SEMANTIC PROCESSING IN NATURAL LANGUAGE UNDERSTANDING

Sigrid Ute Lipka

Department of Psychology
University College,
London

Thesis submitted for
the degree of PhD in Psychology,
University of London.
September 1993.
Acknowledgements

I would like to thank my supervisor John Morton for his great support and guidance. I have greatly benefitted from his gift for clear-sighted theoretical debate. Thanks also go to the colleagues of the MRC Cognitive Development Unit for having provided an intellectual home for many years. Thanks are also due to Nico Preston and Simon Zielinski from the Department of Psychology at UCL; the former for having provided the software used in the first experiment reported here; the latter for helping with occasional interface problems.
ABSTRACT

This thesis investigates the interaction between lexical and syntactic processing. In the first experiment the semantic priming effect was used to test the activation of basic semantic features which were defined on the basis of Jackendoff's (1983) theory. Priming of lexical decisions was studied in Experiment 1 with visually presented nouns. Priming effects were found for an SOA between prime and target of 400 ms but only for associatively related pairs. There was no priming for word pairs which shared basic semantic features but were not associated and did not reflect a highly typical script-based or functional relationship. Experiments 2 and 3 indicated that basic semantic features are processed rapidly during sentence comprehension and are automatically integrated with the sentence. Subjects were slowed when making lexical decisions on semantically violating words in syntactically well-formed sentences. The effects of semantic violations were confirmed in Experiments 4 and 4b using a self-paced reading task, and a self-paced reading task with continuous grammaticality judgements, respectively. Results of Experiment 4 also showed that the effect of pragmatic violations occurred later on in the sentence than did the effect of semantic violations, indicating that the two have different processing origins, each with a different time course. In Experiment 4b, semantic violations were judged to make the sentence ungrammatical more often than pragmatic violations. Using the two self-paced reading paradigms, Experiments 5, 6 and 7 showed that semantic features of subject and object nouns affect the size of the garden path effect when reading syntactically ambiguous sentences in which both a transitive and an intransitive analysis are temporarily possible. The results of the studies allow us to specify a model of the interactions of the semantic subsystem, the parser and the thematic subsystem, which contrasts both with a strong thematic model and a single-analysis model.
# TABLE OF CONTENTS

## CHAPTER 1: OVERVIEW

1.1. Introduction ........................................ 15
1.2. Framework .......................................... 16
1.3. Processing the meaning of words ................. 21
1.3.1. Linguistics and the semantics of words .... 21
1.3.2. The psychological study of word meaning ... 22
1.4 Processing word meaning in sentence contexts: are subjects sensitive to semantic violations in grammatically well-formed sentences? .... 26
1.5. Do semantic and thematic information determine syntactic choices? .... 27
1.6. Outline ............................................ 30

## CHAPTER 2: PROBLEMS IN THE STUDY OF LANGUAGE COMPREHENSION

2.1. The levels problem .................................. 32
2.2. Methodological problems ......................... 36

## CHAPTER 3: THE SEMANTIC PRIMING EFFECT

3.1. Introduction: Semantic relatedness .............. 38
3.2. Semantic versus associative priming: what is the difference? .... 40
3.2.1. Possible mechanisms for priming ......... 50
3.2.2. Summary ....................................... 57
3.3. Lexical ambiguity research ..................... 59
3.3.1. Introduction .................................. 59
3.3.2. Multiple access hypothesis ............... 60
3.3.3. Selective access hypothesis ............... 63
3.3.4. Summary ...................................... 71

## CHAPTER 4: SEMANTIC VERSUS ASSOCIATIVE PRIMING
## OF NOUN PAIRS: EXPERIMENT 1

4.1. Introduction

4.2. Pilot Study

4.3. Method
   - Design
   - Predictions
   - Apparatus
   - Procedure
   - Stimuli
   - Subjects

4.4. Results

4.5. Discussion

## CHAPTER 5: PROCESSING THE MEANING OF NOUNS IN SENTENCE CONTEXTS

5.1. Introduction

5.2. Thematic relations and subcat frames

5.3. Jackendoff's competence theory of thematic relations and semantic selection restrictions

5.4. Psycholinguistic studies of the effect of sentence context on word processing

## CHAPTER 6: SENSITIVITY TO VIOLATIONS OF SEMANTIC SELECTION RESTRICTIONS: EXPERIMENTS 2, 3, 4 AND 4B

6.1. Introduction

6.2. Experiment 2
   - Introduction
   - Pretest

6.2.3. Method
   - Design
   - Task
   - Apparatus
Procedure 122
Stimuli 124
Subjects 128

6.2.4. Results 128
6.2.5. Discussion 136

6.3. Experiment 3: A follow-up: visual degradation and violations of semantic selection restrictions 139
6.3.1. Introduction 139
6.3.2. Pilot study 142
6.3.3. Method 144
6.3.4. Results 144
6.3.5. Discussion 146

6.4. Discussion of Experiments 2 and 3 146
6.4.1. Assumptions about types of information and processing systems involved in sentence priming 147
6.4.2. Review of models to account for sentence priming effects 149
6.4.3. A new post-access linguistic checking model 158

6.5. Experiment 4: The effects of violating semantic and pragmatic information on reading times in a self-paced reading paradigm 160
6.5.1. Introduction 160
6.5.2. Pretests 162
6.5.3. Method 163
Stimuli 163
Design 164
Task 164
Subjects 164
6.5.4. Results 164
6.5.5. Discussion 168

6.6. Experiment 4b: Violating semantic, pragmatic and syntactic information: the effect on decision times in a self-paced reading task with continuous grammaticality judgements 171
6.6.1. Introduction 171
CHAPTER 7: THE USE OF NON-SYNTACTIC INFORMATION IN PROCESSING SYNTACTICALLY AMBIGUOUS SENTENCES

7.1. Introduction
   Overview
   Measures of sentence processing
   Sources of information used in sentence parsing
   Temporary syntactic ambiguity

7.2. Models of processing garden path sentences
   7.2.1. Single-analysis: Frazier's garden path model
   7.2.2. Garrett's multiple-analyses model
   7.2.3. The Capacity Constrained Parsing Model
   7.2.4. Ford's Lexical Preference model

7.3. The role of lexical subcategorisation biases in parsing

7.4. Plausibility and semantic constraints in parsing

7.5. Thematic roles and processing garden path sentences
   7.5.1. Experimental evidence
   7.5.2. Models of the interaction of thematic and syntactic processing
      Rayner, Carlson & Frazier (1983)
      Stowe (1989)
      Carlson & Tanenhaus (1988)
   7.5.3. Empirical problems in Stowe's (1989) experiments
   7.5.4. Summary
CHAPTER 8: THE USE OF SEMANTIC INFORMATION IN PROCESSING GARDEN PATH SENTENCES: EXPERIMENTS 5, 6 AND 7

8.1. Experiment 5: Animacy/inanimacy of the subject noun and the garden path effect: a self-paced reading study

8.1.1. Introduction

8.1.2. Hypotheses

8.1.3. Pretest

8.1.4. Method

Subjects

Apparatus

Stimuli

Design

Task

Procedure

8.1.5. Results

8.1.6. Discussion

8.2. Experiment 6: Animacy/inanimacy of the subject noun and the garden path effect: self-paced reading with continuous grammaticality judgements

8.2.1. Introduction

8.2.2. Hypotheses

8.2.3. Method

Subjects

Task

Design

Stimuli

Procedure and Apparatus

8.2.4. Results

8.2.5. Discussion

Models

1. Frazier's single analysis model
2. Ferreira & Clifton (1986) 295
3. Pritchett's single analysis model 297
4. Multiple-analyses models 298
5. Ford et al.'s (1982) lexical preference model 299
6. Strong thematic models 300
7. My alternative parallel model 301

8.3. Experiment 7: Semantic properties of the direct object noun and the garden path effect: self-paced reading with continuous grammaticality judgements 305

8.3.1. Introduction 305
8.3.2. Hypotheses 306
8.3.3. Method 306
   Design 307
   Stimuli 309
   Subjects 310
8.3.4. Results 310
8.3.5. Discussion 319

CHAPTER 9: CONCLUSIONS 329
9.1. Introduction 329
9.2. Processing the meaning of single words 330
9.3. Processing word meaning in sentence contexts 335
9.4. Word meaning and the choice of syntactic options 344

BIBLIOGRAPHY 352

APPENDICES
A INSTRUCTIONS FOR EXPERIMENT 1 374
B WORD-WORD PAIRS USED IN EXPERIMENT 1 375
C WORD-NONWORD PAIRS USED IN EXPERIMENT 1 379
LIST OF TABLES

Table 4.1.: Experiment 1: percentages of incorrect No-decisions to word targets per condition 89
Table 4.2.: Experiment 1: percentages of incorrect Yes-decisions to nonword targets per condition 89
Table 4.3.: Experiment 1: percentages of incorrect decisions to nonword versus word targets per condition 90
Table 4.4.: Experiment 1: mean response times for correct decisions to word targets per condition 91
Table 5.1.: Examples of complex conceptual constituents in Jackendoff's theory 103
Table 6.1.: Experiment 2: basic semantic categories required by semantic selection restrictions of verbs 126
Table 6.2.: Experiment 2: Error Rates (%) per Stimulus Type & List 129
Table 6.3.: Experiment 2: Mean Reaction Times (ms) in each condition 134
Table 6.4.: Experiment 3: Percentages of false negative responses per condition 144
Table 6.5.: Experiment 3: Percentages of extremely long correct decision times per condition 145
Table 6.6.: Experiment 3: Mean response times for correct decisions on word trials per condition 145
Table 6.7.: Example stimulus set for Experiment 4 161
Table 6.8.: Experiment 4: cell means of average reading times (ms) per Sentence Type per position 165
Table 6.9.: Example stimulus set for Experiment 4b 172
Table 8.1.: Experiment 5: example stimulus set 249
Table 8.2.: Cell means of Experiment 5, position 6 256
Table 8.3.: Cell means of Experiment 5 for responses summed over positions 5 and 6 258
Table 8.4.: Experiment 5: percentage of extremely long responses per condition at position 5

Table 8.5.: Experiment 5: percentage of extremely long responses per condition at position 6

Table 8.6.: Experiment 5: percentage of extremely long responses per condition at positions 5 and 6 for the group of slow subjects

Table 8.7.: Experiment 5: cut-offs and percentages of extremely long responses per condition at the early positions

Table 8.8.: Experiment 6: example stimulus set

Table 8.9.: Number and (in brackets) Percentages of Errors per sentence condition and Position

Table 8.10.: Experiment 6, cell means (in ms) for position 5

Table 8.11: Experiment 6, cell means (geometric means) in ms for pos 5

Table 8.12.: Experiment 6: percentage of extremely long decision times per condition for position 5

Table 8.13.: Experiment 6, cell means in ms on position 5, for responses smaller than 1987 ms

Table 8.14.: Experiment 6, cell means in ms (geometric means) for position 4

Table 8.15.: Experiment 6, cell means at position 6

Table 8.16.: Experiment 7: example stimulus set

Table 8.17: Experiment 7: cell means (in ms) of correct decision times for position 4

Table 8.18: Experiment 7: Number and (in brackets) percentages of No-decisions per condition at position 5

Table 8.19: Experiment 7: cell means (in ms) of correct decision times for position 5

Table 8.20: Experiment 7: percentage of extremely long Yes-decision times per condition on position 5

Table 8.21: Experiment 7: percentage of extremely long Yes-decision times per condition on position 6

Table 6.3.1.: Experiment 2: Mean Reaction Times (ms) for Nonword Targets
LIST OF DIAGRAMS

Diagram 1.1. Framework 17

Diagram 3.1: Schematic representation of entities involved in lexical ambiguity studies 60

Diagram 4.1.: Sequence of events per trials in Experiment 1, Presentation Condition 1 76

Diagram 4.2.: Semantic hierarchy to represent the words 'son' and 'nurse' 78

Diagram 9.1. Information processing model of making lexical decisions on a target preceded by a single prime word 332

Diagram 9.2. Information processing model of making lexical decisions on a target preceded by a sentence 337

Diagram 9.3. Information processing model of reading semantically or pragmatically violated sentences 342

Diagram 9.4. Information processing model of reading garden path sentences 349

Diagram 6.1. Information processing model of making lexical decisions on a target preceded by a sentence 148
LIST OF FIGURES

Figure 6.1. Experiment 2: mean reaction times (ms) per condition 134

Figure 8.1. Experiment 5: mean reaction times per condition at position 6 257

Figure 8.2. Experiment 6: mean reaction times (ms) per condition on Position 5 276

Figure 8.3. Experiment 6: mean reaction times for responses smaller than 1987 ms 282

Figure 8.4. Experiment 7: mean reaction times for correct decisions at position 4 312

Figure 8.5. Experiment 7: mean reaction times for correct decisions for position 5 315
CHAPTER 1: OVERVIEW

1.1. Introduction

This thesis investigates semantic processing in natural language comprehension. It attempts to show which aspects of word meaning are processed automatically when people comprehend single words, and how semantic properties of words affect sentence processing.

There are two different approaches for analyzing semantic processing which I will briefly demonstrate using sentence (1) as an illustration:

(1) The lake was bitten.

One way of analyzing the sentence has to do with linguistic structure. This analysis would produce a description of the sentence as consisting of certain phrases, i.e. an NP, an AUX and a verb, and it would also indicate that it is a passive sentence. This phrasal description is enriched by specific semantic information about the verb and the noun as far as this information is relevant for building up a syntactic representation of the sentence (here of a representation as a passive sentence). In this example, the required semantic information would only involve some abstract semantic features of the noun (eg. Liquid), and semantic selection restrictions of the verb (eg. it requires an object noun that is Solid). In sentence (1) a clash of the semantic selection restrictions of the verb and semantic features of the noun would be noted.

Another way of analyzing the sentence, one that is typical in psychology, focuses on the concepts expressed in the sentence. It would provide a description of people's knowledge about the referents of the words, i.e. their knowledge about the objects, events etc. that words refer to. The oddness of sentence (1) would be accounted for in terms of what subjects know about lakes, and about the act of biting.
This thesis addresses the first type of analysis. It is primarily concerned with the linguistic status of word meaning, and the function of linguistically encoded meaning in language comprehension. It aims to specify the nature of semantic properties of words, and to demonstrate their linguistic relevance in terms of their effects in sentence processing.

1.2. Framework

The work reported here assumes a framework in which a Language Recognition Module (LRM) plays a central part. The LRM is a processing system that is exclusively concerned with the analysis of the linguistic structure of utterances. As such, it is distinct from the central system. Sentence processing cannot be done merely by analyzing people’s conceptual knowledge about the entities referred to by the words in the sentence. Given that an LRM is assumed, the question is how to characterize it, and to determine exactly what degree of autonomy it has.

What does the LRM do? In broad terms, it receives some representation of the words in the input and provides as output a representation of the meaning and relevance of an utterance which is passed on to a central processing system. The operation of the LRM involves several logically distinct types of processes. These are conceptualised as different processing subsystems. Diagram 1.1. below presents a minimal picture of the processing subsystems of the LRM, plus the system responsible for providing the initial representation of the physical stimuli, and the central system.
This thesis only deals with written input. Thus, I assume that there is an initial visual input analysis which produces an abstract visual representation of the words in the input, which is fed into a lexical system (without prejudice to the possibility of involvement of phonological representations. Note also that there might be several different input lexical systems depending on the input modality, see e.g. Morton 1979, 1980; 1989).

The lexical system is concerned with word recognition. I assume that the lexical system is a store of abstract visual representations of words which get activated via input from the visual input analysis and probably also via input from the central systems if there is, for example, a high expectation to encounter a particular word in a sentence. I further assume that each representation of a word in the lexicon also specifies the syntactic category of a word (e.g. noun, determiner, verb, auxiliary, verb, adjective, preposition etc.) and specific types of semantic information, i.e. semantic selection restrictions of verbs (see Chapter 5.3.) and basic semantic features of nouns (see Chapter 4.1.). I will refer to these specifications as lexical entries of words. Note
that this assumption is quite different from the view of the input lexicon in the Logogen model, in which word representations are evidence collectors which do not contain any entries specifying semantic or syntactic features of words. These are assumed to be computed by what is termed the cognitive systems. In my framework, parts of the cognitive systems of the Logogen model are treated as separate subsystems, i.e. as the thematic subsystem, the semantic subsystem and the parser, which will be characterized below.

I assume that upon word recognition in the lexical system, the information represented in the lexical entries of the recognized word is passed on to other components of the LRM.

The temporary linguistic store keeps a record of the ongoing sentence analysis. It receives input from the parser, the semantic subsystem and the thematic subsystem. It is required to explain the operation of the LRM as a whole, in particular the way in which the different subsystems interact during the processing of garden path sentences.

The parser is the processing subsystem which applies syntactic rules or principles to the representation of words which are input from the lexicon. In order to apply syntactic rules, the parser must have access to the lexical entry of a word specifying information about the syntactic category of words (eg. noun, determiner, verb, auxiliary, verb, adjective, preposition etc.), and to the subcat frames of verbs. It constructs phrases (eg. a noun phrase (NP) out of an adjective and a noun) and eventually a complete phrase structure tree, or an equivalent representation, to capture the syntactic structure of a sentence. Experiments 5 to 7 will address whether the parser always pursues a single analysis, or whether, when faced with syntactically ambiguous input, it produces multiple alternative analyses, which are kept in the temporary linguistic store. Analyses in the temporary linguistic store are assumed to have a certain level of activation which is affected by input from any of the three subsystems that feed into it.

The central questions addressed in this thesis concern the semantic subsystem of the
LRM which deals with the semantic information encoded linguistically. The task of such a semantic subsystem is to establish the semantic relationship of the sentence constituents that the parser combined on the basis of syntactic rules or principles (e.g., whether an adjective and a noun making up an NP fit semantically; or whether the head noun of a direct object NP fits the semantic selection restrictions of a verb). It might also compute the semantic relationship between two words in the absence of syntactic information, or possibly independently of syntactic analysis (Stowe, 1989, p.341, discusses the possibility that words are compared in a pairwise fashion with respect to their semantic features before syntactic analysis takes place). Thus, the semantic subsystem would minimally require as input a representation of word-specific semantic information and additionally a representation of the syntactic relationships of the words. The former is assumed to be provided from the lexical system, the latter from the parser. Given that the semantic subsystem requires information from the parser, further questions which will be addressed in several experiments are whether the semantic subsystem provides information back to the parser as to whether or not the syntactic constituents fit semantically, and in what way such information is used by the parser. What is the effect of a semantic mismatch of syntactic constituents? Does the parser build up a syntactic representation of parts of the sentence, and then check the output from the semantic subsystem, or is any syntactic decision suspended until the output from the semantic subsystem has been checked?

In addition to the potential output sent from the semantic subsystem to the parser, I further assume that output specifying the semantic representation of sentence constituents is passed to the temporary linguistic store and from there to the central system. Semantic information also needs to be passed on to the thematic subsystem. The thematic subsystem receives as further input the thematic frame or frames of verbs, which are specified in their lexical entries. Its task is to assign the thematic roles (specified in the thematic frame) to the constituents that are part of the verb phrase and the subject. Thematic roles (which will be characterized more fully in Chapters 5.2. and 7.5) are often referred to as case frames (see Fillmore, 1968). They
are the semantic roles played by the complements of the verb. For illustration, take sentence (1):

(1) He put the book on the shelf.

In this sentence, 'he' has the Agent role, 'the book' has the Patient or Theme role, and 'on the shelf' has the role of Goal. The thematic subsystem might use different types of information in assigning thematic roles, in particular the semantic features of the subject and the nouns of the verb phrase, and the plausibility of a particular assignment given the subject's knowledge about likely actions and events in the world. Experiments 5 and 6 attempt to tease apart these two sources of information. These experiments also test the impact on the parser of thematic assignments computed by the thematic subsystem, in particular with respect to sentences in which two syntactic analyses are temporarily possible.

Any empirical study of the components of the LRM and their interaction faces the problem of how to relate theories developed within different disciplines, i.e. linguistics, psychology and computational linguistics. I will refer to this problem as the levels problem, i.e. the problem of the relationship of different levels of describing and explaining semantic and syntactic aspects of language. Additionally, there are methodological problems of studying human sentence processing arising from the use of a great variety of experimental paradigms. The evidence collected with different methodologies needs to be compared in terms of task demands, sensitivity to different stages of processing, and scope for strategic factors and interindividual differences.

This thesis studies the lexical system and the syntactic, semantic and thematic subsystems by investigating the processing of word meaning in word pairs and in sentence contexts. The question is whether the kind of semantic features that are activated in processing single words are relevant for processing word meaning in sentences. In both areas of the study of word meaning, semantic information has frequently been confounded with other notions. In studying word pairs, I will contrast semantic and associative relations between words. In studying word meaning in
sentence contexts, I will distinguish between aspects of word meaning to do with semantic features of words, and the plausibility of words in context.

