

The results also provide evidence against the garden-path model, because late closure predicts that the NP1 attachment conditions should be more difficult to read than the ambiguous and NP2 attachment condition in both (9) and (10). Although this occurred in (10), it did not occur in (9). Reading times for the NP2 attachment condition in (9b) were longer than for the ambiguous condition, suggesting that readers sometimes reanalysed an initially adopted NP1 attachment in condition (9b). Thus, we can conclude that late closure did not always apply.

The different pattern of results for (9) and (10) is compatible with construal theory. As NP2 is the only NP in the active theta domain in (10), NP2 attachment should always apply. This is consistent with the results: NP2 attachment was easier than NP1 attachment and as easy as the globally ambiguous sentence. In (9), both NP1 and NP2 are in the active theta domain, which is consistent with the finding that the disambiguated conditions did not differ. Finally, because construal theory does not make explicit claims about how the ambiguous condition is resolved, the results from this condition neither confirmed nor disconfirmed the predictions of construal.

Because of the importance of the results from (9), Traxler et al. also tested very similar sentences such as (11). The NP complex in these sentences also contains the preposition *of*. But in contrast to (9), these sentences are disambiguated by the reflexive pronoun *himself/herself* and the pronoun is *who* instead of *that*.

11a. The brother of the colonel who shot himself on the balcony had been very depressed. (ambiguous)

11b. The daughter of the colonel who shot himself on the balcony had been very depressed. (NP2 attachment)

11c. The daughter of the colonel who shot herself on the balcony had been very depressed. (NP1 attachment)

An off-line preference pre-test was conducted to determine whether the materials were biased. This pre-test produced a 70% NP2 attachment bias for the ambiguous condition, which is similar to the bias in (9).

The earliest effects were obtained in first-pass regressions out of the region following the critical word (*on the balcony*), indicating that processing difficulty occurred soon after reading the disambiguating word. More regressions from this region occurred when the relative clause had to be NP1 attached (11c) than when the sentence was globally ambiguous (11a). The number of regressions in the NP2 attachment condition (11c) was also greater than in the ambiguous condition, but this result was only significant by items. A similar pattern occurred in total times on the critical region (*himself/herself*).

These findings are consistent with the results from (9). For materials that are close to being balanced, readers experienced less difficulty for globally ambiguous sentences than for disambiguated sentences. Although the difference between (11a) and (11c) did not quite reach significance in the subject analysis, it corresponds to the pattern found for (9). The pattern of results for such balanced ambiguities is different from that of biased ambiguities such as (10), for which there was no difference in reading times between the ambiguous condition and the condition that was disambiguated toward the preferred NP2 attachment reading.

Clearly, these data are very damaging for competition models such as current constraint-based theories. According to such models, the ambiguous condition should be slower than the disambiguated condition in balanced ambiguities, because competition occurs when there is neither a clear attachment preference nor a plausibility bias toward either analysis. However, exactly the opposite was the case: the ambiguous condition was read faster than the disambiguated conditions.

Because the results from (9) and (11) are also inconsistent with the garden-path theory, which predicts an NP2 preference, and because construal theories does not specify whether reanalysis or competition occurs, an alternative model was proposed by Traxler et al. (1998) and Van Gompel et al. (in press) to account for these data. In Van Gompel et al. (1999; in press) this model was called the *unrestricted race model.* I will describe this model in more detail here.

3.5. The unrestricted race model

The unrestricted race model combines properties from both constraint-based and two-stage models, but is different from both. As in constraint-based theories, there is no restriction on the sources of information that can provide support for the different analyses of an ambiguous structure; hence it is *unrestricted.* In the model, the alternative structures of a syntactic ambiguity are engaged in a race, with the structure that is constructed fastest being adopted (Lewis, 1999; McRoy & Hirst, 1990). Although it is assumed that the processor attempts to construct multiple analyses in parallel, only a single analysis is adopted by the processor[4]. If this analysis is inconsistent with later information, the processor has to reanalyse, and processing difficulty occurs. Thus, the unrestricted race model is in essence a two-stage reanalysis model. In contrast to parallel models, only a single syntactic analysis is ever available to the processor.

The more sources of information that support a syntactic analysis and the stronger this support is, the more likely this analysis will be constructed first. The model claims that when the various sources of information strongly favour one analysis over its alternative (a biased ambiguity), this analysis will nearly always be adopted. In contrast, when two analyses are about equally preferred (a balanced ambiguity), each analysis will be adopted about half the time. A weak bias, however, might lead to one analysis being adopted most, but not all of the time. This is one way the unrestricted

[4] As soon as one analysis is constructed, the processor may either abandon the construction of the other analyses, so that they are never fully constructed, or alternatively, it may continue constructing them (but further ignore them). To my knowledge, there are no data that support one position over the other. Either position could be incorporated in the unrestricted race model.

race model can account for gradations in garden-path effects (e.g., Garnsey et al., 1997; Trueswell et al., 1993; Trueswell et al., 1994).

The model straightforwardly accounts for the pattern of data that Traxler et al. (1998) obtained in their experiments. Because the model is unrestricted, thematic information such as provided by the preposition *of* and *with* are employed in the initial analysis of the sentence. When the preposition is *of* as in (9) and (11), the NP1 and NP2 attachment analyses are about equally supported: the ambiguity is relatively balanced. Because there is no strong preference for either analysis, the unrestricted race model predicts that each analysis will be constructed first on a fair number of trials. When plausibility or gender information subsequently disambiguates the structure towards NP1 attachment, reanalysis occurs on all those trials at which NP2 attachment had been adopted. This causes the processor to slow down. Similarly, when the structure is disambiguated towards NP2 attachment, reanalysis occurs on those trials at which NP1 attachment had been adopted. Thus, this predicts processing difficulty for both disambiguated conditions. In contrast, no reanalysis should occur in the ambiguous condition, because the initially adopted analysis remains possible throughout the sentences, regardless of which analysis was adopted. Taken together, this explains the pattern of results Traxler et al. obtained for (9) and (11). In fact, pre-tests indicated that NP2 attachment was slightly preferred to NP1 attachment. This may explain why the difference between the NP1 attachment and the ambiguous condition was only marginally significant in (11), whereas the difference between the NP2 attachment and ambiguous condition did reach significance.

The pattern of results for a biased ambiguity such as (10) is predicted to be different. Because the preposition *with* favours NP2 attachment of the relative clause, the NP2 attachment analysis will usually beat the NP1 attachment analysis in the race to be constructed first. Therefore, NP2 attachment will be adopted on most of the trials. When the subsequent disambiguation is towards NP1 attachment, reanalysis ensues on all those trials in which the NP2 attachment analysis was adopted, and processing disruption occurs. In contrast, disambiguation toward NP1 attachment results in reanalysis on only a very small proportion of the trials, because NP2 attachment was hardly ever adopted. Reanalysis never occurs in the ambiguous structures, because the initial analysis is always plausible. This explains the results from (10). Presumably, the

NP1 attachment analysis was adopted on only a very small proportion of the trials, so that no difference with the ambiguous condition could be observed.

The unrestricted race model also is consistent with other studies on syntactic ambiguity resolution and it suggests some new ways to look at current data on sentence processing. I will now discuss a few interesting cases in the literature and describe how the unrestricted race model deals with them.

The model works in an incremental fashion. As it encounters each word, it checks whether the syntactic structure built so far is consistent with the new information provided by the word. If it is inconsistent with this information, reanalysis takes place, resulting in processing difficulty. Because the unrestricted race model is a serial model, there are logical restrictions on what information can be used first. Consider the NP/Z ambiguity in (12).

12. While the guests were eating plates were brought in.

In such ambiguities, the NP analysis is preferred, at least for verbs which have a more frequent transitive than intransitive use (e.g., Adams et al., 1998; Clifton, 1993; Frazier & Rayner, 1982; Holmes et al., 1989; Mitchell, 1987; Pickering & Traxler, 1998; Rayner & Frazier, 1987). The semantic implausibility of *eating plates* in fact rules out this analysis in (12). However, the processor does not "know" that the NP analysis is implausible until it has somehow analysed *plates* as the direct object of *eating*. Because syntactic analysis is a prerequisite for semantic analysis, the syntactic structure building has to precede semantic evaluation. In a serial model such as the unrestricted race model, this implies that the processor first has to adopt the implausible NP analysis before plausibility can help it to arrive to the correct, more plausible, Z analysis. This reanalysis process causes processing disruption at *plates*, which is consistent with experimental evidence (Pickering & Traxler, 1998). Thus, the prediction that syntactic processing precedes semantic processing naturally falls out of the architecture of the model. It is not a property that has to be stipulated independently in the way informational encapsulation is in Fodor (1983).

Research suggests that there are similar restrictions on the use of other sources of information. For example, Traxler and Pickering (1996a) investigated sentences such as (13).

13a. I recognised she and her family would be unhappy here.
(incongruent case marking)

13b. I recognised you and your family would be unhappy here.
(congruent case marking)

The verb *recognised* can occur with a sentence complement, as in the ultimately correct analysis in (13), or with a direct object (as in *I recognised her/you*). The case marking of *she* in (13a) makes the direct object analysis impossible, whereas the case of *you* is ambiguous between a direct object and a subject analysis. Traxler et al. observed that first-pass reading times for the region *she and her family* in (13a) were longer than for the region *you and your family* in (13b). This suggests that this case is analogous to (12). As in (12), there is an initial preference to analyse the following noun phrase of the verb (*recognised*) as its direct object rather than as the subject of the following clause, at least for transitive biased verbs. Analogous to the use of semantic plausibility information, we assume that the processor first needs to construct the direct object analysis before it can assess that the case marking of *she* is inconsistent with it. Hence, reanalysis has to take place in (13a), resulting in processing difficulty compared to (13b).

It follows from this discussion that, although the unrestricted race model assumes that all sources of information can be employed immediately, plausibility and case marking information is used in a different way than in constraint-based theories as a result of the logical restrictions inherent in serial models. Because constraint-based theories are parallel models, the analyses can be evaluated in parallel and the most appropriate analysis can be selected. There is no need for reanalysis in constraint-based theories. Instead, they claim that the difficulty at *plates* in (12) and *she and her family* in (13a) is due to competition between the preferred NP analysis and the Z analysis which is supported by plausibility in (12) or case marking in (13a).

Although semantic plausibility information such as in (12) cannot prevent a garden-path from occurring, the unrestricted race model claims that it can be used very rapidly to revise an analysis. This is demonstrated by sentences such as (14) from Pickering and Traxler (1998) (see also section 2.8).

14a. As the woman edited the magazine about fishing amused all the reporters.
(NP analysis plausible)

14b. As the woman sailed the magazine about fishing amused all the reporters.
(NP analysis implausible)

Pickering and Traxler observed that the temporarily ambiguous NP *the magazine about fishing* was more difficult in (14b) than in (14a), but that this pattern was reversed shortly after the definitive disambiguation *amused.* The effect at the NP is consistent with the assumption that the implausibility of the NP analysis forced readers to reanalyse at *the magazine* in (14b), whereas no reanalysis occurred at this point in (14a) because the NP analysis was plausible. At *amused*, the pattern is reversed: no reanalysis is necessary in (14b) any more, whereas readers still have to revise the NP analysis in (14a).

In a similar way, the unrestricted race model can account for other findings of early effects of plausibility. For example, it explains the pattern observed by McRae et al. (1998) for (15) (section 3.3). The unrestricted race model predicts that reanalysis should occur at *arrested* in (15a), whereas it should occur later in the sentence in (15b), consistent with the results from McRae et al.'s self-paced reading study.

15a. The crook arrested by the detective was guilty of taking bribes.
(implausible main clause)

15b. The cop arrested by the detective was guilty of taking bribes.
(plausible main clause)

A number of other studies which manipulated semantic plausibility in structures such as (15) show similar results on the *by*-phrase: reading times are longer when the combination of the initial noun and verb is plausible as a main clause than when it is not (Burgess, Tanenhaus, & Hoffman, 1994; Tabossi, Spivey-Knowlton, McRae, & Tanenhaus, 1994; Trueswell et al., 1994). However, Trueswell et al. did not find evidence for processing difficulty at the verb (e.g., *arrested* in [15a]) when the main clause analysis was implausible. This lack of effect is problematic for the unrestricted race model as well as for constraint-based theories. According to the unrestricted race model, reanalysis should occur at this point, while constraint-based theories predict

competition between the main clause analysis, which is preferred because it is structurally simpler or more frequent, and the reduced relative analysis, which is favoured by plausibility. One plausible explanation is that this simply reflects a type II error. Trueswell et al. only used 4 items and 5 participants per cell, which may not have been enough to pick up a small effect. Other studies (e.g., Clifton & Ferreira, 1986; McRae et al., 1998; Burgess et al., 1994) did observe difficulty at the verb when the main clause analysis was implausible.

Another explanation for the null-effect in Trueswell et al. (1994) lies in the length of their verb region, which was usually very short.[5] This may have had two effects. One possibility is that some readers parafoveally processed the *by*-phrase while fixating on the verb (Burgess, et al., 1994). This may have confounded plausibility effects at the verb with later disambiguation effects due to the disambiguating *by*-phrase. Alternatively, other readers may still have been processing the verb after they had moved to the next region. Given that spillover effects are common in eyetracking research on sentence processing (Pickering & Traxler, 1998; Traxler et al., 1998), it is likely that the effect occurred on the next word(s). Because these words were the disambiguating by-phrase, any effects would have been masked by effects that were due to this disambiguation.

The unrestricted race model claims that, in contrast to plausibility information, discourse information can be employed in the initial analysis of a syntactic structure, because the use of discourse information can precede syntactic structure building. Consider (16), a VP/NP attachment ambiguity with an NP attachment supporting context (section 2.12).

16. The burglar saw that there was a safe with a new lock and a safe with an old lock.

    The burglar blew open the safe with the dynamite.

As soon as the processor reaches the NP *safe*, the context can predict that it needs to be modified in order for the sentence to be felicitous in the context. The processor does not

[5] Trueswell et al. presented a graph which suggested that nouns that were a good patient of the verb produced faster reading times at the verb than nouns that were not. However, they did not provide

need to have reached the ambiguity at *with* in order to hypothesise NP attachment. Hence, context can be used before syntactic principles come into play at *with*. This is consistent with research that shows that VP attachment is preferred to NP attachment in the absence of context, but that this preference is neutralised or even overridden after NP attachment supporting contexts such as in (16) (Altmann & Steedman, 1988; Britt, 1994; Britt et al., 1992). Similar arguments apply for the use of context in other syntactic ambiguities (e.g., Altmann et al., 1992; 1994; 1998).

In contrast to referential theory, the unrestricted race model does not claim that referential context is the only factor that influences initial processing. As in constraint-based theories, other sources of information also have an effect, which explains why context does not always override preferences that exist in the absence of context (e.g., Britt, 1994; Britt et al., 1992). However, because processing difficulty is due to reanalysis in the unrestricted race model, versus competition in constraint-based theories, predictions differ in an interesting way in some cases. For example, constraint-based models predict competition when context conflicts with another constraint. Such a case occurs in the semantically ambiguous sentence (17a), where the context (favouring NP attachment) conflicts with the VP attachment bias. Although the degree of competition obviously depends on how strong both constraints are, the PP attachment ambiguity should be more difficult to resolve than in (17b), where both constraints favour VP attachment. In contrast, the unrestricted race model predicts that (17a) and (17b) are equally difficult to read. In neither case reanalysis should occur. In (17a) readers will adopt the VP attachment on a number of trials, and NP attachment on the other trials. Because later information in the sentence does provide any further information that supports one analysis over the other, the processor will not reanalyse its initial analysis. In (17b), readers will nearly always adopt the VP attachment analysis. Again, because later information does disconfirm the initial analysis, no reanalysis should occur. In future research it would be interesting to test such semantically ambiguous sentences in a discourse context.

---

statistics that allow us to determine whether this effect was significant. Correlations between typicality ratings and first-pass reading times at the verb did not reach significance.

17a. The hunter was chasing a poacher with a rifle and another one with a couple of snares.

When he was near enough, he killed the poacher with the rifle.

17b. The hunter was chasing a leopard.

When he was near enough, he killed the leopard with the rifle.

The unrestricted race model also predicts that frequency information is used immediately. This is consistent with a great deal of studies which showed early influences of frequency information (see section 2.11). It also accounts for recent findings that garden-path effects range on a continuum (e.g., Garnsey et al., 1997; Trueswell et al., 1993). For example, Garnsey et al. (1997) investigated NP/S ambiguities such as (13) in an eyetracking experiment (see also section 2.11).

18a. The art critic wrote the interview had been a complete disaster.

(NP biasing verb)

18b. The bank guard confessed the robbery had been his own idea.

(S biasing verb)

In (18a), the verb *wrote* is biased toward the NP analysis, whereas in (18b) the verb *confessed* is biased towards the S analysis. Garnsey et al. showed that first-pass reading times in the disambiguating region *had been* differed from an unambiguous control sentence in (18a), but not in (18b). Furthermore, the degree of difficulty in the disambiguating region was correlated with the strength of the preference for the direct object or subject analysis. For example, they observed that difficulty in the disambiguating region increased in sentences with S biased verbs as the verb's likelihood of being used with direct objects increased, indicating that garden-path effects ranged on a continuum. The unrestricted race model can readily account for these findings. Because lexical preference information is used initially, it determines the proportion of times that each analysis was initially adopted. The more often that reanalysis is required, the more difficult, on average, a sentence will be. Hence, in sentences with very strong S biasing verbs the subject analysis will be adopted more frequently than in sentences with a somewhat weaker S bias. Consequently, reanalysis in the disambiguating region should occur less often when the S bias is strong than

when it is weak. Similarly, it predicts the S analysis should be adopted more often in (18b) than (18a), explaining the pattern Garnsey et al. obtained for these sentences.

The unrestricted race model does not claim that the proportion of times readers have to revise their initially adopted analysis is the only factor which explains gradations in processing difficulty. Other factors may be involved, too. For example, there is some, although somewhat limited, evidence that the processing cost of adopting an initial analysis may affect degree of difficulty (e.g., Gibson, 1998; Holmes et al., 1987; Kennedy et al., 1989). Even when a structure is unambiguous, structures may differ in the processing cost associated with accessing them or with integrating them into the previously built structure.

A more important factor that affects degree of difficulty in reanalysis is probably that some reanalyses are harder than others. This was demonstrated by Sturt et al. (1999) who showed that reanalysis in NP/S structures such as (19a) was harder than in NP/Z structures such as (19b), even though the preference to take a direct NP was the same for both types of verbs (*saw* and *visited*), as indicated by corpus frequencies (see also section 2.8). This clearly supports the idea that it is not only the proportion of times that reanalysis occurs, but also the type of reanalysis which affects processing difficulty.

19a. The Australian woman saw the famous doctor had been drinking quite a lot. (sentence complement)

19b. Before the woman visited the famous doctor had been drinking quite a lot. (subordinate clause)

3.6. Outlook for the next chapters

In the next chapters, I will report a number of experiments that were designed to further test the unrestricted race model and to contrast it with other current sentence processing models. The experiments did not focus on how various sources of information are used in sentence processing, although they occasionally touch on this issue. As I mentioned in section 3.1, this research has led to somewhat of a stalemate, because both two-stage and constraint-based camps claim to be able to account for the current data. Similarly,

the unrestricted race model is consistent with this type of research, as shown in the previous section. In an attempt to resolve the stalemate, the experiments will focus on an aspect of sentence processing that has often been ignored in previous experimental research, namely the architecture of the processor. The experiments will address this aspect by contrasting reanalysis with competition. This will enable us to distinguish between the unrestricted race model, other two-stage models, and constraint-based theories.

4. Experiment 1: VP/NP attachment ambiguity

4.1. Introduction

Because the pattern of results that Traxler et al. (1998) observed in their study is crucial for the reanalysis/competition debate, it is important to confirm their effects in other experiments. The effects that Traxler et al. obtained for their so-called balanced materials (using the preposition *of*) were relatively weak. For example, the difference between the NP2 attachment and ambiguous condition was only marginally significant in the experiment with gender disambiguation. As suggested in Chapter 3, this may have been due to the fact that the materials were not sufficiently balanced. Readers may have adopted the non-preferred NP1 analysis on a relatively small proportion of trials, so that reanalysis fairly seldom occurred in the NP2 attachment condition. Therefore, experiments with materials that are more closely balanced may provide stronger results. Another explanation of why the effects were relatively weak is that readers may not process relative clauses very deeply. Relative clauses do not contribute to the main proposition of the sentence and may therefore be processed rather shallowly, so that subtle, early effects are difficult to obtain.

In the other experiment using the preposition *of*, which used plausibility as a means to disambiguate, Traxler et al. obtained significant effects only in total times. The effects were not significant in first-pass times or in first-pass regressions, which arguably reflect more immediate processes in reading. As I will argue in section 4.7, there may be several reasons why effects do not always obtain in early measures of processing, and therefore this does not necessarily undermine Traxler et al.'s conclusions. However, in order to convince critics, it would still be more satisfying to find support for the unrestricted race model in such earlier measures.

Traxler et al. investigated only one type of syntactic ambiguity, relative clause attachment to a complex NP. It is unclear whether similar effects also obtain for other types of ambiguities, or whether they are specific to relative clause attachment. For example, Frazier and Clifton (1996) claim that non-primary relations such as relative clause attachment may have a special status in sentence processing. They claim that *construal* applies to non-primary relations. In such cases, an ambiguous phrase is associated into a theta domain, and the processor uses both syntactic and non-syntactic sources of information immediately. This differs from primary relations, that are attached in a determinate fashion following structural processing strategies. Hence, it is unclear whether the pattern of results that was obtained by Traxler et al. (1998) also generalises to structures to which construal does not apply.

Frazier and Clifton (1996) argue that the construal mechanism does not apply to primary relations. Primary relations are attached to a single attachment site in a determinate fashion. Hence, in case of an ambiguity, only one analysis is initially considered. As in the garden-path model, only structural strategies such as minimal attachment and late closure are employed in the initial attachment of primary relations. Other sources of information are only used in the reanalysis stage. This predicts that in ambiguities involving primary relations, one analysis is always the preferred interpretation: as only a single analysis is initially considered, two analyses can never be equally preferred. Clearly, this contrasts with ambiguities involving non-primary relations, because in such cases multiple analyses can exist in the active theta domain, and therefore, these analyses may be equally preferred.

Van Gompel et al. (1999) reported an experiment which investigated VP/NP attachment ambiguities such as (1)[1]. Such ambiguities involve a primary relationship.

1a. The hunter killed the dangerous poacher with the rifle not long after sunset. (ambiguous)

1b. The hunter killed the dangerous leopard with the rifle not long after sunset. (VP attachment)

1c. The hunter killed the dangerous leopard with the scars not long after sunset. (NP attachment)

Frazier and Clifton (1996) claim that VP attachment in these sentences involves a primary relationship, whereas NP attachment is non-primary, and that therefore, VP attachment is initially adopted according to construal theory. If this analysis turns out to be inconsistent with non-syntactic information, reanalysis occurs. This should make sentence (1c) difficult to process, because plausibility information disconfirms the initially adopted VP attachment. In contrast, when plausibility information is consistent with VP attachment, no difficulty should occur. This predicts that both (1a) and (1b) are easy to process, because VP attachment is plausible. It should not matter whether NP attachment is plausible (as in [1a]) or not (as in [1b]), because this analysis will never be considered.

In fact, these predictions are the same as in the garden-path model, though for different reasons. In the garden-path model, VP attachment is initially adopted because this is the simpler analysis according to minimal attachment. There is no preference for arguments over adjuncts, as this information is initially ignored (Clifton et al., 1991).

These predictions were confirmed by Van Gompel et al.'s eyetracking experiment. They showed that VP attachment was easier than NP attachment in

[1] Preliminary data of this study were reported in my Master's thesis (Van Gompel, 1996).

sentences such as (1), as reflected in first-pass regressions and regression path times for the post-critical region (*not long after*), and total times on the critical noun (*rifle/scars*) and the region preceding it (*with the*). This pattern is consistent with the pattern observed in previous research on this type of ambiguity (Britt, 1994; Britt et al., 1992; Clifton et al., 1991; Ni, Crain, & Shankweiler, 1996; Rayner et al., 1983; 1992; Spivey-Knowlton & Sedivy, 1995; but cf. Taraban & McClelland, 1988). Additionally, the experiment provided data about the processing of an ambiguous condition, which had not been investigated before. Both first-pass regressions and regression path times for the post-critical region, and total times on the critical noun and the region preceding it indicated that the ambiguous condition was easier than the NP attachment condition, and did not differ from the VP attachment condition.

However, these results are also consistent with models other than the garden-path model. Pre-tests that Van Gompel et al. conducted showed that the VP attachment analysis was much preferred over the NP attachment analysis. In one pre-test, which they called the *off-line preference pre-test*, participants read the ambiguous conditions and two paraphrase sentences, one which unambiguously paraphrased the VP attachment interpretation, and one which unambiguously paraphrased the NP attachment interpretation. They had to indicate which paraphrase correctly described the meaning they obtained when they first read the sentence. On 78% of the trials they chose the VP attachment paraphrase, indicating that the materials were biased toward VP attachment. In another pre-test, participants had to complete sentence fragments derived from (1) after *with...* On 93% of trials, they completed the sentence as VP attachment, again indicating that the materials were biased toward VP attachment.

The finding of a VP attachment bias in the off-line tasks has important consequences for some models. For example, the unrestricted race model claims that VP attachment should nearly always be adopted initially because of this bias. This should result in the NP attachment condition being difficult to read, because reanalysis should occur on nearly all trials. In the VP attachment condition, reanalysis should hardly ever occur, so that processing difficulty should be negligible. Reanalysis should never occur in the ambiguous condition, where the initial attachment is always plausible. This is consistent with the pattern observed in Van Gompel et al.'s experiment.

The results are also consistent with constraint-based theories. Constraint-based models also claim that reading times for (1) depend on the bias as measured by off-line tasks. The NP attachment condition should be difficult because there is competition between the bias for VP attachment (presumably due to frequency: Spivey-Knowlton & Sedivy, 1995) and plausibility which favours NP attachment. No competition should occur in the VP attachment condition, because both constraints favour VP attachment.

Finally, if one assumes that the VP attachment preference is so strong that it completely overrides any competition due to the fact that both attachment sites are equally plausible, the ambiguous condition should also be easy to process.

4.2. A balanced VP/NP attachment ambiguity: the use of *only*

In order to discriminate between the theories, we need a balanced syntactic ambiguity. I attempted to create a balanced VP/NP attachment ambiguity by including the focus particle *only* before the direct object NP. It was hypothesised that this will make the VP attachment analysis less preferred and will probably result in a balanced ambiguity where there is no preference for either VP or NP attachment. I will now discuss why this may be so.

2a. The hunter killed only the poacher with the rifle not long after sunset. (Ambiguous)

2b. The hunter killed only the leopard with the rifle not long after sunset. (VP attachment)

2c. The hunter killed only the leopard with the scars not long after sunset. (NP attachment)

According to Ni et al. (1996), *only* functions as a form of discourse information. They argued that *only* contrasts an NP (or other constituent) with a set of alternatives: *only* presupposes a contrast set. For example, in (3) good hunters are contrasted with other individuals (presumably bad hunters).

3. Only good hunters can kill leopards.

Ni et al. claimed that *only* affects syntactic ambiguity resolution and interpreted its influence in the context of referential theory (e.g., Altmann & Steedman, 1988; Crain & Steedman, 1985). In their paper, they reported experiments on two syntactic ambiguities that included the word *only*. The first ambiguity was the reduced relative/main clause ambiguity. They used sentences such as (3).

3a. Only businessmen loaned money at low interest were told to record their expenses.

3b. The businessmen loaned money at low interest were told to record their expenses.

Ni et al. argued that in (3a), *only* can set up two different contrast sets. One possibility is that it signals a contrast between businessmen and a set of other individuals who form some kind of contrast with businessmen. This requires the construction of a set of individuals that has not been mentioned in the discourse, and therefore introduces an extra presupposition. Alternatively, the processor can choose to satisfy the presupposition associated with *only* by adopting the reduced relative analysis. This results in a contrast between businessmen who are loaned money and other businessmen. Ni et al. claim that the principle of parsimony (Crain & Steedman, 1985) predicts that the latter is the preferred option, because it requires the parser to partition the set of businessmen that are already part of the discourse model, rather than adding new entities. Hence, they predict that the processor adopts the reduced relative analysis in (3a). This contrasts with (3b), where the principle of parsimony predicts a main clause preference because the definite article *the* presupposes the existence of only one set of businessmen, which makes the modifying reduced relative clause infelicitous (see section 2.12).

Ni et al. conducted an eyetracking experiment to test their predictions. First-pass reading times showed that (3b) was more difficult to process than an unambiguous control containing *who was*, whereas (3a) was no more difficult. This confirmed Ni et al.'s predictions. However, on closer inspection, their data may not be so clear-cut. Ni et al. also observed that both sentences in (3) produced more regressions than their unambiguous counterparts. Because regressions also reflect processing difficulty, this may be taken as evidence that the main clause preference was not completely neutralised. This explanation is also consistent with the word-by-word grammaticality judgement experiment that Ni et al. conducted on the same materials. Although (3a) resulted in faster "grammatical" responses at *were told* than (3b), the responses were still slower than for the unambiguous controls.

Other research also suggests that the effect of *only* is less strong than Ni et al.'s version of the referential theory predicts. In a word-by-word grammaticality judgement and a self-paced reading experiment using exactly the same materials as Ni et al., Clifton, Bock, and Radó (in press) found that both (3a) and (3b) produced difficulty compared to their unambiguous controls. Furthermore, an eyetracking experiment by Paterson, Liversedge, and Underwood (in press) showed that *only* had a relatively weak effect on processing reduced relatives: *only* did not have an effect on first-pass reading times, although it did affect second-pass reading times. A follow-up study (Clayes, Liversedge, & Paterson, 1999) suggested that the effectiveness of *only* depends on the type of relative clause. They found that when the verb in the relative clause was immediately followed by a PP, reduced relative clauses were read equally fast as

unambiguous relative clauses. This contrasted with the reduced relatives in Paterson et al. (in press), in which the verb was followed by a direct object NP. This suggests that *only* does affect ambiguity resolution, but that it does not always override preferences that exist in ambiguities without *only*.

Ni et al. also investigated the use of *only* in VP/NP attachment ambiguities. They used sentences such as (4).

4a. The man painted the doors with new brushes before the festival.
(VP attachment, *the*)

4b. The man painted the doors with large cracks before the festival.
(NP attachment, *the*)

4c. The man painted only doors with new brushes before the festival.
(VP attachment, *only*)

4d. The man painted only doors with large cracks before the festival.
(NP attachment, *only*)

Their prediction for such ambiguities is parallel to that for reduced relatives. When *only* precedes the direct object noun (*doors*), the parser can either contrast doors with other entities, or partition the set of doors into two contrasting sets. Ni et al. claim that the latter option is preferred, and that therefore the modifying NP attachment analysis is preferred. This predicts that NP attachment sentences that include *only* such as (4d) are easier to read than VP attachment sentences such as (4c). This contrasts with (4a) and (4b), in which *only* is replaced by the definite article *the*. Referential theory predicts that in such cases, VP attachment is preferred (e.g., Altmann & Steedman, 1988; see section 2.12).

Ni et al. tested sentences such as (4) in a word-by-word grammaticality judgement and an eyetracking experiment. In the grammaticality judgement experiment, participants had to decide as fast as possible whether the sentence was grammatical up to that point. The results from this experiment were clear and confirmed their predictions: grammatical responses to VP attachment sentences with *the* were faster than for NP attachment sentences with *the*, but the pattern was reversed for sentences with *only*. The results from the eyetracking experiment, which reflect more natural reading processes, were not so clear. First-pass reading times on *large cracks/new brushes* for readers with a large working memory size (measured by a reading span test, see section 3.3) were marginally faster for (4d) than (4c), and (4a) was faster than (4b). This was predicted by their account. First-pass reading times for readers with a small working memory showed a different pattern: both sentences with *the* and with *only* revealed a VP attachment preference, which is inconsistent with their

theory. However, Ni et al. also observed that there were more first-pass regressions from the same region in (4c) than (4d). On the basis of this, they argued that both span groups experienced difficulty with VP attachment sentences that included *only*. However, this conclusion may not be correct. It seems quite possible that readers with a small working memory experienced difficulty with NP attachment in *only* items on some proportion of trials. This would be consistent with the findings for the reduced relative ambiguity. I will therefore conclude that *only* may not exert such a strong influence on syntactic ambiguity resolution as Ni et al. claim it has.

The experiment that is reported here did not aim to investigate the predictions that Ni et al. or other researchers make about the effects of *only*. It simply employed *only* as a means to reduce the strong VP attachment bias that was observed in Van Gompel et al. (1999). The research discussed above suggests that *only* has at least some effect (although it is unclear how much). It should be noted that *only* was employed in a slightly different way from earlier studies. The direct object NP in the materials tested here was singular and contained both *only* and a definite article, as shown in (2). It is not entirely clear how this should affect processing. As in (4), the parser has two choices: it can either contrast the poacher with any other individual, or take the modifying NP attachment reading and contrast the poacher with the rifle with another poacher. However, because the definite article presupposes the existence of only one entity (e.g., one poacher in [2a]) in the discourse, there is no set in the reader's discourse representation that can be partitioned (as in the indefinite noun *poachers*). This contrasts with the situation in (4). The question is what requires adding fewest discourse presuppositions: adding an entity that is very different from a poacher, or adding another poacher, which shares the same meaning as the poacher in the sentence. Presumably, the principle of parsimony predicts the latter option, because it does not require the presupposition that there are other lexical entities in the discourse than the ones that are talked about in the sentence. However, it is not entirely clear whether this prediction directly follows from referential theory. Proponents of this theory have not considered such cases, and may claim that they do not affect ambiguity resolution. This would result in no preference for either VP or NP attachment. Furthermore, given that the effects of *only* often do not appear to be as strong as claimed by Ni et al., it seems likely that *only* plus definite article does not produce an NP attachment bias in sentences such as (2), but rather results in an ambiguity where there is no clear preference for either VP or NP attachment. Such a balanced ambiguity is exactly what we need to contrast the predictions of the garden-path model, construal, constraint-based theories and the unrestricted race model.

Because the predictions of constraint-based models and the unrestricted race model depend on the bias of the materials as measured by off-line tasks, I will first

describe the pre-tests that assessed the bias of the materials before I discuss the predictions of the various models. I will also describe the plausibility pre-test that was conducted, because the experimental materials were selected on the basis of this test.

4.3. Plausibility pre-test

Plausibility was very carefully controlled for in the materials. Previous studies have often manipulated plausibility by intuition alone, with the result that VP-attachment conditions have often included ambiguous sentences where the NP analysis is also plausible (e.g., *The spy saw the cop with binoculars but the cop didn't see him.*, from Rayner et al., 1983). Additionally, even unambiguous sentences can differ in plausibility from each other, and such differences can impact on reading times (e.g., Traxler & Pickering, 1996). In order to rule out the possibility that any effect was due to an irrelevant difference in plausibility, a pre-test was conducted. Furthermore, the materials were constructed in such a way that the point at which analyses became plausible or implausible was localised to a single word (*rifle* or *scars* in [2], repeated below as [5]).

Participants. Eighteen participants were paid to take part in this pre-test. All of them were native-English-speaking students at the University of Glasgow. Nobody participated in more than one experiment or pre-test reported in this chapter or had participated in the biased VP/NP attachment experiment reported in Van Gompel et al. (1999).

Materials. We constructed items like (2), repeated here as (5) (see Appendix 1):

5a. The hunter killed only the poacher with the rifle not long after sunset. (ambiguous)

5b. The hunter killed only the leopard with the rifle not long after sunset. (VP attachment)

5c. The hunter killed only the leopard with the scars not long after sunset. (NP attachment)

The VP attachment condition was semantically disambiguated to the VP analysis; the NP attachment condition was semantically disambiguated to the NP analysis; the Ambiguous condition was semantically ambiguous between the two analyses. All items consisted of a subject NP, a verb, an object NP (containing *only* followed by *the* + N), a

PP (consisting of the preposition *with* and a noun phrase), and a sentence-final adverbial phrase. On the VP analysis, the PP was interpreted as an instrument of the verb; on the NP analysis, it was interpreted as an attribute of the object NP. The object NP and the noun within the PP differed between conditions. The object in the NP-attachment condition was the same as the object in the VP-attachment condition for half of the items, and the same as the object in the ambiguous condition for the other half. The noun within the PP was the same in the ambiguous and VP-attachment conditions, whereas the NP-attachment condition had a different noun. The total length of the adjective plus noun was identical for each version of a given item, as was the length of the noun within the prepositional phrase.

The plausibility study tested 50 sets of sentences such as (6), which were derived from sentences such as (5).

6a. The hunter killed only the poacher by using the rifle.
6b. The hunter killed only the leopard by using the rifle.
6c. The hunter killed only the leopard by using the scars.
6d. Only the poacher with the rifle was killed by the hunter.
6e. Only the leopard with the rifle was killed by the hunter.
6f. Only the leopard with the scars was killed by the hunter.

These sentences were unambiguous paraphrases of item sets like (5a-c), on the VP or the NP analysis. For the VP analysis (6a-c), *with* was replaced with *by using*, and for 4 item sets with *when using* (because *by using* did not appropriately paraphrase the VP attachment reading). We refer to these conditions as the *VP-PP paraphrase conditions.* For the NP analysis (6d-f), the original sentence was converted into a passive. I refer to these conditions as the *NP-PP paraphrase conditions.* Notice that the VP-PP paraphrase of the NP-attachment condition (6c) and the NP-PP paraphrase of the VP attachment condition (6e) derived from analyses that were constructed to have implausible interpretations. All other conditions were derived from analyses that were constructed to be plausible.

Design. Each participant rated all 300 sentences. Half the participants received the items in one random order, and half received the exact reverse order of items. Care was taken that versions of the same item never appeared close to each other.