1.3. Processing the meaning of words

1.3.1. Linguistics and the semantics of words

In much of linguistics, especially transformational grammar, the study of word meaning has long been neglected. Language was often treated mainly as a syntactic entity. Much early psycholinguistic work consequently restricted itself to demonstrating the psychological reality of the syntactic rules developed by linguists (particularly Chomsky, 1957, 1965). The complexity of sentences in terms of number of transformations was used to predict the relative difficulty that different sentence structures would cause during processing (e.g. Miller, 1962). The initial apparent failure in psycholinguistic research to use descriptive syntactic rules for making straightforward predictions about processing difficulty (e.g. Fodor, Bever & Garrett, 1974) led some researchers to believe linguistic theory to be irrelevant for the psychological study of sentence processing. The apparent lack of psychological reality of transformations led some linguists to develop non-transformational linguistic theories which postulated a rich lexical system containing much of the information that transformational grammar assumed had to be generated (e.g. Bresnan, 1978). Note that this controversy about the psychological reality of syntactic rules illustrates the general, still topical problem of the relationship of theories from linguistics, psychology and (latterly) computational linguistics.

More recently, a number of developments have taken place within linguistic theory motivated by the impossibility of describing and explaining certain linguistic phenomena purely in terms of syntactic rules. Consequently, lexical and semantic factors received more interest.
1.3.2. The psychological study of word meaning

Psychological research has provided behavioural evidence for different types of word relations. The so-called semantic priming effect demonstrated a variety of relationships between two words which seem to be activated fast, and facilitate performance in some task to be carried out on the target. The semantic priming effect was originally described by Meyer & Schvaneveldt (1971) and Schvaneveldt & Meyer (1973). Semantic priming refers to the reliable finding that making a lexical decision, or reading out aloud a word (the target) is faster if another word (the prime) which is related in meaning to the target word, is presented very briefly before the target.

The relations between words that created priming are of a variety of types which, however, have not been studied systematically. Free association norms were found to separate out a variety of word relations that lead to priming from other relationships which do not lead to priming. This behavioural definition of semantic relations was taken to correspond to a shared underlying representational format. Relations between words that primed were represented as links between word nodes in a lexical network, and priming was assumed to result from activation spreading along those links. Taking identical behavioural outcome as evidence for identical representations, however, results in the ad hoc and potentially inflationary incorporation of word relations in the lexicon. Thus, when priming was found for word pairs which were not associatively related but which shared a variety of semantic relationships, they, too, were assumed to be represented in the lexicon. This, however, might lead to 'overcrowding' the lexicon with semantic information. Identical behavioural outcome does not force the assumption of identical underlying representations, e.g. as links between word nodes in a lexical network.

The semantic priming effect has been used extensively as a convenient vehicle in lexical ambiguity research (see Chapter 3.3.), which, however, also failed to define semantic relations in a principled way. This research is relevant to the issue of the relationship between the semantic and the syntactic processing subsystems since it indicates that all word senses (which are not properly defined) of a homonym are
activated fast, independently of contextual and syntactic factors. This is relevant because some people have argued that semantic information is not processed fast enough to be, at least in principle, available to the parser.

The definition of semantic relatedness of a word pair (both in word-word priming and lexical ambiguity research) is often not constrained in a principled way (despite the acknowledged fact that semantic relatedness is a complex notion which is difficult to define, Miller & Charles, 1991). To establish the kind of semantic information shared by words, frequent appeal is made, more or less explicitly, to semantic features or properties, i.e. to a componential or decompositional model of word meaning. Semantic relatedness or similarity is then defined in terms of the overlap of semantic properties. However, this does not really solve the problem because semantic properties (features, attributes) are themselves defined only loosely. It is quite widely accepted that the search for a small list of universal semantic features underlying the meaning of all words has not been very successful (Smith & Medin, 1981; Aitchison, 1987). Thus, so far semantic properties have remained "rather elusive notions" (Tabossi, 1988, p.335). Some researchers adopt a purely inductive data-driven approach by trying out whether there is priming for different kinds of semantic relations, starting with words that are near-synonyms. Alternatively, microfeatures of meaning are proposed quite arbitrarily in connectionist models (Hinton & Sejnowski, 1986; Hinton & Shallice, 1991; McClelland & Kawamoto, 1986). The most frequent procedure used to identify semantic properties of words is to ask subjects to list semantic features, rate features, describe features or rate the semantic similarity of words. However, it remains unspecified how subjects understand such tasks, and what they really do in such tasks. Note that these are all off-line tasks, in which subjects bring to bear information and processes that might be quite different to the ones relevant in on-line tasks. Furthermore, many authors note that it is impossible to overcome disagreements among subjects in making judgments regarding the meaning of words (Armstrong, Gleitman & Gleitman, 1983; Fodor, Garrett, Walker & Parkes, 1980; Hayes-Roth & Hayes-Roth, 1977).

The results on priming for semantic versus associative relations have been mixed. A
number of studies fail to find priming for semantically related words which are not associated, whereas others report evidence showing that priming occurred for semantic relations.

Note that if priming only works for associatively related words, it simply reflects "accidents of contiguity" (Fischler, 1977a, p.335), and as such would not be the appropriate method for studying meaning relations between words. Note also that a purely associative priming effect would not reveal anything about the activation of senses of a homonym in lexical ambiguity research. Alternatively, showing that priming does occur for semantically related words which are not associatively related does not force the conclusion that semantic information is stored in the lexicon. Especially given the wide range of semantic relations that, apparently, produce priming for non-associated words (eg. category coordinates, category name plus name of an exemplar, category name plus name of a property of the referent of the category name) one would have to postulate that a vast amount of semantic information is stored in the lexicon (if one takes priming as evidence for word relations stored in the lexicon).

Alternatively, it might be the case that these semantic relations are stored in the central knowledge system, and that they are processed and accessed extremely quickly, especially if the semantic relations are part of highly central or typical information about categories or scripts. On this view, priming effects do not directly reflect spreading activation in the lexicon.

The question is whether semantic relationships which do not operate in the priming tasks might have effects in other tasks. To explore this issue, a theory is needed of what non-associative semantic relations are.

In this thesis, the previous unsatisfactory ad hoc attempts at identifying semantic properties reviewed so far are replaced by a theoretically motivated definition. I derive semantic properties from Jackendoff's (1983) linguistic theory of word meaning. He describes word meaning as a mapping into particular basic ontological categories. A word like 'table', for example, maps into THING, 'lecture' into EVENT, 'colour' into
PROPERTY, etc. These are the most basic distinctions between word meanings. Obviously there is little reason to assume that words would prime each other just by virtue of sharing a mapping function into a particular basic category. Jackendoff, however, also suggests that the main ontological categories preside as the top node over a hierarchical tree of further nodes, eg. THING (the basic concept with the most elaborate hierarchy of subcategories) presides over the categories of ANIMATE THING and INANIMATE THING, and ANIMATE THING, in turn, has its own subcategories, such as HUMAN, ANIMAL, PLANT, etc. Note that these semantic properties probably only capture a small part of a subject’s knowledge of a word. Since semantic properties such as animate, inanimate, liquid, solid etc. are, however, likely candidates for the type of semantic information that might be syntactically relevant, they are possible candidates for the type of semantic information that is represented in the mental lexicon.

Thus, in Experiment 1 semantic properties, defined in this theoretically constrained fashion, are used to create semantically (not associatively) related noun pairs, which share the semantic mapping into sublevels of the basic ontological category THING. In Experiment 1, a lexical decision task is used to test priming for these semantically related noun pairs. I compare this to any priming found for associatively related words, using the same targets, and the same task.

Experiment 1 provides further evidence for evaluating the various models of the semantic priming effect that have been put forward to account for the conflicting results about semantic versus associative priming. The main concern, however, is to use semantic priming as a technique to study the activation of specific kinds of semantic relations between words. These relations are of the kind that has been suggested in linguistic theories as being required to enable syntactic operations.
1.4. Processing word meaning in sentence contexts: are subjects sensitive to semantic violations in grammatically well-formed sentences?

All other experiments reported in this thesis test directly whether the semantic properties derived from Jackendoff's theory are relevant to sentence processing. The experiments contribute to the modularity debate in that they help to specify the degree and time course of the interaction of semantic and syntactic information during comprehension.

Is initial sentence processing purely syntactic, encapsulated from other linguistic and non-linguistic information (eg. Clifton, Frazier & Connine, 1984; Forster, 1979; Frazier, 1987a; Rayner, Carlson & Frazier, 1983)? In such an account, semantic (and other nonsyntactic) information is assumed to be processed more slowly, and only once part of the sentence has been structurally interpreted. The alternative is that sentence processing involves a (possibly word-by-word) incremental process, in which all kinds of information about a word are being processed and used as soon as the word has been accessed, and while syntactic processing of the sentence is ongoing (eg. Crain & Steedman, 1985; Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1987; McClelland, St.John & Taraban, 1989; Tanenhaus, Carlson & Seidenberg, 1985; Tyler, 1989).

This debate obviously relates to the methodological issue introduced above. The central question does not seem to be whether different types of information interact during sentence processing, but when such interactions occur. To tackle this question about the window of interaction, one has to carefully consider the temporal characteristics of the empirical methods employed.

Experiments 2, 3, 4 and 4b were designed to provide evidence to help determine the time course of processing syntactic versus semantic information. Semantic information here was of the type that is required to fulfil semantic selection restrictions of verbs (Chomsky, 1965). Are the semantic requirements of verbs and the semantic properties of the noun (required by the verb) activated rapidly during sentence processing? Furthermore, Experiments 4 and 4b compared the effects of semantic violations to that
of syntactic and pragmatic violations.

The kind of semantic information required by semantic selection restrictions is a good focus for testing the temporal relationship of processing syntactic and semantic information because only a fixed set of very specific semantic information is involved. This seems more likely to be stored in the lexicon than a broad set of diverse semantic properties. Note that verbs in English only require quite specific semantic properties of their subcategorised nouns, if any. Consequently, semantic violations are of a very specific and uncontroversial type, as e.g. sentences such as 'I drank the powder', 'The lake was bitten' demonstrate.

These experiments also touch upon the issue of the relevance of linguistic notions for the study of human sentence processing. In linguistic theory certain types of information have been proposed to be part of a lexical entry, for example information about the syntactic category of a word, its subcat frame and thematic roles, and semantic selections. The assumption in parts of computational linguistics is that a parser would be inefficient if it did not use all types of lexical information since it needs to look up one of them (i.e. the syntactic category of a word) anyway (Weinberg, 1987). It still needs to be established, however, whether human sentence processing uses the linguistically proposed types of information in an 'efficient' way.

Previous research has shown subjects' early sensitivity to semantic violations (Hahne, Pütz, Friederici & Rösler, 1991; Marslen-Wilson, Brown & Tyler, 1988) but the evidence is unsatisfactory because of methodological shortcomings (no neutral baseline condition; in the Hahne et al. study, semantic requirements and predictability are confounded; limited range of semantic properties were tested).

1.5. Do semantic and thematic information determine syntactic choices?

The last set of experiments addresses not the temporal relationship of processing semantic and syntactic information but the nature of the interaction that is possible
between the semantic, the thematic and the syntactic subprocessing systems. Does the parser use information from the other two subsystems in deciding its processing steps?

I will investigate this question by testing the processing of temporarily syntactically ambiguous sentences. Previous research has shown that subjects typically prefer one of two possible syntactic analyses, and that they will have been led up the garden path if later evidence in the sentence is incompatible with the initially preferred option. The crucial question concerns the type of information that the syntactic parser uses to determine the initially preferred analysis. Altmann (1989) points out that a variety of information sources can contribute to the parser's decisions. The question investigated here is whether the parser is affected by semantic and thematic information in making the initial analysis.

The findings in the literature are mixed and represent the whole range of logically possible positions. On the one hand, the parser is assumed to be encapsulated from all non-syntactic information, and in particular it is assumed that "semantic information can't be used predictively to guide the parser" (Fodor, 1983, p. 135). Evidence supporting this position has been provided, for example, by Ferreira & Clifton (1986), Frazier (1987a, 1987b) and Frazier & Rayner (1982).

On the other hand it is maintained that the parser uses nonsyntactic information (such as subcat preferences, semantic information, default thematic assignments, sentence plausibility, previous sentence context) in making its decisions (eg. Altmann, Garnham & Dennis, 1992; Boland, Tanenhaus & Garnsey, 1990; Carlson & Tanenhaus, 1988; Crain & Steedman, 1985; Ford, Bresnan & Kaplan, 1982; Holmes, Stowe & Cupples, 1989; Stowe, 1989; Tanenhaus, Carlson & Trueswell, 1989; Tanenhaus, Garnsey & Boland, 1990; Taraban & McClelland, 1988; Pritchett, 1992).

The range of models that has been proposed to account for syntactic preferences differs in terms of their assumptions about the nature of the operations of the parser. Some authors believe that these are serial in operation (eg. Ferreira & Clifton, 1986; Frazier, 1987a; Pritchett, 1992; Wanner & Maratsos, 1978), others claim they operate
in parallel (Altmann, 1987; Altmann & Steedman, 1988; Bever, Garrett & Hurtig, 1973; Garrett, 1990; Jackendoff, 1987b, 1992; Marcus, 1980). While some authors maintain that multiple outputs would be too cost-intensive in terms of short-term memory load (eg. Frazier), others take multiple outputs as hallmarks of autonomous processes (eg. Seidenberg & Tanenhaus, 1986), claiming that "multiple outputs of modules do not necessarily lead to significantly greater processing load" (Tanenhaus, Carlson & Seidenberg, 1985, p. 375), as demonstrated, for example, by the evidence in favour of multiple exhaustive access to word meaning in research on lexical ambiguity. Frazier & Rayner (1987), however, object to drawing parallels between lexical and syntactic ambiguity, arguing that the former involves accessing prestored representations, whereas the latter requires the computation of representations. Evidence in favour of an exhaustive activation model for lexical ambiguities might not be relevant for predicting multiple outputs from the parser faced with syntactic ambiguity.

The evidence from Experiments 5 to 7 will provide further evidence for assessing the single versus parallel models of syntactic processing. These experiments investigate the impact of semantic and thematic factors in processing the type of garden path studied by Stowe (1989), i.e. sentences of the type:

Even before the police stopped the driver was getting nervous.

The phrase 'the driver' is temporarily syntactically ambiguous in that it can be attached as a direct object of the verb 'stopped', or as the subject of the following clause. Subjects prefer the direct object analysis and consequently get garden pathed when they reach the word 'was' which is incompatible with this analysis.

Sentences such as these will be used to test a variety of semantic manipulations and their effects on the initial choice of the parser. Semantic properties of words will be distinguished from contextual plausibility. Many studies in syntactic ambiguity research treat these two aspects of word meaning interchangeably.
Thus, these studies will provide new evidence to evaluate the various models of garden pathing proposed so far. They will also contribute to the methodological debate about the phases of sentence processing that are being measured by various experimental paradigms, and about the task demands of the different experimental paradigms.

1.6. Outline

Chapter 2 discusses further two of the general problems of psycholinguistic research mentioned in this chapter, i.e.

1. the levels problem: what is the relationship between theories developed in linguistics, psychology and computational linguistics. This is discussed in the context of the different levels of explanation suggested by Marr (1982)

2. the methodological problems: how can evidence about the operation and interaction of different language processing subsystems, which was collected with different types of experimental paradigms be compared? I will briefly discuss differences in task demands and in the sensitivity to various phases of processing.

In Chapter 3 the notion of semantic relatedness is discussed, and the main technique for studying semantic relations between words, i.e. priming, is introduced. Empirical problems in the study of semantic priming are identified, considering research both regarding single word priming and priming of homonyms in sentence contexts.

Chapter 4 presents an experiment using the lexical decision task to compare associative priming with priming for word pairs sharing theoretically justified semantic properties.

Chapter 5 focuses on how word meaning is processed in sentence contexts. It discusses the notions of semantic selection restrictions, subcat and thematic frames, and their characterisation in linguistics. Furthermore, I will review psycholinguistic evidence regarding the temporal relationship of semantic and syntactic processes during sentence
comprehension.

Chapter 6 presents several experiments investigating subjects' sensitivity to violations of semantic selection restrictions in syntactically well-formed sentences. Sensitivity is measured in two ways: as a delay in performing a secondary task, and by response times in two self-paced reading paradigms. Furthermore, one of the experiments compares semantic and pragmatic violations, and one semantic, pragmatic and syntactic violations.

In Chapter 7 the role of various types of non-syntactic information sources in determining syntactic decisions is reviewed with respect to garden path sentences. In particular, different models of the role of semantic, thematic and lexical subcat information are discussed.

In Chapter 8 several experiments are reported which test the influence of word meaning on syntactic processing during garden path sentences in two types of self-paced reading paradigm. These experiments investigate in turn the effect of semantic properties of a subject noun and the semantic properties of a potential direct object noun in determining the choice of an intransitive or transitive syntactic analysis in garden path sentences.

The final Chapter 9 contains concluding remarks. It discusses the findings of the experiments reported here in terms of the framework introduced in this chapter.
CHAPTER 2: PROBLEMS IN THE STUDY OF LANGUAGE COMPREHENSION

The research reported in this thesis reflects two of the prototypical problems in the study of language. These are a) the levels problem, i.e. the relationship between linguistic, psychological and parsing theories; this issue will be discussed briefly against the background of Marr’s (1982) framework; and b) the methodological problem, i.e. how to integrate evidence from the wide range of empirical methods that have been used in the study of language processing. The evidence collected with different methodologies needs to be compared in terms of task demands, sensitivity to different stages of processing and scope provided for strategic factors and interindividual differences to manifest themselves.

2.1. The levels problem

Adequate psychological research, according to Marr (1982), requires explanations at three autonomous levels: the computational, algorithmic and hardware levels. The first level defines what the function is that the system computes, the second specifies how processing is done, and the third characterizes the hardware device in which the process is to be realized physically.

1. At the computational level, a class of phenomena is defined in an abstract analysis as an information-processing task (e.g. in vision, the basic task, according to Marr, is a 2D-to-3D mapping). Analysis at this level provides an abstract formulation of the information-processing task and specifies the basic computational constraints, which hold for any system (human, animal, computer, Martian) faced with the task (in vision, these constraints are seen to result from the structure of the physical world). Boden (1988) compares the first level with Newell & Simon’s notion of an abstract task-analysis which they introduced in their computer modelling of problem solving.

2. At the algorithmic level, the researcher has to specify the computations by means of
which the task is performed that has been abstractly analyzed at level 1. "These processes are defined in terms of a particular system of representation" (Boden, 1988, p.50), which may be connectionist or symbolic. In principle, many different algorithms could carry out the same task.

3. At the hardware level the neural mechanisms that embody the computational and algorithmic functions are specified.

Boden (1988) underlines the interdisciplinary character of this framework and states that the "possibility of a detailed coupling of, and a fruitful interplay between, the three levels promises what cognitive science too often seems to lack: a theoretically integrated "interdisciplinarity"" (Boden, 1988, p. 60).

What is the relationship of the three levels of explanation? With respect to the study of language, what is the relationship between linguistic research, the development of parsers and psycholinguistics?

Linguistic theories can be seen as theories of linguistic competence (in Chomsky's sense), which corresponds to Marr's level of computational theory, as Marr has pointed out himself (Marr, 1982). As competence theories they describe what, in principle, the task of sentence processing consists of, and provide answers in terms of linguistic structures and the abstract rules that generate them. Processing theories (corresponding to Marr's algorithmic theories), on the other hand, address how the competence theory is realized. Are there any constraints that theories at one level impose onto theories at the other level? Marr thought of the two levels as quite independent of each other (and also independent of a third level, the hardware level, i.e. explanations at the level of the brain or computer hardware). Chomsky (1986a) maintains that the three levels are related, and that they are part of one and the same enterprise. Proposals about the complexity of transformations in early transformational grammar - which were put forward as abstract rules of generating sentences - were thought to be open to direct empirical tests measuring the time subjects took to verify sentences of different
complexity. The experimental results (e.g. Fodor, Bever & Garrett, 1974; Slobin, 1966) originally interpreted as demonstrating the inadequacy of Chomsky’s transformational theory have recently, in a curious development in psycholinguistics, been re-evaluated more favourably. Paradoxically, it was because of reconceptualizations of the nature of language processing that Chomsky’s competence theory appeared less questionable than after the original negative experimental results. The re-evaluation was brought about by Berwick & Weinberg’s (1984) criticism of the processing assumptions underlying the Derivational Theory of Complexity (DTC). It was assumed that language comprehension required transformations from surface structure to deep structure which were the inverse of those generating surface structure from deep structure. Furthermore, each transformation was seen to take a particular unit of time. Crucially, it was also assumed that all transformations are applied serially. Berwick & Weinberg argue that it might not have been the theory of grammar that was wrong, but the assumptions regarding processing. They propose a parsing model that assumes parallel processing, while upholding a transformational competence grammar.