Procedure. The participants rated how realistic they thought that the situations described by (6) were, on a 7-point scale, with 7 indicating a very realistic situation, and 1 indicating a very unrealistic situation.

Results. Thirty item sets were selected on the basis of the plausibility pre-test. Table 4.1 shows the mean plausibility rating for each condition, for each analysis. We conducted 2 (Paraphrase type: VP-PP vs. NP-PP paraphrase) x 3 (Condition: Ambiguous vs. VP attachment vs. NP attachment) ANOVAs. Both factors were treated as within subjects and items. The analyses revealed a significant interaction between the two factors: F1 (2, 34) = 99.7, $p < .001$, MSe = 1.12; F2 (2, 58) = 714, $p < .001$, MSe = .26. For the purposes of the experiment, it was crucial that conditions (6c) and (6e) were implausible, but the other conditions were not; Table 4.1 shows that this was indeed the case. Planned comparisons revealed plausibility differences between the two paraphrase conditions for VP attachment (F1 (1, 34) = 85.6, $p < .001$; F2 (1, 58) = 613, $p < .001$) and for NP attachment (F1 (1, 34) = 112, $p < .001$; F2 (1, 58) = 801, $p < .001$). In the ambiguous conditions, the numerical difference between the paraphrases was very small, although the effect was significant by items: (F1 (1, 34) = 1.81, p = .19; F2 (1, 58) = 12.9, $p < .001$).

We analysed the absolute differences between the two paraphrases in the three attachment conditions. Planned comparisons revealed that the difference between the paraphrases was much smaller in the ambiguous condition than in the VP condition (F1 (2, 34) = 28.9, $p < .001$, MSe = 2.32; F2 (2, 58) = 269, $p < .001$, MSe = .39) and the NP condition (F1 (2, 34) = 39.7, $p < .001$, MSe = 2.32; F2 (2, 58) = 371, $p < .001$, MSe = .39). This showed that plausibility difference in the ambiguous condition was much less than in the other conditions.

The mean logarithm of the lemma frequency of the noun within the PP was calculated using the Celex English database (Baayen, Piepenbrock, & Van Rijn, 1993). The logarithm of a word's frequency instead of the raw frequency was used, because word recognition studies have shown that there is a linear relationship between the logarithm of the frequency and response latencies to words (Balota & Chumbley, 1984; Chumbley & Balota, 1984). Word recognition studies have also shown that such response latencies reflect the lemma (type) frequency rather than the word (token) frequency (Baayen, Dijkstra, & Schreuder, 1997; Schreuder & Baayen, 1997).

The Celex database showed that the noun in the NP condition was somewhat more frequent than in the other two conditions: F (1, 29) = 5.02, p = .03, MSe = .29, see Table 4.1. The frequency difference works against the prediction of the unrestricted race model, which claims that the NP condition should be more difficult to read than the ambiguous condition.

Table 4.1

Experiment 1: Item characteristics

	condition		
	ambiguous	VP attachment	NP attachment
Plausibility pre-test:			
Plausibility VP-PP paraphrase	5.8	5.8	2.2
Plausibility NP-PP paraphrase	5.3	2.5	5.9
Log frequency of noun in PP	1.92	1.92	2.23
Off-line preference task:			
Percentage of VP attachment	55.0	94.4	2.8
Completion task:			
Percentage of VP completions	37.2	46.1	37.4

Note. Log frequency is based on the number of occurrences in the Celex database (17.9 million words).

4.4. Off-line preference pre-test

This pre-test investigated the bias of the items as they were used in the eyetracking experiment. The crucial condition to test the bias of the items is the ambiguous condition. Given that the plausibility pre-test found that both attachment sites were equally plausible in this condition, a preference for either the VP or NP analysis must be the result of a bias in the items before the critical noun (i.e., *rifle* in [5a]), independent of plausibility. Testing a semantically ambiguous condition is a commonly used methodology to test for syntactic preferences (e.g., Cuetos & Mitchell; 1988; De Vincenzi & Job, 1993; Frazier & Clifton, 1996; Gilboy et al., 1995; Hemforth et al., in press; Traxler et al., 1998; Van Gompel et al., 1999).

Participants. Eighteen participants[2] from the same population as the plausibility pre-test took part.

Materials. The materials were 30 item set such as (5) that were selected on the basis of the plausibility pre-test. These materials were identical to those used in the eyetracking experiment. Each item was followed by the two paraphrases of the item used in the plausibility pre-test.

Design. Order of the paraphrase conditions was treated as a factor. Therefore, six lists of items were constructed (corresponding to the three lists in the eyetracking experiment), with one version of each item appearing in each list. The items occurred in a single random order in the lists. Three participants were assigned to each of the lists.

Procedure. Participants indicated whether the VP-PP or the NP-PP paraphrase best described the situation described by the sentence. They were asked to choose the paraphrase that corresponded to their first reaction to the sentence.

Results. Table 4.1 shows the mean percentage of VP attachment responses for each condition. ANOVAs with the factor Condition (Ambiguous vs. VP attachment vs. NP attachment) as a within subjects and within items variable were conducted. We predicted that the VP and NP conditions would normally be given VP and NP interpretations respectively. In other words, participants should avoid implausible interpretations. This would indicate that our plausibility manipulations were strong

[2] One additional participant was excluded because she chose the VP-PP paraphrase 100% of the time in the implausible NP attachment condition, a pattern clearly different from any other participant. We presumed that she had failed to understand the task.

enough to override any syntactic bias. Table 4.1 indicates that this was indeed the case (the two paraphrases always added up to 100%). ANOVAs revealed a main effect of condition: F1 (2, 34) = 322, $p < .001$, MSe = 118; F2 (2, 58) = 174, $p < .001$, MSe = 364. Planned comparisons revealed that the percentage of VP-PP attachment responses in the ambiguous condition differed from both the VP-attachment condition (F1 (1, 34) = 118, $p < .001$; F2 (1, 58) = 64.1, $p < .001$) and the NP-attachment condition (F1 (1, 34) = 208, $p < .001$; F2 (1, 58) = 112, $p < .001$), which indicated that the plausibility manipulation worked.

Most importantly, the 55.0% VP-PP paraphrase choice in the ambiguous condition did not differ from chance (F1 (1, 17) = 1.19, $p = .29$, MSe = 190; F2 < 1). This indicates that plausibility and preference information did not favour either the VP or NP analysis. Additionally, participants hardly ever took the implausible analysis in the VP and NP condition: 5.6% and 2.8% respectively. We concluded that there was no preference for either VP or NP attachment in the ambiguous condition, but that the final analysis in the disambiguated conditions was nearly always determined by plausibility.

4.5. Completion pre-test

An alternative means of assessing bias employed by constraint-based theorists is sentence completion (e.g., Garnsey et al., 1997; McRae et al., 1998; Trueswell et al., 1993; but see Clifton et al., 1997 and Pickering et al., 1998 for arguments that it does not reflect on-line preferences).

Participants. Eighteen participants from the same population as before took part.

Materials. The same materials as in the off-line preference study were tested, cut off after the word *with* (e.g., *The hunter killed only the poacher with ...*). In addition, 66 fillers were included, which were cut off at various points so that they appeared similar to the experimental items. These filler fragments were identical to the fillers in the eyetracking experiment.

Design. Three lists corresponding to the lists in the eyetracking experiment were created, with one version of each item in each list. Six participants saw each list.

Procedure. The participants were given a list with sentence fragments and instructed to write down the first grammatical and meaningful continuation that they could think of. Next, they were given a second booklet containing the experimental fragments and two

paraphrase fragments (e.g., *The hunter had...*, indicating VP attachment and *The dangerous poacher had...*, indicating NP attachment), balanced for order, and were told to tick the paraphrase which corresponded to the meaning that they intended when they completed the sentence.

Results. On 0.2% of the trials participants said that they did not know which paraphrase was correct. These trials were excluded from the analyses. Table 4.1 indicates how often participants employed a VP completion. We conducted ANOVAs with the factor Condition (Ambiguous vs. VP attachment vs. NP attachment) as a within subjects and within items variable. The overall number of VP completions was significantly less than chance on the item analysis, but not on the subject analysis ($F1 = 1.62$, $p = .22$, $MSe = 527$; $F2 = 11.44$, $p = .002$, $MSe = 124$). There were no significant differences between the conditions ($F1\ (2, 34) = 2.28$, $p = .12$, $MSe = 204$; $F2\ (2, 58) = 1.68$, $p = .20$, $MSe = 460$). The mean percentage of VP completions, 40.3%, is within the range normally considered balanced in lexical ambiguity resolution (e.g., Rayner & Duffy, 1986), and it was therefore concluded that there was no clear preference for either VP or NP attachment.

4.6. Predictions for the eyetracking experiment

Both the off-line preference pre-test and the completion task indicated that there was no clear bias for either VP or NP attachment for these materials. Thus, the ambiguity can be considered as balanced. This suggests that at least off-line, the focus particle *only* affects the way VP/NP attachment ambiguities are resolved, as similar materials without *only* in Van Gompel et al. (1999) revealed a strong VP attachment preference. However, the use of *only* did not result in a clear NP attachment preference. As mentioned in section 4.2, under certain assumptions this is consistent with referential theory. It is also consistent with those studies that have shown that the effects of *only* are rather subtle (Clifton et al., 1999; Paterson et al., in press) and do not completely override preferences that exist in sentences without *only*.

On the basis of the pre-tests, we can derive predictions for constraint-based theories and the unrestricted race model. Constraint-based theories predict that competition should arise in the ambiguous condition (5a) because the constraints favour the VP and NP attachment analysis to an equal degree: neither does plausibility favour one analysis over the other, nor is there an attachment bias which activates one analysis more than the other. On the other hand, the disambiguated conditions (5b) and (5c) should be relatively easy to read, because plausibility favours only one analysis. The

disambiguated conditions should not differ from each other, because the off-line preference pre-test and completion indicated that VP and NP attachment are equally preferred.

The unrestricted race model also predicts that the VP and NP attachment condition do not differ, but for very different reasons. Because the ambiguity is balanced, each analysis is adopted on about half of the trials. When the VP analysis is adopted initially, but the sentence is disambiguated toward NP attachment, reanalysis occurs, and similarly, reanalysis occurs when the disambiguation is toward VP attachment but the initial analysis was NP attachment. This predicts reanalysis on about half of the trials in both disambiguated conditions, and therefore both should produce processing disruption when averaged over trials. With respect to the ambiguous condition, the prediction of the unrestricted race model is very different from that of constraint-based theories. Because both analyses are plausible, reanalysis should never occur, and therefore the ambiguous condition should be easiest to read.

Finally, the predictions of construal and the garden-path theory are independent of the results from the pre-tests. Both theories predict that the parser initially pursues the VP attachment analysis. In construal theory, this analysis is preferred because of the preference for primary relations, while in the garden-path theory it is preferred because of minimal attachment. Because it has been argued that *only* constitutes discourse information and therefore cannot be used in the initial parse of the sentence (e.g., Clifton et al., 1999), both theories predict that the pattern of results for ambiguities including *only* should be the same as for ambiguities that do not. In other words, the same VP attachment preference should be observed as in the biased VP/NP attachment experiment in Van Gompel et al (1999). Furthermore, because the NP attachment analysis is never considered as long as VP attachment is plausible, the ambiguous condition should be as easy to read as the VP condition.

4.7. Eyetracking as a means to investigate syntactic ambiguity resolution

The predictions of the theories were investigated using eyetracking methodology. This method involves measuring people's eye-movements while they are reading in order to determine where and for how long people fixate their eyes. Because a correct interpretation of eyetracking data is important for the conclusions that one draws, I will side-step for a moment and discuss this method in some detail.

Eyetracking has a number of advantages over other methods to measure reading processes. In the first place, it is more natural than other methods that are used to investigate reading. In eyetracking, people simply read sentences or texts and do not

need to perform a task they are not familiar with while they are reading[3]. Methods such as grammaticality judgement (e.g., Crain & Steedman, 1985; Ferreira & Henderson, 1991; Frazier & Clifton, 1996; Ni et al., 1996), the "stop-making-sense" task (e.g., Altmann, 1999; Boland et al., 1995) and probe-word naming/recognition (e.g., Cloitre & Bever, 1988; MacDonald, 1989; McDonald & MacWhinney, 1995) all involve secondary tasks that readers do not perform during normal reading. This may result in specific reading strategies to which readers do not normally adhere. This is especially a concern for the grammaticality judgement and "stop-making sense" tasks, in which readers have to reflect about the linguistic properties of an utterance in order to be able to tell whether a sentence is grammatical or makes sense. Furthermore, because these tasks do not directly measure reading itself but rather the performance on a secondary task, it is unclear whether they measure processing difficulty (which is the issue of interest in this thesis). For example, the grammaticality judgement and "stop-making sense" tasks may reveal which sentences are either grammatical or make sense, but may not tell us much about whether a sentence is hard to read or not. Similarly, probe word naming or recognition may tell us something about the activation of words, but this does not necessarily correlate with reading difficulty. Other methods such as self-paced reading and RSVP (rapid serial visual presentation), which is often used in ERP reading research (e.g., Mecklinger, Schriefers, Steinhauer, & Friederici, 1995; Van Berkum, Brown, & Hagoort, 1999), do not involve a secondary task. Still, these methods involve tasks that are quite different from normal reading. In RSVP readers cannot control the rate at which they read. Hence, when the reading rate is faster than the participants' reading rate, they may abandon certain processes, and when it is slower, they may engage in additional processes. In self-paced reading, readers have to press a button to proceed through a sentence or text. Because they can only see the text that is within the presentation window, readers may adopt specific strategies to cope with this restriction. In addition, as we have seen in section 2.11, the way sentences are divided into presentation windows greatly affects the way people read.

Eyetracking also provides a more or less continuous record of sentence processing. This differs from methods such as grammaticality judgement and probe naming/recognition, which usually measure at only one point in time. Additionally, eyetracking can tap into early processes in sentence processing. Tasks such as grammaticality judgement, "stop-making sense" and probe-word recognition usually yield very long reaction times, and therefore it is unclear whether they tap into initial processes. A similar problem applies for self-paced reading. This is most obvious when the presentation window consists of entire sentences. When using this

[3] Readers have to perform additional tasks such as calibration checks in eyetracking, but they do not occur *during* reading.

methodology, it is unclear where in the sentence the effect first arose. However, this is true even when sentences are presented word-by-word. Because word-by-word self-paced reading times are usually much longer than in normal reading, reading times for a word may reflect both early and late processes.

In eyetracking, several measures have been developed which tap into different, but often overlapping, time intervals during processing of a sentence. Using several of these measures can give an insight in how processing difficulty develops over time. I will discuss the different measures using the hypothetical eye-movement record in Fig. 2.1. Before calculating the various measures, the materials need to be divided into regions for which the measures are to be calculated. This is one way to determine the time course of processing difficulty: because it can be assumed that readers incrementally process sentences from left to right, the nearer a region is to the beginning of a trial, the earlier the processes it taps into. The example in Fig. 2.1 is divided into 7 regions. The region that I will consider here is region 6 (*not long after*).

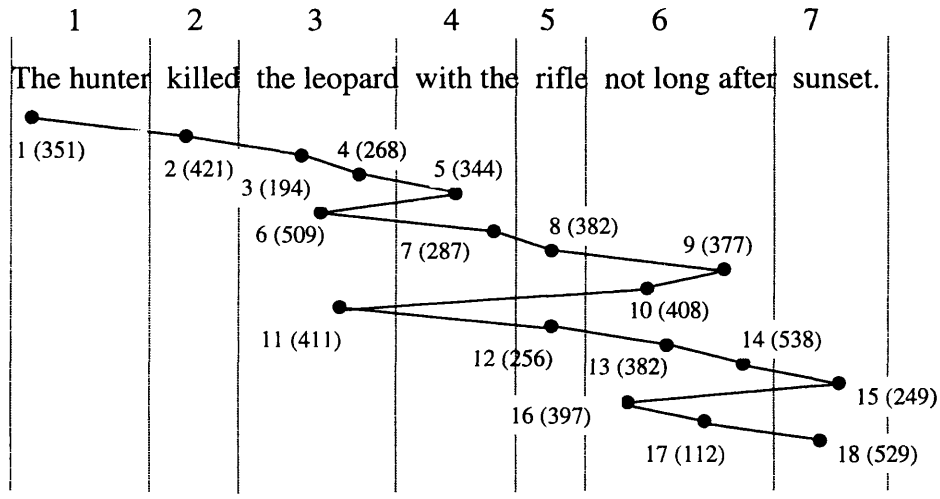

Fig. 4.1. A hypothetical eye-movement record (fixation durations in brackets).

Eyetracking measures can be ranked according to whether they reflect early or late processing. The measure that reflects the earliest processes is the *first fixation duration* in a region. For region 6, this is fixation number 9, which lasts for 377 ms. In this thesis, I will not report the first fixation duration, because it is generally not very sensitive to syntactic processing. First fixation duration is probably too early a measure to tap into syntactic processes. It is much more sensitive to lexical processes (e.g., Rayner & Duffy, 1986; Inhoff, 1984). Studies that obtained first fixation effects in syntactic processing often compared a region which was preceded by short function words with a region that was not preceded by such words (e.g., Ferreira & Henderson, 1990; Garnsey et al., 1997; Trueswell et al, 1993). As argued by Trueswell et al.

(1993), first fixation durations may be shorter in regions following short function words, and therefore the effects may not be due to syntactic processing.

First-pass reading times reflect somewhat later processes in reading. This measure includes all fixations starting with the first fixation in a region until the point of fixation is outside the region. If a reader passes a region before fixating in it, there is no first-pass time. For example, first-pass time for region 6 in Fig. 4.1 includes fixation 9 and 10, and is therefore 785 ms. If there is only one fixation in a region before the region is left, the first-pass reading times and first fixation duration are identical. But because first-pass reading times may also include other fixations, it may capture later processes than first fixation duration. Numerous experiments have shown that syntactic processes can be reflected in first-pass reading times (e.g., Britt et al., 1992; Rayner et al., 1983; Trueswell et al., 1994; Ferreira & Clifton, 1986; Trueswell et al., 1993).

Another early measure of processing is *first-pass regressions.* First pass regressions are defined as all eye-movements that leave a region to the left after a first-pass fixation in this region. An example of a first-pass regression is the eye-movement from fixation 10 to fixation 11. First-pass regressions data are normally presented as the proportion or percentage of trials for which such an eye-movement occurred[4]. Because first-pass reading times necessarily have to precede first-pass regressions, the latter tap into somewhat later processes.

The interpretation of first-pass reading times and regressions is problematic if they show opposite effects (Altmann, 1994; Rayner & Sereno, 1994a; 1994b; see Clifton et al., 1991; and Experiment 4 of Ni et al., 1996 for examples of such cases.). In such cases it is unclear which measure represents the real effect. It is possible that there is a trade-off between the two measures in such cases. If one condition has long first-pass reading times, but few first-pass regressions, this suggests that readers try to recover from processing difficulty in the region where they experienced the difficulty. Hence, they do not go back to preceding parts of the sentence In contrast, if another condition has short first-pass times, but many regressions, this suggests that readers try to recover by going back to preceding parts of the sentence. This may result in short first-pass times, because they do not try to recover in the region where the difficulty was discovered. Therefore, it is unclear which condition presents most difficulty. For this reason, it is important to present both first-pass reading times and regressions.

In contrast to first-pass reading times, there are relatively few studies on syntactic ambiguity resolution which report effects in first-pass regressions. Although there has been some sort of change recently, many papers do not report first-pass regressions (e.g., Britt et al, 1992; Ferreira & Henderson, 1990; Garnsey et al., 1997;

[4] Because first-pass regressions is defined with respect to first-pass reading times, I will exclude trials for which the first-pass reading time is zero (when readers skipped the region) in the calculation of the percentage of first-pass regressions.

Trueswell et al., 1994), or sometimes even deny the relevance of this measure (e.g., Clifton et al., 1991). Hence, it is not always clear whether any effects were obtained in first-pass regressions or not. However, there are a few studies that do report regressions and do find significant effects (e.g., Altmann et al., 1992; Rayner et al., 1983; Pickering & Traxler, 1998; Traxler et al., 1998).

A disadvantage of first-pass regressions is that it only gives a dichotomous distinction: either a regression occurs or not. It does not quantify the degree of difficulty readers experience. When readers experience difficulty, they may go back to an earlier part of the sentence and recover very quickly or alternatively, go back and take a long time to recover. In both cases, the first-pass regressions measure yields the same outcome. Therefore, it is useful to present another measure, which I will call *regression-path times* (e.g., Brysbaert & Mitchell, 1996; Duffy et al., 1988; Konieczny , Hemforth, Scheepers, & Strube, 1997; Traxler, Bybee, & Pickering, 1997). The regression-path time measure includes all fixations in a region until first going past it. For example, in Fig. 4.1, the regression-path time for region 6 includes fixations 9 - 14. Because it includes fixations after a first-pass regression has been made, regression-path time provides a measure of how long it took to recover after such a regression was initiated and therefore the degree of difficulty readers experienced after making a first-pass regression. Regression-path time also includes first-pass time. Hence, this measure can be useful to determine the extent of processing difficulty in cases where first-pass times and first-pass regressions produce opposite patterns. Regression-path time is a later measure than either first-pass reading times or first-pass regressions, because it may include fixations that occur after a first-pass regression has been made.

Few papers have reported regression-path times, and therefore it is somewhat unclear whether it is a sensitive measure for syntactic ambiguity resolution. However, recently there have been a number of studies that show effects in regression-path times (e.g., Brysbaert & Mitchell, 1996; Konieczny et al., 1997; Traxler et al., 1997). These studies often fail to find significant effects in the first-pass measure, and hence it may be that regression-path times are in fact more sensitive. Given that it combines first-pass times and regressions, which are both known to reflect syntactic processing difficulty, this is probably what should be expected.

The final measure is *total reading times*. This is simply the sum of all fixations in a region. For example, the total time for region 6 in Fig. 4.1 consists of fixation numbers 9, 10, 13, 14, 16 and 17. Total time is a late measure of processing, because it includes rereading of a region after the reader has first gone past it.

Although the various eyetracking measures can be ranked according to whether they reflect early or late processes, the real time interval that they tap into depends on how long the reading times for that measure are. For instance, a first-pass reading time

of 300 ms clearly taps into a shorter time interval than a first-pass reading time of 700 ms. Although the intervals start at the same time (when the reader enters the region), the 700 ms time includes processing that is not included in the 300 ms time. Hence, the 700 ms time includes later processes than the 300 ms time. This has important consequences when one compares data from different trials. For instance, if a regression is initiated after a first-pass time of 100 ms on one trial, it captures earlier processes than a first-pass time of 600 ms for another trial. Therefore, if one study finds effects in first-pass times, but these times are very long, while another study finds effects in first-pass regressions and first-pass times are very short, the latter study may in fact show earlier effects. Similarly, regression-path times may reflect earlier processing than first-pass times when one compares between (means of) different trials. For example, if the regression-path time in one study is 300 ms, while the first-pass time in another study is 700 ms, the regression-path time in the former study is likely to represent the earlier measure.

Because reading times are strongly correlated with region length, the reading time duration for a measure depends for a large part on the length of the region it is calculated for. Therefore, reading time measures calculated for long regions generally reflect later processing than short regions. Another consequence of using long regions is that different eye-movement patterns are captured than when the measures are calculated for short regions. For instance, first-pass reading times for a very short region usually include only one fixation. The first-pass reading time for a short region will include few regressions or forward saccades, because they are very likely to fall outside the region. Hence, for short regions, the first-pass measure is very similar to the first fixation duration and therefore taps into a short time interval. But the longer the region, the more forward saccades and regressions it includes. Hence, first-pass times for longer regions are more likely to include spill-over and rereading effects than short regions. The longer a region, the more similar the first-pass measure becomes to the regression-path and total reading measure. If a trial consists of only one region, first-pass, regression-path and total reading time are identical. Except for first fixation times, this line of argument applies for all reading time measures: the longer a region, the more forward and backward saccades it includes. Even first-pass regressions are affected by region length. Short regressive saccades are more likely to be captured when regions are short than when they are long.

In sum, one needs to be cautious when one makes claims about the time course of processing based on eyetracking results from different studies. If a factor has an effect on an early measure such as first-pass times in one study, whereas another factor only has an effect on later measures, this may simply be due to differences in region length. Even when two factors are compared in the same experiment and region length

is controlled for, one needs to be cautious when one shows effects in an earlier measure than the other. For example, number of researchers has claimed to find evidence for the delayed use of one source of information relative to another, because the effects of one factor came out significantly in an earlier measure or an earlier region than the effects of the other (e.g., Ferreira & Henderson, 1991; Paterson et al., in press; Rayner et al., 1992). However, although this is one explanation that is consistent with such a pattern of results, it is not a necessary conclusion. Another explanation is that effects that show up in "late" measures can simply be weak effects rather than late effects (see McElree & Griffith, 1995; 1998 for a similar argument for grammaticality judgement times). This can be illustrated with the graph in Fig. 4.2. This graph shows two factors, A and B, that have an influence on sentence processing (e.g., a structural preference and context). Both factors start influencing the processing of a sentence at the same point in time (point α) and both die out at the same time (point δ). Hence, the time course of the two factors is identical. However, factor A is stronger than factor B: A causes more processing difficulty than B. This is reflected in higher values on the y-axis. This can have an important consequence: the effect of factor A may reach significance in a certain time interval, whereas the effect of B may not. For example, A may reach significance in time interval α-β, but B may not. In the later time window β-γ, both may be significant. In such a case, one cannot conclude that factor A has an earlier impact on processing than factor B. A is simply stronger than B, and because processing difficulty develops gradually (at least when averaged over subjects and items), A reaches significance in earlier time intervals than B. This explanation is particularly likely if later time intervals (e.g., β-γ) show stronger effects of A than B, or if off-line measures show such a difference in strength[5].

[5] Even if this is not the case, this explanation may be correct. Difficulty due to factor B may increase for longer than A, so that it is stronger in later time intervals.

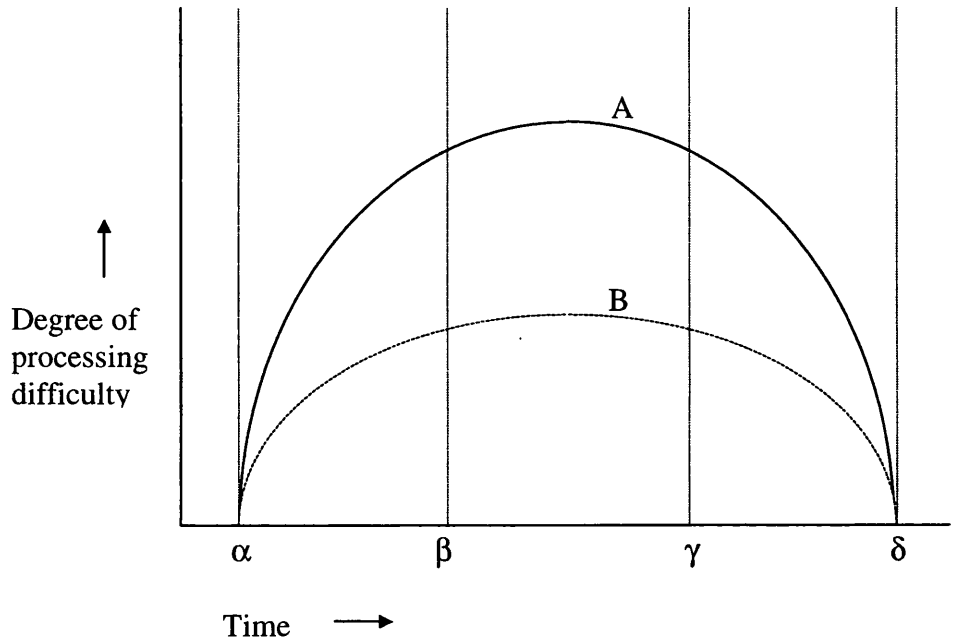

Fig. 4.2. Time course of processing difficulty for two hypothetical factors.

Although eye-movement researchers generally agree that difficulty due to syntactic processing is reflected in elevated reading times, little is known about how exactly syntactic processes affect eye-movement patterns and therefore, how they should be reflected in the various eye-movement measures. Current sentence processing theories do not make predictions about how processing difficulty is reflected in eye-movement patterns and eye-movement measures, while theories of eye-movement control (e.g., Morrison, 1984; Reichle, Pollatsek, Fisher, & Rayner, 1998) largely ignore syntactic processing and mainly focus on how lexical factors affect eye-movement patterns.

There are a number of ways of how elevated reading times can arise. When readers experience processing difficulty, they can in principle do three things (Liversedge, Paterson, & Pickering, 1998). Firstly, they can keep fixating in the region where the sentence was disambiguated (the *critical region*) in order to recover from processing difficulty. This will lead to prolonged first-pass reading times in this region. Indeed, many studies on syntactic ambiguity resolution have found effects in first-pass reading times for the disambiguating region (e.g., e.g., Britt et al., 1992; Rayner et al., 1983; Trueswell et al., 1994; Ferreira & Clifton, 1986; Trueswell et al., 1993). Secondly, readers can go back to an earlier part of the sentence or text in order to check whether they can recover from the difficulty using information that precedes the disambiguation. This will result in more first-pass regressions and longer regression-path times. A number of studies have found effects in these measures (e.g., Altmann et al., 1992; Brysbaert & Mitchell, 1996; Hemforth et al., in press). Finally, readers may

proceed to words following the disambiguation and experience difficulty at this point (e.g., Pickering & Traxler, 1998; Traxler et al., 1998; Van Gompel et al., 1999). There may be two explanations for such effects. One possibility is that the processor expects information following the disambiguation to be useful to recover from the difficulty it experiences and therefore immediately proceeds to this region (the *post-critical region*). If this information is indeed useful, processing of the words following the disambiguation will be relatively easy. But if the words following the disambiguation do not help the processor out, processing difficulty will occur in this region. Again, this can either be reflected in regressions from this region or in longer and more fixations in this region. A second, perhaps more plausible, explanation may be that readers may execute saccades before syntactic processing is finished. For example, Ehrlich and Rayner (1983) argued that there is a lag between higher level linguistic processes such as pronoun resolution and eye-movements, so that effects of such processes may be found after readers have made a forward saccade from the word where processing difficulty occurred. It is quite possible that such spill-over effects also occur in syntactic processing. In fact, such an explanation is suggested by the E-Z reader model of eye-movement control that was recently proposed by Reichle et al. (1998). In this model, saccades are executed before lexical access is complete when accessing a word is slow. Assuming that syntactic processing follows lexical access, this predicts that syntactic processing also lags behind the execution of saccades, especially when syntactic processing is slow as in garden-path sentences. Therefore, it seems likely that effects of syntactic ambiguity resolution show up in the post-critical region. Of course, this is particularly likely when the critical region consists of only one word.

It is currently unclear what factors determine eye-movement patterns in syntactic ambiguity resolution. However, it is likely that the type of syntactic ambiguity, the type of disambiguation, the severity of the garden-path and the position of the disambiguation (e.g., whether it is at the end of a clause or not) all affect the pattern of eye-movements and interact with each other. To my knowledge, there are no studies that have systematically investigated eye-movement patterns in syntactic ambiguity resolution. As I have argued, it is difficult to reach conclusions by making comparisons between studies, because the studies usually differ with respect to region length and reading time durations, which both influence whether an effect in a particular measure reflects early or late processing. Furthermore, measures of eye-movements may also be affected by whether effects are weak or strong.

In conclusion, given our current knowledge of eye-movement patterns, it is a very precarious task to predict in which measures and regions effects will arise in the current experiment. However, it may be possible to derive predictions from previous studies on similar ambiguities. Some studies on VP/NP attachment ambiguities have

observed effects in first-pass times (Britt et al., 1992; Ferreira & Clifton, 1986; Rayner et al., 1983). Unfortunately, these studies differ in important ways from the current experiment and are therefore unlikely to be good predictors of the kind of effects we can expect. Rayner et al.'s first-pass effects were based on very large critical regions including the following clause. Therefore, they are likely to reflect rather late processing and include fundamentally different eye-movement patterns from the current study, which used much shorter critical regions. Ferreira and Clifton used syntactic disambiguation, which provides a clearer error signal than is possible using semantic disambiguation (cf. Pickering & Traxler, 1998). Finally, Britt et al. placed their disambiguating material unambiguously at the end of a clause, at which point reading times normally increase and are likely to incorporate more integrative "wrap-up" processing (e.g., Just & Carpenter, 1980).

A much better basis for predicting the current results may be the biased VP/NP attachment materials in Van Gompel et al. (1999). Except for the difference in bias, these materials were very similar to the balanced materials in the experiment reported here. For the biased VP/NP attachment ambiguities the earliest effect was observed in first-pass regressions out of the post-critical region. Significant effects were also obtained in regression-path times from this region and in total reading times preceding it.

The reason that the earliest effects were obtained in the post-critical region may be because the critical region contained only a single word, which was sometimes quite short. Hence, readers may not have completed syntactic processing before they planned a forward saccade to the next region. Furthermore, the processor may proceed to the following region because it expects that the word following the critical noun may be helpful to recover from processing difficulty. For example, it may make the NP attachment analysis plausible, as shown in (7).

7. The hunter killed the dangerous leopard with the rifle wound.

The fact that the effects were observed in first-pass regressions and regression-path times rather than in first-pass times may be due to the type of ambiguity. After participants read the disambiguating word in the NP attachment condition and experienced difficulty, they may go back to check why VP attachment is implausible, and try to find another attachment site (the direct object NP).

First-pass regression effects for the post-critical region were also found in Experiment 1 and 3 of Traxler et al. This suggests that this type of effect is common for modifier attachment ambiguities with a short disambiguating region. Hence, we can

expect differences between the conditions to show up in the same measures and region in the current experiment.

4.8. Eyetracking Experiment 1

Method

Participants. Thirty-three participants were paid to participate in the experiment. These participants were recruited from the same population as the pre-tests, with the additional restriction that they all had normal vision. Some had taken part in other eyetracking experiments. Three participants were excluded because they had a very high percentage of tracker losses, and three because they answered more than 25% of the questions incorrectly. The analyses were conducted on the remaining 27 participants.

Materials. The experimental materials were the same as in the off-line preference pre-test. The fillers in the completion study were finished off to make complete sentences and used in the eyetracking experiment. These sentences contained a variety of structures unrelated to the experimental materials.

Design. Three lists of items, consisting of ten items from each condition were constructed. Exactly one version of each item appeared in each list. Nine participants were assigned to each list. Each list also contained 66 fillers unrelated to the purpose of the experiment. Thirty items were followed by a question in order to encourage comprehension of the sentences. Half of the questions required a "yes" response and half a "no" response. Participants received feedback when they answered a question incorrectly. The experimental items and fillers were placed in a single random order, with six fillers preceding the first experimental sentence and three fillers following a break in the experiment.

Procedure. Participants' eye movements were recorded with a Fourward technologies Dual Purkinje generation 5.5 eye tracker (see Crane, 1994 for a description). The tracker has an angular resolution of 10' arc. It monitored only the right eye's movements. A computer displayed the materials on a screen 77 cm from the participants' eyes. The screen displayed 3.8 characters per degree of visual angle. The tracker monitored participants' gaze location every millisecond and the software sampled the tracker's output to establish the sequence of eye fixations and their start and finish times.

Each participant was run individually. The experimenter told the participant to read the sentences carefully in order to understand them, but to read at his or her normal rate. Bite bars and head restraints were used to minimise head movements. Next, the participant completed a calibration procedure. Before each item or filler a calibration check was performed, and the calibration was repeated if necessary. Each sentence was presented on a single line. After reading the sentence, the participant pressed a button which led to the presentation of a comprehension statement or the next trial. The experiment, which contained one break, took about half an hour.

Analyses. On average, the participants made 8% errors on the comprehension questions. Prior to all analyses, trials with major tracker losses and trials with fewer than four fixations were excluded. This eliminated 1% of the trials. If a fixation was shorter than 80 ms and within one character space of the previous or next fixation, it was assimilated to this fixation. All remaining fixations shorter than 80 ms were excluded. Following Rayner and Pollatsek (1989), we assume that readers do not extract much information during such short fixations. Fixations longer than 1500 ms were also excluded. These extremely long fixations are presumably due to a major tracker losses.

If a reader skipped a region in a reading time measure, resulting in a zero reading time, it was excluded from the analyses for that measure. In addition, if a reader skipped two or more consecutive regions, the data of the regions following them were also excluded from the analyses on that measure. It was assumed that participants could not have processed the sentence completely if that was the case. The latter procedure removed less than 1% of the data for any analyses.

We divided the experimental items into 7 regions, indicated by slashes below:

8. The hunter/ killed/ only the poacher/ with the/ rifle/ not long after/ sunset.

These regions corresponded to (1) the subject NP, (2) the verb, (3) the object NP, (4) the words *with the*, (5) the noun within the PP (known as the critical noun region), (6) the next words (known as the post-critical region), and (7) the final word of the sentence. The space between words was included with the following word.

Results

Table 4.2 presents the mean reading times and percentage regressions by condition and region. For each eyetracking measure for each region, we conducted one-

way ANOVAs with three levels of Condition (ambiguous vs. VP attachment vs. NP attachment), which was treated as a within subjects and items factor. Table 4.3 presents the results of these analyses. Planned comparisons between levels are reported below.

First-pass times did not produce any significant effects, as in the biased VP/NP attachment experiment in Van Gompel et al. (1999). However, the first-pass regression measure showed clear effects in Region 6 (the post-critical region), with planned comparisons showing that fewer regressions occurred in the ambiguous condition than both the VP condition (F1 (1, 52) = 4.27, p = .044; F2 (1, 58) = 6.23, p = .015) and the NP condition (F1 (1, 52) = 12.1; p < .001; F2 (1, 58) = 13.4, p < .001). The VP and NP conditions did not differ (F1 (1, 52) = 1.99, p = .16; F2 (1, 58) = 1.36, p = .25).