In contrast to the DTC and developments, in which linguistic theory is seen as playing an important role in constraining theories of parsing, some computational linguists maintain that evidence about the operation of the processing system can constrain competence grammars (see e.g. Marcus, 1980), or at least can provide additional evidence for the validity of notions developed in linguistics (e.g. Pritchett, 1992).

Many approaches to language acquisition and sentence processing deny any relevance of the linguistic competence level for the study of language use and its development. Processing strategies and "operating principles" or preferences justified in terms of assumed short-term memory constraints are invoked to explain sentence comprehension (Bever, 1970; Frazier, 1987a, 1987b; Slobin, 1985).

For Marr the first level appears to have been of primary importance, both in his methodological views of scientific psychology (the first level should determine the questions asked at the other levels), and in his own research. He regards level-1 theorising as crucial if AI is to be of use for psychology. Marr in fact attacked AI,
arguing that many computer programs were patched up and made to run by ad hoc
programming tricks, while the programmer did not know what the program actually
did, and why it worked. Such AI modelling was not guided by an abstract analysis of
the task. Boden (1988) defends AI research against Marr's attacks. Only "angelic
scientists" (Boden, ibid., p. 52) might be able always to be guided by theoretical
analysis. She points out that some computer models were very useful for psychological
work in vision research even if researchers only understood much later why a program
succeeded or failed in particular tasks. She also mentions that some AI researchers
themselves stressed the importance of level 1, eg. Newell & Simon in their appeal to
abstract task analyses. However, in her own evaluation of computational models she,
too, emphasizes the necessity of level 1-analyses.

Note that this does not imply any a priori criticism of connectionist computer models,
which are understood by many to work in some holistic unspecifiable way which
cannot, in principle, be made explicit. Boden (1988) stresses that to explain how
systems (connectionist or symbolic) do x (or learn to do x), one has to have an
understanding both of what counts as x-ing, and what the system needs to be able to
do in order to be able to do x. Thus, even a pattern-matching connectionist system
requires a theory of grammar which specifies what is to count as good equilibrium
states of the system.

Boden stresses that level 1 might be necessary but clearly not sufficient in developing
psychological explanations (which Marr never claimed in the first place!). In many
domains there might be fast and computationally less expensive ways of doing a task
which were not predicted on the basis of competence models. These computational
short-cuts, Boden suggests, might result from familiarisation with certain inputs.

In language processing, computational processes, such as the construction of a full
parse tree, might be by-passed by a quick and dirty method of accessing a 'phrasal
lexicon' containing pre-stored interpretations of very familiar input.

Finally, I would like to stress that in the study of language there are two kinds of
algorithmic level, i.e. processing by computer-simulated parsers and processing by humans. The development of computational parsers has been useful, I think, for psycholinguistics, since it introduced a principled discussion of usually implicit assumptions about eg. the time units that transformations take, and the sequentiality or otherwise of mental operations carrying out transformations (eg. Berwick & Weinberg, 1984). However, the study of human processing requires further analysis in terms of performance factors, such as concentration, memory, goals and intentions of an individual. These performance factors influence the way in which the outputs of the linguistic systems are being used by a subject in a particular task. They will be discussed briefly in the next section.

2.2. Methodological problems

The fundamental problem is to decide what counts as relevant empirical evidence. There is by definition no pure measure of competence, even though traditionally (and even contemporarily, eg. Pritchett, 1992) linguists assume that competence can be tested directly by using grammatical judgement tasks. It is still an open question, however, which information from which sources subjects use when performing linguistic judgement tasks. Normally, the output of the parser is not available to conscious access. It needs to undergo some form of intellectual reconstruction, which then becomes available to conscious report.

The problem of measuring internal operations which are not consciously accessible, very fast and possibly occur automatically without conscious control, has been approached with a great variety of experimental paradigms, eg. priming studies, judgement tasks, reaction time paradigms, eye movement measurements. The results often squarely contradict each other. In the reviews to follow I will try to resolve some of the contradictions by specifying the different stages of sentence processing which are tapped by the various experimental paradigms.

The measurement of eye movement is thought by many to be the most sensitive on-
line measure of language processing. Relative to other measures, such as self-paced reading, lexical decision and naming tasks, eye movement data are seen to have greater validity (eg. Rayner and Pollatsek, 1987) since they are measured in a reading situation which is natural (the subject looks at continuous text) and does not involve any secondary activities. Alternative methods, such as presenting text sequentially word-by-word (as in RSVP, and self-paced reading), do not allow the subject to receive parafoveal information, or to move the eyes back to previous parts of the text.

However, even those convinced of the benefits of eye movement data admit that they are not easy to obtain and to interpret. Decisions have to be made, for example, about the units of text to be presented to subjects (eg. words, clauses or whole sentences) and about which parameters should be measured (eg. mean duration of the first fixation on a word; the probability of refixating within a word; single fixation durations for those first fixations which are not followed by refixations; mean gaze duration, usually defined as the "total fixation time on a word prior to the eye moving to another word" (Rayner & Pollatsek, 1987, p. 349); first-pass and second-pass reading times). The interpretation of eye movement data has to take into account the effects of parafoveal preview and the fact that more than one word can be identified during one fixation if the words are short. Furthermore, the relationship between eye movement measures and cognitive processes is complicated, and some eye movements might not be under any cognitive control. Processing might continue after the eye moved on to the next word (in what Rayner & Pollatsek, 1987 call "spill over" effects). Rayner & Pollatsek (1981) further showed that the decision of when to move the eye from a current fixation, and programming the saccade length, are affected by information available on the prior fixation. This evidence for spill-over effects contrasts with the "immediacy hypothesis" and its later weakened version (Carpenter & Just, 1983; Just & Carpenter, 1980). The first version stated that all cognitive processes required to understand a word are completed before the eye moves on. In the weakened version, all processes that can be completed given merely information from the word currently being fixated, are assumed to be completed before the eyes move to the next word.
CHAPTER 3: THE SEMANTIC PRIMING EFFECT

3.1. Introduction: Semantic relatedness

The so-called semantic priming effect was originally described by Meyer & Schvaneveldt (1971) and Schvaneveldt & Meyer (1973). Semantic priming refers to the reliable finding that making a lexical decision, or naming a word (the target) is faster if another word (the prime) is presented very briefly before the target which is related in meaning to the target word. Typically, being related in meaning was used interchangeably with being associatively related. However, it is still unclear whether semantic and associative priming are two distinct phenomena which are subserved by two different mechanisms. There are a number of studies, which I will briefly review below, that attempted to disentangle semantic and associative priming.

One of the challenges of trying to disentangle semantic and associative priming is to define semantic relatedness. This notion is difficult to define (Miller & Charles, 1991). One popular approach is to define it in terms of an overlap of semantic properties or features. However, since semantic properties (features, attributes) are themselves defined only loosely (or simply postulated, as is the case for microfeatures of meaning proposed in connectionist models: Hinton & Shallice, 1991; Hinton & Sejnowski, 1986; McClelland & Kawamoto, 1986), semantic relatedness remains elusive, too.

It is quite widely accepted that the search for a small list of universal semantic features which are necessary and sufficient for defining the meaning of words has not been very successful (eg. Aitchison, 1987; Fodor, Fodor & Garrett, 1975; Fodor, Garrett, Walker & Parkes, 1980; Rosch, 1973; Smith & Medin, 1981).

Even if semantic properties of words could be uncontroversially established, it is not clear whether all of them rather than a subset would always be activated automatically. As was briefly summed up in Chapter 3.2, research on lexical ambiguity investigates whether context can lead to the activation of only one meaning of a homonym.
Some models of word meaning propose that the differential activation of aspects of word meanings is not merely due to contextual factors. Instead, they assume typical or characteristic features as part of the mental representation of word meaning (eg. Barsalou, 1982; Jackendoff, 1983; Mervis & Rosch, 1981; Rosch, 1973; Smith, Shoben & Rips, 1974).

In priming studies, the most frequent procedure to identify semantic properties of words is to ask subjects to generate, rate or describe semantic features, or to rate the semantic similarity of words. However, it remains absolutely unspecified how subjects understand such tasks, and what they really do in them. Note that these are all off-line tasks, in which subjects bring to bear information and processes that might be quite different to the ones relevant in on-line tasks. Furthermore, many authors note that it is impossible to overcome disagreements among subjects regarding the correct decompositions of the meaning of words (Armstrong, Gleitman & Gleitman, 1983; Fodor, Garrett, Walker & Parkes, 1980; Hayes-Roth et al., 1977).

The variety of relations between primes and targets that has been studied illustrates that semantic relatedness is a broad and fuzzy notion. It includes category coordinates (such as dog - lion, eg. by Lupker, 1984; Moss, Ostrin, Tyler & Marslen-Wilson, 1992; Chiarello, Burgess, Richards & Pollock, 1990), category names and exemplars (such as bird - robin, eg. by Neely, Keefe & Ross, 1989) and words and names of properties of the objects they refer to (such as lemon - sour; shelf - wood, eg. by Tabossi 1988). Prime - target relations also cover shared perceptual or conceptual properties of the referents (eg. cherry - ball (is round), and banana - cherry (is edible), Flores d'Arcais, Schreuder & Glanzenborg, 1985; Schreuder, Flores d'Arcais & Glanzenborg, 1984).

In this thesis, the previous unsatisfactory attempts at identifying semantic properties reviewed so far are replaced by a theoretically motivated definition. I derive semantic properties from Jackendoff’s (1983) linguistic theory of word meaning. He describes word meaning as a mapping into particular basic ontological categories. A word like 'table', for example, maps into THING, 'lecture' into EVENT, 'colour' into
PROPERTY, etc. These are the most basic distinctions between word meanings. Obviously there is little reason to assume that words would prime each other just by virtue of sharing a mapping function into a particular basic category. Jackendoff, however, also suggests that the main ontological categories preside as the top node over a hierarchical tree of further nodes, eg. THING (the basic concept with the most elaborate hierarchy of subcategories) presides over the categories of ANIMATE THING and INANIMATE THING, and ANIMATE THING, in turn, has its own subcategories, such as HUMAN, ANIMAL, PLANT, etc. Note that these semantic properties probably only capture a small part of a subject's knowledge of a word. Since semantic properties such as animate, inanimate, liquid, solid etc. are, however, likely candidates for the type of semantic information that might be syntactically relevant, I consider them to be possible candidates for the type of semantic information that is represented in the mental lexicon. Thus, on this view, the lexicon is a linguistic entity, containing only those kinds of semantic information which are relevant for syntactic analysis.

3.2. Semantic versus associative priming: what is the difference?

To try to answer this question is important for a number of reasons. Firstly, if it were found that priming only works for associatively related words, simply reflecting "accidents of contiguity" (Fischler, 1977a, p.335), priming studies would not be the appropriate method for studying meaning relations between words. Secondly, a purely associative priming effect could not count as straightforward evidence showing that two senses of a homonym were (or were not) automatically accessed. Thirdly, a closer investigation of semantic versus associative priming forces a more explicit account of why two words are assumed to be related in meaning.

Research into single word priming has produced mixed results regarding the distinction of semantic versus associative relations. A number of studies fail to find priming for purely semantically related words (Huttenlocher & Kubicek, 1983, for a naming task; Lupker, 1984, for naming tasks; Napps, 1989 and Shelton & Martin, 1992, and for
continuous lexical decisions on sequence of single words), whereas others report
evidence showing priming occurred for semantic relations (Fischler, 1977a, for double
lexical decision task (LDT), i.e. decision whether simultaneously presented prime and
target are both words; Seidenberg, Waters, Sanders & Langer, 1984 for LDT and
naming; Lupker, 1984, for LDT; Chiarello, Burgess, Richards & Pollock, 1990, for
LDT and naming; Neely et al., 1989, for LDT; Schreuder, Flores d'Arcais &
Glanzenborg, 1984, for LDT and naming; Flores d'Arcais, Schreuder & Glanzenborg,
1985, for LDT and naming; Moss et al., 1992 for LDT and continuous lexical
decisions on sequence of single words).

Fischler (1977a) was the first to take literally the prevalent account of lexical priming
effects as semantic priming effects, assumed to be brought about by spreading
activation in a semantic network. However, Fischler's selection of semantically related
stimuli lacks a theory of the principles according to which words are related. Fischler
reports that the initial choice of semantically related words was made intuitively. The
resulting pairs were then used in a free association experiment (the first member of
each semantically related pair (S-pairs) was given as a stimulus). This established that
the second member of each pair was not an associate of the first. Furthermore, the (32)
semantically related pairs, together with 16 associatively related pairs (A-pairs) and 32
unrelated pairs were tested further in a similarity judgement experiment ("judge to
what extent the meaning of the two words seems to you to share any aspects or
characteristics", Fischler, 1977a, p. 337). A seven-point scale was used. S- and A-
pairs were rated as more related than the unrelated pairs, and the words in the A-pairs
were rated as being more similar than words in the S-pairs (mean rating of 5.1 for A-
pairs, and 3.8 for S-pairs). The 16 S-pairs with the highest similarity ratings were
included in the main experiment. Here, the task was a double lexical decision task in
which the subject had to decide whether the simultaneously presented prime and target
were both words or not (stimuli were presented until the subject made a response).
There was significant facilitation of lexical decisions for both the A-pairs (as
compared to their unrelated control pairs) and the S-pairs, as compared to their control
pairs.
Seidenberg et al. (1984, experiment 4) replicated this result with the same stimuli as used by Fischler in lexical decision and naming tasks (here prime and target were presented sequentially, not simultaneously, unlike in Fischler’s experiment. A fixation point was presented for 2.5 ms, followed by the prime which was displayed for 500 ms, and the target which was presented until the subject made a response, but not longer than 2000 ms). Seidenberg et al. pointed out, however, that the semantic priming effect (for non-associated words) was small and might occur for a limited set of words only. However, the associative priming effect was small, too. For both types of relations, priming was greater in the lexical decision task (32 ms for semantically related pairs, 31 ms for associatively related pairs) than in the naming task (11 ms for the semantically related pairs, and 9 ms for the associatively related pairs). Note that the statistical analyses had been done by subjects only. Seidenberg et al. attributed the priming effect for both associative and semantic relations to intralexical priming, i.e. spreading activation among relations between items within the lexical module, with post-access processes assumed to increase the priming effect in the lexical decision task.

Lupker (1984) consistently (his experiments 1 to 3) failed to find a pure semantic priming effect using a naming paradigm. He reported several experiments in which he varied the task (reading aloud prime and target versus reading just the target) and presentation rate (SOA of about 800 ms versus 250 ms). One has to note that he did find a small amount of facilitation for semantic pairs (7 ms in experiment 1 and 2, 6 ms in experiment 3), which was significant in the statistical analysis by subjects, and which he treated as a null effect because it was not significant by items. His semantic pairs were made up of two words from the same semantic category (the categories were body parts, kitchen utensils, furniture, animals, vehicles and clothing). No details are given about how particular category members were selected and combined to form related pairs. Association norms were consulted to ensure that the pairs were not associated.

In his experiment 4, Lupker compared priming for two types of associatively related word pairs. For half of these pairs, the two words were category coordinates (e.g. arm -
leg) and for half the words were from different categories (eg. author - book). For all associatively related pairs, reading aloud the target was facilitated (523 ms in the associated condition versus 541 ms in the unrelated control condition), but facilitation was not increased for pairs that were associated and semantically related (qua being category coordinates). Inspecting Lupker's data from his experiment 4 closely, it becomes apparent that the priming effect for the associated pairs might have resulted from a surprisingly long mean response latency for the unrelated control words for this condition (541 ms). The mean response latencies for the unrelated controls of the semantic (nonassociated) pairs in Lupker's first three experiments were lower, i.e. 527 ms, 525 ms, and 519 ms for experiments 1 to 3, respectively. Lupker does not comment on this. However, it is striking that the mean response time of 523 ms for the associatively related pairs in experiment 4 does not differ much from the mean response times for the semantic (nonassociated) pairs, which were 520 ms in experiment 1, 518 ms in experiment 2 and 513 ms in experiment 3. The priming observed for the associated pairs is not due to shorter responses but to an increase of response latencies for the unrelated control pairs. I cannot even speculate why the latencies for unrelated controls in experiment 4 were higher than in experiments 1 to 3; all unrelated trials were created by randomly combining primes and targets, and subjects for all experiments were recruited from the same pool of undergraduate students.

Lupker (1984) also reported results from two lexical decision experiments: in his experiment 5, using the same stimuli as in the naming experiments (experiments 1 to 3), he did find a significant priming effect for the semantic, non-associated pairs (548 ms versus 574 ms for the control pairs). The prime was presented for 550 ms, followed by a blank screen for 250, followed by the target, which remained on the screen for 750 ms, regardless of the latency of the lexical decision response (which was measured from the onset of the target presentation).

However, in Lupker's experiment 6 (in which the sequence of events per trials was the same as described for experiment 5), associatively related pairs that were of the same semantic category (same stimuli as in experiment 4) again failed to produce greater
facilitation of lexical decisions than associatively related words which were not of the same semantic category. There was a main effect of associative relatedness which did not interact with semantic category membership.

Thus in Lupker's results a pure semantic priming effect is quite elusive. The results do not, however, constitute conclusive evidence against pure semantic priming effects either, since at least in the lexical decision experiment priming was found for non-associated, semantically related word pairs.

Like Lupker, Huttenlocher & Kubicek (1983, experiment 2) also failed to find a priming effect for semantically related word pairs in a naming task. Related words were from the same category, and unrelated pairs were created by randomly combining a prime and a target word. Subjects had to read aloud both prime and target. The prime was presented until the subject initiated the naming response. Voice onset terminated the display of the prime, and after a blank screen of 600 ms, the target word was presented (no details are given about the display duration of the target). The intertrial gap was about 10 seconds. The relatedness effect was not significant in the min F analysis, but it was significant in the by-subjects analysis, and the by-items analysis. The percentage of related trials out of all trials (12.5% versus 87.5%) did not have any effect. Unfortunately, the authors do not present a list of their stimuli, nor do they describe the stimuli except to say that they were selected from published category norms.

Shelton & Martin (1992) claimed to have produced evidence that automatic priming occurs only for associated words. They used a single presentation lexical decision paradigm, which, they claimed, allows a better test of automatic processes than the more common paired lexical decision task. In this paradigm, subjects have to make a lexical decision on each, individually presented word (or nonword). Thus, there is no explicit pairing of primes and targets. Each letter string stayed on the screen until a response was made by pressing one of two buttons. There was a blank of 500 ms after the response before the next word was displayed. The results showed facilitation when a word was preceded by an associatively related word (as compared to response times.
if the preceding word was unrelated: 517 ms versus 553 ms). No such facilitation was found if the target shared semantic features with the preceding word, but was not highly associated with the preceding word. The semantically related pairs shared perceptual or functional attributes and belonged to the same superordinate category. The words in these pairs were not highly salient members of familiar superordinate categories. This pattern of priming for associatively related pairs, but lack of priming for semantically but nonassociatively related pairs was replicated in experiment 4 in the single presentation lexical decision paradigm, when the gap between response and next stimulus was reduced to 200 ms. Napps (1989, experiment 3) also found associative priming using the same paradigm (there are a few differences in procedure from the Shelton & Martin paradigm, which increased the time gap between stimuli in the Napps experiment: Napps presented a row of plus signs as a warning signal for 1000 ms before each word/nonword was presented; furthermore, after each response, either the message ERROR was presented for 1250 ms, or the response time was presented for 1000 ms, followed by a blank screen for 500 ms). Napps failed to find any priming effect for a second set of stimuli, in which pairs of words were designed to be synonyms (in a pilot study, pairs had been rated as related, with a mean rating of 6.24, on a 7-point scale with 7 designating a very strong relationship; the mean rating for associate pairs was 4.95). This result, however, should be interpreted with caution because Napps also reported that subjects made significantly more errors when a word was preceded by a synonym (eg. surplus - excess), or an unrelated word, as compared to being preceded by the identical word (excess - excess). This seems to indicate that there were some intrinsic difficulties either in accessing the stimuli, or in making decisions. Furthermore, subjects were significantly faster in responding to the set of stimuli used in the associative as compared to the synonym condition.

Shelton & Martin also used the more conventional paired presentation lexical decision paradigm (prime presentation of 250 ms, 500 ms interstimulus gap, target presentation of 250 ms, lexical decision on targets). Here, priming was found for both associatively related pairs and for semantically but non-associatively related pairs, in conditions with a high and with a low proportion of related pairs (74% or 26%). There was no difference in the amount of priming for the two types of relatedness. These priming
effects, the authors argue, are due a strategy of checking the relation between prime and target after completed lexical access to the target rather than being brought about by the automatic spreading of activation via shared semantic features. However, this interpretation does not seem to be consistent with the finding of priming when a low proportion of related pairs was used, which is conventionally (eg. den Heyer, Briand & Dannenbring, 1983; Tweedy, Lapinski & Schvaneveldt, 1977) seen as circumventing the problem of inducing postlexical processing strategies. Furthermore, it is not clear why the post-access checking process did not increase the priming in the lexical decision task for the associatively related pairs (compared both to the amount of priming found for the semantically but non-associatively related pairs in the lexical decision task, and compared to the amount of priming for the associatively related pairs in the single presentation lexical decision task), since these pairs should have benefitted from automatic spreading activation and post-access checking.