The pattern for regression-path times was very similar. In Region 6, the condition effect was significant by items and was marginally significant by participants (p = .071). Planned comparisons showed that the ambiguous condition was faster than both the VP condition (F1 (1, 52) = 4.42, p = .041; F (1, 58) = 10.0, p = .002) and the NP condition (F1 (1, 52) = 3.91, p = .053; F (1, 58) = 9.04, p = .004). The VP and NP conditions did not differ (F1 (1, 52) = .016, p = .90; F (1, 58) = .026, p = .87).

Total times showed an effect in Region 4 (*with the*). Planned comparisons showed that the ambiguous condition was read faster than the VP condition (F1 (1, 52) = 6.24, p = .016; F (1, 58) = 5.70, p = .020) and the NP condition (F1 (1, 52) = 5.91, p = .019; F (1, 58) = 5.40, p = .024), which did not differ from each other (Fs < 1). There was a marginally significant effect in Region 3, the direct object (p = .053 by participants and p = .074 by items); the means suggest that the ambiguous condition may have been read faster than the disambiguated conditions. No other effects approached significance by participants and items.

Table 4.2

Experiment 1: Means

	region						
	1	2	3	4	5	6	7
First-pass reading times:							
ambiguous	371	331	520	265	310	448	464
VP attachment	398	315	555	272	317	463	479
NP attachment	405	312	560	286	314	442	491
First-pass regressions out:							
ambiguous		6.2	9.0	6.8	15.6	13.9	59.1
VP attachment		9.8	9.0	7.4	17.1	22.1	56.4
NP attachment		6.5	9.5	9.1	14.3	27.1	61.1
Regression-path times:							
ambiguous	371	363	620	299	360	562	1063
VP attachment	398	374	626	332	379	693	1172
NP attachment	405	342	645	342	377	682	1133
Total times:							
ambiguous	421	425	706	362	398	633	595
VP attachment	469	435	795	409	404	681	597
NP attachment	456	412	764	410	432	674	611

Notes. The regions were defined as follows (delimited by slashes):
The hunter/ killed/ only the poacher/ with the/ rifle/ not long after/ sunset.

First pass, regression path, and total times are reported in ms, first pass regressions as the percentage of saccades leaving the region to the left after a first pass fixation.

Table 4.3

Experiment 1: ANOVA results by region for effect of condition

	F1 (2, 52)	MSe	F2 (2, 58)	MSe
Region 1:				
First pass reading times	2.47	2769	1.91	4664
Regression path times	2.47	2769	1.91	4664
Total reading times	2.47	5451	1.46	
10720				
Region 2:				
First pass reading times	< 1		1.87	1977
First pass regressions out	2.29	53.1	1.52	78.9
Regression path times	1.82	3528	2.14	3999
Total reading times	< 1		< 1	
Region 3:				
First pass reading times	2.95	4006	1.69	7769
First pass regressions out	< 1		< 1	
Regression path times	< 1		< 1	
Total reading times	3.10	16752	2.72	
21394				
Region 4:				
First pass reading times	2.30	1035	1.74	1706
First pass regressions out	< 1		< 1	
Regression path times	1.54	8908	1.20	
13745				
Total reading times	4.05*	5112	3.70*	6567
Region 5:				
First pass reading times	< 1		< 1	
First pass regressions out	< 1		< 1	
Regression path times	< 1		< 1	
Total reading times	1.69	4458	1.85	6743
Region 6:				
First pass reading times	< 1		< 1	
First pass regressions out	6.12**	181.1	7.00**	200.2
Regression path times	2.78	45505	6.37**	
27139				
Total reading times	< 1		1.88	
10018				
Region 7:				
First pass reading times	< 1		< 1	
First pass regressions out	< 1		< 1	
Regression path times	< 1		< 1	
Total reading times	< 1		< 1	

Note. * is significant at the .05 level, ** is significant at the .01 level.

Table 4.4

Experiment 1: Percentage of first-pass regressions for region 6 by condition and item and subject bias

	item bias		subject bias	
	balanced	biased	balanced	biased
ambiguous	11.2	16.5	14.8	13.1
VP attachment	21.4	22.8	24.0	20.2
NP attachment	21.5	32.8	26.2	27.9

Thus, first-pass regressions, regression-path times and total reading times all indicated that the ambiguous condition caused less processing difficulty than either the VP or NP attachment condition. These results are clearly difficult to reconcile with constraint-based theories that incorporate competition. In order to test these models further, additional analyses were conducted. Competition models predict that the more balanced a material is, the stronger competition should be. In order to test this hypothesis, the materials were split into balanced and biased items. The bias of the materials was determined on the basis of the results from the eyetracking study itself. The bias was not determined by the pre-tests, because some researchers have argued that they do not accurately reflect on-line preferences (e.g., Clifton et al., 1997; Pickering et al., 1999). Because first-pass regressions from the post-critical region was the earliest point at which significant effects occurred, balanced materials were defined as those 15 materials for which the difference between the VP and NP attachment in first-pass regressions from region 6 was smallest (this turned out to be a difference of 12.5% or less). The other 15 materials were assigned to the biased group. Table 4.4 presents the means for each condition for each group of materials.

ANOVAs with Condition as a within subjects and items factor and Item bias (Balanced vs. Biased) as a within subjects and between items factor were conducted. They revealed an effect of Bias ($F1\ (1, 26) = 8.44$, $p = .007$, $MSe = 309$; but $F2\ (1, 28) = 1.90$, $p = .18$, $MSe = 423$), which is presumably due to irrelevant differences between the two groups of items (e.g., frequency, length) and, as expected, an effect of Condition ($F1\ (2, 52) = 7.31$, $p = .002$, $MSe = 389$; $F2\ (2, 56) = 6.90$, $p = .002$, $MSe = 203$). No interaction between the two factors was observed ($F1\ (2, 52) = 2.39$, $p = .10$, $MSe = 290$; $F2\ (2, 56) = .60$, $p = .55$, $MSe = 203$), which indicates that the pattern of results was roughly similar for balanced and biased items[6]. Most importantly, when the balanced items were analysed separately, the effect of Condition was still significant: $F1\ (2, 52) = 3.52$, $p = .037$, $MSe = 302$; $F2\ (2, 28) = 6.99$, $p = .003$, $MSe = 80.4$. As can be seen in Table 4.4, even for the balanced items the ambiguous condition produced fewer regressions than either the VP and NP attachment condition. In fact, the pattern is identical to that for all items. These results are therefore inconsistent with competition models, which predict that competition should be especially severe in balanced materials.

In a similar way as items may be either biased or balanced, participants may have either a strong preference for one analysis over the other, or have no clear

[6] It should be noted that the NP attachment condition seemed to produce more regressions than the VP attachment condition in the biased materials, but not in the balanced materials. This was due to the fact that overall, there was a slight (non-significant) preference for VP attachment. As a consequence of removing all balanced materials, the difference increased for the biased materials.

preference. In other words, some participants may be biased, while others are balanced. Competition models might predict that balanced participants experience more competition in the ambiguous condition than biased subjects. In order to test this, participants were split into two groups. Balanced participants were defined as those 13 participants for which the difference between the VP and NP attachment in first-pass regressions from region 6 was smallest (in effect 12.2% or less). Biased participants were all 14 other participants. The means for each group of participants are presented in Table 4.4. ANOVAs with Condition as a within subjects and items factor and Participant bias (Balanced vs. Biased) as a between subjects and within items factor were conducted. They showed an effect of Condition ($F1\ (2, 50) = 5.98$, $p = .005$, $MSe = 186$, $F2\ (2, 58) = 6.15$, $p = .004$, $MSe = 383$), but no effect of Bias ($F1\ (1, 25) = .050$, $p = .83$, $MSe = 676$; $F1\ (1, 29) = 1.41$, $p = .24$, $MSe = 208$) or interaction between the two factors ($F\ (2, 50) = .37$, $p = .69$, $MSe = 186$; $F2\ (2, 58) = .53$, $p = .59$, $MSe = 383$), indicating that biased and balanced participants performed similarly. As shown in Table 4.4, the pattern for the balanced participants was very similar to the overall pattern: the ambiguous condition produced fewer regressions than both the VP and NP attachment condition. Separate analyses of the balanced participants showed an effect of Condition: $F1\ (2, 58) = 3.13$, $p = .051$, $MSe = 294$, $F2\ (2, 24) = 4.97$, $p = .016$, $MSe = 97.4$.

4.9. Discussion

The results clearly showed that the ambiguous condition was easier to process than both disambiguated conditions, which did not differ from each other. This pattern of results was obtained in first-pass regressions, regression-path times and total reading times. The findings differ from those for the biased VP/NP attachment experiment in Van Gompel et al. (1999). In this experiment, the ambiguous and VP attachment condition did not differ, and both produced less difficulty than the NP attachment condition. Although the comparison between this experiment and the current experiment involves slightly different items, it is likely that the different pattern is caused by the word *only*, which did not occur in the biased materials of Van Gompel et al.

The pattern of results provides clear evidence against competition-based models such as most current constraint-based models. Because the materials are balanced and because there is no plausibility difference in the ambiguous condition, competition should occur in the ambiguous condition, resulting in processing difficulty. In contrast, no such competition should occur in the disambiguated conditions, because only one analysis is plausible. In fact, the opposite effect was obtained: the disambiguated conditions were harder than the ambiguous condition.

In theory, there are a number of ways that these results might still be consistent with competition models. One possibility is that a large proportion of materials was biased, with about an equal number of VP and NP attachment biased materials. In such a case, the NP attachment condition would produce difficulty in the VP attachment biased materials (as in Van Gompel et al.'s biased experiment), because there is a competition between plausibility favouring the NP analysis and the bias favouring the VP analysis, and similarly, the VP attachment condition would be difficult in the NP attachment biased materials. Overall, both conditions would produce difficulty. Assuming that the biases are very strong and overrule any competition due to equal plausibility, little competition should occur in the ambiguous condition. If this type of explanation were correct, one would expect biased materials to show little sign of competition, whereas competition should be much stronger for balanced materials. The analyses on the balanced and biased item groups suggest that this is incorrect. Although some materials appeared to be more balanced than others (based on the difference in regressions between the VP and NP attachment condition), there was no interaction between bias and condition. Importantly, even the most balanced materials clearly showed that the ambiguous condition was read faster than the disambiguated conditions. Therefore, although materials differed in strength of bias, there was no evidence that more balanced materials produced more competition than less balanced materials.

Another possibility is that a large percentage of participants had a preference for one analysis over the other, with about an equal number of participants preferring VP and NP attachment. This would result in the same kind of pattern as when a large proportion of materials are biased. If this explanation is correct, little competition should occur for biased participants, whereas competition should be much stronger for balanced participants. Indeed, some participants appeared to have stronger preferences than others. However, the analyses on biased and balanced groups of participants showed that participant bias did not affect the degree of competition. Analyses with the factor participant bias showed no interaction between condition and bias, and the ambiguous condition still produced less difficulty than both disambiguated conditions for the balanced participants.

Finally, it is conceivable that both the item and participant bias might have affected the degree of competition and that their interaction somehow produced the observed pattern of results. Participant and item bias might conspire so that on some proportion of trials, VP attachment was strongly preferred, whereas on another proportion, NP attachment was preferred. This predicts strong competition on trials where neither analysis was preferred and little competition on trials where there is a strong bias. It is not possible to split the data in a similar way as before to test this hypothesis, as the bias differs for every trial and depends on the particular interaction

between participant and item. However, although it is likely that both the participant and item bias influence the preference on an individual trial, it seems implausible that balanced trials caused more competition than biased trials. Even though the item bias effect may be somewhat obscured by the additional participant bias effect, one would still expect to see some influence of item bias. One would expect more competition in balanced items than in biased items. However, the analyses with item bias groups did not show any evidence for this. Similarly, one would expect to see an influence of participant bias, even though this may be somewhat weakened by the item bias effect. But the analyses with participant bias groups did not show such an effect. We can therefore conclude that the results are inconsistent with models incorporating competition such as current constraint-based theories (e.g., MacDonald et al., 1994; McRae et al., 1998; Spivey & Tanenhaus, 1998; Tabor et al., 1997).

The results also pose difficulties for other models. For example, according to the garden-path theory and construal, VP attachment should always initially be adopted. In the garden-path theory this is due to minimal attachment (e.g., Frazier, 1978; 1987a), whereas it is due to a preference for arguments over adjuncts in construal (Frazier & Clifton, 1996). As a result, no reanalysis should occur in the VP attachment condition, and it should be equally easy to read as the ambiguous condition. This prediction was disconfirmed by the current results: the VP attachment condition caused more processing difficulty than the ambiguous condition. Given the results from the biased experiment in Van Gompel et al. (1999), it seems likely that the focus particle *only* forced readers to initially adopt the NP attachment analysis on some proportion of trials. The use of discourse information provided by *only* is difficult to reconcile with the garden-path theory, which claims that non-syntactic information cannot be employed in initial processing. Similarly, construal claims that non-syntactic information is not used for attachment of primary relations such as VP attachment in the current materials.

Ni et al. 's (1996) version of referential theory does claim that *only* has an influence on initial processing. Ni et al. claim that NP attachment is preferred in VP/NP attachment ambiguities containing an indefinite, plural, direct object which is preceded by *only* (e.g., only poachers). As explained in section 4.2, it is unclear what referential theory predicts when the direct object NP is singular and definite (e.g., only the poacher). One possibility which is consistent with the current results is that there is no preference for either NP or VP attachment in such cases. When *only the poacher* is processed, readers have two choices. They can either add a set of non-poachers to the discourse model and contrast the poacher with them, or add a set of poachers without rifles, and set up a contrast between the poacher with the rifle and poachers without rifles. The latter results in NP attachment, whereas the former does not. In both cases, discourse entities have to be added to the discourse model. If one assumes that adding

discourse entities that are lexically different from the poacher (the non-poachers) is no more complex than adding entities that are similar (poachers without rifles), it follows that VP and NP attachment are equally preferred. This is compatible with the data of the current experiment. Clearly, the data do not support a version of referential theory in which *only* forces the processor to always adopt the NP attachment analysis. The NP attachment condition was more difficult to read than the ambiguous condition, which suggests that the VP attachment analysis was adopted on a proportion of trials.

A model that accounts for the data has to assume that the VP and NP attachment analysis are equally preferred. Furthermore, it has to account for the fact that the ambiguous condition was easier to read than the disambiguated conditions. The unrestricted race model does both. It claims that the VP and NP attachment analysis are each adopted about half of the time, because the ambiguity is balanced. Because it assumes unrestricted use of all sources of information in the initial analysis, discourse information provided by *only* has an initial effect on processing. This is consistent with the finding that there was a strong VP attachment preference in materials without *only* in Van Gompel et al. (1999), whereas similar materials with *only* appeared to be balanced. The unrestricted race model assumes a reanalysis mechanism which causes processing disruption. Because the VP and NP attachment analysis are each adopted about half the time, it predicts that reanalysis should occur half the time when the sentence is disambiguated (either toward VP or NP attachment). In contrast, when the sentence is globally ambiguous, no reanalysis should occur, because the initial analysis, regardless of whether it is VP or NP attachment, remains plausible.

In the current study, no significant effects were obtained in first-pass reading times. The earliest effects occurred in first-pass regressions from the post-critical region. This is different from many other eyetracking studies investigating syntactic ambiguity resolution, which did find first-pass effects. Hence, one might object that the current results reflect late processing and do not reveal any insights in the early stages of processing. Therefore, the results might not distinguish between current sentence processing theories, because they are mainly concerned with such early processes. It might further be argued that such "late" processes are specific to balanced ambiguities, because readers may not immediately resolve ambiguities where there is no clear preference for one analysis over the other.

In fact, this is unlikely to be the case. As argued in section 4.7, effects that come out in so-called late measures or regions are not necessarily late effects. Instead, they may simply be weak effects. Possibly, this was the case for the current experiment. Some researchers have argued that VP/NP attachment ambiguities do not give rise to such strong garden-path effects as for example the reduced relative/main clause ambiguity, which usually gives rise to first-pass effects (e.g., Britt et al., 1992; Sturt &

Crocker, 1996). Furthermore, disambiguation by plausibility may not result in such clear reanalysis effects as syntactic disambiguation. A number of other studies using plausibility as a means to disambiguate also found the earliest effects in first-pass regressions (Pickering & Traxler, 1998; Traxler et al., 1998; Van Gompel et al., 1999).

It was also argued in section 4.7 that region length affects in what measures and regions one is likely to find effects. When a region is short, effects are more likely to show up in later measures and regions because measures based on short regions tap into earlier processes than measures based on long regions. The critical region in the current experiment consisted of only a single, and in some cases fairly short, word. Therefore, readers may have executed a forward saccade to the next word before syntactic processing had taken place (Reichle et al., 1998). This is consistent with the fact that the earliest effects were observed in the region following the critical word. Thus, the particular eye-movement pattern in the current experiment is most likely to be the result of a combination of factors, rather than the result of delayed processing.

Furthermore, the experiment on biased VP/NP attachment materials in Van Gompel et al. (1999) also showed the earliest effects in first-pass regressions out of the post-critical region, which shows that this pattern is not specific to balanced ambiguities. There is little reason to believe that the current results (and the results from the biased experiment) reflect late processing stages. Because the first-pass regressions effects were not preceded by earlier effects in first pass, it seems unlikely that the regression results reflect a second stage in processing. In any case, it is unclear how this explanation can save constraint-based theories. Because these models claim that there is only a single processing stage and therefore do not distinguish between early and late processes, it is difficult to see how it can be argued that the current results reflect late processing stages which are different from earlier processing stages.

5. Experiment 2 and 3: Delayed vs. immediate disambiguation

5.1. Introduction

Experiment 1 clearly showed that globally ambiguous sentences were easier to process than disambiguated sentences. This is difficult to reconcile with competition as a mechanism of syntactic ambiguity resolution, as it predicts that globally ambiguous sentences should be more difficult to read than disambiguated sentences. However, there might be one way to reconcile these results with competition. The disambiguation in VP/NP attachment ambiguities occurred after an ambiguous region. For example, in (1) the VP/NP attachment ambiguity arises as soon as *with* is reached, but the disambiguation occurs later, at *rifle*. Both analyses are equally possible and plausible in the ambiguous region *with the*.

1. The hunter killed only the leopard with the rifle not long after sunset.

Possibly, competition occurred in this region. Because there was no experimental manipulation before *rifle*, this kind of competition would have occurred in both the disambiguated and ambiguous conditions, so that no differences could be found in the experiment. If one further assumes that the competition process finished very quickly because one analysis was selected very rapidly, one can explain the reanalysis-like effect after the disambiguation was read. As only one analysis would be available at the disambiguation, difficulty should arise when the disambiguation forces the alternative analysis to be reactivated. Because the ambiguity was balanced, VP and NP attachment should each be selected about half of the time, and therefore difficulty should arise in both disambiguated conditions. No difficulty should occur in the ambiguous condition, because the initially selected analysis remains plausible.

Whether this explanation actually follows from current constraint-based models is uncertain. In most models, syntactic analyses are not selected, but activated. Unlikely or non-preferred analyses are not abandoned completely, but are merely deactivated. For example, in McRae et al.'s (1998) and Spivey and Tanenhaus' (1998) competition-integration model, selection does not take place. The processor moves to the next word in the sentence when one analysis has reached a criterion level of activation as a result of the competition process. This criterion level is lower the more balanced the ambiguity is. The analysis that did not reach the criterion level remains available, although it is less strongly activated. At the subsequent word, competition occurs again until the criterion level is reached. If the word contains no further

information favouring one analysis over the other, the already preferred interpretation will become more strongly activated than in the previous region. This is because the end activation of the previous word is taken as the starting activation level for the competition process for the next word. As a result of this, biases gradually become stronger when more words (which do not support one analysis over the other) are read. However, non-preferred analyses do remain activated.

Assuming that there is absolutely no difference between the activation level of the two analyses at the first word of the ambiguous region *with*, the competition process at *with* will not result in one of the analyses being more activated than the other. The criterion level will be as low as possible, and the processor will move to *the* when the activation for the two analyses is still equal. The same will happen at *the*, and the processor will reach the disambiguation with an equal activation level for the VP and NP attachment analysis. At this point, plausibility comes into play. Because it supports only one analysis in the disambiguated conditions, there will be little competition, but because both are supported in the ambiguous condition, there should be severe competition here, because both analyses remain equally activated.

This situation contrasts with a situation where there is a preference for one analysis at *with*. McRae et al.'s model predicts that this preference gradually grows stronger when more (non-biasing) words are read. Therefore, a small bias at *with* may become so strong at *rifle* that there may be little competition in the ambiguous condition. In the disambiguated conditions, the bias may result in a reanalysis-like effect when the disambiguation is inconsistent with the most activated analysis. Assuming that the VP attachment analysis is preferred on about half of the trials and the NP attachment analysis on the other half, such a reanalysis-like effect would occur about half the time in each of the disambiguated conditions. This may cause the disambiguated conditions to be more difficult than the ambiguous condition[1].

The prediction that can be derived from this is that closely balanced materials should produce more competition in the ambiguous condition and less reanalysis-like effects in the disambiguated conditions than materials that are less closely balanced. This prediction was not supported by the analyses that treated item bias as a factor in Chapter 4. They showed no evidence that competition was less severe for the more balanced items than for the less balanced items. Thus, models that assume that non-

[1] In fact, this is merely one possibility. On some trials, the disambiguated conditions are predicted to be very easy, because the disambiguation is consistent with the initial activation. Therefore, the disambiguated conditions may not be that difficult on average. It is unclear whether this averaged difficulty in the disambiguated conditions should actually be more than in the ambiguous condition. Which conditions produce most difficulty depends on how one sets the criterion level (there is no principled way to do this in the integration-competition model), and how one weighs the preference at *with* and the plausibility information.

preferred analyses remain activated rather than that they are abandoned have difficulty explaining these results.

However, one could easily modify such models so that selection of the most activated analysis takes place at each word. Such models would be able to account for the results from Experiment 1 and therefore these results might still be compatible with competition frameworks. In order to test competition more rigorously, this chapter investigates the syntactic ambiguity in (2), in which the disambiguation (by plausibility on *containing*) is immediate; that is, there is no ambiguous region. The ambiguity involves attachment of the relative clause (*containing lots of junk*) to either the first NP (NP1) or the second NP (NP2) in the main clause (e.g., Brysbaert & Mitchell, 1996; Carreiras & Clifton, 1993; Cuetos & Mitchell, 1988; Gilboy et al., 1995; Traxler et al, 1998). In (2a) NP1 is *the basement* and NPs *the tenement.*

2a. The basement of the tenement containing lots of junk was very gloomy. (globally ambiguous)

2b. The tenement of the landlord containing lots of junk was very gloomy. (NP1 attachment)

2c. The landlord of the tenement containing lots of junk was very gloomy. (NP2 attachment)

These sentences are contrasted with the ambiguity in (3), which is very similar but does contain an ambiguous region. In the example in (3) the ambiguous region is *that used to*. These sentences are similar to those used in Traxler et al. (1998), Experiment 1 and 2, which also contained an ambiguous region. However, Traxler et al.'s materials were either disambiguated by plausibility information on a noun (Experiment 1) or by the gender of a reflexive pronoun (Experiment 2). The sentences in the current experiments are disambiguated by plausibility information on the verb (e.g., *contained*).

3a. The basement of the tenement that used to contain lots of junk was very gloomy. (globally ambiguous)

3b. The tenement of the landlord that used to contain lots of junk was very gloomy. (NP1 attachment)

3c. The landlord of the tenement that used to contain lots of junk was very gloomy. (NP2 attachment)

If competition is a very short-lived process and selection of one analysis takes place very rapidly, (2) and (3) should be processed differently. In (3) no competition should occur at or following the disambiguation *contain* in the ambiguous condition

(3a) because one analysis has been selected in the region *that used to* and the other one has been dropped. Reanalysis should occur in the disambiguated conditions (3b) and (3c), because the initially selected analysis will be inconsistent with the disambiguation on some proportion of trials. Given the similarity of the current materials to those in Traxler et al.'s Experiment 1 and 2, we can expect that the ambiguity is relatively balanced, as was the case in Traxler et al. (respectively 68% and 70% NP2 preference in the off-line pre-tests). Therefore, reanalysis should occur about equally often in the NP1 and NP2 attachment conditions, and there should be little or no difference between the NP1 and NP2 attachment conditions (as in Traxler et al.'s experiments). In contrast, the early competition hypothesis predicts that such a reanalysis effect cannot occur in (2), because there is no ambiguous region where the processor can drop one analysis. Both analyses should be activated at the disambiguating region (*containing*). Because plausibility supports only one analysis in the disambiguated conditions, little competition is predicted for these conditions, whereas in the ambiguous condition competition should be strong, because plausibility does not favour either analysis.

The predictions of reanalysis models are very different. They are the same as for Traxler et al.'s Experiment 1 and 2. The garden-path model predicts initial NP2 attachment due to late closure and therefore the NP1 attachment condition should be more difficult to read than both the globally ambiguous and the NP2 attachment condition. There should be no difference between immediate and delayed disambiguation. However, the predictions of these models were disconfirmed by Traxler et al.'s experiments, because the NP2 attachment condition appeared to be more difficult to read than the ambiguous condition, and it did not differ from the NP1 attachment condition.

Construal (Frazier & Clifton, 1996; Gilboy et al., 1995) predicts that there should be no clear preference for NP1 or NP2 attachment, because the relative clause is associated with the whole NP complex. This was confirmed by Traxler et al.'s experiments. Construal does not make predictions as to how the ambiguous condition is processed. Traxler et al. observed that the ambiguous condition was faster to read than the disambiguated conditions in materials with a delayed disambiguation similar to (3). Although construal was consistent with Traxler et al.'s experiments, it made incorrect predictions for the VP/NP attachment ambiguity of Chapter 4, which involved attachment of a primary relation.

The unrestricted race model has been able to account for all results so far. It correctly predicted the results for both Traxler et al.'s experiments and the results on VP/NP ambiguities. Its predictions for (2) and (3) are identical to those for Traxler et al.'s Experiment 1 and 2, which also involved relative clause attachment to an NP complex containing the preposition *of*. Assuming that the current materials are close to

balanced (as in Traxler et al.), it is predicted that NP1 and NP2 attachment are each adopted about half the time. Reanalysis should occur about half the time when the sentence is disambiguated, because the initial analysis will be incorrect half the time. In contrast, no reanalysis should occur in the ambiguous conditions, because the initial analysis is always plausible. The unrestricted race model predicts the same pattern for (2) and (3). It does not matter whether the disambiguation is delayed or immediate. This is due to the fact that in the unrestricted race model, the plausibility of a syntactic structure can only be evaluated after it has been syntactically analysed (see section 3.5). For example, in (2b) the processor first has to attach the relative clause to NP2 before it can determine that this analysis is implausible. This results in reanalysis. The processor cannot use plausibility information to avoid an initial attachment decision. Therefore, the unrestricted race model predicts that reanalysis should occur both when the disambiguation is delayed and when it is immediate.

Given the similarity of the current experiments to those in Traxler et al., it seems likely that effects will be observed in similar measures and regions. As in Traxler et al., the critical region in the current experiments (the verb in the relative clause: *containing/contain*) was short. Traxler et al. obtained effects in first-pass regressions from the post-critical region (Experiment 2) and in total reading times (Experiment 1 and 2). In Experiment 1 of this thesis (Chapter 4) and the biased VP/NP attachment experiment in Van Gompel et al. (1999), the earliest effects were observed in first-pass regression and regression-path times for the post-critical region. Given that these experiments also disambiguated by plausibility and used a short critical region, we may expect a similar pattern in the current experiment.

5.2. Experiment 2: Delayed disambiguation

Experiment 2 investigated relative clause ambiguities such as (3), which have delayed disambiguation. The results are compared with those from Experiment 3, which used very similar materials with immediate disambiguation.

Plausibility pre-tests were conducted to check the plausibility manipulation in the materials. In addition, a pre-test assessing the acceptability of the NP complex was carried out. As in Experiment 1, off-line preference and completion pre-tests were conducted to assess the bias of the materials, which is important to derive predictions from some of the models.

5.3. Plausibility pre-tests

Two plausibility pre-tests ensured that the plausibility of the materials was carefully controlled. The plausibility manipulation was localised to a single word, the verb in the relative clause (e.g., *contain* in 2).

5.4. Attachment plausibility pre-test

The first plausibility study tested the plausibility of the relative clause as a modifier of either NP1 or NP2.

Participants. Eighteen participants from the same population as the pre-tests in Chapter 4 took part. None of the participants in this chapter took part in more than one pre-test or eyetracking experiment, and nobody had participated in Traxler et al.'s (1998) experiments.

Materials. We constructed materials like (3), repeated here as (4) (see Appendix 2):

4a. The basement of the tenement that used to contain lots of junk was very gloomy. (globally ambiguous)

4b. The tenement of the landlord that used to contain lots of junk was very gloomy. (NP1 attachment)

4c. The landlord of the tenement that used to contain lots of junk was very gloomy. (NP2 attachment)

The NP1 attachment condition was semantically disambiguated towards the NP1 attachment analysis; the NP2 attachment condition was semantically disambiguated towards the NP2 attachment analysis; the Ambiguous condition was semantically ambiguous between the two analyses. All materials consisted of a complex NP (*the* + N + *of* + *the* + N), which was the subject of the main clause, a relative clause, and the rest of the main clause (usually a form of *to be* followed by a predicate). The relative clause consisted of *that* + *has been/had been/used to* + main verb + NP/PP. The complex NP was always the subject of the relative clause. The NPs in the NP complex sometimes differed between conditions of an item. This was necessary in order for the plausibility manipulation to work.

The attachment plausibility pre-test tested 50 sets of questions such as (5), which were derived from sentences such as (4).

5a. Can a basement of a tenement contain something?
5b. Can a tenement of a landlord contain something?
5c. Can a landlord of a tenement contain something?
5d. Can a tenement contain something?
5e. Can a landlord contain something?
5f. Can a tenement contain something?

These questions unambiguously asked the participants to evaluate the plausibility of the NP1 and NP2 attachment analyses for item sets like (4). They always had the form *Can* + subject NP + verb (+ NP). For the NP1 attachment analysis paraphrases (5a-c), the whole NP complex was taken as the subject NP. I will refer to these conditions as the *NP1-relative clause paraphrase conditions.* For the NP2 attachment analysis paraphrases (5d-e), NP2 was taken as the subject. I will refer to them as the *NP2-relative clause paraphrase conditions.* The verb was always the infinitive form of the main verb in the relative clauses of the experimental item sets. If the verb obligatorily subcategorised for a direct object NP, the verb was followed by *something* or *someone*, depending on what was appropriate. Conditions (5c) and (5e) were derived from analyses that were constructed to have implausible interpretations. All other conditions were derived from analyses that were constructed to be plausible. Notice that condition (5d) and (5f) are identical for the item in (5). This was not the case for all items, because the NPs sometimes differed between conditions.

Design. Six lists of items in a single random order were constructed, with each list containing all 300 questions. The conditions were rotated over the items so that each participant saw each condition of every item once. Care was taken that versions of the same item never appeared close to each other.

Procedure. The participants rated how plausible the situation described by the questions was, on a 7-point scale, with 7 indicating very plausible, and 1 very implausible.

Table 5.1

Experiment 2: Item characteristics

	condition		
	ambiguous	NP1 attachment	NP2 attachment
Attachment plausibility pre-test:			
Plausibility NP1 attachment paraphrase	6.3	6.4	1.8
Plausibility NP2 attachment paraphrase	6.5	1.9	6.5
Main clause plausibility	5.0	5.0	4.8
Acceptability of NP complex	5.2	4.8	5.0
Log frequency of nouns:			
noun1	2.3	2.4	2.2
noun2	2.3	2.5	2.4
Length of NP complex in characters:	13.6	13.5	13.4
Off-line preference task:			
Percentage of NP2 attachment	51.1	5.0	89.4
Completion task:			
Percentage of NP2 completions	54.0	55.6	59.4

Note. Log frequency is based on the number of occurrences in the Celex database (17.9 million words).

Results. Thirty item sets were selected on the basis of this plausibility pre-test, the main clause plausibility pre-test, and the NP complex acceptability pre-test. Table 5.1 shows the mean plausibility rating for these materials. ANOVAs with the factors Paraphrase type (NP1-relative clause vs. NP2-relative clause) and Condition (Ambiguous vs. NP1 attachment vs. NP2 attachment) were conducted. Both factors were treated as within subjects and items. The analyses revealed a significant interaction between these factors: $F1\ (2, 34) = 1009$, $p < .001$, $MSe = .063$; $F2\ (2, 58) = 301$, $p < .001$, $MSe = .351$. It was crucial for the purposes of the experiment that (5c) and (5e) were implausible, and that the other conditions were not. This was indeed the case, as shown in Table 5.1. Planned comparisons showed that the two paraphrase conditions for the NP1 and NP2 attachment condition (4b) and (4c) differed significantly (NP1 attachment: $F1\ (1, 34) = 185$; $p < .001$; $F2\ (1, 58) = 308$, $p < .001$; NP2 attachment: $F1\ (1, 34) = 199$; $p < .001$; $F2\ (1, 58) = 331$, $p < .001$), whereas they did not for the Ambiguous condition (4a) ($F1\ (1, 34) = 2.29$; $p = .14$; $F2\ (1, 58) = .35$, $p = .55$).

5.5. Main clause plausibility pre-test

This plausibility study tested the plausibility of the complex NP as the subject of the main clause. For example, for (4a) it tested whether *the basement of the tenement was very gloomy* was plausible or not.

Participants. Eighteen participants from the same population as before took part.

Materials. The study tested 50 item sets like (6a-c), which were derived from item sets such as (4a-c) respectively.

6a. The basement of the tenement was very gloomy.
6b. The tenement of the landlord was very gloomy.
6c. The landlord of the tenement was very gloomy.

The sentences were identical to the experimental items, except that the relative clause was omitted.

Design. Three lists of items in a single random order were constructed, with each list containing all 150 items. The conditions were rotated over the items so that each subject saw each condition of every item once. Care was taken that versions of the same item never appeared close to each other.

Procedure. The participants rated how plausible the situation described by the sentences was on a 7-point scale.

Results. Table 5.1 shows the mean rating for each condition for the 30 selected items. ANOVAs with the factor Condition (Ambiguous vs. NP1 attachment vs. NP2 attachment) as a within subjects and items variable were conducted. No significant differences were observed: F1 (2, 34) = 1.31, p = .28, MSe = .125; F2 (2, 58) = .65, p = .52, MSe = .408.

5.6. NP complex acceptability pre-test

We attempted to match the acceptability of the NP complex as closely as possible, because this may have an influence on reading times for this region and perhaps regions following it.

Participants. Eighteen participants from the same population as before took part.

Materials. The acceptability study tested 50 sentence fragments like (7a-c), again derived from item sets such as (4a-c) respectively.

7a. The basement of the tenement ...
7b. The tenement of the landlord ...
7c. The landlord of the tenement ...

These fragments consisted of the NP complex of the experimental materials followed by dots.

Design. The design was identical to the design for the main clause plausibility pre-test.

Procedure. The participants rated how natural the fragments were as the beginning of a sentence on a 7-point scale.

Results. Table 5.1 shows the mean acceptability ratings for each condition. ANOVAs with the factor Condition (Ambiguous vs. NP1 attachment vs. NP2 attachment) as a within subjects and items variable were conducted on the selected item sets. These analyses showed a significant effect of condition by subjects, but not by items: F1 (2,

34) = 8.43, p = .001, MSe = .078; F2 (2, 58) = .94, p = .40, MSe = 1.16. Although there may be some differences in acceptability of the NP complex, Table 5.1 shows that they were extremely small. It seems unlikely that such differences affect regions following the NP complex.

The log lemma frequency for the nouns in the NP complex was obtained with the Celex database, as shown in Table 5.1. An ANOVA with the factors Noun (noun 1 vs. noun 2) and Condition (Ambiguous vs. NP1 attachment vs. NP2 attachment) as within factors were conducted. Neither factor revealed a significant effect, nor was there an interaction between them (Condition: F (2, 58) = 2.12, p = .13, MSe = .184; Noun: F (1, 29) = 2.79, p = .11, MSe = .180; Interaction: F (2, 58) = 0.91, p = .91, MSe = .652).

The length of the NP complex was also controlled for (see Table 5.1). ANOVAs with Condition as a within items factor revealed no differences: F (2, 58) = .23, p = .79, MSe = .741.

5.7. Off-line preference pre-test

Participants. Eighteen participants from the same population as before took part.

Materials. The materials were 30 item sets such as (4), which were selected on the basis of the plausibility and the acceptability pre-tests. These materials were identical to those in the eyetracking experiment. Each item was followed by two paraphrases that unambiguously described either the NP1 or NP2 attachment analysis. For example, the paraphrase sentences for (4a) were as in (8).

8a. The basement used to contain lots of junk.
8b. The tenement used to contain lots of junk.

Design. Order of the paraphrase conditions was treated as a factor. Therefore, six lists of items were constructed (corresponding to the three lists in the eyetracking experiment), with one version of each item appearing in each list. The items occurred in a single random order in the lists. Three participants were assigned to each of the lists.

Procedure. The procedure was the same as for the off-line preference pre-test in Chapter 4.

Results. Table 5.1 shows the percentage of NP2 attachment responses for each condition. The NP1 and NP2 responses always added up to 100%. ANOVAs with the factor Condition as a within subjects and items variable were conducted. They revealed a significant effect of condition: F1 (2, 34) = 103, $p < .001$, MSe = 313; F2 (2, 58) = 185, $p < .001$, MSe = 290. Planned comparisons showed that the ambiguous condition differed from both the NP1 attachment condition (F1 (2, 34) = 61.1, $p < .001$; F2 (2, 58) = 110, $p < .001$) and the NP2 attachment condition (F2 (2, 34) = 42.3, $p < .001$; F1 (2, 58) = 76.0, $p < .001$).

Crucially, the percentage of NP2 attachment responses (51%) in the ambiguous condition did not differ from chance, indicating that the materials were balanced (F1 (1, 17) = .038, p = .85, MSe = 293; F2 (1, 29) = .063, p = .80, MSe = 296). Comparison with Traxler et al.'s Experiment 1 and 2, which also tested relative clause attachments to NP complexes with the preposition *of*, shows that the current materials are more closely balanced: the percentages of NP2 responses in Traxler et al.'s off-line pre-tests were 68% and 70% respectively. It is not entirely clear why there is such a difference in bias. They may simply be the result of random differences in bias between materials with NP *of* NP complexes.

Participants adopted the implausible interpretation on only a small percentage of trials: 5% in the NP1 attachment condition and 11% in the NP2 attachment condition.

5.8. Completion pre-test

Participants. Eighteen participants from the same population as before took part.

Materials. The same materials as in the off-line preference study were tested, cut off before the disambiguating verb in the relative clause (e.g., *The basement of the tenement that used to ...*). In addition, 68 fillers were included, cut off at various point so that they appeared similar to the experimental items. These filler fragments were identical to the fillers in the eyetracking experiment and appeared in the same order.

Design. Three lists corresponding to the lists in the eyetracking experiment were created, with one version of each item in each list. Six participants saw each list.

Procedure. The procedure was similar to the off-line preference pre-test in Chapter 4. Participants were first given a booklet with sentence fragments that they had to complete. Subsequently, they received a second booklet containing the experimental materials and two paraphrase sentences such as in (8). They were told to tick the

paraphrase that corresponded to the meaning that they intended when they completed the sentence.

Results. The mean percentage of NP2 attachment responses for each condition is presented in Table 5.1. On 0.9% of the trials participants indicated that they did not know which paraphrase was correct. These trials were excluded from the analyses. ANOVAs with the factor Condition as a within subjects and within items variable revealed no significant effects: F1 (2, 34) = .584, p = .56, MSe = .239; F2 (2, 58) = .188, p = .83, MSe = 1033, indicating that the percentage of NP2 completions was similar across conditions. The overall number of NP2 completions did not differ from chance on the subject analysis, but it did on the item analysis: F1 (1, 17) = 1.49, p = .24, MSe = 244; F2 (1, 29) = 11.75, p = .002, MSe = 55. The mean percentage of NP2 completions (55%) is within the range that is normally considered as balanced (e.g., Rayner & Duffy, 1986), and it was therefore concluded that there was no clear preference for either NP1 or NP2 attachment.

5.9. Eyetracking Experiment 2

Method

Participants. Forty participants from the same population as the pre-tests took part. All of them had normal vision. Some of them had taken part in other eyetracking experiments, but none of them had taken part in the Experiments of Traxler et al. (1998). Three participants were excluded because they had a very high percentage of tracker losses, and one because she answered more than 25% of the questions incorrectly. The analyses were conducted on the remaining 36 participants.

Materials. The experimental materials were the same as in the off-line preference pre-test. The fillers in the completion study were finished off to make complete sentences and used in the eyetracking experiment. These sentences contained a variety of structures unrelated to the experimental materials.

Design. Three lists of items, consisting of ten items from each condition were constructed. Exactly one version of each item appeared in each list. Twelve participants were assigned to each list. Each list also contained 68 fillers unrelated to the purpose of the experiment. Forty-eight items were followed by a question in order to encourage comprehension. Half of the questions required a "yes" response and half a

"no" response. Participants received feedback when they answered a question incorrectly. The experimental items and fillers were placed in a single random order, with six fillers preceding the first experimental sentence and three fillers following a break in the experiment.

Procedure. The procedure was the same as in eyetracking Experiment 1 (Chapter 4). All materials were presented on two lines. In the experimental materials, the line break was always before the last or the penultimate word in the sentence.

Analyses. On average, the participants made 10% errors on the comprehension questions. Prior to all analyses, trials with major tracker losses and trials with fewer than four fixations were excluded. This eliminated less than 1% of the trials. Less than 1% of the data was removed as a result of two consecutive regions having zero reading times.

Reading times that were more than 2 standard deviations from both the subject and item mean of the region were also excluded. This never removed more than 2% of the data on any of the analyses.

The materials were divided into five regions, indicated by slashes below:

9. The basement of the tenement/ that used to/ contain/ lots of junk/ was very gloomy./

The regions corresponded to (1) the NP complex, (2) the ambiguous region, (3) the verb in the relative clause (the critical region), (4) the rest of the relative clause (the post-critical region), and (5) the final region, which contained the remainder of the main clause.

Results

Table 5.2 presents the mean reading times and percentage regressions by condition and region. For each eyetracking measure for each region, one-way ANOVAs with three levels of Condition (Ambiguous vs. NP1 attachment vs. NP2 attachment) were conducted. Condition was treated as a within subjects and items factor. Table 5.3 presents the results of these analyses. Planned comparisons between levels are reported below.

Table 5.2

Experiment 2: Means

	region				
	1	2	3	4	5
First-pass reading times:					
ambiguous	994	543	303	571	872
NP2 attachment	949	564	311	565	907
NP1 attachment	958	533	303	572	946
First-pass regressions out:					
ambiguous		7.3	17.6	8.4	53.1
NP2 attachment		5.9	21.1	16.2	61.3
NP1 attachment		7.0	19.3	15.4	57.0
Regression-path times:					
ambiguous	994	590	373	638	1680
NP2 attachment	949	614	419	715	1907
NP1 attachment	958	582	392	728	1891
Total times:					
ambiguous	1332	742	394	732	1165
NP2 attachment	1413	837	448	810	1225
NP1 attachment	1345	797	451	782	1248

Notes. The regions were defined as follows (delimited by slashes):
The basement of the tenement/ that used to/ contain/ lots of junk/ was very gloomy./

First pass, regression path, and total times are reported in ms, first pass regressions as the percentage of saccades leaving the region to the left after a first pass fixation.

Table 5.3

Experiment 2: ANOVA results by region for effect of condition

	F1 (2, 70)	MSe	F2 (2, 58)	MSe
Region 1:				
First pass reading times	3.36*	5794	< 1	
Regression path times	3.36*	5794	< 1	
Total reading times	2.53	26220	< 1	
Region 2:				
First pass reading times	2.11	4897	1.70	3839
First pass regressions out	< 1		< 1	
Regression path times	1.45	7491	< 1	
Total reading times	6.45**	12610	5.66**	11956
Region 3:				
First pass reading times	< 1		< 1	
First pass regressions out	< 1		< 1	
Regression path times	1.68	10444	3.16*	4822
Total reading times	5.55**	6939	5.41**	5148
Region 4:				
First pass reading times	< 1		< 1	
First pass regressions out	4.72*	147	4.31*	130
Regression path times	4.11*	19687	4.44*	16656
Total reading times	3.57*	15359	4.49*	10650
Region 5:				
First pass reading times	2.10	21975	2.69	15712
First pass regressions out	3.10	188	2.84	178
Regression path times	6.48**	86324	3.72*	128353
Total reading times	4.07*	16478	2.60	21664

Note. * is significant at the .05 level, ** is significant at the .01 level.

No effects of first-pass reading times were significant by items and subjects. However, the effect of condition reached significance by subjects in region 1, the complex NP ($p = .041$). The means suggest that ambiguous region may be somewhat more difficult to read than the disambiguated conditions. Although we attempted to control the complex NP across conditions, these reading times suggest that there may still have been some small differences in complexity.

First-pass regressions showed an effect of condition in region 4, the post-critical region. Planned comparisons revealed that the ambiguous condition produced fewer first-pass regressions than the NP1 attachment condition ($F1\ (1, 70) = 6.55$, $p = .013$; $F2\ (1, 58) = 5.83$, $p = .019$) and the NP2 attachment condition ($F1\ (1, 70) = 7.58$, $p = .008$; $F2\ (1, 58) = 7.04$, $p = .010$). The two disambiguated conditions did not differ: $F1\ (1, 70) = .04$, $p = .85$; $F2\ (1, 58) = .06$, $p = .81$. A marginally significant effect was observed in the final region ($p = .057$ by subjects; $p = .066$ by items). Planned comparisons showed that there were fewer regressions in the ambiguous condition than in the NP2 attachment condition: $F1\ (1, 70) = 6.20$, $p = .015$; $F2\ (1, 70) = 5.68$, $p = .020$. The difference between the ambiguous condition and the NP1 attachment condition ($F1\ (1, 70) = 1.59$, $p = .21$; $F2\ (1, 58) = 1.33$, $p = .25$) and between the NP1 and NP2 attachment conditions ($F1\ (1, 70) = 1.52$, $p = .22$; $F2\ (1, 58) = 1.52$, $p = .22$) did not reach significance.

Regression-path times showed a condition effect that was significant by items but not by subjects ($p = .19$) in region 3, the critical region. The means suggest that the ambiguous condition may have been faster to read than the NP2 attachment condition, with the NP1 attachment condition showing reading times in between. Regression-path times showed the same pattern as first-pass regression in the post-critical region. The ambiguous condition was read faster than the NP1 attachment condition ($F1\ (1, 70) = 6.69$, $p = .011$, $F2\ (1, 58) = 7.64$) and the NP2 attachment condition ($F1\ (1, 70) = 5.38$, $p = .023$; $F2\ (1, 58) = 5.50$, $p = .022$). The NP1 and NP2 attachment condition did not differ: $F\ (1, 70) = .09$, $p = .76$, $F2\ (1, 58) = .18$, $p = .68$. A similar pattern was observed in region 5, the next region. Ambiguous was faster than NP1 attachment ($F1\ (1, 70) = 9.20$, $p = .003$; $F2\ (1, 58) = 5.31$, $p = .025$) and NP2 attachment ($F1\ (1, 70) = 10.22$, $p = .002$; $F\ (1, 58) = 5.83$, $p = .019$) and the disambiguated conditions did not differ ($F1\ (1, 70) = .03$, $p = .87$; $F\ (1, 58) = .01$, $p = .91$.

Total reading times also showed a pattern that indicated that the ambiguous condition was faster to read than the disambiguated conditions. The ambiguous condition differed from the NP1 attachment condition in region 2 ($F1\ (1, 70) = 4.10$, $p = .047$; $F2\ (1, 58) = 3.96$, $p = .051$), region 3 ($F1\ (1, 70) = 9.47$, $p = .003$; $F2\ (1, 58) = 8.66$, $p = .005$), region 4 ($F1\ (1, 70) = 3.04$, $p = .086$; $F2\ (1, 58) = 4.04$, $p = .049$) and region 5 ($F1\ (1, 70) = 7.81$, $p = .007$; $F\ (1, 58) = 4.91$, $p = .031$) and from the NP2

attachment condition in region 2 (F1 (1, 70) = 12.83, $p < .001$; F2 (1, 58) = 11.19, $p = .001$), region 3 (F1 (1, 70) = 6.98, $p = .010$; F2 (1, 58) = 7.52, $p = .008$) and region 4 (F1 (1, 70) = 6.88, $p = .011$; F2 (1, 58) = 8.58, $p = .005$). The difference between the two disambiguated conditions did not reach significance in any of the analyses (all ps > .12).

5.10. Discussion

The results from this experiment, which used delayed disambiguation, showed that the ambiguous condition produced less processing difficulty than both disambiguated conditions. As in Experiment 1, the earliest effects occurred in first-pass regressions and regression-path times for the post-critical region. As argued in section 4.9, this pattern is most likely caused by a combination of factors such as region length, type of disambiguation and strength of the garden-path effect and is probably not due to late processing. Significant effects also occurred for regression-path times in the final region, and in total reading times in the regions surrounding the critical verb.

These results resemble the pattern of results from Traxler et al. (1998), Experiment 1 and 2. However, they are more robust. In Traxler et al.'s Experiment 1, which used delayed disambiguation on a noun, the ambiguous condition was read faster than both disambiguated conditions in total reading times on the critical region, but not in earlier measures. In Experiment 2, in which gender provided a late disambiguation, the ambiguous condition produced significantly fewer first-pass regressions than the NP1 attachment condition in the post-critical region, but the difference between ambiguous and NP2 attachment was not significant. A similar pattern occurred in total reading times on the critical region. Traxler et al. did not report regression-path times. The current relative clause attachment experiment is the first to show that the ambiguous condition produces less difficulty than both disambiguated conditions in relatively early measures of processing.

The data support the unrestricted race model, which predicts that the ambiguous condition should be easy to read, because reanalysis does not take place. In contrast, reanalysis should occur half the time in both disambiguated conditions. Because the ambiguity is balanced, readers are predicted to adopt each analysis about half the time. When the sentence is subsequently disambiguated, it will be inconsistent with this initial analysis on about half of the trials, and reanalysis should occur on those trials.

The results are also consistent with construal theory, which predicts little or no difference between NP1 and NP2 attachment, because the relative clause is associated with the whole NP complex. The garden-path theory cannot explain the current results, because it predicts that NP2 attachment is always adopted initially, and therefore there should be no difference between the ambiguous and NP2 attachment condition.

The results are also incompatible with constraint-based theories, which predict that the ambiguous condition should be more difficult to read than the disambiguated conditions because competition occurs in the former, but not in the latter. However, competition models may be saved if one assumes that competition occurred very early on in the ambiguous region, and that one analysis was selected very rapidly. Hence, when readers reached the disambiguation, only one analysis was available. If this analysis was inconsistent with the disambiguation, the alternative, abandoned analysis had to be reactivated, causing processing disruption. Because the ambiguity is balanced, this happened about half the time in the disambiguated conditions. In contrast, the ambiguous condition was easy, because the initially selected analysis remained plausible. This version of competition models was tested in Experiment 3, which employed immediate disambiguation.

5.11. Experiment 3: Immediate disambiguation

Experiment 3 investigated the same materials as in Experiment 2, but the ambiguous region was removed. Because the basic characteristics the materials were the same, no plausibility and acceptability pre-tests were conducted. However, in order to assess whether the materials for Experiment 3 were as balanced as those in Experiment 2, an off-line preference pre-test was conducted. It was not possible to carry out a completion pre-test for the current materials. Because there was no ambiguous region, nothing would have indicated to the participants that the sentences had to be continued with a relative clause, as shown by 10.

10. The basement of the tenement ...

Therefore, participants would probably have completed the sentences with a relative clause on only very few trials. Analyses would have to be conducted on too few observations to produce reliable results.

5.12. Off-line preference pre-test

Participants, design and procedure. They were identical to those for the off-line preference pre-test for Experiment 2.

Materials. The materials were the same as for Experiment 2, but did not have an ambiguous region preceding the verb in the relative clause. The infinitival verb form in the relative clause was replaced by either a present or past participle form. An example item is presented in (11):

11a. The basement of the tenement containing lots of junk was very gloomy. (globally ambiguous)
11b. The tenement of the landlord containing lots of junk was very gloomy. (NP1 attachment)
11c. The landlord of the tenement containing lots of junk was very gloomy. (NP2 attachment)

In the off-line preference pre-test, each item was followed by two paraphrases that unambiguously described either the NP1 or the NP2 attachment analysis. For example, the paraphrase sentences for (11a) were as in (12).

12a. The basement contained lots of junk.
12b. The tenement contained lots of junk.

Results. Table 5.4 shows the percentage of NP2 responses for each condition. The NP1 and NP2 responses always added up to 100%. ANOVAs with the factor Condition as a within subjects and items variable were conducted. They revealed a significant effect of condition: $F1\ (2, 34) = 69.6$, $p < .001$, $MSe = 354$; $F2\ (2, 58) = 92.9$, $p < .001$, $MSe = 442$. Planned comparisons showed that the ambiguous condition differed from both the NP1 attachment condition ($F1\ (2, 34) = 41.8$, $p < .001$; $F2\ (2, 58) = 55.8$, $p < .001$) and the NP2 attachment condition ($F2\ (2, 34) = 28.2$, $p < .001$; $F1\ (2, 58) = 37.7$, $p < .001$).

As in the previous experiment, the percentage of NP2 attachment responses (49%) in the ambiguous condition did not differ from chance: $F1\ (1, 17) = .035$, $p = .85$, $MSe = 317$; $F2\ (1, 29) = .049$, $p = .82$, $MSe = 382$. This indicated that the materials were balanced. Participants adopted the implausible interpretation on 8% of the trials in the NP1 attachment condition and on 18% of the trials in the NP2 attachment condition.

Table 5.4

Experiment 3: Item characteristics

	condition		
	ambiguous	NP1 attachment	NP2 attachment
Off-line preference task:			
Percentage of NP2 attachment	48.9	8.3	82.2

Combined analyses for the off-line pre-tests of Experiment 2 and 3 were conducted with the factor Condition as a within subject and item variable, and the factor Experiment as a between subjects and within items variable. There was no main effect of Experiment ($F1$ (1, 34) = .269, p = .61, MSe = 417; $F2$ (1, 29) = .849, p = .36, MSe = 222), nor an interaction between Experiment and Condition ($F1$ (1, 34) = .752, p = .48, MSe = 334; $F2$ (1, 29) = 1.10, p = .34, MSe = 381), indicating that the off-line pre-tests for both experiments produced very similar results.

5.13. Eyetracking Experiment 3

Method

Participants. Forty-one participants from the same population as before took part. All of them had normal vision. Four participants were excluded because they had a very high percentage of tracker losses, and one because she answered more than 25% of the questions incorrectly. The analyses were conducted on the remaining 36 participants.

Materials. The experimental materials were the same as in the off-line preference pre-test. The fillers were similar to those in Experiment 2, but shortened to make them look similar to the experimental materials.

Design. Identical to the design in Experiment 2.

Procedure. The procedure was the same as in Experiment 2, but all materials were presented on one line.

Analyses. On average, the participants made 12% errors on the comprehension questions. Prior to all analyses, trials with major tracker losses and trials with fewer than four fixations were excluded. This eliminated less than 1% of the trials. Less than 1% of the data was removed as a result of two consecutive regions having zero reading times.

Reading times that were more than 2 standard deviations from both the subject and item mean of the region were also excluded. This never removed more than 2% of the data on any of the analyses.

The materials were divided into four regions, indicated by slashes below:

13. The basement of the tenement/ containing/ lots of junk/ was very gloomy./

The regions corresponded to (1) the NP complex, (2) the verb in the relative clause (the critical region), (3) the rest of the relative clause (the post-critical region), and (4) the final region, which contained the remainder of the main clause.

Results

Table 5.5 presents the mean reading times and percentage regressions by condition and region. For each eyetracking measure for each region, one-way ANOVAs with three levels of Condition (Ambiguous vs. NP1 attachment vs. NP2 attachment) were conducted. Condition was treated as a within subjects and items factor. Table 5.6 presents the results of these analyses. Planned comparisons between levels are reported below.

The only significant first-pass times were observed in region 4, the final region. Planned comparisons showed that the NP1 attachment condition took longer to read than the NP2 attachment analysis: F1 (1, 70) = 6.41, p = .014; F2 (1, 58) = 8.58, p = .005. There was no difference between Ambiguous and NP2 attachment (F1 (1, 70) = .88, p = .35; F2 (1, 58) = .004, p = .95) or between NP1 and NP2 attachment (F1 (1, 70) = 2.53, p = .12; F2 (1, 58) = 3.50, p = .067).

First-pass regressions revealed an effect in region 2, the critical region, although this was only marginally significant in the items analysis (p = .055). The ambiguous condition produced more regressions than the NP2 attachment condition: F1 (1, 70) = 6.87, p = .011; F2 (1, 58) = 6.01, p = .017). No differences were observed between Ambiguous and NP1 attachment (F1 (1, 70) = 2.46, p = .12; F2 (1, 58) = 2.25, p = .14) and NP1 and NP2 attachment (F1 (1, 70) = 1.11, p = .30; F2 (1, 58) = .90, p = .35). In the next region, the post-critical region, there was some suggestion that there were fewer regressions in the ambiguous condition than in the disambiguated conditions, but this effect was not significant by items (p = .13) and only marginally so by subjects (p = .054).

Regression-path times also showed an effect in the critical region (region 2), although this was again only marginally significant by items. Reading times for ambiguous condition in this region were longer than for the NP1 attachment condition (F1 (1, 70) = 6.14, p = .016; but F2 (1, 58) = 3.39, p = .071) and the NP2 attachment condition (F1 (1, 70) = 8.90, p = .004; F2 (1, 58) = 5.07, p = .028), while the disambiguated conditions did not differ (F1 (1, 70) = .26, p = .62; F2 (1, 58) = .06, p = .81). The effect was reversed in the next region, the post-critical region. Ambiguous was faster than NP1 attachment (F1 (1, 70) = 7.15, p = .009; F2 (1, 58) = 5.92, p = .018) and NP2 attachment (F1 (1, 70) = 6.26, p = .015; F2 (1, 58) = 4.78, p = .033). NP1 and NP2 attachment did not differ: (F1 (1, 70) = .03, p = .86; F2 (1, 58) = .06, p = .81).

Table 5.5

Experiment 3: Means

	region			
	1	2	3	4
First-pass reading times:				
ambiguous	965	343	556	916
NP2 attachment	954	348	558	889
NP1 attachment	954	341	560	960
First-pass regressions out:				
ambiguous		13.3	14.0	47.5
NP2 attachment		7.3	20.0	51.0
NP1 attachment		10.0	18.7	48.6
Regression-path times:				
ambiguous	965	428	662	1733
NP2 attachment	954	376	747	1804
NP1 attachment	954	387	754	1705
Total times:				
ambiguous	1363	501	799	1141
NP2 attachment	1353	556	851	1124
NP1 attachment	1332	523	851	1154

Notes. The regions were defined as follows (delimited by slashes):
The basement of the tenement/ containing/ lots of junk/ was very gloomy./

First pass, regression path, and total times are reported in ms, first pass regressions as the percentage of saccades leaving the region to the left after a first pass fixation.

Table 5.6

Experiment 3: ANOVA results by region for effect of condition

	F1 (2, 70)	MSe	F2 (2, 58)	MSe
Region 1:				
First pass reading times	< 1		< 1	
Regression path times	< 1		< 1	
Total reading times	< 1		< 1	
Region 2:				
First pass reading times	< 1		< 1	
First pass regressions out	3.48*	89.4	3.05	92.8
Regression path times	5.10**	6598	2.88	8810
Total reading times	2.79	10554	1.99	11958
Region 3:				
First pass reading times	< 1		< 1	
First pass regressions out	3.05	119	2.13	142
Regression path times	4.48*	20159	3.59*	21337
Total reading times	1.87	16764	1.44	17421
Region 4:				
First pass reading times	3.27*	12985	4.40*	9121
First pass regressions out	< 1		< 1	
Regression path times	1.45	80629	< 1	
Total reading times	< 1		< 1	

Note. * is significant at the .05 level, ** is significant at the .01 level.

Table 5.7

Experiment 2 and 3: ANOVA results for combined analyses

	Experiment				Condition				Experiment x Condition			
	F1 (1, 70)	MSe	F2 (2, 58)	MSe	F1 (2, 140)	MSe	F2 (2, 58)	MSe	F1 (2, 140)	MSe	F2 (2, 58)	MSe
Region 1:												
First pass reading times	< 1		< 1		2.43	6673	< 1		< 1		< 1	
Regression path times	< 1		< 1		2.43	6673	< 1		< 1		< 1	
Total reading times	< 1		< 1		1.82	24945	< 1		1.31	24945	1.11	28436
Region 2:												
First pass reading times	8.65**	8685	42.0**	1507	< 1		< 1		< 1		< 1	
First pass regressions out	6.13*	729	28.8**	127	< 1		< 1		2.79	103	2.79	131
Regression path times	< 1		< 1		< 1		< 1		5.49**	8521	5.00**	7621
Total reading times	8.60**	58943	53.6**	8121	6.85**	8746	6.21**	9953	< 1		< 1	
Region 3:												
First pass reading times	< 1		1.64	3482	< 1		< 1		< 1		< 1	
First pass regressions out	2.85	347	4.74*	183	7.69**	133	5.21**		< 1		< 1	
Regression path times	< 1		1.15	30028	8.59**	19923	6.16**	24411	< 1		< 1	
Total reading times	1.28	148530	6.11*	25149	5.17*	16062	6.62**	17637	< 1		< 1	
Region 4:												
First pass reading times	< 1		< 1		4.21*	17480	5.31**	12909	< 1		1.16	11924
First pass regressions out	1.77	2007	25.6**	117	3.32*	189	2.63	204	< 1		< 1	
Regression path times	< 1		3.13	96840	4.80**	83477	2.45	142393	3.30*	83477	1.68	120365
Total reading times	< 1		10.1**	24690	2.79	13734	1.90	23485	2.42	13734	1.14	23485

Notes. * is significant at the .05 level, ** is significant at the .01 level. The regions were defined as in Experiment 3: the ambiguous region in Experiment 2 was excluded from the analyses.

No other effects approached significance in any of the measures. Total reading times did not show any significant effects. The only region where the factor condition approached significance was in the subjects analysis of region 2 ($p = .069$), but the effect was clearly non-significant in the items analysis ($p = .15$).

5.14. Combined analyses for Experiment 2 and 3

Combined analyses for Experiment 2 (delayed disambiguation) and Experiment 3 (immediate disambiguation) were also conducted. Experiment was treated as a between subjects and within items factor in the ANOVAs. The ambiguous region in Experiment 2 was excluded from these analyses, because Experiment 3 did not contain such a region. The results are summarised in Table 5.7.

An interaction between Experiment and Condition occurred in regression-path times at the critical relative clause verb region (e.g., *contain/containing*), indicating that immediate and delayed disambiguation caused a different pattern of results. The interaction between Experiment and Condition was marginally significant in first-pass regressions for this region ($p = .065$ by subjects; $p = .070$ by items). No interactions occurred in the post-critical region, suggesting that this region was processed similarly in both experiments.

In first-pass reading times, there was a significant effect of Experiment for the verb in the relative clause. Reading times were shorter in Experiment 2 than in Experiment 3. This is most likely due to a length difference between the verbs in the two experiments. The verb in Experiment 2 was an infinitive (e.g., *contain*), whereas the verb in Experiment 3 was a participle (e.g., *containing*) which was often longer. A similar effect was observed in total reading times. First-pass regressions revealed the opposite effect: there were more first-pass regressions in Experiment 2 than 3. This suggests that shorter regions result in more regressions. The item analyses for first-pass regressions and total times in the two final regions showed significant effects of Experiment in the items analyses, but these effects were not significant by subjects. Given that the factor Experiment involved a between subject comparison, these differences are likely to be due to general differences in reading rate between the two groups of participants.

5.15. Discussion

The results of Experiment 3, which employed immediate disambiguation, showed that the ambiguous region caused less difficulty than the disambiguated conditions. This pattern of results was obtained in regression-path times on the post-critical region, the same region where this measure reached significance in the experiment with delayed disambiguation. This suggests that readers process ambiguities with immediate and delayed disambiguation similarly. This was confirmed by the combined analyses of Experiment 2 and 3, which did not show any sign of interaction in the post-critical region.

These results are inconsistent with versions of competition models which assume that competition occurs very early on and that one analysis is selected before the disambiguation when the disambiguation is delayed. Such models predict that sentences with immediate and delayed disambiguation are processed differently. In ambiguities with delayed disambiguation, one analysis is selected during reading the ambiguous region, so that only one analysis is available when the sentence is subsequently disambiguated. If the disambiguation is inconsistent with the selected analysis, this causes a reanalysis-like effect. But in ambiguities with immediate disambiguation, the processor has no chance to select one analysis before the disambiguation. Hence, both analyses are available and one is selected in a process of competition. When both analyses are equally activated, competition should occur, whereas little competition should occur when one analysis is much more activated than its alternative. This predicts competition in the ambiguous condition of the current experiment, because plausibility supports both analyses. In contrast, the disambiguated conditions should be easy because only one analysis is supported. This was disconfirmed by the regression-path times on the post-critical region.

The results from this region are consistent with the unrestricted race model. This model predicts the same pattern of results for immediate and delayed disambiguation. In both cases, plausibility is evaluated after an initial syntactic analysis has been made. Because the ambiguity was balanced, it is predicted that NP1 and NP2 attachment are each adopted about half the time. This leads to reanalysis on about half of the trials in the disambiguated conditions. No reanalysis occurs in the ambiguous condition, because the initial analysis is always plausible.

Unfortunately, regression-path times and first-pass regressions for the critical region complicate the interpretation of the results. Although the effects were only marginally significant by items, both measures suggested that the ambiguous condition was more difficult than the disambiguated conditions in the critical region. If this is taken to reflect a competition process, it is unclear how this can be reconciled with the results from the

following region, which showed the opposite (reanalysis) pattern. It is hard to see how reanalysis and competition can both occur. In order to explain the reanalysis process in the post-critical region, it has to be assumed that the processor initially selected the implausible analysis on a number of trials and had to revise this analysis. However, if competition occurred in the critical region, it is reasonable to assume that the most plausible analysis was selected as a result of this process. It seems unlikely that the plausibility of the two analyses determined whether competition occurred or not, and that at the same time, the processor ignored this plausibility information when it selected one of the analyses. Alternatively, one might hypothesise that competition occurred in the critical region, but that selection did not take place (neither in the ambiguous nor disambiguated conditions). However, if all analyses were available to the processor, it is also unclear how reanalysis can occur. Therefore, it is improbable that reanalysis and competition both occurred.

The fact that the competition-like process did not reach significance by items suggests that it may be due to particular characteristics of some of the materials. It should be noted that in Experiment 2, a similar effect, also significant by subjects only, was observed in first-pass reading times for the first region (the NP complex). The means suggested that the ambiguous condition took longer to read than the disambiguated conditions. It was suggested in section 5.10 that this may have been caused by subtle differences in the NP complex. Clearly, the effect was not due to ambiguity resolution of the relative clause, because the relative clause had not been read at this point. Experiment 3 showed no such effect in region 1, although the items were identical to Experiment 2 until this point. Hence, it seems possible that the effects that were observed in the critical region of the current experiment may in fact have been spill-over effects from the preceding region rather than effects caused by syntactic ambiguity resolution of the relative clause.

The fact that significant effects were found in the critical region casts further doubt on whether the effects truly reflect ambiguity resolution of the relative clause. None of the six previous experiments on the competition/reanalysis issue (3 relative clause attachment experiments in Traxler et al., 1998; the biased VP/NP attachment experiment in Van Gompel et al., 1999; and Experiment 1 and 2 of this thesis) found significant effects in first-pass regressions or regression-path times in the critical region. Nearly all experiments showed the earliest effects in first-pass regressions and regression-path times for the *post-critical* region. The exception is Traxler et al.'s Experiment 1, which showed effects in an even later measure, total times on the critical region. Given the similarities of the materials in these experiments to the current ones with respect to the type of ambiguity and disambiguation, it is surprising to find that effects in the current experiment occurred in the critical region in first-pass regressions and regression-path times. It suggests that caution is

needed in interpreting the current results. They may have been caused by spill-over effects from the preceding region or alternatively, they may simply reflect a spurious effect. In contrast, the reanalysis-like effect in regression-path times for the post-critical region is more consistent with the effects in the other experiments.

However, the results from the current experiment do not enable us to draw clear conclusions. Therefore, immediate disambiguation was investigated in two further experiments. In the next chapter, immediate disambiguation is investigated in very similar materials to the ones tested in the current chapter, but an additional condition was included. In Chapter 7, completely different materials were used.

6. Experiment 4: Processing unambiguous sentences

6.1. Introduction

The experiment that is reported in this chapter served two purposes. Firstly, it was conducted in order to clarify the results that were obtained in Experiment 3 of the previous chapter. This experiment did not enable us to make firm conclusions about how syntactic ambiguities with immediate disambiguation are processed. The current experiment employed materials that were very similar (see [1a-c]) in order to determine whether the results from Experiment 3 could be replicated or whether they instead reflected a spurious effect.

1a. Colin noted that the basement of the tenement containing lots of junk was very gloomy. (Globally ambiguous)

1b. Colin noted that the tenement of the landlord containing lots of junk was very gloomy. (NP1 attachment)

1c. Colin noted that the landlord of the tenement containing lots of junk was very gloomy. (NP2 attachment)

1d. Almost at once Colin noted that the tenement containing lots of junk was very gloomy. (Unambiguous)

Secondly, the experiment further explored the architecture underlying syntactic processing. This was achieved by including a syntactically unambiguous condition. In this condition, only one attachment site exists. For example, the relative clause in (1d) can only be attached to *the tenement*. There is no other attachment site. This differs from the disambiguated conditions (1b) and (1c), which have two attachment sites, but one is ruled out by plausibility information.

Competition models which assume that one analysis is selected very rapidly (even when the ambiguity is balanced) predict that the ambiguous sentence (1a) should be more difficult to process than the disambiguated sentences (1b) and (1c). Because there is no ambiguous region in the disambiguated conditions, the processor cannot select one analysis before the point of disambiguation. Both the NP1 and NP2 attachment analysis are activated. In the ambiguous condition, severe competition should occur between them, because both are equally plausible, and there is no attachment preference (see off-line preference pre-test, section 6.2). In the disambiguated conditions, only one analysis is plausible, and therefore no such competition should occur. The unambiguous condition (1d) should be easiest to read. Because only one analysis is syntactically possible, there can be no competition.

Therefore, the unambiguous condition should be easier than the ambiguous condition. The unambiguous condition should presumably also be easier than the disambiguated conditions. Some competition should occur in the disambiguated conditions due to the parallel activation of the two syntactic analyses. However, assuming that plausibility information activates one analysis much more than the other, this competition may be relatively weak. In fact, all predictions are the same as those made by standard constraint-based competition models which do not assume rapid selection of an analysis. However, we have seen that such models are unable to account for the results from Experiment 1 (Chapter 4) and Experiment 2 (Chapter 5). I will therefore focus on the "rapid selection" versions of competition models here.

The predictions of reanalysis models are different. Because the previous experiments in this thesis ruled out some versions of reanalysis models such as the garden-path model and construal, I will only consider those predictions that are compatible with the results that were obtained so far.

In the unrestricted race model, all possible syntactic structures are involved in a race to be constructed first. When there is only one possible structure as in the unambiguous (1d), this structure always wins the race and is adopted. In contrast, when there are two possible structures as in (1a-c), either structure can win. When a structure is biased, the preferred interpretation is constructed fastest on most occasions and is therefore usually adopted. But when a structure is balanced, each analysis wins the race half the time, and each analysis is therefore adopted on half the trials. Assuming the ambiguity in (1) is balanced (given that the sentences are very similar to the ones in Chapter 5) this results in reanalysis on half the trials in the disambiguated conditions (1b) and (1c).

In the unrestricted race model, the initial selection of a syntactic analysis involves a race between the analyses. The selection process does not involve competition, and therefore the ambiguous condition should be no more difficult to process than the unambiguous condition. In fact, it is possible that the globally ambiguous structure takes less time to read than the unambiguous structure. This results from the fact that in the ambiguous condition, there are two pathways (two syntactic analyses) to arrive at a grammatical structure, whereas there is only one in the unambiguous condition. Hence, if one pathway is slowed down in the ambiguous condition, this has a great effect on the time it takes to arrive at an analysis, because there is no alternative. In contrast, the slow-down of one analysis has no effect on the ambiguous condition, because another analysis that has not been slowed down is available. If we assume that a pathway is sometimes slowed down as a result of fluctuations with respect to how long it takes to construct a syntactic structure, it will affect the unambiguous condition more severely than the ambiguous condition. Hence,

on average the ambiguous condition may be easier to process than the unambiguous condition. Such a finding would be similar to findings from lexical decision studies on lexical ambiguities, which show that lexical decision times to ambiguous words are faster than to unambiguous words, even when their frequency is the same (e.g., Borowsky & Masson, 1996; Kellas, Ferraro, Simpson, 1988; Millis & Button, 1989; Rubenstein, Lewis, & Rubenstein, 1971). Of course, the strength of the ambiguity advantage depends on how great the fluctuations to construct a syntactic analysis are. If the variability is very small, constructing an analysis will usually not be much slower than average. Hence, the ambiguity advantage may be very small and may be undetectable.

The selection of an initial analysis does not necessarily have to involve a race process. Another possibility is that the selection of one analysis involves a process of what could be termed *structural* competition (as opposed to the type of competition that is predicted by constraint-based theories). When a structure is unambiguous, as in the relative clause attachment in (1d), selection of a structure is straightforward, because there is no other structure that can interfere with it. But when two structures are grammatically permissible, as in (1a), the structures can compete with each other. This process of competition can occur at different levels. One possibility is that constructing two analyses causes competition because the two construction processes draw upon the same processing resources. When two analyses compete for the same resources, constructing the analyses takes longer than when only one analysis uses these resources. Alternatively, competition can occur at the selection stage. When two analyses have been constructed, they may compete for selection, whereas no such competition occurs if only one analysis has been constructed.

It is important to note that structural competition is quite different from the kind of competition that is predicted by constraint-based theories. Structural competition is predicted to occur both in the globally ambiguous structure (1a) and in the disambiguated structures (1b) and (1c), because in both cases two syntactic structures are grammatical (although only one is plausible). Because plausibility is only used after the initial selection process, it does not affect the structural competition process (but it does affect constraint-based competition). The notion of structural competition maintains the serial nature of reanalysis models, as only a single analysis is ever selected. Therefore, it also maintains reanalysis as a mechanism of ambiguity resolution. It is predicted that reanalysis occurs in both the NP1 (1b) and NP2 (1c) attachment condition. Because the ambiguity is balanced, each analysis is selected about half the time (after the structural competition process). Hence, when the sentence is disambiguated, the initial analysis is implausible half the time. In contrast, no

reanalysis occurs in the ambiguous condition, because the initial analysis is always plausible.

In short, the structural competition hypothesis predicts that the ambiguous condition causes more difficulty than the unambiguous condition, because syntactic competition occurs in the former, but not in the latter. However, the ambiguous condition is easier than the disambiguated conditions. In both cases, structural competition occurs, but reanalysis occurs in the disambiguated conditions, whereas it does not in the ambiguous condition.