Moss et al. (1992) reported two priming studies to show that not only associative relations between a prime and a target word but also semantic (non-associative) relations of different sorts lead to priming. They interpret this as showing that different aspects of meaning are represented in the lexical entry of a word. In their experiments, the relation between words was either based on category coordination (eg. cat - dog) or functional relations (half of the functional relations were instrumental, such as in 'hammer - nail', half were script-based, as in 'pub - beer'). Half of the pairs for either type of relation was associatively related, half was not associatively related. In a pretest, subjects were given lists of word pairs for rating. These included the final experimental stimuli and distractors such as pairs that were synonymous, that rhymed and that were unrelated. All semantic and associated pairs included in the main experiment were rated as being related. Two priming experiments were run: one with word pairs, one with items presented singly. In the word pair study, both prime and target were presented auditorily, with an ISI of 200 ms. The subject had to make lexical decisions on the targets. Included as stimuli were (related and unrelated) word-word pairs and word-nonword filler pairs. No neutral baseline condition (such as presenting the word 'next', or a warning signal) was used. All items were mixed, rather than blocked in an attempt to discourage subjects from using conscious
strategies. Priming was found for both types of related pairs. The associated pairs led to more priming than the non-associated pairs for both types of semantic relation (category coordination and functional relation).

In the second experiment words were presented singly at a rate of 1000 ms. Subject had to make lexical decisions on each word. This set-up was chosen since it has been argued that it reduces subjects’ use of strategies (see Shelton & Martin, 1992). There is no pairing of prime and target, and yet it has been found, and was found here, too, that making a lexical decision is faster if a word is preceded by a related word. The effects were the same as in the first experiment, with the overall amount of priming reduced, however. It is not clear why their semantically but not associatively related pairs produced priming in this single presentation lexical decision task, whereas Shelton & Martin (1992, see above) did not find priming for such pairs in the same paradigm, using visual rather than auditory presentation. Moss et al.’s results are also in contrast to Lupker’s (discussed above) since they found that priming was greater for semantically related pairs that were also associatively related as compared to semantically but not associatively related pairs.

Moss et al. interpret their findings as showing that different types of information (category coordination, instrumental and script-based relations) are stored in lexical entries and are automatically accessed. They give up the notion of a fixed boundary between lexical and encyclopedic semantic knowledge, and assume that there is in principle no limit to the type of semantic properties that might become activated during word recognition (including properties that are traditionally considered to be encyclopedic in nature), and that might be postulated to be stored in the lexicon.

Thus, superficially, their conclusions contrast with the view that the lexicon only contains minimal semantic information, i.e. merely the types of semantic features that are required in syntactic processing (eg. to fit the semantic selection restrictions of verbs). Tyler and Marslen-Wilson’s interactionist Cohort Model of sentence processing (Marslen-Wilson, 1987; Zwitserlood, 1989) assumes that pragmatic information and a range of semantic and conceptual information is processed as fast as syntactic
information. The priming effects reported by Moss et al. seem to support the view that such conceptual information is accessed fast.

It is not clear, however, if general knowledge can affect the choices made by a syntactic parser (see Chapter 7). Furthermore, Moss et al.'s study leaves open some questions. Firstly, as mentioned above, prime and target pairs of the different types of relatedness were matched for semantic relatedness as established in a pretest. One cannot be sure that subjects did not use 'association' between words as an implicit criterion in their judgements. It might be that the 'semantic' pairs had high co-occurrence and/or predictability even though they were not associated according to free-association norms.

A second question concerns the different status of word relations. Moss et al.'s studies showed that category coordinates and functional pairs lead to the same amount of priming, and that priming for both types of pairs is increased when they are also associatively related. This does not necessarily force the conclusion that all these kinds of information are stored in the lexicon. It might be that script-based information is stored in a central knowledge system, but that it is also processed extremely fast (especially if it is highly central or typical information from a script).

Chiarello et al. (1990) report priming effects for semantic pairs in both the lexical decision task and the naming task. Semantic pairs were category coordinates and the words in the pair were not associated (e.g. pony - deer). Latencies in the naming task were quite slow (around 770 ms in the related condition, and 800 ms in the unrelated condition), which was probably due to the fact that presentation conditions were made difficult: prime and target were presented laterally to either the left or the right visual field, and both were presented for only 100 ms, with a blank screen in between of 500 ms. The authors point out that it might have been because of these difficult stimulus conditions that semantic priming was found in their naming task: because processing of the target word is slowed, activation from the prime word has more time to spread. Thus, they interpret the often found discrepancy between the results obtained by the lexical decision task versus the naming task (i.e. semantic priming effects for the
former but not for the latter) as resulting from differences in the speed with which a
response can be made: under normal presentation conditions, a naming response can be
made faster than a lexical decision response, and thus priming resulting from spreading
activation has time to occur only in the lexical decision task but not in the naming
task.

Schreuder & Flores d’Arcais (1989) summarised the priming evidence form several
experiments conducted in Dutch. In the lexical decision task, they found priming for
perceptually related pairs, such as 'ball - cherry' (mean decision time 681 ms),
conceptually related pairs, such as 'banana - cherry' (670 ms), and pairs related both
perceptually and conceptually, such as 'apple - cherry' (654 ms). Priming was defined
as the difference in decision times in the related conditions versus the unrelated
condition (694 ms), in which the target was randomly paired with an unrelated prime
word. In the naming task, using the same stimuli as in the lexical decision task, no
priming was found for the conceptually related pairs which did not also share
perceptual features (eg. 'banana - cherry'; mean response time 619 ms), but priming
was found for the two other types of stimulus pairs (eg. 'apple - cherry': 600 ms; and
'ball - cherry': 601 ms). In interpreting this pattern of results, the authors refer to the
different response speeds in the naming versus the lexical decision task in the same
way as Chiarello et al. did, pointing out that responses are faster in the naming task
than in the lexical decision task. To explain their effects, they suggest that perceptually
based features of referents of words become available faster than conceptually based
properties. Thus, in the faster naming task the perceptual but not the conceptual
properties would have been activated by the time a response had to be made, whereas
in the slower lexical decision task, both types of properties would have been available.
Flores d’Arcais et al. (1985) showed that if the response times in the naming task were
experimentally increased by degrading the visual quality of the target, priming was
found for the conceptually related pairs (eg. 'banana - cherry').
3.2.1. Possible mechanisms for priming

I think that the final verdict about which kinds of word relations lead to facilitation can not as yet be given. Priming for associatively related pairs has been found more consistently than priming for semantically related pairs, especially when the naming task was used. It is still not clear, however, how priming for associatively related pairs actually works. Why should association norms determine which words can facilitate some performance on other words? And what are the factors determining word associations in the first place?

Can associative pairs be taken as instances of some implicit learning of predictive correlations or contingencies such that the activation of one word leads to the activation of its associates but without any activation of meaning? Activation from the prime could be thought of as spreading along a link that was established on the basis of some 'blind' learning of frequent co-occurrences between the words. The fact that associatively related words share some meaning relationship would be coincidental on this account. Alternatively, priming might occur because associative pairs are related by some prototypical or salient thematical or functional link, i.e. by relationships that require general knowledge (eg. doctor - nurse; bug - spy; mermaid - water), assuming that such prototypical knowledge can be accessed fast.

I would suggest that this issue can not be resolved by demonstrating that association norms are correlated with measures of co-occurrence of stimulus and primary associate in written discourse (as shown by Rapp, 1991; Spence & Owens, 1990). The contentious issue is whether the high degree of co-occurrence frequency, and the strong associative links, are a result of accidents of contiguity, which are best represented in a "nonsemantic lexical network" (Shelton & Martin, 1992, p. 1205) or whether they emerged as a result of knowledge about shared typical characteristics of, or functional relations between the referents of the words.

The notion of unconstrained registration of co-occurrence does not seem to be plausible in principle (see Morton, 1964) since this would entail that every noun, for
example, should have as its primary associate some determiner, which clearly is not the case. It has to be assumed that individuals register and store only some kind of predictive correlations of the occurrence of words, such that only informative links are important, rather than links from one word to many other words. In order to notice such predictive relationships the system has to apply linguistic constraints (such as membership of syntactic categories) onto an otherwise equipotential amorphous mass of contingent words.

Nevertheless, some authors, such as Shelton & Martin (1992), take such correlations between association- and co-occurrence norms as explanatory. They argue that priming for associated pairs resulted from connections in the word name network but not in the conceptual network (to use Collins & Loftus' terms): "associative priming might result from connections between lexical phonological or orthographic representations that have developed on the basis of co-occurrence frequency, rather than from connections at a meaning level" (p.1207). Words that occur together frequently are assumed to be strongly linked in the word name network. Note that in the original Collins & Loftus model (1975), the links between the names of concepts were conceived of as reflecting "phonemic (and to some degree orthographic) similarity" (Collins & Loftus, 1975, p.413), but not co-occurrence frequency. Shelton & Martin argue that the single-presentation lexical decision task, but not the paired presentation lexical decision task, taps the automatic spread of activation along the links in the word name network.

Note that even if one were to regard associative pairs as accidents of contiguity/co-occurrences, one need not rule out that meaning relations play a role both for association norms and textual co-occurrences. Most associates are close in meaning, either because they are of the same category, or because they share a thematic or functional relation. Some associations might be idiosyncratic, but subjects do not tend to give responses that are, say, merely related phonologically (eg. each beginning with the same letter; or words that rhyme: Klang-associations are rare). The strategy of giving responses that are semantically related is not predetermined in the instructions of free association tasks: typically, the instruction is to respond with the first word (or words, in a continuous free association experiment) that comes to mind but no
restrictions are made on what sort of words may be produced (see Woodworth and Schlossberg (1955) and instructions in the various reports of association norms). The very nature of association norms, however, also strongly suggests that the meaning relationships of associates reflect a highly overlearned kind of shared meaning. Association norms per definition stand for the frequency with which individuals produce identical primary associates.

The double nature of associatively related words as reflecting accidental over-learned co-occurrences of words as well as some meaning relation between them has led to two types of accounts for associative priming effects. Priming is accounted for by activation spreading either through a network of words (or their phonological or orthographical representations) or a network of concepts to which the words are related. Shelton & Martin's account discussed above exemplifies the former approach. Seidenberg & Tanenhaus (1986) also propose a lexical network, suggesting that priming occurs for words that are close in the network "in terms of number of intervening nodes, or a distance metric" (Seidenberg & Tanenhaus, 1986, p. 143). Unlike Shelton & Martin, however, they allow for activation to spread not only to highly associated words, but also to strongly semantically related but nonassociated words.

Forster (1979) provided one of the most detailed accounts of priming within a lexical, rather than a conceptual, network. He adopts a strong autonomy position of lexical processing: "the process involved in locating the lexical entry of a word cannot be modified by processing at either the syntactic level or at the message level" (Forster, 1979, p. 56). He distinguishes the lexicon, which contains knowledge about a word, from the lexical processor, which accepts features from the peripheral perceptual systems and accesses the entry in the lexicon that corresponds to the word in the input. Word recognition is an active search process. The lexical processor is like a library catalogue, providing files which allow the search for a lexical entry under different descriptions, i.e. according to orthographic or phonetic properties (and according to semantic and syntactic properties which are used during language production). Associative and semantic priming effects are interpreted as resulting from the semantic
organisation of the lexical entries in the lexicon: "lexical entries for related words are linked together in a network of cross-references" (Forster, 1979, p. 71). Priming for a target word such as 'nurse' is assumed to demonstrate that its lexical entry was accessed not via the access files of the lexical processor, but via a search inside the lexicon along the semantic cross-references starting from the prime word 'doctor'. Forster (1979) assumes that a whole set of semantically related lexical entries would be searched after 'doctor' had been accessed, not just entries corresponding to high associates of the entry for 'doctor'. Thus, priming is an "intralexical effect" (Forster, 1979, p. 72), and the hypothesis of the autonomy of lexical processing is not threatened because no information from other levels of the language processor (syntactic and message levels) is involved. This model is intended as a proposal about the information structures that underlie the priming effect, and "it is quite irrelevant (...) whether the effect is a product of automatic, involuntary processes, or consciously controlled strategies, or a mixture of the two" (Forster, 1979, p. 73).

Lupker (1984) accounts for the associative priming effects by suggesting a modified model of the conceptual, rather than the word name, network. He proposes "direct, associative links" (Lupker, 1984, p.727) in the conceptual network along which activation spreads, thus facilitating lexical access. This account seems to take one step further Collins & Loftus' (1975) assumption that the links in the conceptual network have different criterialities or strengths. Note, however, that in Collins & Loftus model the links represent different kinds of semantic relations between concepts, whereas Lupker's associative links do not represent any semantic relationships but merely the strength of an association between concepts.

In the Collins and Loftus (1975) model (C&L for short), there is a conceptual/semantic network and a lexical network. The lexical network consists of nodes that represent the names of concepts. These nodes are connected by links representing the words' shared phonemic (and orthographic) properties. Furthermore, each name node is connected to one or more of the concept nodes. The concept nodes are represented in the conceptual network, which is also called the semantic network. This is where associative (and potentially semantic) priming is supposed to occur.
The structure of the conceptual network can very briefly be described as follows. Concepts are represented as nodes which are linked to each other according to semantic similarity ("The conceptual (semantic) network is organized along the lines of semantic similarity.", p. 411). The more related two concepts are, the more links exist between them. The number of links to other concepts depends on the number of properties one concept shares with others. The kinds of properties concepts can share is not explicitly defined. In Quillian’s theory (1962, 1967, 1969, quoted in C&L, p. 407), on which C&L base their own model, properties of a concept are represented as "labeled relational links" (C&L, p. 408) from one concept node to another. Different kinds of links are proposed, i.e. superordinate and subordinate links, modifier links, disjunctive sets of links, conjunctive sets of links, and a class of links which allows the "specification of any relationship where the relationship (usually a verb relationship) itself was a concept" (C&L, p. 408). C&L assume that the links have various strengths or 'criterialities'. For example, superordinate links are highly criterial. Links can represent functional properties, e.g. a property of 'chicken' is that people eat them, that chicken are raised on farms and that they are large. In light of current controversial findings regarding priming with semantically but not associatively related words, it becomes crucial to understand the kinds of properties that C&L claimed as the basis for connections between concepts in the network, and to understand how the criteriality or strength of links is determined (both in terms of how the individual builds these up, and in terms of the researcher's theory guiding his hypotheses about links and their strengths).

C & L discuss the role of association when they reinterpret a categorisation experiment by Loftus (1973, quoted in C&L, p. 416) within the spreading activation framework. Subjects were presented with a category name and the name of one instance of the category. Either category or instance name were given first. Relevant to the present discussion is that the words were selected such that they were either mutual close associates, or unidirectional close associates. Note that the association is assumed to hold between concepts, not words. It was found that subjects were fastest if the first item, be it the category or the instance, evoked the second with high frequency (i.e. when the second item was a primary associate of the first). The interpretation is that if
one concept evokes another concept with high frequency, more activation spreads to that concept. It is important, I think, that C&L assume that highly associated concepts are processed faster, but otherwise by the same mechanisms as other related concepts. This is also demonstrated in their discussion of Meyer & Schvaneveldt’s work when they refer to the related words as "semantically similar" (p.419) without remarking on the fact that these were also highly associated words. One can conclude that for C & L semantic priming can legitimately be demonstrated with associatively related words. There is no concern yet with distinguishing associative from semantic word relations since both are represented by a conceptual network, the only difference being that activation spreads faster among concepts standing for associatively related words.

The lack of a purely semantic priming effect (eg. Shelton & Martin, 1992) seems to suggest that C&L’s conceptual network might have to be modified, or restricted, to comprise only nodes between those concepts that are associatively related. Alternatively, the network as originally proposed could be left unchanged; finding a lack of semantic priming would merely underline C&L’s assumption that activation spreads faster along those links that hold between associatively related concepts as opposed to those links holding between semantically related concepts. Note that this view presupposes that free association norms reflect relationships between concepts, rather than implicitly learned predictive correlations of words (as eg. Shelton & Martin, 1992, assume).

Lexical priming effects have also been interpreted in a different framework altogether. Rather than accommodating them in some kind of network account or other (either a conceptual or a word-node network) and invoking spreading activation as the mechanism for (pre-access) priming, lexical priming effects have been discussed in the context of post-access checking procedures, in which the target and prime are semantically compared during the decision stage of making lexical decisions. This kind of interpretation would allow a modular view of lexical access, but a nonmodular view of post-access processes. Such post-access interpretations have usually been invoked to explain why a semantic priming effect is more likely to be found in lexical decision tasks than naming tasks (eg. Lupker, 1984; Shelton & Martin, 1992).
Forster (1990) characterizes post-access processes as procedures checking whether the just-accessed word (the target) stands in some meaningful relation with the context (the prime). Any relationship, such as thematic or episodic, can be used as evidence during this checking process (in fact, as STROOP effects indicate, subjects cannot ignore even irrelevant information). If the result is negative, a close orthographic check is made on the target to confirm that the word accessed in the lexicon is in fact identical with the target. This check increases the response latency in the lexical decision task. If there is a meaningful relation between target and context, only a fast and superficial orthographic check needs to be made. Forster suggests that "the semantic priming effect does not reflect properties of the lexical processor but does reflect properties of the decision system" (p. 127), which uses "any kind of relationship between the prime and the target (...) whether it be semantic, associative, formal, or even episodic" (p. 128). While this account seems plausible for priming effects obtained with the lexical decision task, it is not clear how to apply it to priming effects found with the naming task (e.g. Chiarello et al., 1990; Schreuder & Flores d'Arcais, 1989; Seidenberg et al., 1984). Forster (1990) seems to be in two minds about this problem, stating on the one hand that "the fact that the target word is meaningfully related to the context tells us nothing about how to pronounce it" (p. 106), and on the other that "the naming task does involve at least one decision, and that is the decision that the correct entry has been accessed" (p. 127).

Thus, in Forster's (1990) account, lexical priming does not provide evidence about processes which facilitate access to a word (or about the proposed semantic organisation of the lexicon, as proposed by Forster, 1979), but instead it sheds light on central, decision-making processes that occur rapidly after the target has been accessed. What would make lexical priming evidence so remarkable, it seems to me, is that it shows how very fast these post-decision checking procedures operate.

This account, however, does not appear to fit easily with the findings from lexical ambiguity research (see Chapter 3.3.). The evidence favouring multiple access demonstrates that post-access meaningfulness checking procedures can not have been completed when the target occurs at the offset of the prime. Context, which in this
case is not a preceding word but a sentence, has a meaningful relation to only one sense of the ambiguous word, yet both senses prime their targets.

Schreuder & Flores d'Arcais (1989) give a unified account of semantic priming regardless of the experimental task. Remember that they (and Chiarello et al., 1990) found a semantic priming effect both in the lexical decision task, and the naming task, provided that response speed in the naming task was experimentally reduced (either by presenting the target in a visually degraded format, or by presenting it laterally to one visual field, and reducing display times). Schreuder & Flores d'Arcais (1989) suggest that semantic priming effects are mediated by shared semantic properties. Semantic priming is attributed to activation feeding forward from word nodes (via their concepts) to semantic properties, raising the level of activation of the nodes representing the properties. These properties represent a subjects' encyclopedic knowledge about the referents of the words, encompassing perceptual, functional and abstract knowledge (eg. a referent’s superordinate category). Since properties can be shared by several concepts, activation of one property can feed back to other concepts, which in turn would activate the word nodes in the lexicon to which they are connected, thus facilitating processing for target words sharing semantic properties with the prime.

3.2.2. Summary

Reliable priming has been found for associatively related word pairs, and more controversially, for semantically related pairs in lexical decision tasks and naming tasks. The mechanisms underlying priming are not yet fully understood. The range of meaning relations between primes and targets is considerable, and has not been guided by a theory of word meaning. The creation of semantically related pairs typically involves operational definitions, relying on experimenters’ intuitions and off-line rating or judgement studies.

The question of how relevant these priming effects are for other aspects of linguistic
processing has been neglected. It is frequently assumed that in linguistic processing, semantic analysis is secondary to syntactic analysis (e.g. Pylyshyn, 1983, p. 150: "we do not make decisions about meaning until after the syntactic analysis has been performed", emphasis in the original). For example, if a choice has to be made in case of structural ambiguities (i.e. a choice between multiple syntactic representations for one input string), syntactic parses are assumed to occur automatically, not constrained by 'meaning' (or pragmatics) (see Chapter 7). However, if some aspects of word meaning are activated fast and automatically, as shown by lexical priming effects, then the general statement that 'meaning' is not immediately taken into account seems to be misleading.

Experiment 1 was designed to establish further evidence of rapid meaning activation in the lexical decision paradigm when semantically related pairs are defined in a theoretically-constrained fashion. It aims to investigate the operation of a putative semantic submodule which rapidly and automatically outputs a representation of word meaning in terms of the basic ontological concepts which make up the structure of a word.

Later experiments will provide further ways of showing activation/processing of meaning aspects of individual words, by examining whether there are syntactic processes that require some information regarding some meaning aspect of individual words in the sentence. These meaning aspects, however, might turn out to be different from the meaning aspects activated in free association experiments or lexical priming. Experiments 2, 3 and 4 attempt to show that lexical semantic information is processed rapidly in sentences with violations of selectional restrictions of verbs. Experiments 5, 6 and 7 investigate whether lexical semantic information can be used to influence the initial choices of the syntactic parser.
3.3. Lexical ambiguity research

3.3.1. Introduction

This chapter briefly reviews the main findings of lexical ambiguity research which sheds further light on the relationship of different processing systems involved in sentence processing. Overall, the findings indicate that all word senses of a homonym (such as 'bug', 'port') are activated fast and independently of contextual and syntactic factors, which seems to be relevant in the light of claims that semantic information is not processed fast enough to be available to the parser. Processing lexical ambiguities has often been compared to processing syntactic ambiguity, especially in order to undermine the claim that maintaining multiple outputs of a parser are too cost-intensive in terms of short-term memory load. The findings that multiple meanings of lexically ambiguous words are output supported the view that "multiple outputs of modules do not necessarily lead to significantly greater processing load" (Tanenhaus, Carlson & Seidenberg, 1985, p. 375). However, as Frazier & Rayner (1987) point out, the similarities in processing lexical and syntactic ambiguities might be limited, given that the former involves accessing prestored representations, whereas the latter requires the computation of multiple representations.

In lexical ambiguity research, the semantic priming effect is used as the main tool to detect which meaning or meanings of the ambiguous word are activated. Subjects have to perform a lexical decision or a naming task on a word related to one of the meanings of the ambiguous word (eg. after the ambiguous word 'bug', either 'ant' or 'spy' is presented). Performance on this word should be facilitated due to priming from the activated meaning/s of the ambiguous word. This is usually studied with a cross-modal paradigm, in which the context sentence and the ambiguous noun are presented auditorily and the word for the naming or lexical decision task is presented visually. This technique was pioneered by Conrad (1974), Swinney (1979) and Tanenhaus, Leiman & Seidenberg (1979). These studies presuppose that there is semantic priming, and focus on the investigation of whether or not sentential context can function as a sort of prime, in the sense that a sentence which is related to one of
the meanings of the ambiguous noun might facilitate access to this meaning of the ambiguous word. Since the focus of these studies is on context effects, not much attention is given to the semantic priming effect itself, which here merely serves as a welcome instrument to measure meaning activation of the ambiguous noun.

Diagram 3.1. shows in summary form the entities involved in ambiguity studies.

---

**Lexical decision**

or naming task

---

..... sentence .... ambiguous noun . TARGET ... sentence ..

---

**Diagram 3.1**: Schematic representation of entities involved in lexical ambiguity studies

The timing of presenting the target is crucial. There is no debate that after some time following the presentation of the ambiguous noun only the contextually-appropriate meaning is activated (and performance is facilitated only on the target related to that meaning). Controversy exists regarding the nature of meaning activation prior to this point in time: is there multiple or selective (autonomous or context-dependent) activation of the meaning of homonyms?

### 3.3.2. Multiple access hypothesis

A number of studies converge in their findings that all senses are activated rapidly for a short period. Apart from the speed of sense activation, the fact that context does not affect the initial access is regarded as evidence for a modular view of sense activation. Neither the syntactic structure of a sentence preceding the prime (Seidenberg,
Tanenhaus, Leiman & Bienkowski, 1982; Tanenhaus, Leiman & Seidenberg, 1979) nor the previous textual context (Swinney, 1979) led to the selection of only the contextually appropriate sense. Onifer and Swinney (1981) showed that even the less dominant meaning of an ambiguous noun is initially activated despite a context which was biased towards the more dominant meaning. There is some variation in the details about the time when context begins to have an influence. Swinney (1979) showed that context has an effect after 750 ms or less after the occurrence of the ambiguous noun: at that time, only the contextually appropriate meaning is activated. Till, Mross & Kintsch (1988) found that the inappropriate sense becomes deactivated after 400 msec. In the Tanenhaus et al. study (1979), a syntactic bias (such as the sentence 'John began to watch') led to the selective facilitation of the correct target ('look' as opposed to 'time') after an initial period of 200 ms, during which both targets were equally facilitated.

Gernsbacher and Faust (1991, G & F for short) replicated Swinney’s classic finding with a different technique, using a within-modality priming paradigm. The subjects’ primary task was to read short sentences, presented one word at a time, with a 150 ms gap between each word. The words were presented in the centre of a VDU screen. 150 ms after the prime word disappeared, the target word was presented in capital letters at the top of the screen. The subject had been instructed to make a rapid lexical decision for a word occurring in this position. After the lexical decision was made, the remaining words of the sentence were presented in the same way as the words before the prime word. Two testing points were introduced, one immediate and one delayed, which were created by varying the presentation rate of the sentences. At the fast rate, a 5 letter word appeared for 233 ms, and at the slow rate, for 700 ms. In the first experiment reported in G & F (1991), the prime words were ambiguous because they could be read as nouns or verbs (eg. punch). In the second experiment, the prime word was ambiguous since it had two possible noun readings (eg. quack). The results are the same in both experiments: they replicate the classic findings, i.e. at the immediate test point, both the contextually appropriate and the inappropriate meanings were activated, whereas at the delayed test point, only the contextually appropriate one was activated. G & F also attempted to evaluate different mechanisms put forward to explain how the
inappropriate meanings become less activated after the delay. One such mechanism is 'compensatory inhibition', i.e. the suggestion that multiple meanings share a fixed amount of activation, and that at the late point in processing, when activation for the appropriate meaning is increased, activation for the inappropriate one has to decrease because there is only this limited, fixed amount of activation available. Other mechanisms involve the decay due to lack of stimulation, and suppression from the activated meaning. They tried to decide between these proposals in their second experiment. This time, the prime was a noun which was ambiguous due to the two possible noun readings it has (eg. 'quack'). Primes were part of a sentence which was biased to one or the other of the meanings of quack (eg. 'Pam was diagnosed by a quack...' and 'Pam heard a sound like a quack...'). In the Neutral Condition, the preceding sentence did not bias either reading of the noun (eg. 'Pam was annoyed by the quack...'). The target words for this example were 'doctor' and 'duck'. The control condition against which facilitation was compared was the RT to 'duck' or 'doctor' following an unrelated prime (eg. 'pupil'). (the same Control procedure had been used in Experiment I). The crucial manipulation in experiment 2 is the Neutral Sentence Condition (sentences are not biased to either meaning of the ambiguous word). How activated will the meanings of the ambiguous word be at the late test? The two explanations of the decrease in activation of the inappropriate meaning at the late test point make different predictions about the neutral sentences. According to the decay proposal, inappropriate meanings decay because they do not receive any stimulation from the sentence context. In the neutral sentences, neither meaning of the prime word receives any contextual stimulation: appropriate and inappropriate meanings should both decay and be less activated at the late test point. This does not seem to be a very likely suggestion because it entails that context is necessary to maintain activation of meaning. It would seem to be very inefficient of a system to allow for such rapid decay of the meaning of a particular word due to lack of support from the previous context since it would give up information that might be very valuable in processing the very next word. On the other hand, according to the suppression hypothesis, inappropriate meanings are less activated because the context suppresses their activation. For the crucial test case of the Neutral Sentence Condition, this implies the prediction that appropriate and inappropriate meanings should remain as activated as
they were at the early test point since the neutral context does not suppress either meaning. G & F's data clearly show that in neutral sentences in the late condition both meanings of the ambiguous noun are activated. Thus the suppression hypothesis is supported: context suppresses the inappropriate meaning after a short delay.

As a further, stronger test of the suppression hypothesis one could introduce an Anomalous Condition, in which the sentential context rules out both meanings of the prime (eg. 'Pam was eaten by the quack'). The prediction would be that at the early test point, lexical decision on both 'doctor' and 'duck' would be facilitated relative to a control condition (since 'quack' produced facilitation in G & F's Neutral Sentences). At the late test point, no facilitation should be evidenced: context should have suppressed activation for either meaning of 'quack'.

3.3.3. Selective access hypothesis

In this section, I will review studies that proclaim to have found evidence for selective access to one meaning of a homonym. If these results are valid, they would indicate that the strong view of early meaning activation as an independent, context-insensitive process is too general.

As an alternative interpretation of Swinney's results, Glucksberg, Kreuz and Rho (1986) propose backward priming effects. A critical fact of the Swinney study is that the visually-presented target word (on which the subject has to make the lexical decision) follows the ambiguous auditory prime word (which is preceded by a sentence fragment) immediately at the offset of this word. Glucksberg et al. (1986) consider it likely that the visual target might be "available" at the same time as the auditory prime given that it takes some time to process the auditory stimulus. Thus what has been designed to be the target might well serve as a "biasing context for the ambiguous prime" (Glucksberg et al., 1986, p. 326); prime and target are psychologically simultaneous and it has been shown that it is under these circumstances that lexical decisions are facilitated (as in Meyer & Schvaneveldt, 1971). Glucksberg et al. (1986)
support their line of reasoning by an experiment of their own which is in fact a replication of the Onifer and Swinney (1981; O & S for short) study, using the same 48 sentences as were used in O & S’s experiment 2. Sense activation was measured as the slowing down of a correct no-decision to a nonword, derived from a real word which would have been related to one of the meanings of the ambiguous prime (eg. for the prime 'state', the two nonword targets 'condry' and 'conbishun' were derived from the words 'country' and 'condition', each assumed to be related to one of the meanings of 'state'). The logic of using this methodology is that no backward priming should occur from a nonword target to the prime. Using these targets, Glucksberg et al. (1986) found evidence for selective access to the contextually appropriate meaning of the ambiguous prime, at an interstimulus interval of zero. However, the result is difficult to interpret: as Glucksberg et al. point out themselves, making a no-decision in a lexical decision task takes longer than a yes-decision, and even more so if the nonwords are similar to real words (which was, of course, the point of choosing the nonword targets in the first place). It might be that in avoiding backward priming by choosing nonword targets a new problem was introduced, that of allowing extra time for processing the ambiguous prime word. The activation of all the senses of the ambiguous prime might exist for an extremely short period only, which might have been exceeded in Glucksberg et al.'s experiment. A study by Burgess, Tanenhaus & Seidenberg (1989) introduces further doubts about the Glucksberg et al. (1986) study. Burgess et al. (1989) show that Glucksberg et al.'s interference effects probably resulted from differences between nonword targets. They themselves found no selective interference for nonwords 'related' to one sense of an ambiguous prime when the properties of the target words were controlled for. Burgess et al. presented nonword targets following a related prime without a preceding sentence context (eg. responses to the nonword 'weign', which is similar to the real word 'weigh', were not any longer when it followed the related prime 'scale' than when it followed the unrelated prime 'shark'). Note that interference was measured by comparing response times to a nonword target following a 'related' prime with response times on the same target following an unrelated prime. Glucksberg et al. (1986), however, assessed interference by comparing response times to nonwords like 'weign' following the related prime to response times for a different target, eg. 'winge' following the same prime. Thus,
Glucksberg et al.'s interference effects might have resulted from differences between the nonword targets. They do not constitute conclusive evidence that multiple access to the different senses of a homonym found in other studies is an artefact of backward priming.

Burgess et al. (1989) suggest that the kind of sentence context used to constrain the meanings of the ambiguous word is a crucial variable in detecting exhaustive versus selective access. If the sentence contains a word "semantically related to or associated with one sense of an ambiguous word" (Burgess et al., 1989, p. 629), as was the case in Seidenberg et al.'s study (1982), selective access is likely to be found. Selective access has also been reported when the sentence context highlights salient feature of the dominant sense of an ambiguous word (Tabossi 1988; Tabossi, Colombo & Job, 1987). Tanenhaus, Dell & Carlson (1987) conclude that "whether or not certain linguistic processing decisions can be carried out independent of context depends on the nature of the context" (Tanenhaus et al., 1987, p. 106). However, there are contexts that seem to be highly specific, strongly constraining only one meaning of the ambiguous word (such as 'John began to ', followed by 'watch') which do not lead to selective access (Seidenberg et al., 1982; Tanenhaus et al., 1979).

Blutner & Sommer (1988) conducted a study in German, using a paradigm very similar to Swinney's. A crossmodal (visual-auditory) paradigm with lexical decision was used to assess whether context influenced the initial activation of the meanings of an ambiguous noun. Their ambiguous nouns were equibiasied (i.e. both senses were equally frequent). SOA's were 0 and 350 ms. A new factor was the focus in the test.
sentences, which was manipulated by an introductory question of the following kind, which was also presented auditorily:

a) Which scenery disappointed the visitors?
or b) Which visitors did the scenery disappoint?

The test sentence for this example was:

The scenery with the mast disappointed the visitors to the gallery.

Note that in German the word for 'mast' ('Mast') is ambiguous, meaning either 'stake' or 'animal fodder'.

The introductory question a) focused on the noun that was lexically ambiguous in the following sentence, whereas question b) did not focus on it.

The related, contextually appropriate target was 'stake', the related, contextually-inappropriate one, 'fodder'. Blutner & Sommer (1988) found, just as Swinney and others, that at the late SOA, only the contextually-appropriate sense of the ambiguous noun was activated. Focus did not affect this pattern at all. At the zero SOA, however, focus was crucial: when the ambiguous noun was focused, the 'classical' finding was replicated: both contextually appropriate and inappropriate senses of the prime were activated. When the ambiguous noun was not in focus, neither of the meanings was activated.

This study shows that processing word meaning is affected if the word is not in a sentence position which was focused through a preceding question. The authors interpret their findings as showing a "breakdown of the modular architecture in the case of reduced attention" (Blutner & Sommer, 1988, p. 365). However, the results surely also show a breakdown of an interactive architecture in the case of reduced attention. The point is that lexical processing was prevented temporarily while the listener's attention was drawn away from the material to be processed. It seems to me
that the modular view would only have been threatened if context selectively affected
the activation of one of the senses of an ambiguous noun. Since in the present study
neither of the senses was activated, all one can conclude is that if the listener’s
attention has been distracted away from the ambiguous noun, access to the meanings
of that noun is delayed. This merely underlines the fact that input processes only
operate if attention is focused on the input.

Tabossi’s work is widely quoted as showing that context can in fact lead to fast
selective activation of word meaning if the context is highly biased to the dominant
meaning of the ambiguous noun. How justified is this interpretation?

Tabossi, Colombo & Job (1987) report two experiments using the Swinney type cross-
modal lexical priming paradigm. Ambiguous nouns which had one dominant and one
secondary sense (established through judgements of 12 subjects) were embedded in
sentences which biased either the dominant sense (experiment 1) or the secondary
sense (experiment 2) of the ambiguous word. The target word was chosen to refer to a
property specific to the referent of one or the other of the readings of the ambiguous
noun (eg. for the prime 'port', the targets were 'safe' and 'red'). For sake of brevity,
targets such as 'safe' will be called dominant targets, and targets such as 'red'
secondary targets. Only ISI’s of zero were studied. It was found that if the sentence
biased the dominant sense of the prime (the 'harbour' sense in the above example), the
contextually related, dominant target was faster than both the inappropriate, secondary
target and unrelated control targets. The latter two kinds of target did not differ. If
context was biased towards the secondary meaning of the ambiguous word (the 'drink'
sense of the above example), both contextually appropriate, secondary targets and
contextually inappropriate dominant targets were faster than unrelated targets.

These results show that the secondary meanings of ambiguous nouns become activated
only if they are contextually biased, whereas the dominant meanings become activated
regardless of context. This conflicts with Onifer & Swinney’s (1981) finding that
inappropriate senses were activated regardless of context, i.e. even when the sentence
bias was towards the dominant sense of the prime. In both studies, targets were placed
immediately at the offset of the ambiguous prime. Tabossi et al. (1987) suggest that this discrepancy might be due to differences in the targets used in the two studies. Their own targets were related to a specific aspect of one of the senses of the ambiguous word, and unrelated to any aspects of the other sense (the selection of targets was based on 12 judges who were "requested to produce relevant semantic aspects of both meanings for each ambiguity" (Tabossi et al., 1987, p. 163). The words that were finally chosen as targets had been mentioned by at least nine of the judges). The aim was "to make sure that the most characteristic aspects of the word's meanings were tested" (Tabossi et al., 1987, p. 163). Onifer & Swinney's targets, so Tabossi et al. (1987), were less specific and thus might have provided a less sensitive test of activation of one particular sense of the ambiguous word. Tabossi et al. (1987) claim that most of Onifer & Swinney's targets (disregarding the control targets) "were associated to either meaning of the ambiguities" (Tabossi et al., 1987, p. 166).

This evaluation, however, can be ruled out by a detailed analysis of Onifer & Swinney's (1981; O & S, for short) materials. A close check of their primes and targets clearly indicates that there are only a few targets which might be seen to be related to both senses of the ambiguity. I will list the cases where the targets are indeed quite close to both senses of the prime, giving the prime first, then the target related to the dominant sense of the prime, and, lastly, the target related to the secondary sense of the prime. In O & S's first experiment, 'scale' - 'weight', 'fish'; 'cold' - 'hot', 'flu'; 'plane' - 'propeller', 'wood' might have been doubtful (since fish is often weighed, 'fish' could be seen as related to the measurement-sense of 'scale'; a flu you often get when you are cold, therefore, 'flu' might be related to the temperature-sense of 'cold', and finally, some aeroplanes are made out of wood so that 'wood' could be seen as related to the aircraft-sense of 'plane'). These three examples are the only ones for which one could possibly argue that the target is not specific enough to one particular sense of the prime. They form only 19% of all the prime-target sets. In O & S's second experiment, only 12.5% of the prime - target sets were questionable since the targets designed to measure activation of one sense of the prime could also be seen to be, albeit remotely, related to the other sense (for example: 'tank': 'war', 'container' (the tank as weapon could be seen as a kind of container,
too); 'cold': 'chill', 'sick' ('sick' is not only related to 'cold' as disease but also to 'cold' in the temperature sense since you can become sick in the cold); 'scale': 'weigh', 'fish' (as for experiment 1); 'toast': 'bread', 'glasses' ('glasses', even though designed to measure activation of the secondary meaning of 'toast', can be seen as related to the dominant sense as well since glasses are thematically related to toast in the bread sense); 'beach'/ 'beech': 'sand', 'tree' (trees can be found on beaches); 'hair'/'hare': 'head', 'rabbit' (there are long-haired rabbits). Thus, analyzing O & S's stimuli in detail, one has to conclude that the majority of their targets specifically related to one of the senses of the ambiguous prime, contrary to Tabossi et al. (1987)'s claims.

Another difference in the stimulus materials which has not been mentioned in the literature is that O & S's dominant and secondary targets seem to be more closely related to the relevant sense of the prime than is the case for Tabossi et al. (1987)'s stimuli. They typically chose adjectives denoting a property of the Thing that one of the senses of the prime mapped into, whereas O & S used close associates, or synonyms. It might be that associations are so strongly related to the primes that they pick up any activation of the senses of the prime, regardless of context. A later study by Tabossi (1988), however, rules out this possibility. She found selective access with highly associated targets if (and only if) context referred to a central aspect of the dominant meaning of the prime.

Another noteworthy difference between the two studies which has not been discussed in the literature is the number of control targets used. O&S have two different control targets: what I will refer to as Control 1, i.e. control targets matched in word frequency and number of letters to the dominant targets, and Control 2, matched in these properties to the secondary targets. The mean word frequency was 83 and 84 for dominant targets and Control 1, respectively, but a much lower 25 and 26 for the secondary targets and Control 2. Tabossi et al. (1987) only use one control target against which both the dominant and the secondary targets are compared. In O & S the Control 2 targets yield much longer RTs than Control 1 targets (in both experiments). If one compares RTs for the secondary and dominant targets (when the sentential bias
is towards the dominant prime meaning) it turns out that they are not equal (706 ms vs. 734 ms in Experiment 1 and 693.5 and 726.1 in Experiment 2). O & S’s finding of facilitation for the inappropriate secondary target in a sentence biased towards the dominant meaning only obtains if one compares the dominant and secondary targets with their respective matched controls (i.e. Control 1 or Control 2). It is striking to note the RT differences for Control 1 versus Control 2: 727.4 ms for Control 1 targets, and 760.4 ms for Control 2 targets (this is for sentences with a sentence bias to the dominant meaning, at a zero ISI). If the secondary targets had not been matched in frequency and word length to their particular controls, no effect of activation of the secondary prime meaning would have been found. It is not clear whether Tabossi et al. (1987) would have found the same effect if they had matched their targets to particular controls.