6.2. Off-line preference pre-test

Participants. Eighteen participants from the same population as the previous experiments took part[1].

Materials. Two of the 30 materials that were used for Experiment 3 (Chapter 5) were discarded in order to make the number of materials a multiple of four (there were four conditions). The materials were the same as for Experiment 3, except for the following. An unambiguous condition was constructed by removing NP1 and *of* from the ambiguous condition. Because this made the unambiguous condition shorter than the other conditions, a clause was added to the beginning of all conditions (e.g., *Colin noted that* ...). This clause contained some more words in the unambiguous condition, so that the unambiguous condition became similar in length. Overall, the length of the unambiguous condition was the same as the ambiguous condition. The final region of a few materials was slightly changed in order to fit at least two words following the relative clause on the first line in the eyetracking experiment. An example of an item is given in (1), repeated here as (2). See Appendix 3 for the complete list of materials.

2a. Colin noted that the basement of the tenement containing lots of junk was very gloomy. (Globally ambiguous)
2b. Colin noted that the tenement of the landlord containing lots of junk was very gloomy. (NP1 attachment)
2c. Colin noted that the landlord of the tenement containing lots of junk was very gloomy. (NP2 attachment)

[1] Two additional participants were excluded. One chose the (implausible) NP1 attachment paraphrase on 90% of trials in the NP2 attachment condition, and the other chose the (implausible) NP2 attachment paraphrase on 100% of NP1 attachment trials. These patterns were clearly different from the other participants, and we assumed that these participants had failed to understand the task.

2d. Almost at once Colin noted that the tenement containing lots of junk was very gloomy. (Unambiguous)

In the off-line preference pre-test, each item was followed by two paraphrases that unambiguously described either the NP1 or NP2 attachment analysis, as in Experiment 3. The unambiguous condition (2d) was not used in the off-line preference study, because there was only one possible analysis in this condition.

Design. The off-line study contained three conditions: the ambiguous, NP1 attachment and NP2 attachment. Order of paraphrase condition was treated as a factor. Therefore, six lists of items were constructed, with one version of each item appearing in each list. Because there were 28 items and 3 conditions, each participant saw 10 items in one condition, and 9 in the other two. The items occurred in a single random order in the lists. Three participants were assigned to each of the lists.

Procedure. Identical to those for the off-line preference pre-test for Experiment 2 and 3 (Chapter 5).

Results. Table 6.1 shows the percentage of NP2 responses for each condition. The NP1 and NP2 responses always added up to 100%. ANOVAs with the factor Condition as a within subjects and items variable were conducted. They revealed a main effect of condition: $F1\ (2, 34) = 89.2$, $p < .001$, $MSe = 313$; $F2\ (2, 54) = 144$, $p < .001$, $MSe = 300$. Planned comparisons indicated that the ambiguous condition differed from both the NP1 attachment condition ($F1\ (1, 34) = 52.3$, $p < .001$; $F2\ (1, 54) = 83.4$, $p < .001$) and the NP2 attachment condition ($F1\ (1, 34) = 38.2$, $p < .001$; $F2\ (1, 54) = 61.5$, $p < .001$).

The 56% NP2 attachment percentage in the ambiguous condition did not differ from chance ($F1\ (1, 17) = 1.73$, $p = .21$, $MSe = 190$; $F2\ (1, 27) = 1.84$, $p = .19$, $MSe = 270$), indicating the ambiguity was balanced. Participants took the implausible analysis on very few occasions: 14% in the NP1 attachment condition, and 7% in the NP2 attachment condition. This indicated that the final analysis in the disambiguated conditions was nearly always determined by plausibility.

Table 6.1

Experiment 4: Item characteristics

	condition		
	ambiguous	NP1 attachment	NP2 attachment
Off-line preference task:			
Percentage of NP2 attachment	56.0	92.5	13.8

6.3. Eyetracking Experiment 4

Method

Participants. Thirty-five participants from the same population as before took part. All of them had normal vision. Three participants were excluded because they had a very high percentage of tracker losses. The analyses were conducted on the remaining 32 participants.

Materials. The experimental materials were the same as in the off-line preference pre-test. In addition, 68 fillers containing a variety of structures, unrelated to the purpose of the experiment, were included.

Design. Four lists of items, consisting of seven items from each condition were constructed. Exactly one version of each item appeared in each list. Eight participants were assigned to each list. Except for this, the design was the same as the design used in Chapter 5.

Procedure. The procedure was the same as in the eyetracking experiments of the previous chapter.

Analyses. On average, the participants made 8% errors on the comprehension questions. Prior to all analyses, trials with major tracker losses and trials with fewer than four fixations were excluded. This eliminated less than 2% of the trials. Less than 1% of the data was removed as a result of two consecutive regions having zero reading times.

Reading times that were more than 2 standard deviations from both the subject and item mean of the region were also excluded. This never removed more than 1% of the data on any of the analyses.

The materials were divided into four regions, indicated by slashes below:

3. Colin noted that the basement of/ the tenement/ containing/ lots of junk/ was very gloomy./

The regions corresponded to (1) the beginning of the sentence until *of*, (2) the NP preceding the relative clause, (3) the verb in the relative clause (the critical region), (4) the rest of the relative clause (the post-critical region), and (5) the final region, which contained the remainder of the main clause. In the current experiment, region 1 was not

controlled between conditions. In the unambiguous condition, the contents of this region were fundamentally different from the other three conditions in terms of both semantics and syntactic structure. In order to check whether these differences caused any reading time differences on subsequent words, NP2 was defined as a separate region (in contrast to the region definition in Experiment 2 and 3). If no differences are observed in the NP2 region, we can be reasonably confident that any effects further downstream are not caused by irrelevant differences in region 1.

Results

Table 6.2 presents the mean reading times and percentage regressions by condition and region. For each eyetracking measure for each region, one-way ANOVAs with three levels of Condition (ambiguous vs. NP1 attachment vs. NP2 attachment) were conducted. Condition was treated as a within subjects and items factor. Table 6.3 presents the results of these analyses. Planned comparisons between levels are reported below.

No significant effects were obtained in first-pass reading times. First-pass regressions showed a significant condition effect in region 4, the post-critical region. Planned comparisons revealed that the NP2 attachment condition produced more regressions than the ambiguous condition ($F1\ (1, 93) = 9.19$, $p = .003$; $F2\ (1, 81) = 8.32$, $p = .005$), the unambiguous condition ($F1\ (1, 93) = 4.20$, $p = .043$; but $F2\ (1, 81) = 3.56$, $p = .063$) and the NP1 attachment condition ($F1\ (1, 93) = 5.09$, $p = .026$; $F1\ (1, 81) = 4.58$, $p = .035$). The ambiguous condition did not differ from the unambiguous condition ($F1\ (1, 93) = .97$, $p = .33$; $F2\ (1, 81) = 1.00$, $p = .32$) nor from the NP1 attachment condition ($F1\ (1, 93) = .60$, $p = .44$; $F\ (1, 81) = .55$; $p = .46$). The effect of condition for first-pass regressions from the final region was significant by subjects, but not by items ($p = .12$).

Regression-path times revealed an effect in the post-critical region (region 4), although this was only marginally significant by items ($p = .069$). The NP2 attachment condition was read more slowly than the ambiguous condition ($F1\ (1, 93) = 6.94$), $p = .010$; $F2\ (1, 81) = 5.09$, $p = .027$), the unambiguous condition ($F1\ (1, 93) = 6.59$, $p = .012$; $F1\ (1, 81) = 5.49$, $p = .022$) and the NP1 attachment condition ($F1\ (1, 93) = 5.06$, $p = .027$; $F2\ (1, 81) = 3.92$, $p = .054$). The ambiguous condition did not differ from the unambiguous condition ($F1\ (1, 93) = .01$, $p = .95$; $F2\ (1, 81) = .01$, $p = .93$) or the NP1 attachment condition ($F1\ (1, 93) = .15$, $p = .70$; $F2\ (1, 81) = .08$, $p = .78$). The effect of condition in the final region was significant by subjects, but not by items ($p = .30$).

Table 6.2

Experiment 4: Means

	region				
	1	2	3	4	5
First-pass reading times:					
ambiguous	1206	430	373	550	836
NP1 attachment	1162	434	348	567	815
NP2 attachment	1180	420	356	572	812
unambiguous	1196	425	358	556	803
First-pass regressions out:					
ambiguous		14.6	12.1	13.1	58.8
NP1 attachment		11.9	9.5	15.9	60.6
NP2 attachment		13.6	8.4	23.4	65.8
unambiguous		10.7	9.4	16.7	51.9
Regression-path times:					
ambiguous	1206	537	438	696	1901
NP1 attachment	1162	488	408	706	2050
NP2 attachment	1180	487	421	776	2117
unambiguous	1196	496	422	693	1788
Total times:					
ambiguous	1724	702	539	773	1079
NP1 attachment	1685	690	581	878	1115
NP2 attachment	1837	713	601	866	1094
unambiguous	1630	667	540	789	1020

Notes. The regions were defined as follows (delimited by slashes):
Colin noted that the basement of/ the tenement/ containing/ lots of junk/ was very gloomy./

First pass, regression path, and total times are reported in ms, first pass regressions as the percentage of saccades leaving the region to the left after a first pass fixation.

Table 6.3

Experiment 4: ANOVA results by region for effect of condition

	F1 (3, 81)	MSe	F2 (3, 93)	MSe
Region 1:				
First pass reading times	1.03	13792	< 1	
Regression path times	1.03	13792	< 1	
Total reading times	3.59*	65969	2.02	102414
Region 2:				
First pass reading times	< 1		< 1	
First pass regressions out	< 1		< 1	
Regression path times	2.23	7835	1.90	7868
Total reading times	< 1		< 1	
Region 3:				
First pass reading times	2.22	1932	< 1	
First pass regressions out	< 1		< 1	
Regression path times	1.03	6902	< 1	
Total reading times	2.17	11689	1.02	25864
Region 4:				
First pass reading times	< 1		< 1	
First pass regressions out	3.35*	181	3.01*	173
Regression path times	3.14*	15257	2.45	18211
Total reading times	4.10*	22863	3.00*	25484
Region 5:				
First pass reading times	< 1		< 1	
First pass regressions out	4.18	262	2.02	457
Regression path times	3.30*	241216	1.25	506010
Total reading times	1.93	30148	1.31	35497

Note. * is significant at the .05 level, ** is significant at the .01 level.

A significant effect also appeared in region 4 in total reading times. The ambiguous condition was faster to read than the NP1 attachment (F1 (1, 93) = 8.22, p = .005; F2 (1, 81) = 6.00, p = .017) and NP2 attachment condition (F1 (1, 93) = 6.24, p = .014; F2 (1, 81) = 4.58, p = .035), but was no faster than the unambiguous condition (F1 (1, 93) = .23, p = .63; F2 (1, 81) = .17, p = .68). The two disambiguated conditions did not differ either (F1 (1, 93) = .14, p = .71; F2 (1, 81) = .10, p = .76).

6.4. Discussion

First-pass regressions and regression-path reading times for the post-critical region showed that the NP2 attachment condition was more difficult to read than the other conditions, which suggests an NP1 attachment preference. This is somewhat surprising, given that the off-line preference pre-test showed that the ambiguity was relatively balanced (56% NP2 attachment preference). The pattern of total reading times for the post-critical region was more consistent with the pre-test: the disambiguated conditions took equally long to read, but both took longer than the globally ambiguous and unambiguous condition. This pattern of results may seem to require an explanation in which the processor initially adopts the NP1 attachment analysis, and later revises it to an NP2 attachment analysis. However, such an explanation would imply that such a reanalysis also occurred in the ambiguous condition, which is not supported by the data: the ambiguous condition took no longer to read than the unambiguous condition. Instead, it is more likely that the difficulty simply took longer to develop in the NP1 attachment condition than in the NP2 attachment condition.

The current experiment was partly conducted to clarify the results of Experiment 3 (Chapter 5). Both experiments used immediate disambiguation in very similar relative clause attachment ambiguities. The only difference between the materials in the two experiments was that a clause was added to the beginning of the sentence and that an unambiguous condition was included in the current experiment. Experiment 3 revealed conflicting results. It showed that the disambiguated conditions took longer to read than the ambiguous condition in regression-path times for the post-critical region, but it showed the opposite pattern in regression-path times and first-pass regressions for the critical region. However, there were reasons to question the results from the critical region. The effects did not reach significance by items, which suggests that the pattern may have been caused by only a few items which behaved differently from the rest. In addition, previous experiments on similar ambiguities always obtained the earliest effects in the post-critical region. This suggests that the effects in the critical region may have been spurious.

The current experiment supports this conjecture. No significant effects occurred in the critical region and there were no measures that showed that the ambiguous condition was more difficult to process than the disambiguated conditions at any point. The current results confirm the pattern that was observed in the post-critical region in Experiment 3. In both experiments, the disambiguated conditions took longer to read than the ambiguous condition. The only difference between the studies is that this pattern occurred somewhat later in the current experiment. The pattern was obtained regression-path times in Experiment 3, whereas it was obtained in total times in the current experiment. Possibly, this difference was caused by the addition of a clause at the beginning of the materials. As a result of this, the relative clause in the current experiment was syntactically more deeply embedded than in the experiments in Chapter 5, and hence it may have been processed more superficially. This may have led to weaker and therefore later effects.

The current results indicate that disambiguated sentences are more difficult to read than globally ambiguous sentences even when the disambiguation is immediate. This provides evidence against versions of competition models which claim that competition occurs very early on and that one analysis is selected very rapidly. Such accounts can explain the reanalysis pattern in ambiguities with delayed disambiguation. They predict that one analysis is selected in the ambiguous region preceding the disambiguation and that therefore only one analysis is available at the disambiguation. If this analysis is inconsistent with the disambiguation, the alternative has to be reactivated, which results in processing difficulty. However, such models cannot explain the pattern of results in the current experiment. Because the disambiguation was immediate, there was no ambiguous region in which a single analysis could be selected. Hence, both analyses should be available to the processor. These analyses should compete in the ambiguous condition, because they are equally plausible, and there is no preference for either NP1 or NP2 attachment. In contrast, no such competition should occur in the disambiguated conditions, because plausibility supports only one analysis. In fact, the opposite pattern was obtained in the current experiment: the ambiguous condition was easier to read than the disambiguated conditions. This pattern is similar to the pattern that was obtained for ambiguities involving delayed disambiguation, indicating that both types of ambiguity are processed similarly. Furthermore, competition models are also inconsistent with the results from the unambiguous condition. Because there are no competing analyses in this condition, it is predicted to be much easier than the ambiguous condition, where there should be severe competition. However, there was no evidence that the unambiguous condition was any faster than the ambiguous condition in any of the measures.

The results are consistent with some versions of reanalysis models. They support models that predict that the ambiguities were processed in two stages. In the first stage, one analysis was selected. Because the ambiguity was balanced, the NP1 and NP2 attachment analysis were each selected on about half of the trials. In the second stage, the plausibility of the initial analysis was checked. If plausibility information disconfirmed the initial analysis, reanalysis occurred and processing difficulty ensued. This resulted in reanalysis on about half the trials in the disambiguated conditions. In contrast, reanalysis never occurred in the ambiguous condition, because the initial analysis was always plausible.

Assuming this type of reanalysis model, the results from the unambiguous condition shed light on the kind of processes that are involved in the first stage of processing, in which the initial analysis is selected (independent of plausibility). According to the structural competition model, competition occurs in the first stage of processing when there are two possible grammatical structures. This competition could arise because two analyses of an ambiguous structure have to share the same processing resources, or because the selection process involves competition. Ambiguous syntactic structures should be more difficult to process than unambiguous structures, because structural competition occurs in the former, but not in the latter, where there is only one possible analysis. However, the results do not support this prediction: the ambiguous and unambiguous did not differ. If structural competition exists at all, it does not appear to increase processing difficulty to the same extent that reanalysis does.

The results from the unambiguous condition are more consistent with a race model such as the unrestricted race model. In this model, selection of the initial analysis involves a race. The analysis that is constructed first is selected. There is no competition in the model, and therefore the ambiguous condition should be no slower than the unambiguous condition, as was borne out by the data. In section 6.1, the possibility was raised that the ambiguous condition might in fact be faster than the unambiguous condition. This prediction follows from the greater probability that the single analysis of an unambiguous structure is slow as a result of fluctuations in the time it takes to construct an analysis relative to the probability that both analyses of an ambiguous structure are slow. The current data showed no evidence for such an ambiguity advantage. This may simply be because the fluctuations in the time it takes to construct a syntactic analysis are not very big and therefore the effect could not be detected. The ambiguity advantage may be very subtle, as the ambiguity effect in word recognition studies. It has been noted that the word ambiguity advantage is often not significant by items (e.g., Clark, 1973). Furthermore, it occurs in lexical decision, but not in naming (Borowsky & Masson, 1996).

In sum, the findings indicate that processing difficulty is caused by reanalysis. There is no evidence that other processes cause difficulty in syntactic ambiguity resolution. Competition of the kind that is predicted by constraint-based theories did not occur. In addition, there was no evidence that structural competition or fluctuations in the time it takes to construct an analysis affects the complexity of ambiguity resolution.

7. Experiment 5: A subject-object word-order ambiguity in German

7.1. Introduction

The experiments that investigated the reanalysis/competition issue have so far investigated either relative clause attachment ambiguities (Traxler et al., 1998; This thesis, Chapter 5 and 6) or PP attachment to a VP or NP (Van Gompel et al., 1999; Chapter 4). The relative clause and VP/NP attachment ambiguity constitute similar types of ambiguities. In both cases, the structure is ambiguous because a phrase (the relative clause or PP) can be attached to two phrase markers in the preceding part of the sentence. In addition, although construal (Frazier & Clifton, 1996) claims that PP attachment to a VP constitutes a primary relation, whereas relative clause attachment does not, other researchers have noted that the type of VP attachment that was tested in Chapter 4 (involving the preposition *with*) has at least some characteristics of adjunct attachment (e.g., Schütze & Gibson, 1999; Sedivy & Spivey-Knowlton, 1994): on the basis or some argument/adjunct tests, it could be argued that both the VP/NP attachment and relative clause ambiguity involve ambiguous adjunct attachment. Therefore, if it is indeed true that adjuncts are processed differently from arguments (Frazier & Clifton, 1996; Liversedge, Pickering, Branigan, & Van Gompel, 1998; Schütze & Gibson, 1999), there may be reason to assume that the reanalysis pattern that was obtained in the experiments that have been discussed so far does not occur for ambiguities that involve argument attachment.

For this reason, the current experiment tested whether the reanalysis pattern also occurred in a completely different type of ambiguity that involved syntactically ambiguous arguments. In addition, it investigated a different language, German, in order to see whether the reanalysis pattern also replicates in other languages than English. The experiment investigated word-order ambiguities such as (1).

1a. Er wollte wissen, welches Huhn gestern die Ente verscheucht hat. (Ambiguous)
He wanted to know, which chicken [subject/object] yesterday the duck chased-away has.
"He wanted to know which chicken chased away the duck yesterday."
(subject-first interpretation)
"He wanted to know which chicken the duck chased away yesterday."
(object-first interpretation)

1b. Er wollte wissen, welcher Hahn gestern die Ente verscheucht hat. (Subject-first)
He wanted to know, which rooster [subject] yesterday the duck chased-away has.
"He wanted to know which rooster chased away the duck yesterday."

1c. Er wollte wissen, welchen Hahn gestern die Ente verscheucht hat. (Object-first)
He wanted to know, which rooster [object] yesterday the duck chased-away has.
"He wanted to know which rooster the duck chased away yesterday."

In the ambiguous condition (1a), the grammatical roles of the NP arguments in the subordinate clause (*welches Huhn* and *die Ente*) are ambiguous. The first NP (NP1) can either receive the subject role and the second NP (NP2) the object role (the subject-first interpretation) or alternatively, NP1 can have the object role and NP2 the subject role (the object-first interpretation). The disambiguated conditions (1b) and (1c) employ immediate disambiguation on the wh-pronoun *welcher/n*. In (1b) the nominative the case marking of the wh-pronoun *welcher* disambiguates the first NP *welcher Hahn* toward the subject role and consequently, the second NP has to be the object. In (1c), the accusative case marking of *welchen* disambiguates the first NP toward the object role.

Frazier (1987b; Clifton & Frazier, 1989; Frazier & Clifton, 1989; Frazier & Flores d'Arcais, 1989) proposed that the active filler strategy predicts parsing preferences for word-order ambiguities such as (1) (see also section 2.4). The active filler strategy assumes a transformational grammar, in which phrases are generated in a base position (the gap or trace), and are subsequently moved by a transformation to another position in the sentence (the filler) (e.g., Chomsky, 1981; 1986). According to the active filler strategy, a gap is posited as early in the sentence as possible, and therefore sentences with an early gap are easier to process than sentences with a late gap. Because the basic word order in German is assumed to be S(ubject) O(bject) V(erb), subject-first word order is preferred to object-first order. Similar preferences are also predicted for Dutch, which is closely related to German and is also argued to have basic SOV word order. For example, the active filler strategy predicts a preference for the Dutch subject-first sentence in (2a) over the object-first sentence in (2b). In

sentence (2a) the first NP *de patient* is the subject of the clause, because it agrees in number with the verb *bezoekt*. In (2b), the second NP *de dokters* agrees with the verb, and therefore this NP is the subject (and the first NP the object).

2a. De patient bezoekt de dokters. (subject-first)
The patient [subject] visits the doctors [object].
"The patient visits the doctors."

2b. De patient bezoeken de dokters. (object-first)
The patient [object] visit the doctors [subject].
"The doctors visit the patient."

As shown in (3), the subject gap due to moving *de patient* (indicated by ___) in the subject-first sentence occurs earlier than the object gap in the object-first sentence[1]. Because the gap is posited as early as possible, the subject-first analysis is predicted to be easiest.

3a. De patient bezoekt ___ de dokters. (subject-first)

3b. De patient bezoeken de dokters ___. (object-first)

Similar predictions are made for other types of word-order ambiguities. The structure in (1) involves movement of the wh-phrase to clause-initial position, regardless of whether the wh-phrase is a subject or object. If it is a subject, the gap is earlier than if it is an object, as shown in (4). Hence, a subject-first preference is predicted for this structure as well.

4a. (=1b) Er wollte wissen, welcher Hahn ___ gestern die Ente verscheucht hat. (subject-first)

4b. (=1c) Er wollte wissen, welchen Hahn gestern die Ente ___ verscheucht hat. (object-first)

More recently Frazier and Clifton (1996) claimed that the active filler strategy is superseded by the *minimal chain principle* that was proposed by De Vincenzi (1991; cf. Clifton & De Vincenzi, 1990). As the active filler strategy, this principle assumes a transformational grammar such as Government and Binding theory (Chomsky 1981; 1986). In such a grammar, a moved constituent is co-indexed with the trace(s) or gap(s) that it leaves behind in the tree structure. This co-indexation relation is called a *chain*.

[1] The sentences involve an additional movement of the verb. Because this movement is not predicted to affect the complexity of the sentence, the gap position is omitted here.

The minimal chain principle makes use of this notion. It states that unnecessary chain members are avoided and that required chain members are posited as soon as possible. The claim that unnecessary chain members are avoided deals with parsing preferences such as for (5).

5. Ha chiamato Gianni.
Has called Gianni.
pro Has called Gianni: "He/she has called Gianni." (null-pronoun subject)
t_i Has called $Gianni_i$: "Gianni has called." (sentence-final subject)

This Italian sentence is ambiguous between an interpretation in which a null-pronoun is the subject of *ha chiamato*, and an interpretation in which the sentence final NP is the subject. The minimal chain principle claims that the null-pronoun is the preferred subject, because this does not involve a chain between two co-indexed elements in the sentence: the *pro* which represents the null-pronoun is not co-indexed with another constituent. In contrast, the sentence-final subject interpretation involves, according to Government and Binding theory, a chain between *Gianni* and its trace t_i. Because this interpretation includes one more chain member than the null-pronoun subject interpretation, it is disfavoured.

The minimal chain principle further states that required chain members are postulated as soon as possible. In other words, this predicts that a trace or gap is posited as early in the sentence as possible. In fact, this is a different way of phrasing the active filler strategy. Therefore, the minimal chain principle makes the same predictions as the active filler strategy for ambiguities that involve two possible gap locations such as the word-order ambiguities under discussion.

The predictions of the active filler strategy and the later minimal chain principle were supported by Frazier and Flores d'Arcais (1989), who tested sentences such as (2). They observed that grammaticality judgement times for subject-first sentences such as (2a) were shorter than for object-first sentences such as (2b). Other studies have confirmed the subject-first preference in main clauses in both Dutch (e.g., Kaan, 1997; Read, Kraak, & Boves, 1980) and German (e.g., Hemforth, 1993). The pattern obtains both when the first NP is a definite (as in [2]) as well as when the noun is preceded by a wh-pronoun (Schlesewsky et al., in press), although the preference appears to be stronger for definite NPs (Kaan, 1997).

The subject-first preference is also found in relative clauses such as the Dutch sentences in (6).

6a. Karl hielp de mijnwerker die de boswachters vond. (subject-first)
Karl helped the miner who [subject] the foresters found.
"Karl helped the miner who found the forester."

6b. Karl hielp de mijnwerker die de boswachters vonden. (object-first)
Karl helped the miner who [object] the foresters found.
"Karl helped the miner who the forester found."

In (6a) the relative pronoun *die* (who) is the subject of the relative clause (*die de boswachters vond*) because it agrees in number with the verb, and *de boswachters* is the object. In (6b) the subject-object order is reversed: *die* is the object, and *de boswachters* is the subject. A number of studies have shown that the subject-first order is preferred in such structures, both in Dutch (Frazier, 1987b; Van Gompel, 1995) and German (Mecklinger et al., 1995; Friederici, 1998; Schriefers et al., 1995). This is consistent with both the active filler strategy and the minimal chain principle.

However, there is some controversy with respect to whether gaps are psychologically real or not. Pickering and Barry (1991; Pickering, 1993) argued that they are not. For example, they considered sentences such as (7).

7a. Which box did you put the very large and beautifully decorated wedding cake bought from the expensive bakery in ___.

7b. In which box did you put the very large and beautifully decorated wedding cake bought from the expensive bakery ___.

In both sentences, the gap should be in sentence-final position according to standard transformational grammars, and the moved element (*which box* in [7a] and *in which box* in [7b]) is in sentence-initial position. Therefore, there is no reason why gap-filling theories should predict differences in complexity for both sentences. However, Pickering and Barry claim that (7a) is in fact more difficult than (7b), and argue that this is inconsistent with gap-filling accounts. Instead, they proposed that it is the distance between *(in) which box* and its subcategoriser that determines processing difficulty. The subcategoriser in (7a), *in*, is away far from *which box*, and therefore the sentence is hard to process. In contrast, the subcategoriser *put* in (7b) is near to *in which box*, and therefore this sentence is easy.

Further support for this argument comes from Traxler and Pickering (1996b), who investigated sentences like (8). In these sentences, the gap position (indicated by ___) for the NP *the garage/the pistol* follows *man*. The NP is either a plausible direct object of *shot* (in [8b]), or not (in [8a]). Traxler and Pickering showed in an eye-movement experiment that the plausibility effect in a relative clause such as (8) occurred

at the verb *shot* and not at the gap position. This suggests that *garage* and *pistol* are directly associated to the subcategorising verb *shot* and that the gap position does not play a part in the parsing of a relative clause sentence.

8a. That's the garage with which the heartless killer shot the hapless man ___ yesterday afternoon. (implausible filler)

8b. That's the pistol with which the heartless killer shot the hapless man ___ yesterday afternoon. (plausible filler)

However, Gibson and Hickok (1993) and Crocker (1994) have argued that the processor may be able to predict a gap position before it has been reached and that effects of gap position can therefore occur before the actual gap has been reached. In fact, the assumption of gaps depends entirely on the particular grammar that one presupposes. Although some grammars assume the existences of gaps (e.g., Bresnan, 1982; Chomsky, 1981; 1986; Gazdar, Klein, Pullum, & Sag, 1985), other grammars do not (e.g., Hudson, 1990; Kaplan & Zaenen, 1988; Moortgat, 1988; Pollard & Sag, 1994, Chapter 9; Steedman, 1987).

One model in which predictions for word-order ambiguities in Dutch and German do not depend on the existence of gaps is Gorrell (1996). Although Gorrell claims that gaps do exist, word order preferences derive from a different principle, the principle of simplicity. According to this principle, the parser does not build unnecessary nodes in the tree structure. In the grammar that he assumes, object-first structures usually (but not always) require more nodes than subject-first structures, and therefore they are predicted to be more difficult (see Crocker, 1994 for a similar proposal). For example, Fig. 1a and 1b show examples of tree structures for (9), which is the ambiguous version of (2) (because both NPs can agree with the verb). Fig. 7.1a shows the subject-first analysis, whereas Fig. 7.1b shows the object-first analysis.

9. De patient bezoekt de dokter.
The patient [subject/object] visits the doctor [object/subject].
"The patient visits the doctor./The doctor visits the patient."

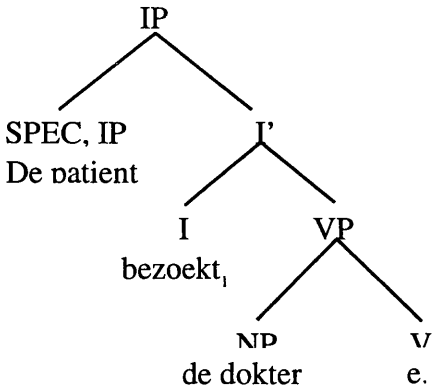

Fig.7.1a. Subject-first tree

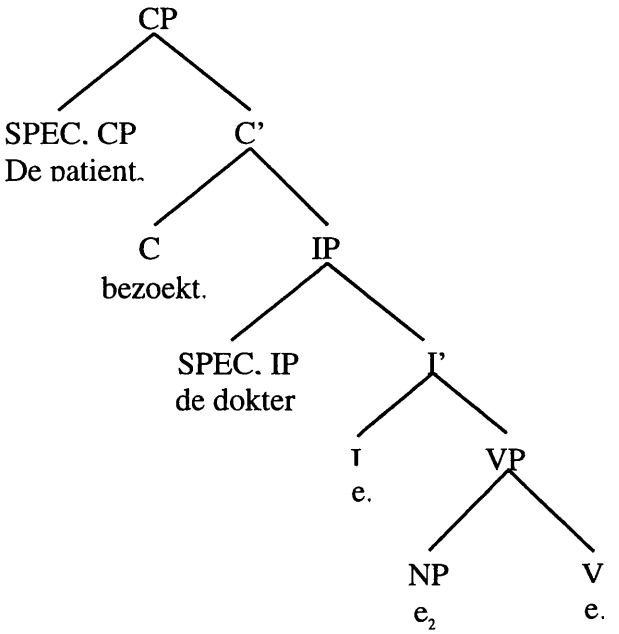

Fig 7.1b: Object-first tree

As can be seen, the subject-first tree involves the postulation of fewer nodes than the object-first tree. Therefore, the subject-first analysis is predicted to be preferred.

Gorrell makes an interesting predictions for the ambiguity in (1). He claims (following Travis, 1991) that the wh-phrase occupies the SPEC, CP position regardless of whether it is a subject or object. As a result, the subject-first and object-first analysis have the same number of nodes, and they should not differ in complexity. Because the phrase structure assumptions for the subject-first analysis are different in (1) and (2), ambiguities such as (1) should be processed differently from (2).

Kaan (1997) investigated word-order ambiguities with clause-initial wh-phrases in Dutch. In a completion study, participants had to complete sentence fragments such as (10). In (10), the wh-phrase *welke student* (which student) can either be the subject or object of the subordinate clause.

10. Ik vroeg me af welke student ...
"I wondered which student ..."

In nearly all cases (96%), participants completed the sentences such that the wh-phrase was the subject. However, in a word-by-word grammaticality judgement task and self-paced reading study using sentences such as (11), Kaan observed an object-first preference.

11a. Ik vroeg welke vrienden hem in het dorp aan de kust hadden opgezocht.
I asked which friends him [subject] in the village at the coast had visited.
"I asked which friends had visited him in the village at the coast."

11b. Ik vroeg welke vrienden hij in het dorp aan de kust hadden opgezocht.
I asked which friends he [object] in the village at the coast had visited.
"I asked which friends he had visited in the village at the coast."

In these sentences, case marking on the pronoun *hij/hem* disambiguates the ambiguity. In the self-paced reading experiment, the object-first preference was observed at *aan* and *de*, and was marginally significant on the word following the pronoun (*hij/hem*). Although these results are also inconsistent with Gorrell's model, because it predicts no preference for either analysis, the object-first preference does suggest that word-order ambiguities with wh-phrases are processed differently from ambiguities that do not contain such a phrase. In a further self-paced reading experiment Kaan observed that the object-first preference was affected by the length of the region between the pronoun and the verb (*hadden*). When this region was short, the object-first advantage was neutralised, suggesting that there was no initial preference for either subject or object-first. This would be consistent with Gorrell's model. However, another experiment by Kaan also suggested that word-order preferences depend on whether the second NP (e.g., *hem/hij* in [11]) is a definite NP or a pronoun. In this experiment, which used a region with intermediate length, no preference for either word order was observed when this NP was a pronoun, but a marginally significant subject-first preference occurred when the NP was a definite NP.

In sum, it is not entirely clear what preference will obtain for the ambiguity in (1) which was tested in the current experiment. Although nearly all studies on subject-object ambiguities in German and Dutch show a subject-first preference, the results from Kaan (1997) suggest that there may not be a strong preference for either subject or object-first in sentences with wh-phrases, or that there may even be an object-first preference. However, Kaan's completion data also indicated that there is at least an off-line subject-first preference. Because the predictions of both constraint-based competition models and the unrestricted race model depend on the bias of the materials, I will first describe the off-line pre-tests, before outlining their predictions.

7.2. Plausibility pre-test

Although plausibility was not used as a way to disambiguate the items in this study, a plausibility study was conducted in order to check whether there were any differences in plausibility between the conditions and between the two syntactic analyses of a condition.

Participants. Twenty native German-speaking students from the University of Freiburg (Germany) participated in this pre-test.

Materials. Items like (12) were constructed (see Appendix 4).

12a. Der Knecht auf dem Bauernhof mußte sich um das Federvieh kümmern.
Er wollte wissen, welches Huhn gestern die Ente verscheucht hat,
um sie voneinander zu trennen. (Ambiguous)
The farm hand from the farm must himself after the poultry look.
He wanted to know, which chicken [subject/object] yesterday the duck chased-away has, in order to separate them.
"The farm hand from the farm had to look after the poultry.
He wanted to know which chicken chased away the duck yesterday,
in order to separate them" (subject-first interpretation)
"The farm hand from the farm had to look after the poultry.
He wanted to know which chicken the duck chased away yesterday,
in order to separate them" (object-first interpretation)

12b. Der Knecht auf dem Bauernhof mußte sich um das Federvieh kümmern.
Er wollte wissen, welcher Hahn gestern die Ente verscheucht hat,
um sie voneinander zu trennen. (Subject-first)
The farm hand from the farm must himself after the poultry look.
He wanted to know, which rooster [subject] yesterday the duck chased-away has, in order to separate them.
"The farm hand from the farm had to look after the poultry.
He wanted to know which rooster chased away the duck yesterday,
in order to separate them" (subject-first interpretation)

12c. Der Knecht auf dem Bauernhof mußte sich um das Federvieh kümmern.
Er wollte wissen, welchen Hahn gestern die Ente verscheucht hat,
um sie voneinander zu trennen. (Object-first)
The farm hand from the farm must himself after the poultry look.
He wanted to know, which rooster [object] yesterday the duck chased-away has,
in order to separate them.
"The farm hand from the farm had to look after the poultry.
He wanted to know which rooster the duck chased away yesterday,
in order to separate them" (object-first interpretation)

Because the experimental sentence (e.g., *Der Knecht wollte wissen, welches Huhn gestern die Ente verscheucht hat*) could not always be presented on one line, we replaced the first NP (e.g., *Der Knecht*) with a pronoun to make it shorter and added a context sentence which contained this NP. In addition, a clause following the experimental item was constructed in order to make the whole item more natural as a small text.

The line with the experimental sentence consisted of a main clause and a subordinate clause, forming an indirect question. In the Subject-first condition, gender of the first noun in the subordinate clause (e.g., *Hahn*) was always masculine. Because *welcher* is the nominative masculine form of the wh-pronoun *welche*, the NP *welcher Hahn* is disambiguated toward the subject of the clause. Consequently, the word order in the subordinate clause is subject-object. Similarly, because *welchen* is the accusative masculine form, the NP is disambiguated toward the object of the clause in the Object-first condition and therefore the clause has object-subject word order. There is no ambiguous region: the disambiguation is immediate. In the Ambiguous condition the gender of the noun (e.g., *Huhn*) was always neutral. Because *welches* is ambiguous in case marking between nominative and accusative form, the NP is ambiguous between a subject and object reading. Because the case marking of the second NP (e.g., *die Ente*) is also ambiguous, and because the number of the auxiliary (e.g., *hat*) does not disambiguate either, the sentence remains ambiguous throughout. In order to avoid any confounds with length differences in the eyetracking experiment, the neutral and masculine nouns were equated for length. The noun was always followed by an adverb (usually a temporal adverb), a second NP (always with feminine gender), the main verb of the clause and an auxiliary (usually *hat*).

For the plausibility pre-test, 30 sets of paraphrase sentences such as (13) were derived from items such as (12) by turning the critical main clause into a passive.