Tabossi (1988) systematically varied both the sentence context and the kind of target, using the same materials as Tabossi et al. (1987). Context was always biased towards the dominant meaning, but it either referred to a specific feature of the dominant sense of the prime (eg. "The violent hurricane did not damage the ships which were in the port, one of the best equipped along the coast.") or it did not ("The man had to be at five o’clock at the port for a very important meeting."). The latter type of context, so Tabossi (1988), is the same kind as used by O & S. Targets were either associates of the primes (as established through free association data from ten Italian subjects), which did not refer to any features of either meaning of the primes; or targets named a central feature of either the dominant or the secondary meaning of the prime (this is the type of target used by Tabossi et al., 1987). Selective access was found with either kind of target, but only if the context referred to a central feature or aspect of the dominant meaning. In the less specifically biasing context (of the O & S type), both appropriate and inappropriate targets were facilitated relative to the controls, thus demonstrating multiple access to the meanings of the ambiguous prime.

As other researchers have also done, Tabossi defined semantic features of the prime operationally as the outcome of a production experiment in which twelve native Italian speakers had been asked to "produce relevant semantic aspects of both meanings for
each ambiguity" (Tabossi et al., 1987, p. 163). In addition, another six judges had to indicate what aspect of the meaning of the ambiguous word each sentence made them think of. While this is a thorough empirical way of creating stimuli, the notion of semantic features remains in a theoretical void. Tabossi (1988) is aware of this. She admits that "semantic features are rather elusive notions", "not easy to define", but "extensively used in lexical semantics research" (Tabossi, 1988, p. 335). She claims that the information needed to characterize word meaning can be established empirically "without committing oneself to any specific theory of how the lexicon is mentally represented" (p. 336).

Generally, her results are accepted, even by people who support a multiple access view. They acknowledge that "context" needs to be studied more carefully. One should add that semantic features need to be studied more carefully, too.

3.3.4. Summary

In sum, lexical ambiguity research has focused mainly on whether sentence context can influence the activation of word meaning. The target words of the secondary task were chosen to reflect one or the other of the senses of the ambiguous prime word. For most experimenters it seemed to have been intuitively clear which words would be close to the sense of another word. Usually some near synonyms are chosen (eg. state - country, state - condition). However, semantic similarity is a complex phenomenon, as was discussed in Chapter 3.1.

The findings show that word processing involves activation of all the senses attached to a word, irrespective of syntactic constraints or constraints imposed through the meaning of the overall sentence, unless the sentence not only biases one reading of an ambiguous word, but specifically refers to a characteristic feature of that meaning. Lexical ambiguity research demonstrates that access to the prime, and activation of some meaning aspects, is very fast, if not instantaneous, since priming effects were found at 0 ISI.
CHAPTER 4 : SEMANTIC VERSUS ASSOCIATIVE PRIMING OF NOUN PAIRS: EXPERIMENT 1

4.1. Introduction

Experiment 1 was designed as a first test of a model of word meaning derived from Jackendoff (1983). A hierarchy of basic semantic features is proposed, and the aim in Experiment 1 is to test whether these features are activated fast enough during word processing to lead to priming effects. In this section, I will briefly describe his theory as far as it is relevant to discussing the meaning of nouns.

Jackendoff (1983) offers a competence theory of word meaning, that is he is concerned with describing the structure of mentally represented semantic information, rather than with how it is stored or how it is processed. He describes word meanings as the mapping of a word into a basic ontological category: "the meaning of a lexical item of any major syntactic category (noun, verb, adjective, adverb, preposition) is a function of zero or more arguments that maps into a conceptual constituent of one of the major ontological categories" (Jackendoff, 1983, p. 11). When a word does not subcategorize any other phrases (eg. 'dog', or the verb 'rain') its meaning is "a constant, that is a complete conceptual constituent" (Jackendoff, 1983, p. 110). Jackendoff assumes a set of such basic ontological categories, which are mutually exclusive primitives of meaning: THING, PLACE, DIRECTION, ACTION, EVENT, MANNER, AMOUNT. They constitute the major most abstract categories with which the world is represented by the human mind. Jackendoff's (1983) starting point for identifying these basic concepts is via pragmatic anaphora and wh-questions: 'I bought that' : 'that' refers to an object in the world, which is mentally represented by the basic concept THING; 'Your coat is here': 'here' refers to a PLACE; 'What happened': 'what' refers to an EVENT; 'What did she do?': 'what' refers to an ACTION. The basic concepts correlate with major syntactic categories: noun phrases typically map into the basic category THING (eg. 'table', 'house'), but they can also map into EVENT (eg. 'earthquake', 'lecture'), AMOUNT (eg. 'mile') or PROPERTY (eg. 'brilliance' or 'beauty'). Note that even though the sense of a word such as 'beauty' is characterized
as a mapping function into the basic concept of PROPERTY, the word may be used to refer to an object rather than a property (e.g. somebody might want to refer to a particularly nice apple by saying 'what a beauty!'). Clearly, sense and reference are not the same. Prepositional phrases map into PLACE ('on', 'in', 'under' something), or PATH ('toward', 'from' something). Verb phrases map into ACTION, EVENTS, or STATES. Adverbial phrases map onto MANNER and AMOUNT, and adjectival phrases map onto PROPERTY. Describing word meaning as a function mapping into basic ontological concepts obviously only captures highly abstract aspects of word meaning. Jackendoff (1983) suggests that the basic ontological categories form the most abstract level of a hierarchy. He suggests that within the major conceptual categories there are further "distinctions such as solid versus liquid, human versus animal, and so on" (Jackendoff, 1987a, p. 385). These finer distinctions are mutually exclusive, as are the major categories. I will focus on those categories which represent the kind of semantic information that some verbs require as part of their semantic selection restrictions (see Chapter 5.3.), e.g. animate versus inanimate, liquid versus solid, concrete versus abstract. They are likely candidates for the kind of semantic information that is represented in the mental lexicon, and that is activated fast during the processing of words.

In Jackendoff’s competence theory, conceptual structure is isomorphic with semantic structure, consisting of the same set of primitives and principles of combination: "word meanings are expressions of conceptual structure" (Jackendoff, 1983, p. 110). Semantic structure is the part of conceptual structure expressed by language. The mapping of lexical items into conceptual structure is based on preference rules, which specify necessary conditions (e.g. 'tiger' must at least map into THING), plus centrality conditions specifying a focal value for variable attributes (e.g. specifying focal values for the height-width ratio for cups versus bowls), and typicality conditions (which are typical but subject to exception, e.g. 'stripedness' in tigers).

In Experiment 1 this characterisation of the structure of word meaning is tested by investigating whether there is priming for word pairs which map into the same basic categories. I will focus here on basic concepts that form part of Jackendoff’s
"necessary" conditions, eg. the concept THING rather than 'stripedness' as part of the meaning of 'tiger'. These are the concepts that represent the type of semantic information that might be syntactically relevant, and that is required by semantic selection restrictions (Jackendoff, 1987a).

In Experiment 1 a semantic-priming-with-lexical-decision paradigm was used as a test of rapid access to semantic knowledge. The choice of words which make up the semantically-related stimuli and control stimuli was based on the proposed hierarchy of semantic features. Thus, stimulus selection was stringent and theoretically motivated. In this respect, Experiment 1 fills a gap that the review of previous semantic priming studies in Chapter 3 has pinpointed. The majority of the existing studies suffered from the lack of a clear definition of what constitutes semantic relations between words. Consequently, the choices of stimuli in those studies were mainly guided by impressionistic and incoherent criteria. Often, words were called semantically-related when in fact they were associatively related.

In Experiment 1, a target word, mapping into THING, is preceded by three different primes: associative primes, i.e. words for which the target is a high associate; semantic primes, i.e. words that map into the same basic concept as the target word, and unrelated control primes, which are words that map into another basic concept different from THING, and which have no associative relationship with the target.

Thus, the word pairs in Experiment 1 followed the pattern illustrated here:

bug - insect : associative relation
fish - insect : semantic relation
honesty - insect : control relation

Experiment 1 is only a first test of the model of semantic decomposition in two respects: first, only words mapping into one of the basic concept, namely THING, were selected as the semantically-related stimuli; second, in Experiment 1 the semantic feature into which both words map is only one level up in the semantic feature.
hierarchy. If this initial test fails, then the even stronger claim that words sharing semantic features only at the higher levels of the hierarchy has to be ruled out. However, if Experiment 1 succeeds, one can proceed to explore at which levels of the ontological hierarchy priming occurs. There might be facilitation between words that do not share any of the lower level, more specific basic concepts of a hierarchy but only the most abstract one, eg. there might be facilitation from a word that maps into ANIMATE THING to a word that maps into INANIMATE THING, since both map into THING.

Several assumptions underlie the use of a semantic priming study as a test of how semantic knowledge is processed, as was discussed in Chapter 3. Specific to the proposed test of the model of semantic decomposition in Experiment 1 is the assumption that the mapping from words into ontological concepts is activated rapidly during the first-pass semantic analysis (see Pylyshyn, 1980). If making lexical decisions is assumed to reflect lexical access processes, priming effects could be interpreted as indicating that words in the lexicon are connected along links representing shared basic semantic features, with activation spreading from the prime word to other words sharing the same basic semantic features. Alternatively, if priming effects are taken to reflect postaccess checking processes, one would assume that word meaning is passed on to the semantic processing system which registers whether a target word shares basic semantic features with a prime word, and, if this is the case, speeds up the lexical decision process.

As a method to test whether there is priming for associatively related words and for semantically related words, the following kind of semantic priming paradigm was chosen. Pairs of words, without any sentence context, were employed. Using word pairs allows one to use a within-modality paradigm rather than a cross-modality one, which has the advantage of being an easier task for subjects. A visual rather than auditory presentation mode was chosen. Presentation of prime and target was always sequential (for justification see section on Pilot Study) rather than simultaneous. Two time intervals were chosen: the prime was presented for either 75 or 250 ms. 75 ms was the display time chosen by Chiarello et al. (1990), and is among the shortest
display times reported in the literature (not taking into account the subliminal priming studies). The aim in the present study was to test whether there is automatic access to semantic features, and thus a very short display of the prime seemed to be in order. A slightly longer display time of 250 ms was also introduced, which is similar to Gernsbacher & Faust's (1991) fast condition (for a 5-letter prime word, they allowed for a display of 233 ms). In either case there followed a blank screen for 150 ms, before the target was displayed which remained on the screen till the subject made a response. The sequence of events per trial is schematised in diagram 4.1. below:

**Diagram 4.1.: Sequence of events per trials in Experiment 1, Presentation Condition 1**

<table>
<thead>
<tr>
<th>Pres. Cond. 1:</th>
<th>+</th>
<th>500 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>prime</td>
<td></td>
<td>75 ms</td>
</tr>
<tr>
<td>blank screen</td>
<td></td>
<td>150 ms</td>
</tr>
<tr>
<td>target</td>
<td></td>
<td>till response button is pressed</td>
</tr>
<tr>
<td>blank screen</td>
<td></td>
<td>2000 ms</td>
</tr>
</tbody>
</table>

In Presentation Condition 2, the sequence of events was the same, except that the prime was presented for 250 ms. Subjects had to make a lexical decision on the target. In 50% of the trials, the target was a nonword. Half of the nonwords were pseudohomophones (eg. laryer), and half were nonpseudohomophones (eg. stex). Thus, the factor Stimulus Relation had five levels (word-pseudohomophone, word-nonword, associatively related word pairs, semantically-related word pairs and unrelated word pairs). These five levels will be addressed in short as PS-pairs, NW-pairs, A-pairs, S-pairs and U-Pairs). Stimulus Relation was a within factor in both the by-subjects and the by-items design. Stimulus pairs were created such that each word target occurred once each in an A-pair, an S-pair and a U-pair to allow the comparison of the priming effects within items. To avoid any unnecessary risk of noncomparable subject groups (and to keep the number of subjects within reasonable bounds), it was decided to design Stimulus Relation as a within-factor, with each subject being presented stimuli
pairs at each of the 5 levels of Stimulus Type. In order to avoid a subject seeing the same target more than once, which might make the subject aware of our aim of comparing their reaction to targets as a function of the preceding word, three parallel lists of stimulus pairs were created, making up the factor List. Each list contained the same PS- and NW-pairs, but the word pairs were distributed across the three lists such that the pairs that had the same target word were in different lists. This implied that List was a between-subjects factor. Presentation Condition had to be a between factor, too, for the same reason: subjects should not see the same target more than once.

It was predicted that making a lexical decision about a word is faster for the A- and S-pairs as compared to the U-pairs (this will be called the Semantic and the Associative Hypothesis in short). All word pairs (A-, S, and U-pairs) should lead to faster response times than the PS- and NW- pairs. This expected pattern of results should also be reflected in the number of false decisions subjects make, i.e. PS- and NW-pairs are expected to attract more false responses than the word pairs. List should not make a difference.

Let me illustrate the predictions regarding the word relations tested in Experiment 1 with some word pairs taken from the actual set of stimuli. Take the pairs 'doctor - nurse', 'son - nurse' and 'baptism - nurse'. The words in the last pair are unrelated, with the prime mapping into a different ontological concept than the target (EVENT versus THING). The time to make a lexical decision on the second member of this pair provides a baseline against which to measure any priming from other kinds of primes. 'Doctor - nurse' are associatively related to each other. For this pair, priming of a lexical decision on the second word was predicted on the basis of the great number of previous studies reviewed in Chapter 3 which found priming for associatively related words. Crucial in Experiment 1 is the 'son - nurse'- pair: the prime 'son' is not associatively related to its target and yet it is predicted to lead to priming since it maps into the same feature (i.e. HUMAN) as the target. As can be seen from diagram 4.2. below, 'son' and 'nurse' do not share the same feature at the lowest level of abstraction (eg. they do not both map into ADULT). The similarity in meaning is due to a shared feature which is one level up the hierarchy.
4.2. Pilot Study

Several aspects of the presentation conditions were tested out in the pilot study.

A simultaneous Presentation Condition was compared to a sequential condition. In the simultaneous condition, prime and target occurred at the same time, one above the other centralised on the screen. Subjects were instructed to first read the prime, and then make a lexical decision on the target. After questioning the pilot subjects in detail after the experiment, it was decided not to include this condition in the actual experiment since it allowed for too many strategic effects. One pilot subject reported not to have looked at the upper string at all. Furthermore, it is impossible to know how long a subject looked at the upper string, if at all. Since the question is whether semantic features are rapidly activated, it was important to control the maximum length of time subjects could look at the prime. Thus, the sequential presentation condition was chosen, in which the prime appeared, for either 75 or 250 ms, then disappeared, and was followed by the target.

The target was presented one line below the prime. It was decided not to present prime and target on the same line to avoid flicker effects.

The location of a fixation point was also tested in the pilot study. It was finally decided to display a small cross (+) for 500 ms centralised on the same line on which

---

Diagram 4.2.: Semantic hierarchy to represent the words 'son' and 'nurse'

```
THING
  /\  
HUMAN
 /   
| CHILD| ADULT
|     |
```
the prime would appear. The pilot study had indicated that the + should be on the same line as where the prime was going to appear, to make sure subjects were focusing on the location where the prime would appear, rather than placing the cross one line above and one character to the left to where the prime would appear, as in Fischler (1977a). Various other possibilities had also been tested: + centralised, but 1 line above prime location; 1 line up, 1 char. to left to prime location; + centralised on same line as prime location with offset of + coinciding with onset of prime. The latter is what was chosen for the actual experiment). When the fixation point was presented one line above where the prime was going to appear, pilot subjects reported they were not aware of having looked at the prime. Presenting the fixation cross on the same line as where the prime would be displayed makes this more likely. In Fischler (1977a) this problem did not arise since her subjects had to look at both words of a stimulus pair presented simultaneously to perform the task of deciding whether both words were real words. In this study, pairs were displayed simultaneously one above the other, and either the word or nonword could appear at the top position.

4.3. Method

Design

The experiment had a mixed design. The factors were Stimulus Relation, Presentation condition, and List and Itemgroup. Stimulus Relation had 3 levels: the relations between prime and target were either associative, semantic or unrelated. Presentation condition had two levels: the prime was displayed for either 75 ms or 250 ms. Presentation condition was a between-groups factor in the by-subjects analysis, and a within-groups factor in the by-items analysis. Stimulus Relation was a within-groups factor in both designs. To prevent subjects from seeing the same target more than once, each member of a triplet was assigned to a different list, using a Latin Square. As a result, there was a between-subjects factor List with three levels, and a between-items factor Itemgroup with 3 levels (representing 3 subsets of triplets which differed in whether the A-, or S- or U-pairs were assigned to list 1, 2 or 3). Subjects were
assigned to the different stimulus lists at random. The two between-groups factors in the by-subjects design, Presentation Condition and List, were factorially combined. Thus, the by-subjects design was a mixed 2 (Presentation condition) x 3 (List) x 3 (Stimulus Relation) design. The by-items design was a 3 (Itemgroups) x 2 (Presentation Condition) x 3 (Stimulus Relation) mixed design.

The dependent variables were whether subjects made a correct or a false response in the lexical decision task, and the time in milliseconds to make the decision. This time was measured by a millisecond clock in the computer which was triggered by the display of the target, and stopped when the subject pressed a response key. The target remained on the screen till the subject made a response.

Predictions

For the critical word-word pairs it was predicted that correct lexical decisions would be faster for targets in the semantically- and associatively related pairs as compared to the unrelated pairs. This effect was predicted to exist for both presentation conditions. Furthermore, it was predicted that correct lexical decision on word targets would be faster than on nonword targets. As regards the dependent variable 'correctness of lexical decision', it was predicted that there would be more incorrect decisions on the nonword targets, and more on the pseudohomophonic nonwords as compared to the nonpseudohomophonic ones. No significant differences in the number of incorrect responses was predicted for the three levels of the factor Stimulus Relation, which only existed for the word-word pairs.

Apparatus

The experiment was completely controlled by a BBC microcomputer. Two response boxes were also used which were linked to the computer through the userport in the keyboard. The response keys were fixed on top of two separate plastic boxes, 7 square
centimetres each. The boxes could be moved around. The keys themselves were round, with a diameter of 2.5 cm each. Next to each key on top of the boxes was a white sticker with the letter Y or N (for yes or no) written in black. The no-key was red, and the yes-key black.

**Procedure**

Subjects were tested individually in experimental cubicles in the Department of Psychology. After they arrived they were greeted by the Experimenter and were paid their reward. Subjects were assigned randomly to the experimental conditions. Each subject was led individually by the Experimenter to one of the cubicles, where the subject was asked to sit down in front of the computer. The Experimenter then explained that all the instructions would be presented by the computer, which would prompt the subject to press certain keys to proceed. The subject was told that at first there would be questions about his or her age and sex, and whether English was the subject’s mother tongue. The subject was told to use the keys of the ordinary computer keyboard to answer these questions. The subject was also told that one of the questions on the screen was to indicate which list they belonged to. The Experimenter told the subject which list s/he belonged to and left a yellow sticker with a number (either 1, 2 or 3) next to the keyboard to remind the subject. It was then pointed out to the subject that after having answered the questions, there would be some instructions on the screen telling the subject exactly what his or her task was. The Experimenter stressed that in the actual experiment the subject would have to press one of the two response buttons, and not any keys of the computer keyboard. The subject was encouraged to handle the response boxes and to try out pressing the key. Subjects were instructed to use either index and middle finger of their preferred hand or the index fingers of both hands for making the responses. They were told that they could move the boxes around if they wished, and that they would have time to try out pressing the response buttons during some practice trials. Subjects were encouraged to ask any questions they might have and, once they were dealt with, the subjects were left alone in the cubicle for the experiment.
All subjects received the same instructions. These were presented as paragraphs on the screen, each ending with the instruction to 'press any key to continue'. The full instructions can be found in Appendix A. Subjects were told that if they noticed they had made a mistake and pressed the wrong response button, they could press the correct button immediately afterwards. This second response was not recorded, however. The instruction was worded like this to avoid interference from subjects’ thinking about having made an incorrect response.

Instructions and stimuli were presented as white letters on a black background on the computer screen which was placed in front of the subject at a comfortable viewing distance. Responses (which key was pressed, time to press the key) were recorded automatically. Each stimulus pair subtended a visual angle of approximately 1 degree horizontally and .39 degree vertically.

All words were presented on the screen written in lower case, since there is evidence that words written in lower case are easier to analyze visually than words written in upper case. Lower case provides useful familiar shape information about a word as a whole (see Parkin & Underwood, 1983; Underwood & Bargh, 1982). There is no evidence that upper case has any beneficial effect.

First, two familiarisation pairs were presented before the actual list (mirror - door; bucket - doam), to make the subjects familiar with pressing the response buttons. After a break, and when the subjects had pressed 'any key to continue' the series of trials began. The first 10 trials were start-up trials and were not included in the analyses.