13a. Die Ente ist von dem Huhn verscheucht worden.
The duck is by the chicken chased-away been.
"The duck has been chased away by the chicken."
(Ambiguous, subject-first paraphrase)

13b. Die Ente ist von dem Hahn verscheucht worden.
The duck is by the rooster chased-away been.
"The duck has been chased away by the rooster."
(Disambiguated, subject-first paraphrase)

13c. Das Huhn ist von der Ente verscheucht worden.
The chicken is by the duck chased-away been.
"The chicken has been chased-away by the duck."
(Ambiguous, object-first paraphrase)

13d. Der Hahn ist von der Ente verscheucht worden.
(Disambiguated, object-first paraphrase)
The rooster is by the duck chased-away been.
"The rooster has been chased-away by the duck."

These sentences are unambiguous paraphrases of the subject-object-verb relationships in the experimental materials. Sentences (13a-b) are paraphrases of the subject-object order interpretations whereas (13c-d) are paraphrases of the object-subject order interpretations. Sentences (13a) and (13c) were derived from the Ambiguous condition, and (13b) and (13c) from the disambiguated conditions. In contrast to the previous experiments, all paraphrases were derived from analyses that were constructed to be plausible.

Design. Four lists of items in a single random order were constructed, with each list containing all 120 paraphrase sentences. The conditions were rotated over the items so that each participant saw each condition of every item once. Care was taken that versions of the same item never appeared close to each other.

Procedure. The participants rated how plausible the situation described by the sentences was, on a 7-point scale, with 7 indicating very plausible, and 1 very implausible.

Table 7.1

Experiment 5: Item characteristics

	condition		
	ambiguous	subject-first	object-first
Plausibility pre-test:			
Plausibility subject-first paraphrase	5.4	5.5	5.5
Plausibility object-first paraphrase	5.2	5.4	5.4
Log frequency of noun following *welche*	1.79	1.87	1.87
Off-line preference task:			
Percentage of subject-first	87.3	96.8	8.7
Completion task:			
Percentage of subject-first completions	94.5		

Note. Log frequency is based on the number of occurrences in the Celex database (5.4 million words).

Results. Twenty-one item sets were selected on the basis of the plausibility pre-test. Table 7.1 shows the mean plausibility ratings for the selected materials. ANOVAs with the factors Paraphrase type (subject-first vs. object-first) and Condition (Ambiguous vs. Disambiguated) were conducted. Both factors were treated as within subjects and items. No significant effects were obtained for either Condition (F1 (1, 19) = 1.57, p = .23, MSe = .155; F2 (1, 20) = 2.82, p = .11, MSe = .093), Paraphrase type (F1 (1, 19) = 1.88, p = .19, MSe = .127; F2 (1, 20) = 1.20, p = .29, MSe = .219) or the interaction between the two (F1 (1, 19) = 2.30, p = .10, MSe = .060; F2 (1, 20) = 1.07, p = .31, MSe = .169). This indicated that the subject-first and object-first interpretations were equally plausible and that the plausibility of the conditions did not differ either.

The log for the lemma frequency of the first noun in the subordinate clause was calculated using the Celex German database. See Table 7.1 for the means. An ANOVA with Condition (Ambiguous vs. Disambiguated) as a within factor revealed no significant difference: F (1, 20) = .26, p = .62, MSe = .227. The log frequencies of *welcher, welchen and welches* are 2.02, 1.90 and 1.84 respectively.

7.3. Off-line preference pre-test

Participants. Eighteen participants[2], recruited from the German-speaking community at the University of Glasgow, took part in the experiment. All participants in the current pre-test, the completion pre-test and the eyetracking experiment were native speakers of German who had grown up in a German-speaking region (Germany, Austria or the German-speaking part of Switzerland). The participants in the off-line preference pre-test had been abroad for one day to 4:6 years, with an average of 9 months. Nobody participated in more than one study reported in this chapter.

Materials. The materials were 21 items like (12), selected on the basis of the plausibility pre-test. These materials were identical to those used in the eyetracking experiment. Each item was followed by the two paraphrases of the item that were used in the plausibility pre-test.

Design. Order of the paraphrase conditions was treated as a factor. Therefore, six lists of items were constructed (corresponding to the three lists in the eyetracking experiment), with one version of each item appearing in each list. The items occurred in a single random order in the lists. Three participants were assigned to each of the lists.

[2] An additional participant was excluded from the analyses because he chose the subject-object paraphrase on all trials, a pattern clearly different from all other participants. We presumed that he had not understood the task.

Procedure. Participants indicated whether the subject-first or object-first paraphrase best described the situation described by the sentence. As in the other preference pre-tests, they were asked to tick the option that corresponded to their first reaction to the sentence.

Results. Table 7.1 presents the percentage of times participants chose the subject-first paraphrase. The subject-first and object-first paraphrases always added up to 100%. As in the previous experiments, ANOVAs with the factor Condition (Ambiguous vs. Subject-first vs. Object-first) as within subjects and within items variables were conducted. A significant effect of this factor occurred: $F1\ (2, 34) = 259$, $p < .001$, $MSe = 162$; $F2\ (2, 40) = 343$, $p < .001$, $MSe = 143$. Planned comparisons showed that the ambiguous condition differed from both the subject-first condition ($F1\ (1, 34) = 5.04$, $p = .031$; $F2\ (2, 40) = 6.67$, $p = .014$) and the object-first condition ($F2\ (2, 34) = 342$, $p < .001$; $F1\ (2, 40) = 453$, $p < .001$).

The percentage of subject-first responses in the ambiguous condition (87%) was different from chance: $F1\ (1, 17) = 151$, $p < .001$, $MSe = 82.9$, $F2\ (1, 20) = 118$, $p < .001$, $MSe = 118$. This indicated that the materials were biased. Ungrammatical interpretations were adopted on only a small number of trials: 3% in the subject-first condition and 9% in the object-first condition. This indicated that case marking nearly always determined the final interpretation of the disambiguated conditions.

7.4. Completion pre-test

Participants. Fourteen participants from the same population as the off-line preference pre-test took part. The time they had spend abroad was between 4.5 months and 5 years, with a mean of 1:5 years.

Materials. Because disambiguation was immediate in the disambiguated conditions (at *welcher/n*), cutting off the experimental materials before the disambiguation would have resulted in very few completions starting with *welches/r/n* (cf. section 5.11). Therefore, it was decided to use the Ambiguous condition of the materials, and cut it off after the first NP in the subordinate clause (e.g., *Der Knecht auf dem Bauernhof mußte sich um das Federvieh kümmern. Er wollte wissen, welches Huhn ...*). Participants could either analyse this NP as a subject or object, and complete the sentence accordingly. Thus, the current completion study differs from those in Chapter 4 and 5, where the sentences

were cut off before the critical region. However, it is similar to the completion study in Kaan (1997), in which the sentence fragments also ended with the wh-phrase.

In addition to the experimental materials, 59 unrelated filler items were included, cut off at various points. The filler fragments were identical to the fillers in the eyetracking experiment.

Design. Because the items appeared only in the Ambiguous condition all participants received the same list. In this list, the materials appeared in a random order.

Procedure. The participants had to complete the sentence fragments in a booklet. They were asked to write down their first grammatical and meaningful response to the fragments. The booklets were scored by the experimenter.

Results. Table 7.1 presents the mean percentage of trials that the completion indicated that the ambiguous NP was taken as the subject of the subordinate clause. One trial was excluded from the analyses because it was ambiguous between subject-object and object-subject word order. All other trials were either grammatically unambiguous or (in a few cases) clearly disambiguated by semantic plausibility. ANOVAs revealed that the 95% of subject-first responses differed significantly from chance: $F1\ (1, 13) = 1161$, $p < .001$; $MSe = 12.0$; $F2\ (1, 20) = 273$, $p < .001$, $MSe = 76.3$. This confirmed the conclusion from the off-line preference study that there is a strong preference for the subject-first interpretation. The result is consistent with the completion study that was conducted by Kaan (1997) on very similar sentence fragments in Dutch. She found a 96% subject-first preference.

7.5. Predictions

On the basis of the off-line preference and completion pre-test, which both showed a strong subject-first preference, we can derive predictions for constraint-based theories and the unrestricted race model. Because there is a subject-first bias, both theories predict that the subject-first condition should be easier to process than the object-first condition. In constraint-based theories, the object-first condition should be hard to process, because there is a competition between the subject-first analysis, which is supported by the subject-first bias, and the object-first analysis, which is supported by case marking. It is unclear how strong this competition is. In some models (e.g., McRae et al., 1998; Spivey & Tanenhaus, 1998), it should not be very strong, because the object case marking is nearly 100% effective (as indicated by the off-line preference

pre-test). In the unrestricted race model, the object-first condition should be more difficult than the subject-first condition because the processor nearly always initially adopts the subject-first analysis. Therefore, reanalysis should occur on nearly all trials in the object-first condition, because the case marking of *welchen* is inconsistent with it. In contrast, the subject-first condition should be easy, because the initial subject-first analysis is consistent with case marking.

The off-line studies indicated that the subject-first bias was very strong. Therefore, there may be very little difference between the subject-first condition and the ambiguous condition. Constraint-based theories predict that in the ambiguous condition, the subject-first analysis is much more strongly activated than the object-first analysis as a result of the subject-first bias, even though the ambiguous case marking of *welches* equally supports both analyses. Therefore, there may be very little competition between the two analyses in the ambiguous condition, and it may not be read more slowly than the subject-first condition, where both case marking and the subject-first bias support the same analysis. However, if a difference between the ambiguous and subject-first condition can be observed, the ambiguous condition should always be the slower one, because it is predicted that that there should be more competition in this condition than in the subject-first condition.

The unrestricted race model also predicts very little difference between the ambiguous and subject-first condition. Because of the strong bias, the processor will nearly always initially adopt the subject-first analysis and as a result, reanalysis should occur on very few occasions in the subject-first condition. Therefore, it may not be more difficult than the ambiguous condition. If there is any difference, however, the ambiguous condition should be slower, because reanalysis should never occur here, whereas reanalysis may occur on some proportion of trials in the subject-first condition.

Previous experiments on biased ambiguities have shown that ambiguous conditions were as easy to read as conditions where the bias was consistent with the disambiguation. Traxler et al. (1998) observed this pattern in relative clause attachment ambiguities in which the NP complex contained *with*. Van Gompel et al. observed a similar pattern in VP/NP attachment ambiguities such as (14): First-pass regressions and regression path times for *not long after* indicated that the ambiguous condition was as easy to read as the VP attachment condition, and that the NP attachment condition was more difficult.

14a. The hunter killed the dangerous poacher with the rifle not long after sunset. (ambiguous)

14b. The hunter killed the dangerous leopard with the rifle not long after sunset. (VP attachment)

14c. The hunter killed the dangerous leopard with the scars not long after sunset. (NP attachment)

However, there was some suggestion that the VP attachment condition was also somewhat more difficult to read than the ambiguous condition. Regression-path times on the final word (*sunset*) were longer for the VP attachment condition than the ambiguous condition. Although this effect occurred at the end of the sentence and may therefore mainly reflect clause wrap-up effects, it may be taken as evidence that readers initially adopted the VP attachment analysis on a small proportion of trials, and had to reanalyse in the VP attachment condition.

A similar pattern was observed by Scheepers (1997), who investigated biased subject-object word order ambiguities in German. He tested sentences such as (15).

15a. Die gebildete Frau mochte den klugen Freund, so wurde erzählt.
(ambiguous NP1, subject-first)
The intellectual woman [subject/object] liked the clever friend [object], it was said.
"The intellectual woman liked the clever friend, it was said."

15b. Die gebildete Frau mochte der kluge Freund, so wurde erzählt.
(ambiguous NP1, object-first)
The intellectual woman [subject/object] liked the clever friend [object], it was said.
"The clever friend liked the intellectual woman, it was said."

15c. Der gebildete Mann mochte den klugen Freund, so wurde erzählt.
(subject-first case marked NP1)
The intellectual man [subject] liked the clever friend, it was said.
"The intellectual man liked the clever friend, it was said."

15d. Den gebildeten Mann mochte der klugen Freund, so wurde erzählt.
(object-first case marked NP1)
The intellectual man [object] liked the clever friend, it was said.
"The clever friend liked the intellectual man, it was said."

15e. Man erzählte, daß die gebildete Frau den klugen Freund mochte.
(ambiguous NP1, subject-first)
One said, that the intellectual woman [subject/object] the clever friend [object] liked.
"It was said that the intellectual woman liked the clever friend."

15f. Man erzählte, daß die gebildete Frau der klugen Freund mochte.
(ambiguous NP1, object-first)
One said, that the intellectual woman [subject/object] the clever friend [object] liked.
"It was said that the clever friend liked the intellectual woman."

15g. Man erzählte, daß der gebildete Mann den klugen Freund mochte.
(subject-first case marked NP1)
One said, that the intellectual man [subject] the clever friend liked.
"It was said that the intellectual man liked the clever friend."

15h. Man erzählte, daß den gebildeten Mann der klugen Freund mochte.
(object-first case marked NP1)
One said, that the intellectual man [object] the clever friend liked.
"It was said that the clever friend liked the intellectual man."

In (15a) and (15b), the first NP *die gebildete Frau* is ambiguously case marked by the definite article *die.* It can either be a subject or object. Both sentences are disambiguated by case marking of the second NP *der/den kluge(n) Freund.* Sentence (15a) is disambiguated toward subject-first, and (15b) toward object-first by the second NP. In (15c) and (15d), the first NP is unambiguously case marked by *der/den.* In (15c), it is subject case marked, and in (15d) object case marked. The manipulation in (15e-h) is the same as in (15a-d), but the word order ambiguity is in a subordinate clause instead of in a main clause.

First-pass times at the first NP showed that the object-first conditions took longer to read than the subject-first conditions, indicating that there was a subject-first preference. This was the case both for main clauses and subordinate clauses, although the difference was more pronounced for subordinate clauses. Interestingly, the ambiguous conditions took less time to read than the subject-first conditions, although this effect did not reach significance in the standard first-pass measure. However, when trials with first-pass regressions from the first NP region were removed (Altmann et al, 1992; Altmann, 1994; Rayner & Sereno, 1994a; 1994b), the first-pass difference between the ambiguous and subject-first conditions did reach significance[3]. This

[3] Note that the regression-contingent measure was the same as the standard first-pass measure for (15a-d), because the critical region in (15a-d) was the first region. Regressions could only occur in (15e-h).

suggests that readers may have adopted the non-preferred object-first analysis on a small proportion of trials. However, it should be noted that the results were observed in a not commonly used measure. Furthermore, Scheepers did not control for word length (although the use of length corrected reading times may have removed any differences due to this factor), or for frequency or plausibility. Therefore, it is important to replicate the observed pattern using more tightly controlled materials and a more commonly used eyetracking measure.

Scheepers's study showed the earliest effects in first-pass reading times at the first NP1. We may expect a similar pattern in the current study, because the type of ambiguity is similar (though the current experiment used wh-phrases instead of definite NPs). However, this is not the type of pattern that is predicted by so-called *head-driven parsers* (e.g., Ford et al., 1982; Gorrell, 1995; MacDonald et al., 1994; Pritchett, 1988; 1992). Unless additional assumptions are made (e.g., Gorrell, 1996), such models assume that syntactic structure building does not occur until the head of a phrase has been reached. Because the verb is normally considered as the head of a clause, structure building in verb final constructions such as the ambiguity tested in the current experiment cannot occur until the verb is reached. When NP1 and NP2 are read, they remain loose NP nodes until the verb is processed. NP1 and NP2 are not analysed as either the subject or object of the clause until the verb, and no effect of word order can be observed until this point. As a result, head-driven parsers predict that no effects occur before the verb.

7.6. Eyetracking Experiment 5

Method

Participants. Thirty-seven participants took part. They were recruited from the same population as the off-line preference and completion pre-test, with the additional restriction that they did not wear glasses. The time they had spent abroad was between one day and 6:6 years, with a mean of 8 months. Two participants were excluded because they had a very high percentage of tracker losses, and two because they answered more than 20% of the questions incorrectly. The analyses were conducted on the remaining 33 participants.

Materials. The materials were the same as in the off-line preference pre-test. The fillers were the same as in the completion study, finished off to make complete sentences.

Design. Three lists of items, consisting of seven items from each condition were constructed. One version of each item appeared in each list. Eleven participants were assigned to each list. Each list also contained 59 filler sentences. The experimental and filler items were placed in a single random order, with 5 fillers preceding the first experimental sentence, and 3 fillers following a break in the experiment. All items were followed by a comprehension question, with half of them requiring a "yes" answer and half a "no" answer.

Procedure. The procedure was the same as in the experiments in the preceding chapters. All experimental materials were presented on three lines. The context sentence was presented on the first line, the experimental sentence (e.g., *Er wollte wissen, welches Huhn gestern die Ente verscheucht hat*) on the second line, and the final clause on the third line.

Analyses. On average, the participants made 10% errors on the comprehension questions. As a result of major tracker losses, 3% of the trials were discarded. Less than 1% of the data was removed as a result of two consecutive regions having zero reading times.

Because initial analyses on the raw reading times only showed weak effects, length-adjusted reading times were calculated (e.g., Ferreira & Clifton, 1986; Trueswell et al., 1994). This method adjusts both for length differences between items and for individual differences in reading speed, which generally results in more power. In order to calculate the length-adjusted reading times, twenty-one filler sentences were divided into regions for which first-pass times were calculated. Using these first-pass times, regression equations with length of a region as a predictor of first-pass reading times were calculated for each participant. This resulted a predicted reading time for each participant for each region length. Subsequently, the length-adjusted reading times for the experimental materials were calculated by subtracting the predicted reading time from the actual reading time for the particular region of the particular trial.

The materials were divided into 7 regions, indicated by slashes below:

16. Er wollte wissen,/ welches/ Huhn/ gestern/ die Ente/ verscheucht hat,/
um sie voneinander zu trennen./

The regions corresponded to (1) the main clause, (2) the wh-pronoun (the critical region), (3) the first noun (N1, the post-critical region), (4) the adverb, (5) the second NP (NP2), (6) the verbs, and (7) the final clause. The first line of the materials was not analysed, because it preceded the experimental manipulation.

Table 7.2

Experiment 5: Means

	region						
	1	2	3	4	5	6	7
First-pass reading times:							
ambiguous	687 (211)	291 (-15)	356(-32)	303 (-40)	412 (-11)	698 (227)	1732 (616)
subject-first	687 (215)	317 (12)	344 (-49)	319 (-27)	425 (-3)	662 (190)	1797 (671)
object-first	683 (208)	319 (14)	351 (-40)	339 (-4)	407 (-19)	684 (214)	1733 (618)
First-pass regressions out:							
ambiguous	1.4	4.6	17.7	7.6	10.5	37.6	64.0
subject-first	3.2	3.6	18.8	10.8	9.1	37.8	64.1
object-first	1.9	6.6	26.7	14.2	11.7	49.8	68.2
Regression-path times:							
ambiguous	698 (221)	331 (25)	415 (26)	348 (5)	468 (46)	1130 (659)	3438 (2322)
subject-first	706 (234)	343 (38)	427 (34)	396 (50)	491 (64)	1144 (672)	3504(2377)
object-first	704 (229)	368 (63)	456 (65)	462 (119)	500 (74)	1136 (966)	3532 (2417)
Total times:							
ambiguous	875 (397)	473 (168)	568 (188)	541 (200)	738 (317)	974 (502)	2067 (944)
subject-first	901 (430)	513 (208)	576 (195)	566 (221)	725 (302)	949 (478)	2074 (953)
object-first	916 (442)	687 (382)	645 (265)	680 (338)	709 (364)	1040 (568)	2051 (932)

Notes. The regions were defined as follows (delimited by slashes):
Er wollte wissen,/ welches/ Huhn/ gestern/ die Ente/ verscheucht hat./
um sie voneinander zu trennen.

First pass, regression path, and total times are reported in ms, with length-corrected times in brackets. First pass regressions is the percentage of saccades leaving the region to the left after a first pass fixation.

Table 7.3

Experiment 5: ANOVA results by region for effect of condition

	F1 (2, 64)	MSe	F2 (2, 40)	MSe
Region 1:				
First pass reading times	< 1		< 1	
First pass regressions out	1.50	22.6	1.11	18.7
Regression path times	< 1		< 1	
Total reading times	< 1		< 1	
Region 2:				
First pass reading times	4.33*	1780	3.38*	1549
First pass regressions out	1.01	77.9	1.17	51.1
Regression path times	1.32	9462	1.63	5722
Total reading times	15.6**	23305	22.3	11853
Region 3:				
First pass reading times	< 1		< 1	
First pass regressions out	3.27*	263	2.98	201
Regression path times	1.46	8050	1.40	6921
Total reading times	2.70	16554	3.22	12067
Region 4:				
First pass reading times	5.56**	2078	6.12**	1000
First pass regressions out	1.79	208	2.75	95.4
Regression path times	5.07**	21939	5.63**	12408
Total reading times	13.5**	12901	12.3**	9337
Region 5:				
First pass reading times	< 1		< 1	
First pass regressions out	< 1		< 1	
Regression path times	< 1		< 1	
Total reading times	< 1		1.20	17334
Region 6:				
First pass reading times	< 1		< 1	
First pass regressions out	5.90**	286	6.73**	155
Regression path times	7.96**	119876	8.64**	75043
Total reading times	1.94	31664	1.63	28760
Region 7:				
First pass reading times	< 1		< 1	
First pass regressions out	< 1		< 1	
Regression path times	< 1		< 1	
Total reading times	< 1		< 1	

Note. * is significant at the .05 level, ** is significant at the .01 level.

Results

Table 7.2 presents the mean reading times and percentage regressions by condition and region. For each eyetracking measure for each region, one-way ANOVAs with three levels of Condition (ambiguous vs. subject-first vs. object-first) were conducted. Condition was treated as a within subjects and items factor. Table 7.3 presents the results of these analyses. Planned comparisons between levels are reported below.

In contrast to the previous experiments, significant effects were obtained in first-pass reading times. In the critical region (*welches/r/n*), first-pass reading times for the ambiguous condition were faster than for the subject-first condition (F1 (1, 64) = 6.79, p = .011; F2 (1, 40) = 4.85, p = .033) and for the object-first condition (F1 (1, 64) = 6.18, p = .016; F2 (1, 40) = 5.28, p = .027). No difference was observed between the two disambiguated conditions (F1 (1, 64) = .02, p = .90; F2 (1, 40) = .01, p = .92). No significant first-pass effect occurred at the next region (N1), but at region 4 (the adverb) the object-first condition was slower than the both ambiguous condition (F1 (1, 64) = 10.9, p = .002; F2 (1, 40) = 11.9, p = .001) and subject-first condition (F1 (1, 64) = 4.29, p = .042; F2 (1, 40) = 5.09, p = .030). The difference between the ambiguous and subject-first condition did not reach significance: F1 (1, 64) = 1.51, p = .22; F2 (1, 40) = 1.41, p = .24. No significant first-pass effects occurred in the other regions.

In first-pass regressions, the condition effect reached significance in the post-critical region (N1). There were more regressions in the object-first condition than in the ambiguous condition (F1 (1, 64) = 4.98, p = .029; F2 (1, 40) = 4.28, p = .045) and subject-first condition (F1 (1, 64) = 4.84, p = .031; F2 (1, 40) = 4.63, p = .038) and no difference between the ambiguous and subject-first conditions (F1 (1, 64) < .01, p = .98; F2 (1, 40) = .01, p = .93). The same pattern was observed in first-pass regressions from region 6 (the verbs): the object-first condition produced more regressions than the ambiguous condition (F1 (1, 64) = 8.93, p = .004; F2 (1, 40) = 10.66, p = .002) and subject-first condition (F1 (1, 64) = 8.77, p = .004 F2 (1, 40) = 9.49, p = .004) with no difference between ambiguous and subject-first conditions (F1 (1, 64) < .01, p = .98; F2 (1, 40) = .03, p = .85).

Regression-path times showed the same pattern as first-pass times. The ambiguous condition was read faster than the object-first condition in region 4 (F1 (1, 64) = 10.1, p = .002; F2 (1, 40) = 11.1, p = .002) and region 6 (F1 (1, 64) = 12.3, p < .001; F2 (1, 40) = 13.3, p < .001). The subject-first condition was also faster than the object-first condition in region 6 (F1 (1, 64) = 11.6, p = .001; F2 (1, 40) = 12.6, p = .001), but the effect was only marginally significant in region 4 (F1 (1, 64) = 3.30, p = .074; F2 (1, 40) = 4.00, p = .052). The ambiguous and subject-first condition did not differ in either region (region 4: F1 (1,

64) = 1.84, p = .18; F2 (1, 40) = 1.78, p = .19; region 6: F1 (1, 64) < .01, p = .92; F2 (1, 40) = .01, p = .92).

Total reading times showed that the object-first condition took longer to read than the ambiguous (F1 (1, 64) = 27.5, p < .001; F2 (1, 40) = 39.7, p < .001) and subject-first condition (F1 (1, 64) = 18.3, p < .001; F2 (1, 40) = 25.6, p < .001) in the critical region. There was no difference between the ambiguous and subject-first condition (F1 (1, 64) = .92, p = .34; F2 (1, 40) = 1.54, p = .22). The same pattern was observed in the post-critical region, but the condition effect was only marginally significant (p = .075 by subjects; p = .51 by items). The effect reached significance again in region 4. Again, the object-first condition was read more slowly than the ambiguous condition (F1 (1, 64) = 23.3, p < .001; F2 (1, 40) = 21.5, p < .001) and the subject-first condition (F1 (1, 64) = 16.6, p < .001; F2 (1, 40) = 14.8, p < .001), with no difference between the latter two conditions (F1 (1, 64) = .56, p = .46; F2 (1, 40) = .62, p = .43). No other effects reached significance in the measures in any of the regions.

In order to check whether the subject-first preference depended on how long the participants had spent abroad, a correlational analysis was conducted. This analysis correlated the difference between the subject-first and object-first condition in first-pass reading times for regions 2-4 with the time the participants had spent abroad. No significant effect was observed in any of the regions (all rs < .14, ps > .22).

7.7. Discussion

The eyetracking experiment showed the earliest effects in first-pass reading times at the critical region. This result is quite different from the other experiments in this thesis, which observed the earliest effects in first-pass regressions and regression-path times at the post-critical region. However, the results are similar to those in Scheepers (1997), who also investigated subject-object word order ambiguities. This suggests that eye-movement patterns for such ambiguities are different from those for relative clause and VP/NP attachment ambiguities (see further section 8.2).

The condition effect occurred long before the verb at the end of the clause was reached. This disconfirms the predictions of head-driven parsers, which claim that syntactic structure building cannot take place before the head of a phrase is reached. Because the verb is the head of the clause, the NPs cannot be attached until this point, and therefore no effect of condition can occur until the verb is reached. Interestingly, an effect of condition occurred even before the head noun (N1) of the determiner *welche* was

reached. This suggests that the processor can assign case marking to an NP before the noun is reached.

The data showed a clear advantage for the subject-first condition over the object-first condition. This preference was not correlated with the time the participants had spent abroad in an English speaking environment. Therefore, it is unlikely that the subject-first preference that they displayed was due to the fact that the object-first word order is extremely infrequent in English. The subject-first preference is consistent with the off-line preference and completion pre-test, which both showed a strong subject-first preference. The earliest eyetracking measure in which this effect was obtained was in first-pass regressions from the post-critical region. The same effect also occurred in first-pass times at the adverb and in regression-path and total times in various regions. The subject-first preference is inconsistent with Gorrell's (1996) prediction that subject-first and object-first word order should be equally preferred. The results also differ from Kaan's (1997) experiments, who showed an object-first experiment in most of her experiments with wh-phrases. However, in most of Kaan's experiments, the second NP was a pronoun, which may have caused the object-first preference that was observed later in her sentences. The one experiment that investigated second NPs that were definite found a marginally significant subject-first preference.

Although there was a strong preference for the subject-first analysis, first-pass reading times for the critical region *welches/r/n* showed that the subject-first condition took longer to read than the ambiguous condition. This suggests that on some proportion of trials, readers initially adopted the object-first interpretation and had to reanalyse this analysis in the subject-first condition. This ambiguous vs. subject-first effect was not very strong: no other measures showed a significant difference between the ambiguous and subject-first analysis. In fact, this is what is predicted by the unrestricted race model. Because there was a strong subject-first preference, the object-first analysis should be adopted on very few trials, and therefore reanalysis should occur on few trials in the subject-first condition. The finding is consistent with Scheepers (1997), who also showed a relatively small first-pass reading time advantage for ambiguous conditions over subject-first conditions. The results are also consistent with Van Gompel et al. (1999), who also showed a rather small difference between the ambiguous condition and the condition that was disambiguated toward the preferred analysis. However, they differ from the biased experiment in Traxler et al. (1998), who did not find such a difference. Possibly, the effect was too small to be detected in this study. Because relative clauses are not part of the main proposition of the sentence, they may be processed relatively shallowly, and therefore any ambiguity effects may be relatively weak.

The faster reading times for the ambiguous condition compared to the subject-first condition are inconsistent with constraint-based theories. These theories predict that if there are any differences in reading times, the ambiguous condition should always be slower, because the ambiguous case-marking causes competition, whereas case marking supports the preferred analysis in the subject-first condition. The present experiment employed immediate disambiguation. Therefore, the results also disconfirm versions of competition models which assume that selection of one analysis occurs if an ambiguous region precedes the point of disambiguation. The finding that no competition occurred even when the disambiguation was immediate is consistent with the results from Experiment 4 (Chapter 6), which also employed immediate disambiguation. It further supports the conjecture that the competition-like effect in the critical region in Experiment 3 (Chapter 5) was a spurious effect.

Although the active filler strategy (e.g., Frazier, 1987b) and the minimal chain principle (De Vincenzi, 1991) correctly predict the subject-first preference that was observed, these theories also have difficulty explaining the pattern of first-pass reading times at *welches/r/n*. Both theories predict that readers should always initially adopt the subject-first analysis. The difference between the ambiguous and subject-first condition suggests that this was not the case.

In fact, the first-pass reading times at *welches/r/n* did not show a difference between the subject-first and object-first condition. This is rather surprising, given that there was such a strong preference for the subject-first analysis in later measures and in the off-line pre-tests. One possible explanation is that it may be more difficult to reanalyse from object-first to subject-first than the other way around. Readers may have adopted the object-first analysis on only very few trials, but if they did, reanalysis to subject-first was extremely difficult. In contrast, the subject-first analysis was adopted very often, but reanalysis to object-first was relatively easy, so that on average, reading times were no longer than in the object-first condition. This explanation seems somewhat unlikely, however. Later measures did not provide any evidence that reanalysis from object-first to subject first was difficult, as the subject-first condition did not differ from the ambiguous condition.

Alternatively, it may be that the off-line studies did not reflect initial preferences for the materials, but instead reflected later preferences. Initially, neither word order may be preferred, but a subject-first preference might develop. This late subject-first preference may be reflected in both the off-line studies and the later eyetracking measures, whereas the earlier no-preference may be reflected in first-pass times at *welches/r/n*. Indeed, some researchers have argued that off-line measures such as completions and the off-line

preference task do not accurately predict initial processing preferences (e.g., Clifton et al., 1997; De Vincenzi & Job, 1995; Pickering et al., 1999). However, other studies have shown that off-line preference do correctly predict on-line preferences (e.g., McRae et al., 1998; Garnsey et al., 1997; Trueswell et al., 1993).

Finally, it is possible that the null-difference between the ambiguous condition and the subject-first condition merely reflects a Type II error. This would be more consistent with the other experiments that investigated the competition/reanalysis question, in which the preferences as measured by the off-line studies were consistent with the preferences in the eyetracking experiments. It would also be more in line with other experiments on subject-object word order ambiguities (e.g., Scheepers, 1997; Konieczny, 1996) which found an early subject-first advantage.

8. Conclusions

8.1. Summary of the findings

The experiments in this thesis were conducted to investigate the architecture underlying syntactic processing. Specifically, they investigated whether syntactic ambiguity resolution involves a process of competition, as claimed by most current constraint-based theories, or a process of reanalysis, a claimed by two-stage theories. This issue was addressed by comparing globally ambiguous syntactic structures in which there was no preference for one analysis over the other (balanced ambiguities) with disambiguated structures. Competition models (e.g., MacDonald et al., 1994; McRae et al., 1998; Spivey & Tanenhaus, 1998; Tabor et al., 1997) predict that competition should occur in globally ambiguous structures that are balanced, because two syntactic structures are equally activated. In contrast, no such competition should occur in disambiguated structures, because the disambiguating information (e.g., plausibility) supports only one analysis. Hence, competition models predict that balanced, globally ambiguous structures are more difficult to process than disambiguated structures. Two-stage models make different predictions. They claim that reanalysis of an initially adopted structure causes processing difficulty. Basically, there are two types of two-stage models. One type of model such as the garden-path model (e.g., Frazier, 1987; Frazier & Rayner, 1982; Rayner et al., 1983) predicts that one analysis of an ambiguity should always be preferred over the alternative(s). In the garden-path model, structural information determines the preferred analysis, and this analysis is adopted in the first stage of processing. If the initial analysis is incompatible with non-structural information, reanalysis occurs. But if non-structural information is consistent with the structurally preferred analysis, reanalysis should not occur, because the structurally non-preferred analysis is not adopted. The second type of two-stage model claims that either analysis of an ambiguous structure can be adopted first. In the unrestricted race model that was proposed in section 3.5 (see also Van Gompel et al., 1999; in press) the number of times that an analysis is adopted depends on the bias of the ambiguity. If there is a strong preference for one analysis (a biased ambiguity), this analysis is nearly always adopted. But if there is no preference (a balanced ambiguity), each analysis is adopted about half the time. This predicts that for the latter type of ambiguity, reanalysis should occur on about half the trials regardless of which way the sentence is disambiguated. All two-stage theories predict that globally ambiguous sentences should be easy to process. Because the initial analysis cannot be disconfirmed by disambiguating information, reanalysis should never occur in ambiguous sentences.

Experiment 1 (Chapter 4) and Experiment 2 (Chapter 5) both confirmed the predictions of the unrestricted race model. Experiment 1 investigated a balanced VP/NP attachment ambiguity, and Experiment 2 a relative clause attachment ambiguity. In both experiments, the ambiguous condition caused less processing difficulty than the disambiguated conditions, which did not differ from each other. This pattern of results is incompatible with competition models, which predict that the ambiguous condition should be more difficult to read than the disambiguated conditions. It is also difficult to reconcile with models such as the garden-path model, which claim that the processor should always adopt the same analysis. Instead, the results indicated that it adopted each analysis about equally often.

In Chapter 5, the possibility was raised that competition models might be able to account for the results in these experiments by assuming that selection of one analysis can take place very rapidly. Because both Experiment 1 and 2 had an ambiguous region preceding the disambiguation, it might have been the case that competition occurred in the ambiguous region, but that by the time the disambiguation was reached, one analysis had been selected. If the disambiguation was inconsistent with the selected analysis, the alternative had to be reactivated, and this might have resulted in processing difficulty. Experiment 3 (Chapter 5), 4 (Chapter 6) and 5 (Chapter 7) tested this hypothesis by using ambiguities that did not have an ambiguous region. Experiment 3 and 4 employed relative clause attachment ambiguities, while Experiment 5 investigated a word-order ambiguity in German. Although the results from Experiment 3 were somewhat equivocal, Experiment 4 and 5 showed that the ambiguous conditions caused less processing difficulty than the disambiguated conditions. Hence, they disconfirmed the early competition hypothesis.

The conclusion that can be drawn from these experiments is that the processor works in two stages. In the first stage, the processor selects one analysis. In the second stage, this information is checked against the disambiguating information (e.g., plausibility or case marking). If this information disconfirms the initial analysis, reanalysis occurs and processing difficult ensues. Experiment 4 (Chapter 6) investigated the selection process in the first stage by using unambiguous relative clause attachment sentences. It was hypothesised that the first-stage selection process might involve a form of competition which was termed structural competition. Under this hypothesis, structural competition is predicted whenever there are two possible syntactic structures (even if only one is plausible). If this were the case, the ambiguous condition should be more difficult than the unambiguous condition. Alternatively, the initial selection process might involve a race, as predicted by the unrestricted race model. This predicts that the ambiguous condition should be no more difficult than the unambiguous condition. This latter prediction was confirmed by Experiment 4: the results showed

that ambiguous sentences were no more difficult to process than unambiguous sentences, while both were easier than disambiguated sentences.

Finally, Experiment 5 (Chapter 7) contrasted reanalysis with competition using a subject-object word-order ambiguity in German. This type of structure was very different from the other ambiguities tested in this thesis. In addition, the disambiguation was by means of case marking rather than by plausibility as in the other experiments. The results from this experiment were again consistent with the unrestricted race model. Because the ambiguity was biased toward the subject-first analysis, the condition that was disambiguated toward the object-first analysis was more difficult to read than the subject-first condition. The object-first condition was also more difficult to read than the ambiguous condition. In addition, reading times at the disambiguation suggested that the subject-first condition was also somewhat more difficult to read than the ambiguous condition. This suggests that the non-preferred object-first reading was initially adopted on a small proportion of trials.

The experiments in this thesis suggest that syntactic ambiguity resolution is fundamentally different from lexical ambiguity resolution. As described in section 3.2, several eyetracking studies on lexical ambiguity resolution have found that balanced ambiguous words (in which both meanings are about equally frequent) are more difficult to process than unambiguous words that are controlled for frequency when they occur in neutral contexts (e.g., Duffy et al., 1988; Rayner & Duffy, 1986). This suggests that competition occurs between the meanings of balanced words. Furthermore, biased words (in which one meaning is much more frequent than the other) that are preceded by a context that supports the infrequent meaning take longer to read than unambiguous words (e.g., Duffy et al., 1988). This may be taken as evidence that both context and frequency activate the meanings of an ambiguous word. When context and frequency support different meanings, competition occurs. Current constraint-based theories (e.g., MacDonald et al., 1994; Trueswell et al., 1994; Trueswell, 1996) claim that syntactic ambiguity resolution is a form of lexical ambiguity resolution, and that therefore, competition should also occur in syntactic ambiguity resolution. The experiments in this thesis provide strong evidence that this claim is incorrect, because their results were inconsistent with competition as a mechanism of syntactic ambiguity resolution. Hence, for current constraint-based theories to be plausible theories of syntactic ambiguity resolution, at least two fundamental assumptions have to be abandoned. They would have to give up competition as a mechanism of syntactic ambiguity resolution, and they would have to abandon the assumption that lexical and syntactic ambiguity resolution involve similar processes. Although this may be possible in principle (e.g., Jurafsky 1996; Pearlmutter

& Mendelsohn, 1999) it leads to a fundamentally different model of syntactic ambiguity resolution than is currently assumed by constraint-based theories.