Subjects were given no feedback during the experiment.
Stimuli

Each list consisted of 138 trials: 10 start up trials which were not analyzed; of the remaining 128 trials, half were nonwords (of which 50% were pseudohomophones, and 50% nonpseudohomophones). Filler and nonword trials were identical in the three stimulus lists. The positions of the word-word pairs in the three lists were identical to allow later comparison within triplets. The order had been determined at random, but it was constrained by two criteria: the first ten pairs in the lists were all word pairs; there were not more than three adjacent pairs that were either all word pairs or all nonword pairs. Once the three lists had been compiled in this fashion, the order remained the same for each subject.

The five types of stimuli used in this experiment will be described in some detail here.

1. Word-word pairs

These pairs were designed as triplets, and then distributed across the three stimuli lists which were given to different subjects. Each triplet consisted of a target preceded by an associatively related prime, a semantically related prime and an unrelated prime. 64 triplets were used, of which 36 were analyzed, and the rest were filler items.

1.1. Semantically related word-word pairs

Primes and targets were related by sharing semantic features within the THING-hierarchy, i.e. ANIMAL, INDOOR THING, FOOD, PERCEPTION, HUMAN, NATURAL KIND, PLANT, ABSTRACT THING. Note that prime and target did not map into the same 'narrow' concept. Primes and targets were not highly associated according to free association norms, and were not linked by strong inferences. Neither were there any functional or script-based links.

1.2. Associatively related word-word pairs

83
Primes and targets had a strong **associative** link. These primes should strongly facilitate the corresponding targets. They were included as a test condition to check whether the experimental set-up was sensitive to priming effects. The words in the associatively related pairs shared a variety of semantic relationships, such as coordination (eg. 'silver - gold'), part-whole relationships ('collar - shirt'), functional relationships ('priest - church'), antonyms ('hatred - love') and near synonyms ('wrath - anger').

### 1.3. Unrelated word-word pairs

The unrelated pairs served to provide a control condition. Prime and target mapped into different ontological concepts, they were not associated and there was no inferential link between them. The target always mapped into THING, and the prime mapped into one of the other basic ontological concepts, i.e. ACTION, EVENT, AMOUNT, and PROPERTY.

### 2. Word-Nonword pairs

This type of stimulus pair made up 50% of the trials, i.e. there were 64 word-nonword pairs, 36 of which were included in the analyses, and the rest were filler trials. Nonword targets were included so that subjects could not build up an expectation that all trials would have real-word targets. Half of the prime words mapped into THING, the other half into various other basic categories. Pseudohomophone targets were derived from words which did not share any semantic relationship with the prime. It is known that lexical decision on non-words takes longer than on real words. The purpose here was not to demonstrate this effect yet again, but merely to make sure subjects expected to have to make both positive and negative lexical decisions.

There were also 10 practice pairs. Half of these were word-word pairs, and half word-nonword pairs. The word-word pairs were unrelated, i.e. there were no associative or semantic relations between them. The targets mapped into various different basic
concepts, i.e. THING, ACTION and PROPERTY.

Appendix B gives a full list of the word-word pairs included in the analyses, and Appendix C lists the nonword trials that were included in the analyses. Note that the word-word pairs are arranged here as triplets, but that in the actual experiments, a subject only saw one pair of the triplet.

Creation of stimuli

All words were at least of medium frequency of occurrence, according to the Francis & Kucera (1982) norms.

Creation of the word-word pairs

To select stimuli for the associatively-related pairs, the association norms by Miller (1970) and Shapiro & Palermo (1968) were scanned. Stimulus - response pairs were selected for inclusion in Experiment 1 if they fitted the following criteria:

1. the response was the primary associate of the stimulus. While the Miller norms give the five most frequent responses to a stimulus, the Shapiro & Palermo - list only gives the primary associate for each stimulus, which in some cases has a very low probability of occurrence. Responses that had a probability of less than .25 in the Shapiro & Palermo norms led to the exclusion of the stimulus - response pair for this study. Furthermore, if the studies listed by Shapiro & Palermo did not agree in the primary associate to a stimulus word, that pair was not included in the present study.

2. both words were unambiguous nouns. Nouns with a great degree of polysemy were excluded, as were nouns that could also be read as verbs.

3. both nouns mapped into THING rather than any other ontological concept.
4. words were not more than three syllables long.

5. stimulus and response word did not form a phrase (such as mutton - chop; husband - wife; army - navy; hand - foot; man - woman; soldier - sailor; bread - butter; needle - thread). One can assume that such pairs form a special class of associates. The decision about whether two words formed a phrase or not was made in a pilot study by three native speakers. One simple test that these subjects were asked to do in assessing whether a given word pair formed a phrase or not was to check whether the words appear to be equally familiar in either direction. If this is not the case, the two words are likely to form a phrase (eg. compare soldier - sailor to sailor - soldier).

To select words for the semantically-related pairs, the prime word of each associatively-related pair was replaced with another noun. It was determined into which semantic features each target word could be decomposed, and another word was chosen that was related to the target by virtue of sharing the same semantic feature in the semantic hierarchy. This word served as the prime in the S-pairs unless the resulting pair was a combination of superordinate and subordinate words (or vice versa). In this case, the pair was dropped since it involves a proposition (eg. bird - robin) and the subject's task is likely to become a verification task rather than a lexical decision task. The pair was also dropped if a check of the norms published by Miller (1970) and Shapiro & Palermo (1968) revealed the target to be among the associates of the prime. (Some primes that were included in Experiment 1 were not mentioned as stimulus words in these norms.)

Finally, to design the primes for the unrelated pairs, the associatively related prime in each pair was replaced by a noun that mapped into a different ontological concept than the second word of the pair, i.e. into ACTION, EVENT, AMOUNT, PROPERTY, and STATE. It was ensured that the words of these unrelated pairs were not associatively-related to each other (by checking the Miller, 1970, and Shapiro & Palermo, 1968, norms).
Most primes in the unrelated pairs were abstract words, whereas most primes in the associatively-related and the semantically-related pairs were concrete words. One might argue that since abstract words have been shown to take longer processing time in various cognitive tasks (see eg. Schwanenflugel & Shoben, 1983), recognition of abstract words might also be slower. Therefore, the predicted lack of priming in the unrelated pairs might have nothing to do with the prime mapping onto a different ontological concept than the target, but might be entirely due to the slower or incomplete recognition of the abstract prime. This possibility is not very likely to be relevant in Experiment 1, however, since the presentation rates of 75 ms or 250 ms (plus 150 ms gap before the display of the target word) should undermine such effects. Concreteness effects occur at shorter time intervals than provided in my presentation rates for primes. Furthermore, in Experiment 1 words are presented clearly, with no masking or degrading. It seems quite likely that under these conditions recognition, even of abstract prime words, will be complete. Furthermore, all words were of comparable word frequency, and they had all been judged as equally familiar by three pilot subjects who did not take part in the main experiment.

Selection of non-words

Half of the targets were non-words. Only 'legal' non-words were used, i.e. those that were possible words according to the graphemic, phonemic and orthographic rules of English. They were generated by replacing or adding one letter of a noun. None of the nouns that the non-words were derived from were used as real words in the experiment.

Half of the non-words were pseudohomophones (eg. 'trupe'). It is known that non-words that sound the same as real words take longer to reject in lexical decisions than non-words that do not sound like a real word (Rubenstein, Lewis and Rubenstein, 1971). Coltheart, Besner, Jonasson & Davelaar (1979) have shown that visually-presented orthographically regular non-words are recoded into a phonological format using grapheme-phoneme correspondences. If, as is the case for pseudohomophones,
the phonological representation of the non-word is identical to the phonological representation of a real word (eg. as in 'brane'), correctly rejecting the non-word in the lexical decision task is slowed down, relative to the time needed to make a No-response on the lexical decision task to non-words that do not sound like real words (eg. 'rolt'). Pseudohomophonic nonwords were introduced into the present study to prevent subjects from making a lexical decision purely on the basis of phonological decoding.

Subjects

66 subjects took part in the experiment. They were either undergraduate or postgraduate students of the department, and were recruited through notes placed in the student common rooms or by personal appeal made at the end of some lectures by the Experimenter. They were paid £1 for their participation in this experiment which took about 15 minutes per subject.

4.4. Results

Data from 36 word-word pairs, 36 word-nonword pairs, and 66 subjects entered the analyses. Each subject saw 12 word-word pairs in the associated, the semantically-related and the unrelated condition. For each word-word pair there were responses from 11 subjects in each of the 6 conditions resulting from the factorial combination of the factors Presentation Condition and Stimulus Relation.

Out of the total of all responses to word-word pairs (2376 responses), in 3.4% cases the decision to the target word was an incorrect No-response. Collapsed over all conditions, the error rate per subject and per item was always below 20%. The percentages of incorrect No-decisions out of all responses per condition are presented in table 4.1. below.
Friedman tests on arcsine-transformed proportions of incorrect No-decisions per subject and item for each condition were not significant.

Table 4.1: Experiment 1: percentages of incorrect No-decisions to word targets per condition

<table>
<thead>
<tr>
<th></th>
<th>associative</th>
<th>semantic</th>
<th>unrelated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cond. 1</td>
<td>3%</td>
<td>5%</td>
<td>2%</td>
</tr>
<tr>
<td>Cond. 2</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
</tr>
</tbody>
</table>

The percentages of incorrect Yes-decisions to the 36 nonword targets, separately for the pseudohomophones and nonpseudohomophones is given in table 4.2. below.

Table 4.2: Experiment 1: percentages of incorrect Yes-decisions to nonword targets per condition

<table>
<thead>
<tr>
<th></th>
<th>pseudohomophones</th>
<th>nonpseudohomophones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cond. 1</td>
<td>9.9%</td>
<td>7.2%</td>
</tr>
<tr>
<td>Cond. 2</td>
<td>10.3%</td>
<td>5.9%</td>
</tr>
</tbody>
</table>

The difference between the types of Nonword was significant in the by-subjects analysis for Presentation Condition 2 (Wilcoxon on arcsine-transformed error proportions: Z=-2.31, p=.02, 2-tail), but not in the by-item analysis.

The percentages of incorrect decisions after collapsing all word-nonword trials on the one hand, and all word-word trials (associatively related, semantically related and
unrelated word-word trials) on the other, are presented in table 4.3. below:

**Table 4.3.:** Experiment 1: percentages of incorrect decisions to nonword versus word targets per condition

<table>
<thead>
<tr>
<th></th>
<th>nonword targets</th>
<th>word targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cond. 1</td>
<td>8.6%</td>
<td>3.5%</td>
</tr>
<tr>
<td>Cond. 2</td>
<td>8.1%</td>
<td>3.3%</td>
</tr>
</tbody>
</table>

For both Presentation Conditions, chi-square tests indicated that Decision was not independent of word versus nonword status (all p's <.0001).

Only response times for correct decisions to word-word pairs were entered into the following analyses.

Response times were logtransformed and subject- and item-means calculated and entered into analyses of variance (without logtransformation, the resulting subject-means violated the homogeneity of variance assumption). The geometric means (based on subject-means) per condition are presented in table 4.4. below.
Table 4.4: Experiment 1: mean response times for correct decisions to word targets per condition

<table>
<thead>
<tr>
<th>Cond.</th>
<th>List 1</th>
<th>assoc</th>
<th>List 2</th>
<th>assoc</th>
<th>List 3</th>
<th>assoc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>503</td>
<td>515</td>
<td>505</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>526</td>
<td>547</td>
<td>541</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>515</td>
<td>510</td>
<td>541</td>
<td>515</td>
<td>524</td>
<td>529</td>
</tr>
</tbody>
</table>

There was a significant main effect for Stimulus Relation [F1(2,120)=6.33, p=.002; F2(2,66)=4.08, p=.021; but Min F'(2,145)=2.48, p<.10]. In the by-subjects analysis, the interaction List x Stimulus Relation was significant [F1(4,120)=2.75, p=.031]. The main effect for Presentation Condition was significant in the by-items analysis, but not the by-subjects analysis [F2(1,33)=27.04, p<.0005]. In the by-items analysis, the interactions Itemgroup x Relation [F2(4,66)=6.40, p=.0005] and Itemgroup x Presentation Condition x Relation [F2(4,66)=10.40, p<.0005] were also significant.

Next, separate analyses per Presentation Condition were conducted.

For Presentation Condition 1, no effects were significant in the by-subjects analysis. In the by-items analysis, the interaction Itemgroup by Stimulus Relation was significant [F2(4,66)=4.50, p=.003. Planned comparisons revealed that for some items, the associative and the unrelated conditions differed significantly [F(2,33)=6.68, p=.004],
while the semantic and the unrelated conditions did not significantly differ. Post-hoc paired t-tests per Itemgroup showed that for Itemgroup 1, response times were significantly faster in the associative than the unrelated condition \(t(11)=3.28, p<.004, 1\text{-tail}\). There were no other significant effects.

For Presentation Condition 2, there was a significant main effect for Stimulus Relation \([F1(2,60)=4.54, p=.015; F2(2,66)=4.93, p=.010; \text{but } \text{MinF'}(2,125)=2.36, p<.10]\). Planned comparisons revealed that the associative and the unrelated conditions differed significantly \([F1(1,30)=7.99, p=.008; F2(1,33)=8.29, p=.007; \text{MinF'}(1,63)=4.07, p<.05]\), and that the semantic and the unrelated conditions did not significantly differ. Post-hoc paired t-tests indicated that responses were faster in the associative than the semantic conditions \([\text{by-subjects: } t(32)=1.99, p<.03, 1\text{-tail}; \text{by-items: } t(35)=2.02, p<.025, 1\text{-tail}]\). In the by-subjects analysis, no further effects were significant. In the by-items analysis, the interaction Itemgroup x Relation was significant \([F2(4,66)=11.90, p<.0005]\). For this interaction, planned comparisons revealed that the difference between the associative and the unrelated conditions was significant \([F2(2,33)=19.58, p<.0005]\), as was the difference between the semantic and the unrelated conditions \([F2(2,33)=12.51, p<.0005]\). Paired t-tests per Itemgroup showed that in Itemgroup 1, the associative condition and the unrelated condition differed significantly (geometric means: 461 ms versus 538 ms; \(t(11)=8.23, p<.0005, 2\text{-tail}\)). In Itemgroup 3, however, this difference was just significant, with the mean response times in the associative condition being higher than in the unrelated condition \(t(11)=2.24, p=.046, 2\text{-tail}\). For the comparisons of the semantically related versus the unrelated conditions, a similarly mixed picture arose: for Itemgroup 1, this difference was significant \(t(11)=5.87, p<.0005, 2\text{-tail}\), with the mean in the semantically related condition lower than in the unrelated condition, but in Itemgroup 2 \(t(11)=2.46, p=.031, 2\text{-tail}\), the means were in the opposite direction. Post-hoc paired t-tests further indicated that the difference between the associative (461 ms) and the semantic condition (496 ms) was significant in Itemgroup 1 \(t(11)=4.19, p=.002, 2\text{-tail}\). There were no further significant effects.
4.5. Discussion

The results are consistent with other findings in that they show the relative difficulty of making lexical decisions for nonwords as opposed to real-words. Lexical decisions to nonwords were more often incorrect than they were to words. As other studies found, too, pseudohomophones fared worse than nonpseudohomophonic nonwords.

The main results of this study concern the priming effects as a function of the nature of the relation between prime and target. In the fast presentation condition, where the prime was presented for only 75 ms, there were no reliable priming effects, except for an indication that, for some items, responses in the associative condition were faster than in the unrelated control condition. In the slower presentation condition, in which the prime was presented for 250 ms, there was a reliable priming effect for the associatively related pairs over the control condition, but not for the semantically related pairs. Furthermore, responses in the associatively related condition were on average faster than in the semantically related condition. The results of Experiment 1 showed that an associative but not a semantic relationship was crucial in producing priming effects.

Why was there no reliable priming of lexical decisions for semantically related pairs? In contrast to other studies, which found priming for semantically but non-associatively related pairs, the semantic pairs in the present study consisted of words which were designed not to be very typical exemplars of categories and not to reflect a prototypical functional relationship (as in Moss et al., 1992, or Shelton & Martin, 1992). As was argued in Chapter 3, such relationships reflect a wide range of knowledge rather than specific semantic features of the kind that are required by semantic selection restrictions of verbs. The aim of Experiment 1 was to investigate whether semantic relationships based solely on the sharing of such specific semantic features facilitate performance. The lack of an effect cannot simply be attributed to problems with the lexical decision paradigm used in Experiment 1, since reliable priming effects were found for the associatively related pairs.
However, the lexical decision task might not have encouraged the subjects to engage in any linguistic processing of the words, and consequently, no fast semantic analyses took place. Note that this would not affect the associatively related pairs, since the lexical decision task is bound to pick up associative relations between words precisely because it might well be a task that is a-linguistic. The associatively related pairs might have yielded reliable priming effects because they allowed the subject to make use of overlearned and highly typical relations between words. Processing the associatively related words might have by-passed both the semantic subsystem and the central systems, instead exploiting established connections between words. As discussed in Chapter 3, it remains an open question as to how the associative relations are acquired. Simple appeals to co-occurrence patterns have been shown to be inadequate. One has to postulate some ability of the system to register predictive correlations between the occurrences of words. This ability indicates that the system, even though it has specialised linguistic subsystems, is able to work on linguistic stimuli merely using a general pattern-matching ability.

Two ways might be used to encourage subjects to analyze the semantic features of words. First, one could embed the lexical decision task into sentence contexts, with the target being presented at the end of a sentence fragment. It might be that the proposed semantic component of the language recognition module only becomes active when the syntactic parser, which constitutes another component of this module, is active, too. This seems plausible from a functional point of view which would suggest that semantic decomposition serves the purpose of providing representations necessary to accomplish the overall function of the language recognition module, which, as in my model, is to transform linguistic input into a format that can be fed to the central systems. Furthermore, some aspects of syntactic analysis are not only structure-dependent but require word-specific semantic information (e.g. to fulfill the semantic selection restrictions of verbs, see next chapter), which would have to be activated rapidly to be used while the analysis of a sentence is on-going.

Thus, while Experiment 1 clearly showed that there was no priming of lexical decisions when a word was preceded by a prime with which it shared basic semantic
features (whereas such priming was found for associatively related pairs), further tasks will have to be used to establish which role such abstract semantic features play in language processing. Making lexical decisions on a word preceded by another word might not require the processing of the semantic features of these words. In the final chapter, the results from Experiment 1 will be discussed again in the context of the results of the following experiments which focus on the processing of word meaning in sentence contexts.
CHAPTER 5 : PROCESSING THE MEANING OF NOUNS IN SENTENCE CONTEXTS

5.1. Introduction

Experiment 1 did not find any reliable priming for making a lexical decision on a target word which was related to its preceding prime word solely by virtue of sharing some basic semantic features. Priming was only found if the target was a high associate of the prime. One of the questions that arises out of these findings is whether semantic features, which according to some linguistic theories are syntactically required, might only be activated in linguistic, rather than single word processing. Two types of syntactically relevant semantic information will be addressed in the next experiments: experiments 2, 3 and 4 investigate subjects' sensitivity to violations of semantic selection restrictions which verbs place on their subcategorised complements, and experiments 5, 6 and 7 focus on the role of semantic features in assigning thematic relations and the consequences for processing syntactically ambiguous (garden path) sentences.

In the following section, I will characterize the notions of thematic and subcategorisation frames as they are being used in linguistics and psycholinguistics. I will then describe in some detail Jackendoff's (1987a) linguistic theory of thematic relations and semantic selection restrictions, which provides a principled definition of grammatically relevant semantic information, i.e. the type of abstract semantic category information that determines which words can instantiate the subcategorised syntactic constituents of a verb. In the final section of this chapter, I will discuss psycholinguistic evidence for the effect of semantic violations during sentence processing, in relation to some classic sentence priming studies as well as to studies of the effect of violations of semantic selection restrictions.
5.2. Thematic relations and subcat frames

Recently linguists focused increasingly on word meaning as part of a competence theory of language. To some extent this is because lexical information, of which information about word meaning is a part, has come to play a larger role in theories of grammar (see eg. Bresnan, 1978, 1982; Chomsky 1981, 1986a, 1986b; publications grown out of the "Lexicon Project" at MIT, see Levin, 1985). In the psychological study of human language processing, too, the representation and processing of lexical information has been the centre of quite diverse research programmes (see eg. contributions in Marslen-Wilson, 1989).

In particular, thematic relations (also called thematic or theta roles) have become the centre of attention. In non-technical terms thematic relations can be described as the semantic roles played by the syntactic constituents which a verb requires to express its meaning. They are also called arguments, or the "semantic-relational elements" which are "necessary to spell out the participants in the events" (Fisher, Gleitman and Gleitman, 1991, p. 335). For example, the meaning of the verb 'put' involves an agent causing an object to move to a certain location. The syntactic constituents playing the thematic roles are noun phrases (NPs for short), eg. agent may be expressed by 'John', the moving object by 'milk', and the location by 'the fridge'. However, not only NPs receive thematic roles but PP may do so as well (eg. in 'The light changed from red to green', 'green' is Goal, see Jackendoff, 1987a, p. 380).