The present results are consistent with claims that lexical and syntactic ambiguity resolution are different, as argued by proponents of two-stage models (e.g., Frazier, 1989; Rayner & Morris, 1991; Traxler et al, 1998; Van Gompel et al., 1999; in press). Competition may occur in lexical ambiguity resolution, because the meanings of ambiguous words are activated in parallel (e.g., Swinney, 1979; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982), but in syntactic ambiguity resolution syntactic structures are constructed in a serial fashion, and therefore no competition can arise. Instead, syntactic ambiguity resolution involves reanalysis when the initially adopted analysis is inconsistent with information that is used later in sentence processing.

The experiments in this thesis do not support all two-stage models. They showed that in balanced ambiguities, both disambiguated conditions were more difficult to process than the ambiguous conditions, with no difference between the disambiguated conditions. This is difficult to reconcile with models such as the garden-path model, which predict that the structurally least complex analysis is first adopted and should be preferred. Instead, the results from the balanced ambiguities suggested that the processor adopted each analysis about half the time, and that reanalysis occurred about equally often when the ambiguity was disambiguated toward one analysis or the other.

Such a pattern for balanced ambiguities is predicted by the unrestricted race model. In the unrestricted race model, the possible analyses of a syntactic ambiguity are engaged in a race to be constructed first. The analysis that is first constructed is selected. In balanced ambiguities, syntactic and non-syntactic sources of information support two analyses about equally, and therefore, each analysis is predicted to win the race about half the time. When the sentence is disambiguated, this results in reanalysis on about half of the trials. Consequently, both disambiguated conditions in balanced ambiguities cause difficulty. Both are more difficult to process than ambiguous (and unambiguous) sentences, where reanalysis does not occur. This pattern contrasts with that for biased ambiguities. Here, the preferred analysis is nearly always adopted initially, and the condition that is disambiguated toward the preferred analysis should be little or no more difficult than the ambiguous condition.

8.2. Late processing?

Is there an alternative way to account for the data? One possibility that has to be considered was mentioned in section 4.9. It might be argued that the results of the

eyetracking experiments in this thesis were obtained in relatively late measures of processing and that for this reason, the current experiments do not allow us to make conclusions about initial processing. The earliest results for Experiment 1-4 consistently occurred in first-pass regressions or regression-path times for the post-critical region. This contrasts with a number of other studies on syntactic ambiguity resolution, which found the earliest effects in first-pass regressions at the critical, disambiguating region (e.g., Ferreira & Clifton, 1986; Rayner et al., 1983; Trueswell et al., 1993; 1994). However, the critical region of nearly all eyetracking studies that found effects in first-pass reading times was longer than one word. This is different from the critical regions in the present experiments, which always consisted of a single, often relatively short, word. As was argued in section 4.7, eyetracking measures that are based on long regions tap into later processing than measures that are based on short regions, and therefore the first-pass effects in other studies may reflect processes that are no earlier than the effects that were observed in experiments of this thesis. It is plausible to assume that the effects in the post-critical region in Experiment 1-4 were obtained because readers made a forward saccade from this region before they had had time to evaluate the plausibility of the syntactic analysis. For example, Reichle et al.'s (1998) model of eye-movement control stipulates that forward saccades to the next word can occur even before lexical access has taken place. Given that evaluating the plausibility of a syntactic analysis requires syntactic structure building and accessing world knowledge in addition to lexical access, it actually seems unlikely that the effects could have occurred in the critical region. In the light of this, the effects are in fact quite early. The high percentage of regressions for the disambiguated conditions in the post-critical region of Experiment 1-4 suggests that readers discovered the implausibility of their analysis in the region following the disambiguating word and immediately went back to recover from their misanalysis.

Another difference between Experiment 1-4 and most (though not all) studies that obtained first-pass effects is the type of ambiguity under investigation. Studies that found effects in first pass usually investigated ambiguities that were different from the VP/NP and relative clause attachment ambiguities tested in Experiment 1-4. As argued by some researchers (e.g., Britt et al., 1992; Rayner et al., 1992; Sturt & Crocker, 1996), ambiguities such as the reduced relative/main clause ambiguity which usually show first-pass effects may show stronger garden-path effects than for example the VP/NP attachment ambiguity. In addition, first-pass effects are often observed in cases of syntactic disambiguation, whereas the sentences in Experiment 1-4 were disambiguated by plausibility. Syntactic information arguably represents a stronger indication that a syntactic analysis is incorrect than plausibility information, and furthermore, it also seems likely that syntactic information can be used more quickly than plausibility

information, which requires evaluation of world-knowledge. Taken together, this suggests that the studies that found first-pass effects may have used ambiguities that produced stronger effects than the ambiguities in Experiment 1-4. This may explain why effects were found in so-called later measures in Experiment 1-4. As argued in section 4.7, such measures may reflect weak processes rather than late processes.

Interestingly, Experiment 5 showed a very different pattern of eye-movements. Significant effects were observed in first-pass times at the critical region, which consisted of a single short word as in the other experiments. However, even though the effects occurred very early on, there was no evidence for competition whatsoever. In fact, the opposite was the case. Hence, the evidence against competition does not entirely depend on the first-pass regressions and regression-path effects in Experiment 1-4. The ambiguity of Experiment 5 was completely different from the VP/NP and relative clause attachment ambiguities tested in the other experiments. Firstly, the disambiguation was by case marking instead of plausibility. Because case marking categorically ruled out one of the analyses in the disambiguated conditions, it may have led to stronger, earlier effects than disambiguation by plausibility, which does not provide such a categorical cue. Secondly, the ambiguity involved ambiguous arguments instead of ambiguous adjuncts in the other experiments. Because arguments constitute the main proposition of the sentence, this may have caused readers to process the critical part of the sentence more deeply than in the case of adjuncts, which are not part of the main proposition. Again, this may have led to stronger and earlier effects in Experiment 5. Finally, the pattern of eye-movements in Experiment 5 may be different from that in the other experiments because in the word-order ambiguity, information preceding the disambiguation was not very informative for recovering from a misanalysis. The main clause that preceded the disambiguation at the wh-pronoun did not provide any cue with respect to whether a subject-first or object-first analysis had to be adopted. Consequently, there was little incentive for readers to make a regression from the wh-phrase. Instead, it appears that readers attempted to recover from the misanalysis either by staying at the wh-phrase, or by proceeding the following regions, resulting in first-pass effects in these regions. This contrasts with the findings from the VP/NP and relative clause attachment experiments, where first-pass regressions and regression-path times indicated that readers went back to earlier regions after reading the critical region. This makes sense for these ambiguities. Because the ambiguous phrase had to be attached to one of two attachment sites that preceded the critical word, readers may have checked the preceding part of the sentence in order to recover from their misanalysis.

Thus, there are several possible factors which may have caused the eye-movement patterns in Experiment 1-4 to differ from other eyetracking studies on

syntactic ambiguity resolution. Unfortunately, our present knowledge about eye-movement patterns during syntactic ambiguity resolution is too limited to determine which of these factors caused the particular pattern in these experiments. It is possible that all of them conspired to give the observed outcome. However, it should be noted that the explanations of the eye-movement pattern in Experiment 1-4 does by no means require the assumption that the effects reflected late processing. In fact, they suggest the opposite. Even if they did reflect late processing, they would be inconsistent with competition in the way it is incorporated in current constraint-based theories. Because constraint-based theories assume that there is only a single stage in syntactic processing, it is unclear how the reanalysis pattern in the current experiments could have been preceded by an earlier competition pattern. In constraint-based theories, there is no mechanism that can explain how an early competition process flipped around and changed into a reanalysis pattern later on.

8.3. Do readers resolve syntactic ambiguities?

Another way to explain the results from the current experiments might be to assume that readers simply did not resolve the ambiguous conditions in the experiments, which caused them to be easy. This interpretation seems very unlikely, for a number of reasons. Note first that the processor clearly did resolve the disambiguated conditions, because they showed processing difficulty. Hence, one has to assume that the disambiguated conditions were processed differently from the ambiguous conditions. Such a distinction between the ambiguous and disambiguated conditions is inconsistent with most delay and underspecification models (e.g., Berwick & Weinberg, 1984; Gorrell, 1995; Marcus, 1980; Marcus et al., 1983; Sturt & Crocker, 1996; Weinberg, 1993), because they assume that plausibility (and probably case marking) does not influence whether delay or underspecification occurs.

Assuming that the processor resolved the ambiguity in the disambiguated conditions, the processor cannot simply have failed to analyse the ambiguous sentences at all. In order to determine whether a sentence was globally ambiguous or not, it must have constructed both analyses in both the ambiguous and disambiguated conditions.

There are still two ways of not settling on a single analysis. First, in the ambiguous condition the processor might maintain both analyses in parallel for at least as long as the sentence is read. In the disambiguated conditions, the less plausible or ungrammatical analysis is dropped. In order to explain that the disambiguated conditions of balanced ambiguities are more difficult to read than the ambiguous conditions, one would further have to make the assumption that abandoning an analysis

is more costly than maintaining two analyses. Regardless of whether such an assumption is plausible (there is currently no evidence that supports it) this account cannot explain the results in this thesis in a straightforward way. It would need an additional mechanism to explain why abandoning the non-preferred analysis in biased ambiguities causes little or no difficulty. Although this is possible in principle, it would make such an account much more complex. Secondly, a more severe problem with this account is that it predicts increased memory load due to parallel processing, which should presumably result in processing difficulty in the ambiguous sentences at some point. This predicts that ambiguous sentences are more difficult to process than unambiguous sentences. However, Experiment 4 did not show any evidence for this.

Alternatively, the processor might abandon both analyses in ambiguous sentences, as soon as it finds out that both analyses are equally plausible. This may make the ambiguous conditions easier to read than the disambiguated conditions, in which the processor keeps at least one analysis. However, this explanation is unlikely for two reasons. Firstly, it predicts that globally *implausible* syntactic ambiguities such as (1) are also easy to process.

1. The hunter killed only the leopard with the rainbow.

Because both analyses are equally (im)plausible, both should be abandoned, and therefore such sentences should be as easy to read as globally plausible sentences. Although there is currently no data on this, this seems highly unlikely. Secondly, this account predicts that in ambiguities with delayed disambiguation (e.g., Experiment 1 and 2), both analyses are abandoned before the disambiguation is reached. For example, in the VP/NP attachment ambiguity of Experiment 1, the processor would abandon both analyses as soon as *with* is reached, because both analyses are equally likely at this point. Because no analysis is available either in the ambiguous or the disambiguated conditions at the critical word, it is predicted that all conditions should be equally difficult to read. This was clearly not the case. In order to explain the differences between the conditions that occurred due to the plausibility manipulation at the critical word, one would have to make the additional assumption that the processor recomputed both analyses on the critical noun. In fact, it would have to constantly abandon and recompute analyses until the sentence is finally disambiguated. This account would be very uneconomical, and does not correspond well with any proposals about parsing. But it also makes the wrong prediction. As long as a sentence remains ambiguous, recurrent computations must continue and will produce reading difficulty. In contrast, these computations do not occur in unambiguous sentences, and therefore

unambiguous sentences should be easier than globally ambiguous sentences, contrary to the findings of Experiment 4.

It appears that explanations that are based on the assumption that readers did not resolve the ambiguous conditions have difficulty to explain the current results in any straightforward way. Such explanations do not appear to be good alternatives to the unrestricted race model, which straightforwardly accounts for the results. However, it should be noted that even if alternative explanations based on the assumption that readers did not resolve the ambiguous conditions could somehow be made consistent with the data, such explanations would be very different from the assumptions that are made by current theories of sentence processing. For example, models based on such explanations would not involve the notion of competition, and they would therefore be fundamentally different from most current constraint-based theories. Hence, the experiments in this thesis would still provide evidence against competition.

8.4. Conclusion

In conclusion, the experiments of this thesis shed new light on the architecture underlying syntactic processing. The experiments indicate that syntactic ambiguity resolution does not involve competition, as claimed by most constraint-based theories which are currently influential in the field. This suggests that syntactic ambiguity resolution is different from lexical ambiguity resolution, where evidence for competition has been found.

The experiments also provided evidence against some two-stage models such as the garden-path model and construal, because the processor does not always adopt the same analysis initially. Instead, the experiments reported in this thesis indicate that the processor can adopt either analysis. This is consistent with the unrestricted race model. In this model, the proportion of times that an analysis is initially adopted is determined by the bias of an item. The stronger the preference for an analysis, the more often it is adopted. The unrestricted race model claims that processing difficulty is due to reanalysis: when an initial analysis has to be revised, processing difficulty occurs.

9. References

Abney, S.P. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research*, **18**, 129-144.

Adams, B.C., Clifton, C. Jr., & Mitchell, D.C. (1998). Lexical guidance in sentence processing? *Psychonomic Bulletin & Review,* **5,** 265-270.

Allopenna, P.D., Magnuson, J.S., & Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, **38,** 419-439.

Altmann, G.T.M. (1994). Regression-contingent analyses of eye movements during sentence processing: Reply to Rayner and Sereno. *Memory and Cognition,* **22,** 286-290.

Altmann, G.T.M. (1999). Thematic role assignment in context. *Journal of Memory and Language,* **41,** 124-145.

Altmann, G.T.M., Garnham, A., & Dennis, Y.I.L. (1992). Avoiding the garden path: Eye-movements in context. *Journal of Memory and Language*, **31**, 685-712.

Altmann, G.T.M., Garnham, A., & Henstra, J.A. (1994). Effects of syntax in human sentence parsing: Evidence against a structure-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **20,** 209-216.

Altmann, G.T.M., & Steedman, M.J. (1988). Interaction with context during human sentence processing. *Cognition*, **30**, 191-238.

Altmann, G.T.M., van Nice, K.Y., Garnham, A., & Henstra, J.A. (1998). Late closure in context. *Journal of Memory and Language,* **38,** 459-484.

Baayen, R.H., Dijkstra, T., & Schreuder, R. (1997). Singular and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language,* **37,** 94-117.

Baayen, H.R., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database (CDROM).* Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J.D. Fodor, & F. Ferreira (Eds.) *Reanalysis in sentence processing* (pp. 1-46). Dordrecht, The Netherlands: Kluwer.

Balota, D.A., & Chumbley, J.I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance,* **10,** 340-357.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing.* Cambridge, MA: Cambridge University Press.

Berwick, R.C., & Weinberg, A. (1984). *The grammatical basis of linguistic performance: Language use and acquisition.* MIT Press: Cambridge, MA.

Bever, T.G. (1970). The cognitive basis for linguistic structures. In J.R. Hayes (Ed.) *Cognition and the development of language* (pp. 279-362). New York: Wiley.

Boland, J., Tanenhaus, M.K., Garnsey, S.M., & Carlson, G.N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language,* **34,** 774-806.

Borowsky, R., & Masson, M.E.J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **22,** 63-85.

Branigan, H.P., Pickering, M.J., Liversedge, S.P., Stewart, A.J., & Urbach, T.P. (1995). Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research,* **24,** 489-506.

Bresnan, J. (1982). *The mental representation of grammatical relations.* 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., Perfetti, C.A., Garrod, S.C., & Rayner, K. (1992). Parsing in context: Context effects and their limits. *Journal of Memory and Language*, **31**, 293-314.

Brysbaert, M., & Mitchell, D.C. (1996). Modifier attachment in sentence parsing: Evidence from Dutch. *The Quarterly Journal of Experimental Psychology,* **49A,** 664-695.

Burgess, C., Tanenhaus, M., & Hoffman, M. (1994). Parafoveal and semantic effects on syntactic ambiguity resolution. In *Proceedings of the 16th annual meeting of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Carlson, G.N. & Tanenhaus, M.K. (1988). Thematic roles and language comprehension. In W. Wilkins (Ed.), *Thematic relations (Syntax and semantics 21)* (pp. 263-288). San Diego, CA: Academic Press.

Carreiras, M. & Clifton, C., Jr. (1993). Relative clause interpretation preferences in Spanish and English. *Language and Speech,* **36,** 353-372.

Chumbley, J.I., & Balota, D.A. (1984). A word's meaning affects the decision in lexical decision. *Memory and Cognition,* **12,** 590-606.

Chomsky, N. (1957). *Syntactic structures.* The Hague, The Netherlands: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Chomsky, N. (1981). *Lectures on Government and Binding.* Dordrecht, The Netherlands: Foris.

Chomsky, N. (1986). *Barriers.* Cambridge MA: MIT Press.

Clark, H.H. (1973). The language-as-a-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior,* **12,** 335-359.

Clayes, E.L., Liversedge, S.P., & Paterson, K.B. (1999). Only *only guides parsing of "long", not "short", relative clauses.* Manuscript submitted for publication.

Clifton, C., Jr. (1993). Thematic roles in sentence parsing. *Canadian Journal of Experimental Psychology,* **47,** 222-246.

Clifton, C., Jr, Bock, J., & Radó, J. (in press). Effects of the focus particle *only* and intrinsic contrast on comprehension of reduced relative clauses. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process.* Oxford: Elsevier.

Clifton, C., Jr., & De Vincenzi, M. (1990). Comprehending sentences with empty elements. In D.A. Balota, G.B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 265-283). Hillsdale, NJ: Erlbaum.

Clifton, C., Jr., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In G.N. Carlson, & M.K. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 273-317).

Clifton, C. Jr., Frazier, L. & Connine, C. (1984). Lexical expectations in sentence comprehension. *Journal of Verbal Language and Verbal Behavior,* **23,** 696-708.

Clifton, C. Jr., & Ferreira (1989). Ambiguity in context. *Language and Cognitive Processes,* **4,** SI 77-103.

Clifton, C. Jr., Kennison, S.M., & Albrecht, J.E. (1997). Reading the words *her, his, him*: Implications for parsing principles based on frequency and on structure. *Journal of Memory and Language,* **36,** 276-292.

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.

Cloitre, M., & Bever, T.G. (1988). Linguistic anaphors, levels of representation, and discourse. *Language and Cognitive Processes,* **3,** 293-322.

Cowart, W. & Cairns. H.S. (1987). Evidence for an anaphoric mechanism within syntactic processing: Some reference relations defy semantic and pragmatic constraints. *Memory and Cognition,* **15,** 318-331.

Crain, S., & Steedman, M.J. (1985). On not being led up the garden-path: The use of context by the psychological processor. In D. Dowty, L. Kartunnen, & A. Zwicky (Eds.) *Natural language parsing* (pp. 320-358). Cambridge: Cambridge University Press.

Crane, H.D. (1994). The Purkinje image eyetracker, image stabilization and related forms of stimulus manipulation. In D.H. Kelly (Ed.), *Visual science and engineering: Models and applications* (pp. 15-89). New York, NY: Marcel Dekker.

Crocker, M.W. (1994). On the nature of the principle-based sentence processor. In C. Clifton, Jr, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 245-266). Hillsdale, NJ: Erlbaum.

Cuetos, F., & Mitchell, D.C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition*, **30**, 73-105.

Cuetos, F., Mitchell, D.C., & Corley, M.M.B. (1996). Parsing in different languages. In M. Carreiras, J. Garcia-Albea, & N. Sabastian-Galles (Eds.), *Language processing in Spanish* (pp. 145-187). Hillsdale, NJ: Erlbaum.

Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Language and Verbal Behavior,* **22,** 75-87.

De Vincenzi, M. (1991). *Syntactic parsing strategies in Italian.* Dordrecht, The Netherlands: Kluwer.

De Vincenzi, M. & Job, R. (1993). Some observations on the universality of the late closure strategy: Evidences from Italian. *Journal of Psycholinguistic Research,* **22,** 189-206.

De Vincenzi, M. & Job, R. (1995). An investigation of late closure: The role of syntax, thematic structure, and pragmatics in initial and final interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **21,** 1-19.

Dopkins, S., Morris, R.K., & Rayner, K. (1992). Lexical ambiguity and eye fixations in reading: A test of competing models of lexical ambiguity resolution. *Journal of Memory and Language,* **31,** 461-476.

Duffy, S.A., Morris, R.K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language,* **27**, 429-446.

Eberhard, K.M., Spivey-Knowlton, M.J., Sedivy, J.C., & Tanenhaus, M.K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research,* **24,** 409-436.

Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Language and Verbal Behavior,* **22,** 75-87.

Farrar, IV, W.T. & Kawamoto, A.H. (1993). The return of "visiting relatives": Pragmatic effect in sentence processing. *The Quarterly Journal of Experimental Psychology,* **46A,** 463-487.

Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, **25**, 348-368.

Ferreira, F., & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**, 555-568.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, **30,** 725-745.

Ferreira, F., & Henderson, J. M. (1993). Reading processes during syntactic analysis and reanalysis. *Canadian Journal of Psychology,* **47,** 247-275.

Fodor, J.A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Fodor, J.A., Bever, T.G., Garrett, M.F. (1974). *The Psychology of Language.* New York, NY: McGraw-Hill.

Fodor, J.D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry,* **9,** 427-473.

Fodor, J.D., & Frazier, L. (1980). Is the human sentence parsing mechanism an ATN? *Cognition,* **8,** 417-459.

Fodor, J.D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research,* **23,** 407-434.

Fodor, J.D., & Inoue, A. (1994). Attach anyway. In F.D. Fodor., & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 101-141). Dordrecht, The Netherlands: Kluwer.

Ford, M., Bresnan, J., & Kaplan, R.M. (1982). A competence-based theory of syntactic closure. In J. Bresnan (Ed.) *The mental representation of grammatical relations,* (pp. 727-796). Cambridge, MA: MIT Press.

Forster, K. (1979). Levels of processing and the structure of the language processor. In W.E. Cooper & E.C.T. Walker (Eds.) *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27-81). Hillsdale, NJ: Erlbaum.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies.* Ph.D. dissertation, University of Connecticut. West Bend, IN: Indiana University Linguistics Club.

Frazier, L. (1987a). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 559-586). Hillsdale, NJ: Erlbaum.

Frazier, L. (1987b). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory,* **5,** 519-560.

Frazier, L. (1989). Against lexical generation of syntax. In W.D. Marslen-Wilson (Ed.) *Lexical representation and process* (pp. 505-528). Cambridge, MA: MIT Press.

Frazier, L. (1990). Parsing modifiers: Special purpose routines in the human sentence processing mechanism? In D.A. Balota, G.B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 303-330). Hillsdale, NJ: Erlbaum.

Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research*, **24**, 437-468.

Frazier, L. & Clifton, C. Jr (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, **4,** 93-126.

Frazier, L., & Clifton, C., Jr. (1996). *Construal.* Cambridge, MA: MIT Press.

Frazier, L, & Clifton, C., Jr. (1998). Sentence reanalysis, and visibility. In J.D. Fodor, & F. Ferreira (Eds.) *Reanalysis in sentence processing.* Dordrecht, The Netherlands: Kluwer.

Frazier, L. & Flores d'Arcais, G.B. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language,* **28,** 331-344.

Frazier, L. & Fodor, J.D. (1978). The sausage machine: A new 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 Psychology*, **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.

Friederici, A.D. (1998). Diagnosis and reanalysis: Two processing aspects the brain may differentiate. In J.D. Fodor, & F. Ferreira (Eds.) *Reanalysis in sentence processing* (pp. 177-200). Dordrecht, The Netherlands: Kluwer.

Garnsey, S.M., Pearlmutter, N.J., Myers, E., & Lotocky, M.A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, **37,** 58-93.

Gazdar, G., Klein, E., Pullum, G.K., & Sag, I.A. (1985). *Generalized phrase structure grammar.* Cambridge, MA: Harvard University Press.

Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown.* Unpublished doctoral dissertation. Carnegie Mellon University.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition,* **68,** 1-76.

Gibson, E., & Hickok, G. (1993). Sentence processing with empty categories. *Language and Cognitive Processes,* **8,** 147-161.

Gibson, E., Pearlmutter, N., Canseco-Gonzalez, E., & Hickok, G. (1996). Recency preference in the human sentence processor. *Cognition,* **59,** 23-59.

Gibson, E., & Schütze, C.T. (1999). Disambiguation preferences in noun phrase conjunction do not mirror corpus frequency. *Journal of Memory and Language,* **40,** 263-279.

Gibson, E., Schütze, C.T., & Salomon, A. (1996). The relationship between the frequency and the processing complexity of linguistic structure. *Journal of Psycholinguistic Research,* **25,** 59-92.

Gilboy, E., Sopena, J-M., Clifton, C, Jr., & Frazier, L. (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). Parsing theory and phrase-order variation in German V2 clauses. *Journal of Psycholinguistic Research,* **25,** 135-156.

Hemforth, B. (1993). *Kognitives Parsing: Repräsentationen und Verarbeitung grammatischen Wissens.* Sanct Augustin: Infix Verlag.

Hemforth, B., Konieczny, L., & Scheepers, C. (in press). Syntactic attachment and anaphor resolution: Two sides of relative clause attachment. In M.W. Crocker, M.J. Pickering, & C. Clifton, Jr. (Eds.), *Architectures and mechanisms for language processing.* Cambridge: Cambridge University Press.

Hindle, D. & Rooth, M. (1993). Structural ambiguity in lexical relations. *Computational Linguistics,* **19,** 103-120.

Holmes, V.M., Kennedy, A., & Murray, W.S. (1987). Syntactic structure and the garden path. *Quarterly Journal of Experimental Psychology,* **39A,** 277-293.

Holmes, V.H., Stowe, L., & Cupples, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language,* **28,** 668-689.

Hudson, R.A. (1990). *English Word Grammar.* Oxford, UK: Blackwell.

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., & Fodor, J.A. (1995). Information-paced parsing of Japanese. In R. Mazuka, & N. Nagai (Eds.) *Japanese sentence processing.* (pp. 9-63). Hillsdale, NJ: Erlbaum.

Juliano, C., & Tanenhaus, M.K. (1996). 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. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review,* **99,** 122-149.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science,* **20,** 137-194.

Kaan, E. (1997). *Processing subject-object ambiguities in Dutch.* Doctoral dissertation. University of Groningen.

Kaplan, R.M., & Zaenen, A. (1988). Long-distance dependencies as a case of functional uncertainty. In M. Baltin & A. Kroch (Eds.), *Alternative conceptions of phrase structure* (pp. 17-42). Chicago, University of Chicago Press.

Kellas, G., Ferraro, R.F., & Simpson, G.B. (1988). Lexical ambiguity and the time-course of attentional allocation in word recognition. *Journal of Experimental Psychology: Human Perception and Performance,* **14,** 601-609.

Kennedy, A., & Murray, W.S. (1984). Inspection for words in syntactically ambiguous sentences under three presentation conditions. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **10,** 833-849.

Kennedy, A., Murray, W. S., Jennings, F., & Reid, C. (1989). Parsing complements: Comments on the generality of the principle of minimal attachment. *Language and Cognitive Processes*, **4,** SI 51-76.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition,* **8,** 15-47.

Kimball, J. (1975). Predictive analysis and over-the-top parsing. In J. Kimball (Ed.), *Syntax and Semantics 4.* New York, NY: Academic Press.

King, J., & Just, M.A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language,* **30,** 580-602.

Konieczny, L. (1996). *Human sentence processing: A semantics-oriented parsing approach.* IIG-Berichte 3/96.

Konieczny, L., Hemforth, B., & Scheepers, C. (in press). Head position and clause boundary effects in reanalysis. In L. Konieczny & B. Hemforth, (Eds.), *Aspects of German parsing.* Dordrecht, The Netherlands: Kluwer.

Konieczny, L., Hemforth, B., Scheepers, C., & Strube, G. (1997). The role of lexical heads in parsing: evidence from German. *Language and Cognitive Processes*, **12,** 307-348.

Koster, J. (1975). Dutch as an SOV language. *Linguistic Analysis,* **15,** 111-136.

Lewis, R. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research,* **25,** 93-115.

Lewis, R. (1999). *Attachment without competition: a computational model of race-based parsing in a limited working memory.* Paper presented at the 12th annual CUNY conference, New York, NY.

Liversedge, S.P., Paterson, K.B., & Pickering, M.J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance and scene perception* (pp. 55-75).

Liversedge, S.P., Pickering, M.J., Branigan, H.P., & Van Gompel, R.P.G. (1998). Processing arguments and adjuncts in isolation and context: The case of *by*-phrase ambiguities in passives. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **24,** 461-475.

MacDonald, M.C. (1989). Priming effect from gaps to antecedents. *Language and Cognitive Processes, **4,** 35-56.*

MacDonald, M.C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, **9**, 157-201.

MacDonald, M.C., Just, M.A., & Carpenter, P.A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology,* **24,** 56-98.

MacDonald, M.C., Pearlmutter, N.J., & Seidenberg, M.S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, **101**, 676-703.

Marcus, M. (1980). *A theory of syntactic recognition for natural language.* Cambridge, MA: MIT Press.

Marcus, M.P., Hindle, D., & Fleck, M.M. (1983). D-theory: Talking about talking about trees. In *Proceedings of the 21st annual meeting of the association for computational linguistics* (pp. 129-136).

Marslen-Wilson, W.D. (1975). Sentence perception as an interactive parallel process. *Science*, **189,** 226-228.

Marslen-Wilson, W.D., & Tyler, L.K. (1975). Processing structure of sentence perception. *Nature,* **257,** 784-786.

Marslen-Wilson, W. & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition,* **8,** 1-71.

McClelland, J.L., St. John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed approach. *Language and Cognitive Processes*, **4,** SI 287-335.

McDonald, J.L., & MacWhinney, B. (1995). The time course of anaphor resolution: effects of implicit verb causality and gender. *Journal of Memory and Language*, **34,** 543-566.

McElree, B., & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **21,** 134-157.

McElree, B., & Griffith, T. (1995). Structural and lexical constraints on filling gaps during sentence comprehension: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **24,** 432-460.

McRae, K., Spivey-Knowlton, M.J., & Tanenhaus, M.K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, **38,** 283-312.

McRoy, S.W., & Hirst, G. (1990). Race-based parsing and syntactic disambiguation. *Cognitive Psychology,* **14,** 313-353.

Mecklinger, A., Schriefers, H., Steinhauer, K, & Friederici, A.D. (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. (1962). Some psychological studies of grammar. *American Psychologist,* **17,** 748-762.

Millis, M.L., & Button, S.B. (1989). The effect of polysemy on lexical decision time: Now you see it, now you don't. *Memory and Cognition,* **17,** 141-147.

Mitchell, D.C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In Coltheart, M. (Ed.), *Attention and performance XII* (pp. 601-618). Hillsdale, NJ: Erlbaum.

Mitchell, D.C., & Brysbaert, M. (1998). Challenges to recent theories of crosslinguistic variation in parsing: Evidence from Dutch. *Sentence processing: A crosslinguistic perspective (Syntax and Semantics 31).* (pp. 313-335). San Diego: Academic Press.

Mitchell, D.C., & Corley, M.M.B. (1994). Immediate biases in parsing: Discourse effects or experimental artefacts? *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **20,** 217-222.

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., Corley, M.M.B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research,* **24,** 469-488.

Mitchell, D.C.,. & Holmes, V.M. (1985). The role of specific information about the verb in parsing sentences with local structural ambiguity. *Journal of Memory and Language,* **24,** 542-559.

Moortgat, M. (1988). *Categorial investigations: Logical and linguistic aspects of the Lambek Calculus.* Dordrecht, The Netherlands: Foris.

Morrison, R.E. (1984). Manipulation of stimulus onset display in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance,* **10,** 667-682.

Murray, W.S., & Liversedge, S.P. (1994). Referential context effects on syntactic processing. In C. Clifton, Jr, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 359-388). Hillsdale, NJ: Erlbaum.

Ni, W., Crain, S., & Shankweiler, D. (1996). Sidestepping garden paths: assessing the contributions of syntax, semantics and plausibility in resolving ambiguities. *Language and Cognitive Processes*, **11**, 283-334.

Paterson, K.B., Liversedge, S.P., & Underwood, G. (in press). The influence of focus operators on syntactic processing of "short" relative clause sentences. *Quarterly Journal of Experimental Psychology.*

Pearlmutter, N.J., & MacDonald, M.C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language,* **34,** 521-542.

Pearlmutter, N.J., & Mendelsohn, A. (1999). *Serial versus parallel sentence comprehension.* Unpublished manuscript.

Perfetti, C. (1990). The cooperative language processors: Semantic influences in an autonomous syntax. In D.A. Balota, G.B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 205-230). Hillsdale, NJ: Erlbaum.

Pickering, M.J. (1993). Direct association and sentence processing: A reply to Gorrell and to Gibson and Hickok. *Language and Cognitive Processes*, **8,** 163-196.

Pickering, M.J., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, **6,** 229-259.

Pickering, M.J., & Traxler, M.J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24,** 940-961.

Pickering, M.J., & Traxler, M.J., & Crocker, M.W. (1999). *Ambiguity resolution in sentence processing: Evidence against likelihood.* Manuscript submitted for publication.

Pollard, C., & Sag, I.A. (1987). *Information-based syntax and semantics.* Stanford, CA: CSLI.

Pollard, C., & Sag, I.A. (1994). *Head-driven phrase structure grammar.* Stanford, CA: CSLI and University of Chicago Press.

Pritchett, B. (1988). Garden path phenomena and the grammatical basis of language processing. *Language,* **64,** 539-576.

Pritchett, B. (1992). *Grammatical competence and parsing performance.* Chicago: University of Chicago Press.

Pynte, J. (1998). The time course of attachment decisions: Evidence from French. In D. Hillert (Ed.) *Sentence processing: A crosslinguistic perspective (Syntax and Semantics 31).* (pp. 227-245). San Diego: Academic Press.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, **22**, 358-374.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, **14**, 191-201.

Rayner, K., & Frazier, L. (1987). Parsing temporarily ambiguous complements. *Quarterly Journal of Experimental Psychology*, **39,** 657-673.

Rayner, K., & Frazier, L. (1989). Selection mechanisms in lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15**, 779-790.

Rayner, K., Garrod, S., & Perfetti, C.A. (1992). Discourse influences during parsing are delayed. *Cognition,* **45,** 109-139.

Rayner, K., & Morris, R.K. (1991). Comprehension processes in reading ambiguous sentences: Reflections from eye movements. In G.B. Simpson (Ed.) *Understanding word and sentence* (pp. 175-198). Amsterdam: Elsevier.

Rayner, K., Pacht, J.M., & Duffy, S.A. (1994). Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words: Evidence from eye fixations. *Journal of Memory and Language,* **33,** 527-544.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Englewood Cliffs, NJ: Prentice Hall.

Rayner, K., & Sereno, S.C. (1994a). Regressive eye movements and sentence parsing: On the use of regression-contingent analyses. *Memory and Cognition,* **22,** 281-285.

Rayner, K., & Sereno, S.C. (1994b). Regression-contingent analyses: A reply to Altmann. *Memory and Cognition,* **22,** 291-292.

Rayner, K., Sereno, S.C., Morris, R.K., Schmauder, A.R., & Clifton, C., Jr. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, **4,** SI 21-49.

Read, C., Kraak, C., & Boves, L. (1980). The interpretation of ambiguous who-questions in Dutch: The effect of intonation. In W. Zonneveld & F. Weerman (Eds.), *Linguistics in The Netherlands 1977-1979* (pp. 389-410).

Reichle, E.D., Pollatsek, A., Fisher, D.L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review,* **105,** 125-157.

Rubenstein, H., Lewis, S.S., & Rubenstein, M.A. (1971). Homographic entries in the internal lexicon: effects of systematicity and relative frequency of meanings. *Journal of Verbal Learning and Verbal Behavior,* **10,** 57-62.

Scheepers, C. (1997). Menschliche Satzverarbeitung: Syntaktische und thematische Aspekte der Wortstellung im Deutschen. Unpublished doctoral dissertation, University of Freiburg.

Schlesewki, , M., Fanselow, G. , Kliegl, R., & Krems, J. (in press). Preferences for grammatical functions in the processing of locally ambiguous wh-questions in German. In L. Konieczny & B. Hemforth, (Eds.), *Aspects of German parsing.* Dordrecht, The Netherlands: Kluwer.

Schmauder, R. & Egan, M.C. (1998). The influence of semantic fit on on-line sentence processing. *Memory and Cognition,* **26,** 1304-1312.

Schriefers, H., Friederici, A.D., & Kühn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language,* **34,** 499-520.

Schreuder, R., & Baayen, R.H. (1997). How complex simplex words can be. *Journal of Memory and Language,* **37,** 118-139.

Schütze, C.T., & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language,* **40,** 409-431.

Sedivy, J., & Spivey-Knowlton, M. (1994). The use of structural, lexical, and pragmatic information in parsing attachment ambiguities. In C. Clifton, Jr, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 389-413). Hillsdale, NJ: Erlbaum.

Sedivy, J.C., Tanenhaus, M.K., Chambers, C.G., & Carlson, G.N. (1999). Achieving incremental semantic interpretation through contextual information. *Cognition,* **71,** 109-147.

Seidenberg, M.S., Tanenhaus, M.K., Leiman, J.M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology,* **14,** 489-537.

Spivey-Knowlton, M., & Sedivy, J. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition,* **55,** 227-267.

Spivey, M.J. & Tanenhaus, M.K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **24,** 1521-1543.

Spivey-Knowlton, M., Trueswell, J., & Tanenhaus, M. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology,* **37,** 276-309.

Steedman, M. (1987). Combinatory grammars and parasitic gaps. *Natural Language and Linguistic Theory,* **5,** 403-439.

Stowe, L.A., Tanenhaus, M.K., & Carlson, G.N. (1991). Filling gaps on-line: Use of lexical and semantic information in sentence processing. *Language and Speech,* **34,** 319-340.

Sturt, P. (1997). *Syntactic reanalysis in human language processing.* Unpublished doctoral dissertation. University of Edinburgh.

Sturt, P., & Crocker, M.W. (1996). Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes,* **11,** 449-494.

Sturt, P., Pickering, M.J., & Crocker, M.W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language,* **40,** 136-150.

Swinney, D.A. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior,* **20,** 120-136.

Tabor, W., Juliano, C., & Tanenhaus, M.K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes,* **12,** 211-271.

Tabossi, O., Spivey-Knowlton, M.J., McRae, K., & Tanenhaus, M.K. (1994). Semantic effects on syntactic ambiguity resolution: Evidence for a constraint-based resolution process. In C. Umilta & M. Moscovitch (Eds.), *Attention and Performance XV.* Hillsdale, NJ: Erlbaum.

Tanenhaus, M.K., & Carlson, G.N. (1989). Lexical structure and language comprehension. In W.D. Marslen-Wilson (Ed.) *Lexical representation and process* (pp. 529-561). Cambridge, MA: MIT Press.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of Visual and linguistic information in spoken language comprehension. *Science,* **268,** 1632-1634.

Tanenhaus, M.K. & Trueswell, J.C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Speech, language, and communication (The handbook of perception and cognition 11)* (pp.217-262). San Diego CA: Academic Press.

Taraban, R., & McClelland, J.R. (1988). Constituent attachment and thematic role assignment in sentence processing: Influence of content-based expectations. *Journal of Memory and Language*, **27**, 597-632.

Taraban, R., & McClelland, J.R. (1990). Parsing and comprehension: A multiple-constraint view. In D.A. Balota, G.B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 231-263). Hillsdale, NJ: Erlbaum.

Townsend, D.J. & Bever, T.G. (1982). Natural units of representation interact during sentence comprehension. *Journal of Verbal Learning and Verbal Behavior,* **21,** 688-703.

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.

Traxler, M.J., Bybee, M.D., & Pickering, M.J. (1997). Influence of connectives on language comprehension: Eye-tracking evidence for incremental interpretation. *The Quarterly Journal of Experimental Psychology,* **50A,** 481-497.

Traxler, M.J., & Pickering, M.J. (1996a). Case-marking in the parsing of complement sentences: Evidence from eye movements. *Quarterly Journal of Experimental Psychology,* **49A,** 991-1004.

Traxler, M.J., & Pickering, M.J. (1996b). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language,* **35,** 454-475.

Traxler, M.J., Pickering, M.J., & Clifton, C, Jr. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, **39,** 558-592.

Trueswell, J. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, **35**, 566-585.

Trueswell, J., Tanenhaus, M.K., & Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, **33**, 285-318.

Trueswell, J., Tanenhaus, M.K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **19**, 528-553.

Tyler, L.K., & Marslen-Wilson, W.D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, **16**, 683-692.

Van Berkum, J.J.A., Brown, C.M., & Hagoort, P. (1999). Early referential effect in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language,* **41,** 147-182.

Van Gompel, R.P.G. (1995). *The processing of subject and object relative clauses in Dutch.* MA thesis, University of Nijmegen.

Van Gompel, R.P.G. (1996). *The use of plausibility information in verb/noun assignment ambiguities: Evidence against interactive approaches.* MA thesis, University of Nijmegen.

Van Gompel, R.P.G., Pickering, M.J., & Traxler, M.J. (1999). *Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models.* Manuscript submitted for publication.

Van Gompel, R.P.G., Pickering, M.J., & Traxler, M.J. (in press). Unrestricted race: A new model of syntactic ambiguity resolution. In: A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process.* Oxford: Elsevier.

Wanner, E. (1980). The ATN and the sausage machine: Which one is baloney? *Cognition,* **8,** 209-225.

Warner, J. & Glass, A.L. (1987). Context and distance-to disambiguation effects in ambiguity resolution: Evidence from grammaticality judgements of garden path sentences. *Journal of Memory and Language,* **26,** 714-738.

Weinberg, A. (1993). Parameters in the theory of sentence processing: Minimal commitment theory goes east. *Journal of Psycholinguistic Research,* **22,** 339-364.

Zagar, D., Pynte, J., & Rativeau, S. (1997). Evidence for early closure attachment on first-pass reading times in French. *The Quarterly Journal of Experimental Psychology,* **50A,** 421-438.

10. Appendices

10.1 Appendix 1: Experimental items from Experiment 1

The old man poked only the chap with the umbrella a number of times.
The old man poked only the soil with the umbrella a number of times.
The old man poked only the chap with the raincoat a number of times.

The brute frightened only the scoundrel with the penknife late at night.
The brute frightened only the greyhound with the penknife late at night.
The brute frightened only the scoundrel with the freckles late at night.

The farmer injured only the shepherd with the staff a few days ago.
The farmer injured only the sheepdog with the staff a few days ago.
The farmer injured only the sheepdog with the spots a few days ago.

The tourists followed only the naturalist with the jeep during the trip.
The tourists followed only the rhinoceros with the jeep during the trip.
The tourists followed only the rhinoceros with the limp during the trip.

The old man pointed at only the girl with the stick after some time.
The old man pointed at only the star with the stick after some time.
The old man pointed at only the girl with the dress after some time.

The tourist saw only the walker with the telescope early in the afternoon.
The tourist saw only the badger with the telescope early in the afternoon.
The tourist saw only the badger with the infection early in the afternoon.

The tribesman knocked down only the enemy with the cudgel the other day.
The tribesman knocked down only the tiger with the cudgel the other day.
The tribesman knocked down only the tiger with the injury the other day.

The peasant damaged only the shed with the tractor after the hard work.
The peasant damaged only the pane with the tractor after the hard work.
The peasant damaged only the shed with the windows after the hard work.

The explorer found only the hiker with the compass after the hurricane.
The explorer found only the coast with the compass after the hurricane.
The explorer found only the coast with the harbour after the hurricane.

The intruder struck only the hero with the sword during the attack.
The intruder struck only the lock with the sword during the attack.
The intruder struck only the hero with the beard during the attack.

The hunter killed only the poacher with the rifle not long after sunset.
The hunter killed only the leopard with the rifle not long after sunset.
The hunter killed only the leopard with the scars not long after sunset.

Jonathan spotted only the bag with the binoculars after a long search.
Jonathan spotted only the fox with the binoculars after a long search.
Jonathan spotted only the bag with the wristwatch after a long search.

The vandal scratched only the lorry with the aerial after the football match.
The vandal scratched only the paint with the aerial after the football match.
The vandal scratched only the lorry with the faults after the football match.

The African hunter aimed at only the rival with the spear during the pursuit.
The African hunter aimed at only the zebra with the spear during the pursuit.
The African hunter aimed at only the zebra with the wound during the pursuit.

The young ruffian kicked only the girl with the trainers a number of times.
The young ruffian kicked only the ball with the trainers a number of times.
The young ruffian kicked only the girl with the pigtails a number of times.

The adventurer found only the bag with the charts after a long time.
The adventurer found only the sea with the charts after a long time.
The adventurer found only the sea with the whales after a long time.

The man touched only the tramp with the knife after some hesitation.
The man touched only the stain with the knife after some hesitation.
The man touched only the tramp with the fever after some hesitation.

The angry man battered only the boy with the crowbar last week on Sunday.
The angry man battered only the cat with the crowbar last week on Sunday.
The angry man battered only the cat with the stripes last week on Sunday.

The handyman repaired only the toolbox with the pliers right after lunch.
The handyman repaired only the lantern with the pliers right after lunch.
The handyman repaired only the lantern with the damage right after lunch.

The police tracked down only the terrorist with the bulldogs after the chase.
The police tracked down only the narcotics with the bulldogs after the chase.
The police tracked down only the terrorist with the dynamite after the chase.

Peter yelled at only the protester with the megaphone for a long time.
Peter yelled at only the racehorse with the megaphone for a long time.
Peter yelled at only the protester with the moustache for a long time.

A coastguard rescued only the tourist with the motorboat during the patrol.
A coastguard rescued only the dolphin with the motorboat during the patrol.
A coastguard rescued only the dolphin with the poisoning during the patrol.

The trooper destroyed only the tank with the projectile during the fight.
The trooper destroyed only the wall with the projectile during the fight.
The trooper destroyed only the tank with the camouflage during the fight.

The criminal shot only the cop with the pistol yesterday late at night.
The criminal shot only the dog with the pistol yesterday late at night.
The criminal shot only the dog with the collar yesterday late at night.

The caretaker cleaned only the pail with the brush after a while.
The caretaker cleaned only the suit with the brush after a while.
The caretaker cleaned only the pail with the holes after a while.

The man repaired only the toolbox with the screwdriver as quickly as possible.
The man repaired only the printer with the screwdriver as quickly as possible.
The man repaired only the printer with the malfunction as quickly as possible.

The man deterred only the thug with the shotgun just before dawn.
The man deterred only the wolf with the shotgun just before dawn.
The man deterred only the thug with the tattoos just before dawn.

The teenager washed only the basin with the sponge early in the morning.
The teenager washed only the cheek with the sponge early in the morning.
The teenager washed only the cheek with the pimple early in the morning.

The outlaw wounded only the guard with the crossbow during the night.
The outlaw wounded only the horse with the crossbow during the night.
The outlaw wounded only the guard with the whiskers during the night.

The burglar stabbed only the guy with the dagger during the night.
The burglar stabbed only the dog with the dagger during the night.
The burglar stabbed only the dog with the collar during the night.

10.2 Appendix 2: Experimental items from Experiment 2 and 3

The altar of the temple disintegrating/that has been disintegrating over the years was very important.
The priest of the temple disintegrating/that has been disintegrating over the years was very important.
The temple of the priest disintegrating/that has been disintegrating over the years was very important.

The rudder of the boat sawn/that had been sawn from the wood looked very impressive.
The rower of the boat sawn/that had been sawn from the wood looked very impressive.
The boat of the native sawn/that had been sawn from the wood looked very impressive.

The section of the field sown/that had been sown early this spring pleased the lord.
The farmer of the land sown/that had been sown early this spring pleased the lord.
The land of the farmer sown/that had been sown early this spring pleased the lord.

The basement of the tenement containing/that used to contain lots of junk is very gloomy.
The landlord of the tenement containing/that used to contain lots of junk is very gloomy.
The tenement of the landlord containing/that used to contain lots of junk is very gloomy.

The tower of the castle decaying/that has been decaying due to acid rain is very old.
The lord of the castle decaying/that has been decaying due to acid rain is very old.
The castle of the lord decaying/that has been decaying due to acid rain is very old.

The stopper of the decanter broken/had been broken in the accident was very old.
The collector of the porcelain broken/had been broken in the accident was very old.
The porcelain of the collector broken/had been broken in the accident was very old.

The collar of the jacket sewn/that has been sewn in a very skilful way looks quite stylish.
The designer of the jacket sewn/that has been sewn in a very skilful way looks quite stylish.
The shirt of the tailor sewn/that has been sewn in a very skilful way looks quite stylish.

The rudder of the vessel corroding/that had been corroding under the paint troubled the sailor.
The builder of the yacht corroding/that had been corroding under the paint troubled the sailor.
The yacht of the playboy corroding/that had been corroding under the paint troubled the sailor.

The jacket of the suit worn/that had been worn during the meeting looked very unfashionable.
The maker of the suit worn/that had been worn during the meeting looked very unfashionable.
The suit of the judge worn/that had been worn during the meeting looked very unfashionable.

The region of the country neighbouring/that used to neighbour the poor nation is quite wealthy.
The tyrant of the country neighbouring/that used to neighbour the poor nation is quite wealthy.
The country of the tyrant neighbouring/that used to neighbour the poor nation is quite wealthy.

The diamond of the collection stolen/that had been stolen in the raid is very beautiful.
The owner of the diamond stolen/that had been stolen in the raid is very beautiful.
The diamond of the millionaire stolen/that had been stolen in the raid is very beautiful.

The base of the statue hewn/that had been hewn from the granite was very strong.
The sculptor of the statue hewn/that had been hewn from the granite was very strong.
The statue of the queen hewn/that had been hewn from the granite was very strong.

The tunic of the costume woven/that had been woven in the mill looked very beautiful.
The wearer of the costume woven/that had been woven in the mill looked very beautiful.
The costume of the actor woven/that had been woven in the mill looked very beautiful.

The garage of the cottage teeming/that used to be teeming with hundreds of mice was very ugly.
The owner of the cottage teeming/that used to be teeming with hundreds of mice was very ugly.
The cottage of the farmer teeming/that used to be teeming with hundreds of mice was very ugly.

The shrubbery of the park containing/that used to contain many beautiful plants had been really pleasant.
The gardener of the park containing/that used to contain many beautiful plants had been really pleasant.
The garden of the duke containing/that used to contain many beautiful plants had been really pleasant.

The brother of the reviewer phoning/that used to phone from the United States is very weird.
The article of the scientist phoning/that used to phone from the United States is very weird.
The reviewer of the article phoning/that used to phone from the United States is very weird.

The client of the lawyer laughing/that had been laughing at the joke looked very smart.
The mansion of the lawyer laughing/that had been laughing at the joke looked very smart.
The lawyer of the company laughing/that had been laughing at the joke looked very smart.

The bodyguard of the governor retiring/that will be retiring after the troubles is very rich.
The province of the governor retiring/that will be retiring after the troubles is very rich.
The governor of the province retiring/that will be retiring after the troubles is very rich.

The assistant of the professor smiling/that was smiling at the student is quite excellent.
The department of the professor smiling/that was smiling at the student is quite excellent.
The professor of the department smiling/that was smiling at the student is quite excellent.

The accomplice of the burglar forgiven/that had been forgiven by the victims was soon discovered.
The revolver of the burglar forgiven/that had been forgiven by the victims was soon discovered.
The burglar of the bungalow forgiven/that had been forgiven by the victims was soon discovered.

The servant of the prince strolling/that has been strolling through the gardens is very old.
The palace of the prince strolling/that has been strolling through the gardens is very old.
The servant of the palace strolling/that has been strolling through the gardens is very old.

The colleague of the admiral talking/that will be talking to the woman is leaving soon.
The battleship of the admiral talking/that will be talking to the woman is leaving soon.
The captain of the battleship talking/that will be talking to the woman is leaving soon.

The companion of the courier waving/that has been waving to the pedestrian is very reliable.
The motorbike of the courier waving/that has been waving to the pedestrian is very reliable.
The rider of the motorbike waving/that has been waving to the pedestrian is very reliable.

The patient of the surgeon walking/that had been walking down the corridor is quite old.
The limousine of the surgeon walking/that had been walking down the corridor is quite old.
The surgeon of the hospital walking/that had been walking down the corridor is quite old.

The colleague of the rector drinking/that has been drinking the pleasant wine is well-known.
The office of the rector drinking/that has been drinking the pleasant wine is well-known.
The rector of the university drinking/that has been drinking the pleasant wine is well-known.

The apprentice of the joiner smoking/that used to smoke in the canteen looked quite strange.
The equipment of the joiner smoking/that used to smoke in the canteen looked quite strange.
The repairman of the agency smoking/that used to smoke in the canteen looked quite strange.

The sister of the survivor woken/that had been woken by the turmoil was in a bad state.
The lifebelt of the survivor woken/that had been woken by the turmoil was in a bad state.
The survivor of the disaster woken/that had been woken by the turmoil was in a bad state.

The guard of the prisoner beaten up/that had been beaten up during the interrogation is in a state.
The cell of the prisoner beaten up/that had been beaten up during the interrogation is in a state.
The guard of the prison beaten up/that had been beaten up during the interrogation is in a state.

The advisor of the mayor driven/that had been driven to the meeting had a lot of problems.
The village of the mayor driven/that had been driven to the meeting had a lot of problems.
The mayor of the village driven/that had been driven to the meeting had a lot of problems.

The admirer of the painter humming/that has been humming the familiar tune is very nice.
The picture of the painter humming/that has been humming the familiar tune is very nice.
The painter of the picture humming/that has been humming the familiar tune is very nice.

10.2 Appendix 3: Experimental items from Experiment 4

Ed assumed that the altar of the temple disintegrating over the years was very important.
Ed assumed that the priest of the temple disintegrating over the years was very important.
Ed assumed that the temple of the priest disintegrating over the years was very important.
Almost everybody assumed that the temple disintegrating over the years was very important.

The man gathered that the rudder of the boat sawn from the wood looked very impressive.
The man gathered that the rower of the boat sawn from the wood looked very impressive.
The man gathered that the boat of the native sawn from the wood looked very impressive.
The man and his friends gathered that the boat sawn from the wood looked very impressive.

The man said that the section of the field sown early this spring pleased the lord.
The man said that the farmer of the land sown early this spring pleased the lord.
The man said that the land of the farmer sown early this spring pleased the lord.
The man from the village said that the field sown early this spring pleased the lord.

Colin noted that the basement of the tenement containing lots of junk was very gloomy.
Colin noted that the landlord of the tenement containing lots of junk was very gloomy.
Colin noted that the tenement of the landlord containing lots of junk was very gloomy.
Almost at once Colin noted that the tenement containing lots of junk was very gloomy.

The guide said that the tower of the castle decaying due to acid rain was very old.
The guide said that the lord of the castle decaying due to acid rain was very old.
The guide said that the castle of the lord decaying due to acid rain was very old.
The other day the guide said that the castle decaying due to acid rain was very old.

Sue thought that the stopper of the decanter broken in the accident was very old.
Sue thought that the collector of the porcelain broken in the accident was very old.
Sue thought that the porcelain of the collector broken in the accident was very old.
Margaret apparently thought that the decanter broken in the accident was very old.

Pam thought that the collar of the jacket sewn in a very skilful way looked quite stylish.
Pam thought that the designer of the jacket sewn in a very skilful way looked quite stylish.
Pam thought that the shirt of the tailor sewn in a very skilful way looked quite stylish.
Kate certainly thought that the jacket sewn in a very skilful way looked quite stylish.

Eric saw that the rudder of the vessel corroding under the paint troubled the sailor.
Eric saw that the builder of the yacht corroding under the paint troubled the sailor.
Eric saw that the yacht of the playboy corroding under the paint troubled the sailor.
Eric could clearly see that the vessel corroding under the paint troubled the sailor.

Simon admitted that the jacket of the suit worn during the meeting looked very unfashionable.
Simon admitted that the maker of the suit worn during the meeting looked very unfashionable.
Simon admitted that the suit of the judge worn during the meeting looked very unfashionable.
My friend Alisdair admitted that the suit worn during the meeting looked very unfashionable.

Neil read that the region of the country neighbouring the poor nation was quite wealthy.
Neil read that the tyrant of the country neighbouring the poor nation was quite wealthy.
Neil read that the country of the tyrant neighbouring the poor nation was quite wealthy.
Some time ago Neil read that the country neighbouring the poor nation was quite wealthy.

Harry mentioned that the diamond of the collection stolen in the raid was very beautiful.
Harry mentioned that the owner of the diamond stolen in the raid was very beautiful.
Harry mentioned that the diamond of the millionaire stolen in the raid was very beautiful.
The other day Patrick mentioned that the collection stolen in the raid was very beautiful.

The boy reckoned that the base of the statue hewn from the granite was very strong.
The boy reckoned that the sculptor of the statue hewn from the granite was very strong.
The boy reckoned that the statue of the queen hewn from the granite was very strong.
The nice little boy reckoned that the statue hewn from the granite was very strong.

The tailor thought that the tunic of the costume woven in the mill looked very beautiful.
The tailor thought that the wearer of the costume woven in the mill looked very beautiful.
The tailor thought that the costume of the actor woven in the mill looked very beautiful.
The tailor definitely thought that the costume woven in the mill looked very beautiful.

Joe noted that the garage of the cottage teeming with hundreds of mice was very ugly.
Joe noted that the owner of the cottage teeming with hundreds of mice was very ugly.
Joe noted that the cottage of the farmer teeming with hundreds of mice was very ugly.
Some time ago Joe noted that the cottage teeming with hundreds of mice was very ugly.

Max said the brother of the reviewer phoning from the United States was a bit weird.
Max said the article of the scientist phoning from the United States was a bit weird.
Max said the reviewer of the article phoning from the United States was a bit weird.
Matthew and Richard said the reviewer phoning from the United States was a bit weird.

The clerk heard that the client of the lawyer laughing at the joke looked very smart.
The clerk heard that the mansion of the lawyer laughing at the joke looked very smart.
The clerk heard that the lawyer of the company laughing at the joke looked very smart.
The other day the clerk heard that the lawyer laughing at the joke looked very smart.

I read that the bodyguard of the governor retiring after the troubles was very rich.
I read that the province of the governor retiring after the troubles was very rich.
I read that the governor of the province retiring after the troubles was very rich.
I read quite recently that the governor retiring after the troubles was very rich.

Ben agrees that the assistant of the professor smiling at the student is quite good.
Ben agrees that the department of the professor smiling at the student is quite good.
Ben agrees that the professor of the department smiling at the student is quite good.
Sebastien certainly agrees that the professor smiling at the student is quite good.

The police said the accomplice of the burglar forgiven by the victims was soon discovered.
The police said the revolver of the burglar forgiven by the victims was soon discovered.
The police said the burglar of the bungalow forgiven by the victims was soon discovered.
The new police officer said that the burglar forgiven by the victims was soon discovered.

Mike asked if the servant of the prince strolling through the gardens was very old.
Mike asked if the palace of the prince strolling through the gardens was very old.
Mike asked if the servant of the palace strolling through the gardens was very old.
Michael asked a few times if the prince strolling through the gardens was very old.

John wonders why the colleague of the admiral talking to the woman is leaving so soon.
John wonders why the battleship of the admiral talking to the woman is leaving so soon.
John wonders why the captain of the battleship talking to the woman is leaving so soon.
The captain of the ship wonders why the admiral talking to the woman is leaving so soon.

Peter asked if the companion of the courier waving to the pedestrian was very reliable.
Peter asked if the motorbike of the courier waving to the pedestrian was very reliable.
Peter asked if the rider of the motorbike waving to the pedestrian was very reliable.
Peter asked a number of times if the courier waving to the pedestrian was very reliable.

Liz thinks that the patient of the surgeon walking down the corridor is quite old.
Liz thinks that the limousine of the surgeon walking down the corridor is quite old.
Liz thinks that the surgeon of the hospital walking down the corridor is quite old.
Katherine seems to think that the surgeon walking down the corridor is quite old.

Zoe heard that the colleague of the rector drinking the pleasant wine is well-known.
Zoe heard that the office of the rector drinking the pleasant wine is well-known.
Zoe heard that the rector of the university drinking the pleasant wine is well-known.
Zoe heard several days ago that the rector drinking the pleasant wine is well-known.

Jody thought the apprentice of the joiner smoking in the canteen looked quite strange.
Jody thought the equipment of the joiner smoking in the canteen looked quite strange.
Jody thought the repairman of the agency smoking in the canteen looked quite strange.
Jennifer and her friend thought the joiner smoking in the canteen looked quite strange.

It was reported that the sister of the survivor woken by the turmoil was in a bad state.
It was reported that the lifebelt of the survivor woken by the turmoil was in a bad state.
It was reported that the survivor of the disaster woken by the turmoil was in a bad state.
It was reported without delay that the survivor woken by the turmoil was in a bad state.

It was rumoured that the advisor of the mayor driven to the meeting had a lot of problems.
It was rumoured that the village of the mayor driven to the meeting had a lot of problems.
It was rumoured that the mayor of the village driven to the meeting had a lot of problems.
It was rumoured for a long time that the mayor driven to the meeting had a lot of problems.

Jane thought that the admirer of the painter humming the familiar tune was very nice.
Jane thought that the picture of the painter humming the familiar tune was very nice.
Jane thought that the painter of the picture humming the familiar tune was very nice.
Jessica immediately thought that the painter humming the familiar tune was very nice.

10.4 Appendix 4: Experimental items from Experiment 5

Der Herausgeber war ziemlich verärgert über den neuesten Klatsch im Geschäft.
Er wollte wissen, welches Fotomodell kürzlich die Journalistin ausgelacht hat, damit er den Streit schlichten konnte.
Der Herausgeber war ziemlich verärgert über den neuesten Klatsch im Geschäft.
Er wollte wissen, welcher Showmaster kürzlich die Journalistin ausgelacht hat, damit er den Streit schlichten konnte.
Der Herausgeber war ziemlich verärgert über den neuesten Klatsch im Geschäft.
Er wollte wissen, welchen Showmaster kürzlich die Journalistin ausgelacht hat, damit er den Streit schlichten konnte.

Die Tante war äußerst neugierig. Sie war erstaunt, als
sie hörte, welches Mädchen nachts die Dame angerufen hat,
weil sie dachte, dies sei ziemlich unverschämt.
Die Tante war äußerst neugierig. Sie war erstaunt, als
sie hörte, welcher Rentner nachts die Dame angerufen hat,
weil sie dachte, dies sei ziemlich unverschämt.
Die Tante war äußerst neugierig. Sie war erstaunt, als
sie hörte, welchen Rentner nachts die Dame angerufen hat,
weil sie dachte, dies sei ziemlich unverschämt.

Hans kam gerade vom Krankenhaus zurück, und hatte viel zu erzählen.
Er berichtete, welches Baby heute Mittag die Pflegerin angelächelt hat,
und wann seine Frau aus dem Krankenhaus zurückkommen wird.
Hans kam gerade vom Krankenhaus zurück, und hatte viel zu erzählen.
Er berichtete, welcher Arzt heute Mittag die Pflegerin angelächelt hat,
und wann seine Frau aus dem Krankenhaus zurückkommen wird.
Hans kam gerade vom Krankenhaus zurück, und hatte viel zu erzählen.
Er berichtete, welchen Arzt heute Mittag die Pflegerin angelächelt hat,
und wann seine Frau aus dem Krankenhaus zurückkommen wird.

Die Kinder sagten, daß sie das Kinderbuch sehr spannend fanden.
Es verriet nicht, welches Küken schließlich die Katze bemerkt hat,
und zwar bis zum Schluß des Kapitels.
Die Kinder sagten, daß sie das Kinderbuch sehr spannend fanden.
Es verriet nicht, welcher Vogel schließlich die Katze bemerkt hat,
und zwar bis zum Schluß des Kapitels.
Die Kinder sagten, daß sie das Kinderbuch sehr spannend fanden.
Es verriet nicht, welchen Vogel schließlich die Katze bemerkt hat,
und zwar bis zum Schluß des Kapitels.

Viele Leute kauften kürzlich die schrill aufgemachte Illustrierte. Sie zeigte auf der Titelseite, welches Staatsoberhaupt neulich die Fürstin geküßt hat, und berichtete darüber auf Seite Drei.
Viele Leute kauften kürzlich die schrill aufgemachte Illustrierte. Sie zeigte auf der Titelseite, welcher Premierminister neulich die Fürstin geküßt hat, und berichtete darüber auf Seite Drei.
Viele Leute kauften kürzlich die schrill aufgemachte Illustrierte. Sie zeigte auf der Titelseite, welchen Premierminister neulich die Fürstin geküßt hat, und berichtete darüber auf Seite Drei.

Der Knecht auf dem Bauernhof mußte sich um das Federvieh kümmern. Er wollte wissen, welches Huhn gestern die Ente verscheucht hat, um sie voneinander zu trennen.
Der Knecht auf dem Bauernhof mußte sich um das Federvieh kümmern. Er wollte wissen, welcher Hahn gestern die Ente verscheucht hat, um sie voneinander zu trennen.
Der Knecht auf dem Bauernhof mußte sich um das Federvieh kümmern. Er wollte wissen, welchen Hahn gestern die Ente verscheucht hat, um sie voneinander zu trennen.

Der Junge war ziemlich neugierig und fragte seinen Freund ständig aus. Zum Beispiel fragte er ihn, welches Kind gestern Abend die Frau gesehen hat, doch sein Freund sagte, es ginge ihn nichts an.
Der Junge war ziemlich neugierig und fragte seinen Freund ständig aus. Zum Beispiel fragte er ihn, welcher Gast gestern Abend die Frau gesehen hat, doch sein Freund sagte, es ginge ihn nichts an.
Der Junge war ziemlich neugierig und fragte seinen Freund ständig aus. Zum Beispiel fragte er ihn, welchen Gast gestern Abend die Frau gesehen hat, doch sein Freund sagte, es ginge ihn nichts an.

Wenn er nicht schlafen konnte, blickte der kleine Junge oft aus das Fenster.
Er hat gesehen, welches Mädchen abends die Spaziergängerin angesprochen hat,
kümmerte sich aber nicht weiter darum.
Wenn er nicht schlafen konnte, blickte der kleine Junge oft aus das Fenster.
Er hat gesehen, welcher Schüler abends die Spaziergängerin angesprochen hat,
kümmerte sich aber nicht weiter darum.
Wenn er nicht schlafen konnte, blickte der kleine Junge oft aus das Fenster.
Er hat gesehen, welchen Schüler abends die Spaziergängerin angesprochen hat,
kümmerte sich aber nicht weiter darum.

Der Besucher konnte sich nicht mehr an alle Gesichter erinnern. Er
war nicht sicher, welches Brautpaar morgens die Pastorin begrüßt hat,
weil dies schon ziemlich lange her war.
Der Besucher konnte sich nicht mehr an alle Gesichter erinnern. Er
war nicht sicher, welcher Bräutigam morgens die Pastorin begrüßt hat,
weil dies schon ziemlich lange her war.
Der Besucher konnte sich nicht mehr an alle Gesichter erinnern. Er
war nicht sicher, welchen Bräutigam morgens die Pastorin begrüßt hat,
weil dies schon ziemlich lange her war.

Der Vorsitzende ist ziemlich schlecht informiert worden. Daher
wußte er nicht, welches Mitglied seinerzeit die Besucherin befragt hat,
was für einige Verwirrung sorgte.
Der Vorsitzende ist ziemlich schlecht informiert worden. Daher
wußte er nicht, welcher Sprecher seinerzeit die Besucherin befragt hat,
was für einige Verwirrung sorgte.
Der Vorsitzende ist ziemlich schlecht informiert worden. Daher
wußte er nicht, welchen Sprecher seinerzeit die Besucherin befragt hat,
was für einige Verwirrung sorgte.

Das gemeine Volk war an den Ereignissen bei Hofe stets sehr interessiert,
und fragte sich, welches Mannequin seinerzeit die Prinzessin eingeladen hat,
und warum darüber nichts berichtet wurde.
Das gemeine Volk war an den Ereignissen bei Hofe stets sehr interessiert,
und fragte sich, welcher Politiker seinerzeit die Prinzessin eingeladen hat,
und warum darüber nichts berichtet wurde.
Das gemeine Volk war an den Ereignissen bei Hofe stets sehr interessiert,
und fragte sich, welchen Politiker seinerzeit die Prinzessin eingeladen hat,
und warum darüber nichts berichtet wurde.

Der Intendant des berühmten Schauspielhauses war ziemlich erstaunt,
als er erfuhr, welches Blasorchester angeblich die Dirigentin beeindruckt hat,
denn er hielt von beiden nicht sehr viel.
Der Intendant des berühmten Schauspielhauses war ziemlich erstaunt,
als er erfuhr, welcher Konzertverein angeblich die Dirigentin beeindruckt hat,
denn er hielt von beiden nicht sehr viel.
Der Intendant des berühmten Schauspielhauses war ziemlich erstaunt,
als er erfuhr, welchen Konzertverein angeblich die Dirigentin beeindruckt hat,
denn er hielt von beiden nicht sehr viel.

Die Reiter ließen ihre Pferde über Nacht im Stall. Am nächsten Morgen erzählte
der Stallbursche, welches Fohlen mittlerweile die Stute gewittert hat,
und daß die Tiere deswegen ziemlich unruhig gewesen sind.
Die Reiter ließen ihre Pferde über Nacht im Stall. Am nächsten Morgen erzählte
der Stallbursche, welcher Hengst mittlerweile die Stute gewittert hat,
und daß die Tiere deswegen ziemlich unruhig gewesen sind.
Die Reiter ließen ihre Pferde über Nacht im Stall. Am nächsten Morgen erzählte
der Stallbursche, welchen Hengst mittlerweile die Stute gewittert hat,
und daß die Tiere deswegen ziemlich unruhig gewesen sind.

Der Chef der Reinigungsfirma hat neulich einen neuen Dienstplan aufgestellt,
der festlegt, welches Dienstmädchen übermorgen die Putzfrau unterstützen wird.
Alles scheint gut durchorganisiert zu sein.
Der Chef der Reinigungsfirma hat neulich einen neuen Dienstplan aufgestellt,
der festlegt, welcher Hilfsarbeiter übermorgen die Putzfrau unterstützen wird.
Alles scheint gut durchorganisiert zu sein.
Der Chef der Reinigungsfirma hat neulich einen neuen Dienstplan aufgestellt,
der festlegt, welchen Hilfsarbeiter übermorgen die Putzfrau unterstützen wird.
Alles scheint gut durchorganisiert zu sein.

Die Tageszeitung hatte heute die internationale Umweltkonferenz zum Schwerpunkt,
und berichtete, welches Staatsoberhaupt gestern die Ministerin kritisiert hat,
was für einigen Wirbel in Diplomatenkreisen sorgte.
Die Tageszeitung hatte heute die internationale Umweltkonferenz zum Schwerpunkt,
und berichtete, welcher Wissenschaftler gestern die Ministerin kritisiert hat,
was für einigen Wirbel in Diplomatenkreisen sorgte.
Die Tageszeitung hatte heute die internationale Umweltkonferenz zum Schwerpunkt,
und berichtete, welchen Wissenschaftler gestern die Ministerin kritisiert hat,
was für einigen Wirbel in Diplomatenkreisen sorgte.

Ein Journalist der Bild-Zeitung sorgte letzte Woche für einiges Aufsehen, als
er aufdeckte, welches Mitglied seinerzeit die Partei unterstützt hat,
was nicht unbedingt zur Glaubwürdigkeit der Partei beitrug.
Ein Journalist der Bild-Zeitung sorgte letzte Woche für einiges Aufsehen, als
er aufdeckte, welcher Direktor seinerzeit die Partei unterstützt hat,
was nicht unbedingt zur Glaubwürdigkeit der Partei beitrug.
Ein Journalist der Bild-Zeitung sorgte letzte Woche für einiges Aufsehen, als
er aufdeckte, welchen Direktor seinerzeit die Partei unterstützt hat,
was nicht unbedingt zur Glaubwürdigkeit der Partei beitrug.

Der Chefredakteur war sehr zufrieden. Sein Magazin machte große Schlagzeilen,
als es aufdeckte, welches Fotomodell neulich die Prominente beleidigt hat.
Der Skandal sorgte für reißenden Absatz.
Der Chefredakteur war sehr zufrieden. Sein Magazin machte große Schlagzeilen,
als es aufdeckte, welcher Kameramann neulich die Prominente beleidigt hat.
Der Skandal sorgte für reißenden Absatz.
Der Chefredakteur war sehr zufrieden. Sein Magazin machte große Schlagzeilen,
als es aufdeckte, welchen Kameramann neulich die Prominente beleidigt hat.
Der Skandal sorgte für reißenden Absatz.

Die Öffentlichkeit war sehr überrascht über die Berichte in den Zeitungen.
Die Journalisten schrieben, welches Team kürzlich die Leiterin gelobt hat,
doch stellte sich dies später als falsch heraus.
Die Öffentlichkeit war sehr überrascht über die Berichte in den Zeitungen.
Die Journalisten schrieben, welcher Chor kürzlich die Leiterin gelobt hat,
doch stellte sich dies später als falsch heraus.
Die Öffentlichkeit war sehr überrascht über die Berichte in den Zeitungen.
Die Journalisten schrieben, welchen Chor kürzlich die Leiterin gelobt hat,
doch stellte sich dies später als falsch heraus.

Die Empfangsdame von dem vornehmen Hotel war sehr umsichtig.
Sie fragte, welches Paar soeben die Dienerin gesucht hat,
und versuchte zu helfen.
Die Empfangsdame von dem vornehmen Hotel war sehr umsichtig.
Sie fragte, welcher Ober soeben die Dienerin gesucht hat,
und versuchte zu helfen.
Die Empfangsdame von dem vornehmen Hotel war sehr umsichtig.
Sie fragte, welchen Ober soeben die Dienerin gesucht hat,
und versuchte zu helfen.

Der neue Regisseur war erstaunt darüber, wieviel Tratsch es in Hollywood gab.
Niemand wußte genau, welches Mannequin unlängst die Schauspielerin besucht hat,
doch jeder hatte seinen Senf zu der Geschichte beizutragen.
Der neue Regisseur war erstaunt darüber, wieviel Tratsch es in Hollywood gab.
Niemand wußte genau, welcher Millionär unlängst die Schauspielerin besucht hat,
doch jeder hatte seinen Senf zu der Geschichte beizutragen.
Der neue Regisseur war erstaunt darüber, wieviel Tratsch es in Hollywood gab.
Niemand wußte genau, welchen Millionär unlängst die Schauspielerin besucht hat,
doch jeder hatte seinen Senf zu der Geschichte beizutragen.

Der Pfarrer des kleinen Dorfes war einigermaßen unangenehm überrascht, als
er erfuhr, welches Schulkind vorgestern die Dorfbewohnerin genervt hat,
denn er kannte diese Dorfbewohnerin sehr gut.
Der Pfarrer des kleinen Dorfes war einigermaßen unangenehm überrascht, als
er erfuhr, welcher Verkäufer vorgestern die Dorfbewohnerin genervt hat,
denn er kannte diese Dorfbewohnerin sehr gut.
Der Pfarrer des kleinen Dorfes war einigermaßen unangenehm überrascht, als
er erfuhr, welchen Verkäufer vorgestern die Dorfbewohnerin genervt hat,
denn er kannte diese Dorfbewohnerin sehr gut.