Note that thematic roles are only assigned to those syntactic constituents that a verb absolutely requires to express its meaning, or, more technically, to the syntactic constituents which are specified in the subcategorisation frame (subcat frame for short) of a verb. Using 'put' again as an example, constituents such as 'in the morning' do not receive a thematic role even though they contribute to the full sentence meaning by adding temporal information, eg. 'In the morning John put the milk into the fridge'. The phrase 'in the morning', however, is not subcategorised by the verb, and hence does not receive a thematic role. (In terms of X-bar theory, subcategorised constituents are added to the verb at the V' level, whereas temporal adjuncts and other modifiers
are added at V" or V"'). Thus, the lexical entry of each verb (and of some prepositions and some nouns) contains the **subcat frame**, in which the required syntactic constituents are specified, and the **thematic frame** which indicates the thematic roles played by the subcategorised syntactic constituents. Originally, Chomsky (1981) postulated that each thematic role (called theta-role in his theory) had to be expressed by one and only one noun phrase, or, to put it the other way around, that each subcategorised noun phrase in a sentence stood for one, and only one, theta role. Jackendoff (1987a) argues that this claim (the "theta-criterion") need not necessarily hold (see discussion below).

Subcat and thematic frames have also been intensely studied in recent psycholinguistic work. Some developmental psychologists proposed that the assumed close relationship between a verb's subcategorisation frame/s (i.e. the contexts of syntactic constituents a verb can correctly appear in) and the verb's argument structure (the thematic roles such as eg. "object that undergoes motion", "path", "agent causing motion") can be exploited in language acquisition. Fisher, Gleitman & Gleitman (1991) set out to show that this close relationship does in fact exist for adults. They report experimental evidence indicating that the closer the perceived meaning of two verbs, the greater the overlap between their subcategorisation frames. One group of subjects had to rate the naturalness or acceptability of verbs in a series of subcategorisation frames. Another subject group was presented with triples of verbs and had to indicate which one was least similar in meaning to the other two. The verbs were from the fields of cognition, perception, motion and location, and also included symmetrical verbs such as 'meet'. The clusters of verbs resulting from the subcategorisation ratings positively correlated with the clusters resulting from the similarity task (eg. verbs that take prepositional phrases tend to express the abstract semantic elements of position in or motion through space; verbs that take sentence complements are cognition- and perception-verbs; verbs that subcategorise three NPs typically express transfer of objects or ideas). The authors also report evidence that subtle semantic differences between verbs **within** one semantic domain are predictive of differences in acceptable subcategorisation frames (eg. verbs of perception). It is worth pointing out that the correlation between syntactic frames and argument structures of verbs is an indicator of subjects' accessible
conscious judgements. A relationship based on rating performance is uninformative with respect to how syntactic and semantic information might interact in automatic, consciously inaccessible processing.

Thematic frames do not of course provide exhaustive information about verb meaning. In particular, there are meaning aspects which do not involve differences in the arguments of the verb and thus have no syntactic consequences. For example, the particular manner of motion (as in the actions of sliding versus rolling) does not affect the verbs’ argument structure: the arguments for both verbs are moving object, path, and agent causing motion. Since subcategorisation frames are the syntactic consequences of arguments (but not other meaning aspects of verbs), different motion verbs have the same subcategorisation frames.

Gleitman and colleagues (eg. Fisher et al., 1991; Landau & Gleitman, 1985) proposed that subcategorisation frames guide acquisition of word meaning. They argued that the acquisition of verb meaning is seriously underdetermined by observational evidence, and that a child would not be able to figure out the full meaning of many verbs if the only available evidence came from observing the scenes that a verb typically describes. For example the differences between 'look' and 'see', 'give' and 'take' etc., which express different perspectives on essentially the same observed event, could never be noticed, let alone figured out, on the basis of observation alone. Since such verbs, however, occur in different subcategorisation frames, the language-learning child can use this linguistic evidence to realise that there are meaning differences, and to establish what they are. For example, 'look' versus 'see' express perceptual exploration versus perceptual achievement, a semantic difference which might be signalled to the child by the verbs’ sharing subcategorisation frames with two different groups of verbs, i.e. perceptual verbs and cognition verbs respectively.

While much of psycholinguistics and experimental psychology focussed on the processing of nouns (see review of the semantic priming effect in Chapter 3), the study of lexical processing has recently been extended to encompass verbs, eg. by investigating the activation of subcat frames and argument structures of verbs (Shapiro
and colleagues, eg. Shapiro, Zurif & Grimshaw, 1987, 1989). The evidence so far speaks clearly for exhaustive context-independent preliminary activation of all the verb's argument structures. This research has stimulated some debate, eg. Schmauder (1991) reported failure to replicate Shapiro’s results. Shapiro, Brookins, Gordon & Nagel's (1991) present further experiments which, so they claim, rule out the criticisms and confirm the modularity claim.

Verb argument structure has also been investigated in adult aphasic subjects. Shapiro & Levine (1990) showed that Broca aphasics, but not Wernicke aphasics, activate multiple argument structure possibilities immediately after encountering a verb. Thematic role assignment by Broca’s aphasics was also studied by Friederici and Frazier (1992).

In the next section I will present Jackendoff’s description of the semantic specifications of verbs and nouns and his account of thematic relations and semantic selection restrictions.

5.3. Jackendoff’s competence theory of thematic relations and semantic selection restrictions

In his account of semantic selection restrictions Jackendoff describes the 'fitting' of semantic requirements of verbs with the semantic features of nouns as a process of fusion. The meaning of verbs is characterised as a function and argument structure (see Chapter 4.1.). The meaning structure of the verb has to "fuse" with the meaning specified in the lexical entries of the heads of the subcategorised constituents of the verb, which fill the argument positions of the verb (Jackendoff, 1987a, p. 383). Jackendoff (1987a) distinguishes between two kinds of characteristics that have to match in order to create a well-formed fusion resulting in a well-formed sentence: first, the major conceptual category of the head of a subcategorised constituent has to fit the one required by the argument. Second, there must also be correspondence in more detailed features than major conceptual category: within the major conceptual
categories there are finer distinctions (or "markers", Jackendoff, 1987a, p. 386), like solid/gas/liquid or animal versus human for the THING category (see Chapter 4.1.). Take, for example, the sentence

* Harry drank powder.

This sentence cannot receive a well-formed reading because there is a clash of the marker +SOLID, which is part of the noun's lexical entry, and the requirement of the verb to have a direct object that is +LIQUID.

The next example is ill-formed because of the incompatibility of the major conceptual category features. The object noun is +PROPERTY, whereas the verb requires a constituent with the major conceptual category feature +THING:

* Harry drank sincerity.

Incorrect sentences such as these have traditionally been discussed as violations of semantic selection restrictions. Jackendoff treats them as examples of violations during argument fusion, a notion which will become more transparent in the next section which gives a brief summary of his theory. I will present the fundamentals of Jackendoff's (1987a) linguistic theory in a psychological framework, elaborating the psychological equivalents of his notions. The questions I will come back to is how the linguistic processing system computes the required semantic category features for the nouns that fill the argument positions, and how clashes ('mis-fusions' or, more traditionally, violations of semantic selection restrictions) affect sentence processing.

Jackendoff sees grammar as the mental representation of linguistic knowledge. He assumes three autonomous levels of grammar, i.e. phonological structures, syntactic structures and conceptual/semantic structures. Jackendoff explicitly develops a framework that "eliminates syntactocentrism" (Jackendoff, 1987a, p. 374) in designing the three levels as autonomous (this is in contrast for example to the Chomskian Government and Binding theory in which the syntactic level is prominent). Each of the
three levels is characterised by primitives, and principles of combining the primitives. Furthermore, at each level there is a set of formation rules, which generates the well-formed structures possible at that level. The three levels are linked via correspondence rules. No independent level is assumed to exist for the lexicon. Instead, Jackendoff assumes that the lexicon is part of the correspondence rules: a lexical item is a "small-scale correspondence between well-formed fragments of phonological, syntactic, and conceptual structure" (Jackendoff, 1987a, p. 372). Note that these rules are not to be equated with processing rules: Jackendoff's account is at the level of competence theory, not at the level of algorithmic or psychological theory.

I will now focus on the conceptual/semantic structures (CS for short) and their links with syntactic structures (SS for short).

CS is the conceptual structure "in which thought is couched, the 'Language of Thought' in the sense of Fodor (1975)" (Jackendoff, 1987a, p. 374). It is at this level that Jackendoff defines thematic relations. The formation rules for CS are based on a vocabulary of primitive basic conceptual categories, introduced in Chapter 4.1. (Jackendoff also calls them the major conceptual categories), which areThing, Event, State, Action, Place, Path, Property and Amount. These primitives can be expanded into more complex conceptual constituents via formation rules which reformulate the original basic conceptual categories as functions with one or more argument positions. Place, for example, is expanded by a Place function (eg. under, top of) which has one argument position filled with the argument Thing, and Path is expanded by a Path function (eg. to/toward/away from (Thing)). Table 5.1. below gives some examples.
Table 5.1: Examples of complex conceptual constituents in Jackendoff’s theory

<table>
<thead>
<tr>
<th>Basic constituent</th>
<th>Complex constituent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place --&gt;</td>
<td>[Place-Function (Thing)]</td>
</tr>
<tr>
<td></td>
<td>'under the table'</td>
</tr>
<tr>
<td>Path --&gt;</td>
<td>[Path-Function (Thing)]</td>
</tr>
<tr>
<td></td>
<td>'towards the table'</td>
</tr>
<tr>
<td>Path --&gt;</td>
<td>[Path-Function (Place)]</td>
</tr>
<tr>
<td></td>
<td>'from under the table'</td>
</tr>
<tr>
<td>Event --&gt;</td>
<td>[Event GO (Thing, Path)]</td>
</tr>
<tr>
<td></td>
<td>'he walked into the room'</td>
</tr>
<tr>
<td>Event --&gt;</td>
<td>[Event STAY (Thing, Place)]</td>
</tr>
<tr>
<td></td>
<td>'he stayed in the kitchen'</td>
</tr>
<tr>
<td>Event --&gt;</td>
<td>[Event CAUSE (Thing, Event)]</td>
</tr>
<tr>
<td></td>
<td>'he drank the beer' (he caused something to go down his throat)</td>
</tr>
<tr>
<td>State --&gt;</td>
<td>[State BE (Thing, Place)]</td>
</tr>
<tr>
<td></td>
<td>'the dog is in the park'</td>
</tr>
<tr>
<td>State --&gt;</td>
<td>[State ORIENT (Thing, Path)]</td>
</tr>
<tr>
<td></td>
<td>'the sign points to New York'</td>
</tr>
</tbody>
</table>

Note that some arguments decompose further into functions with their own arguments. Thus, for example, a full reading of the Go-Event is as follows:

[Event GO ([Thing], [Path TO ([Place IN ([Thing])])))]

The main correspondence rule governing the interface with SS (syntactic structures) states that every syntactic constituent (apart from expletives like 'there' and 'it', as in 'There are some books on the table' and 'It was raining') corresponds to a conceptual constituent of some major conceptual category, but that not every conceptual
constituent making up the meaning of the whole sentence necessarily has to correspond to a syntactic constituent. The latter can be demonstrated for example with verbs which 'incorporate' Path and Place functions: these functions are not expressed by a syntactic constituent, eg. a preposition, as is demonstrated in: 'John entered the room' (compare this with 'John ran into the room', in which the conceptual constituent Path has to be syntactically expressed by a preposition). Note also that 'John entered' means that John went into something, whereas 'John ran' means that John traversed some unspecified trajectory. These observations which Jackendoff discusses at length seem to violate Chomsky's (1981) 'theta-criterion'. The theta-criterion is a principle defining a necessary condition for well-formedness at Logical Form (LF):

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

The theta-criterion is implied to hold at all syntactic levels (LF, d-structure and s-structure) as a consequence of the Projection Principle. This principle expresses the "intuitive idea" that "representations at each of the three syntactic levels are projections of lexical properties" (Chomsky, 1981, p. 39). An example of how the theta-criterion works at the level of d-structure is that it affects movement of constituents, which can only be to a non-theta position. Jackendoff doubts that the theta-criterion holds necessarily. Some counterexamples were given above. He maintains that no rigid correspondence could hold between syntactic constituents and theta-roles. The examples undermine the 'biuniqueness'-aspect (an NP can only have one theta-role, and a theta-role can only be expressed by one NP), which is usually understood to be the main stipulation of the theta-criterion. Jackendoff, however, also points out that Chomsky (1981) himself mentions counterexamples to biuniqueness (which Chomsky credits to Jackendoff (1972), see Chomsky, 1981, p. 139, FN 14). Jackendoff claims that the primary purpose of introducing the theta-criterion in G&B was not to enforce biuniqueness, but to ensure that an NP does not acquire additional theta-roles in the process of derivation (which would go against the Projection Principle). In any event, the biuniqueness assumption clearly fails in two ways: firstly, there are NPs that have more than one theta-role, as in the case of transaction verbs such as buy and sell: these verbs mean that two actions occur at the same time ('buy', for example, implies the
action of giving away something for money and that of receiving something after
having given money). In the case of 'She bought a book', the subject NP is both Goal
(since she received a book) and Source (since money went from her to someone else).
Secondly, several NPs can hold one single theta-role as in 'The box has books in it':
'it' is coreferential with 'the box' and thus does not have a separate theta-role.

After this discussion of the rule requiring each syntactic constituent to correspond to a
conceptual constituent, let us turn briefly to some "subsidiary" correspondence rules
suggested by Jackendoff. They determine (language-specifically) which syntactic
category can express which conceptual constituent. As discussed in Chapter 4.1., Noun
Phrases (NPs for short) in English can express a Thing (eg. 'horse'), an Event (eg.
'earthquake'), a Property (eg. 'redness'), an Amount ('mile') etc. Note that this wide
range makes a mis-fusion of an NP with the semantic requirements of verb-arguments
more likely. PPs can express a Place ('in the house'), or a Path ('toward the house')
or, in idioms, a Property ('in luck', 'out of your mind'), and Sentences express Events
('John ran into the house') or States ('John was very happy').

Let us demonstrate correspondence rules in an example sentence, such as 'John ran
into the house'. The sentence as a whole corresponds to an Event-function. As shown
above, this function has arguments some of which are functions themselves, i.e.

Event [Go ([Thing1], [Path TO ([Place IN ([Thing2])]))]]

The main verb ('ran') corresponds to the Event function Go. The subject of the
sentence ('John') corresponds to the first argument of Go, Thing 1, and the PP ('into
the room') to the second argument of Go, i.e. Path, which itself is a function (the To-
function) with Place as its argument. Place itself is composed of a Place-function (In),
which is syntactically realised as the preposition 'into', and a Thing ('the room').

Let us use the verb 'run' to demonstrate how a verb's argument structure and the
syntactic constituents match. In conceptual structure, the meaning of 'run' is
represented as an Event-function with two arguments, the Thing that is moving and the
Path along which the movement occurs:

run: [Event GO (Thing, Path)]

As mentioned above, this representation can be expanded into:

run:
[Event GO ([Thing1], [Path TO ([Place IN ([Thing2]))]])]

The link between the meaning representation of 'run' and the syntactic constituents which 'run' subcategorises is formally represented by coindexing. The arguments of the verb and the syntactic constituents corresponding to them in a particular sentence receive the same index. The first argument (Thing1) corresponds to the subject of the sentence, here 'John'. The second argument (Path), is coindexed with the PP ('into the room'). Note that the second argument only optionally corresponds to a syntactic constituent since 'John ran' is a correct sentence. Note also that 'John ran' is understood to mean that John ran along some, if unspecified, Path: even if syntactically unexpressed, the argument Path is present at CS and contributes to the meaning of the sentence.

This summary of the primitives and combinatory rules of CS, and the correspondence rules linking CS to SS provide the background for understanding Jackendoff's account of thematic relations and semantic selection restrictions.

The definition of thematic relations is couched in terms of argument positions in CS. The thematic role Theme, for example, paraphrased as 'Thing in motion or being located' is the first argument of the Event-functions GO, STAY, BE and ORIENT. In the sentence 'He ran to the house', 'he' has the role of theme, since 'he' fills the Thing-slot of the argument structure of the verb 'run'(run: Go(Thing,Path)). Theme need not be a Thing, but may be an Event, too, as in 'He postponed the lecture', since 'postpone' means that somebody causes an event to be in motion, as it were. The thematic role Source is the argument of the Path-function FROM. Goal is the argument of the Path-function TO, and Agent is the first argument of the Event-function
CAUSE. Thus, Jackendoff suggests that the labels of thematic roles should be understood as shorthand means of referring to particular structural configurations in CS.

Let us return to Jackendoff’s account of semantic selection restrictions of verbs in terms of argument fusion, which is notated by coindexing. Argument fusion is the combination of semantic information in the verb’s argument structure with the semantic structure of the nouns that correspond to the verb’s arguments in the actual sentence. Jackendoff proposes two conditions for successful fusion corresponding to the two types of information that arguments provide. Take for example the verb 'drink'. The major conceptual categories required by its arguments are represented in its CS as follows (presented here in a reduced format without indices and expansion of Path):

drink:
[Event CAUSE ([Thing 1], [Event GO ([Thing 2],[Path]))]].

In addition, it is part of the meaning of 'drink' that the Thing that 'goes' somewhere (i.e. Thing 2) is of a special subclass of Thing, i.e. it has to be a liquid rather than a solid or a gas. In Jackendoff's notation this semantic requirement is expressed by adding the marker LIQUID to the conceptual constituent in question, thus leading to the following representation of the meaning of 'drink':

[Event CAUSE ([Thing1],
[Event GO ([Thing2 LIQUID],[Path]))]].

If the argument [Thing LIQUID] is fused (coindexed) with a noun whose conceptual structure does not match with the semantic marker of the verb’s argument, an ungrammatical sentence results, which is exactly of the kind traditionally referred to as a violation of selectional restrictions, eg.

* 'He drank the powder.'

The meaning of 'powder' (Thing, solid) does not incorporate the required semantic
subcategory 'liquid'. Semantic selectional restrictions are part of verb meaning. A verb specifies semantic information about two aspects of its argument/s. Firstly, about the membership of the argument in one of the mutually exclusive major conceptual categories (Thing, Property, Place, Event, etc.), and secondly, about membership in one of the equally mutually exclusive lower-level conceptual categories such as Solid, Liquid, Gas, Inanimate, Animate, Human or Animal as sublevels within Thing. A violation of selectional restrictions can occur either because of a mismatch of the lower-level conceptual categories of the verb’s argument (the indexed argument) and the noun corresponding to it (as in *'Harry drank the powder'), or also because of a mismatch of the major conceptual categories (as in * 'Harry drank sincerity'). Note that no violation of thematic roles is involved, since 'sincerity' is Theme in the same way as eg. 'beer' would have been, i.e. it is the first argument of an Event function.

If an argument is coindexed with a syntactic constituent, the semantic specification of the verb serves as selectional restrictions of the noun that can fill the coindexed syntactic constituent. If the argument is not coindexed, the semantic information specified in the argument serves as an implicit argument. Further above I gave the example of the verb 'enter' whose CS contains a non-indexed Path-function: it does not allow a syntactic constituent to express the Path-function (*'He entered into the room'). 'enter' is an example in which an argument is (obligatorily) not indexed because the argument is incorporated in the verb (another example is 'to butter', in which the first argument of the GO-function, the Theme (Thing 2), is not indexed: [Event CAUSE ([Thing 1], [Event GO ([Thing 2], [Path TO ([Place])])]), yielding 'John buttered the bread', but *'John buttered the bread with butter'). Other arguments are optionally not indexed, as in 'drink': both 'John drank' and 'John drank the wine' are grammatical.

5.4. Psycholinguistic studies of the effect of sentence context on word processing

I will first describe some of the classic studies of sentence priming. Although they did not focus on violations of selection restrictions, they often used an unrelated or so-
called incongruous sentence condition which included targets that fitted the sentence syntactically but not semantically. These studies produced an overall priming effect and significant facilitation and, under certain conditions, inhibition as well. I will briefly discuss these results here. Their interpretations will be reviewed in the section 6.4., when I compare various models in the light of the findings of Experiments 2 and 3.

Schuberth & Eimas (1977) found facilitation and inhibition in a lexical decision task. Targets were placed in sentence-final position. The factor of greatest interest to the present discussion is the 'congruity' of target and sentence. Examples of their congruous sentences are 'The puppy chewed the bone' (the target 'bone' is a high frequency word), or 'The puppy chewed the sock' (target is a low frequency word), whereas incongruous sentences were 'The puppy chewed the hour', and 'The puppy chewed the campus'. Unfortunately, the authors do not present their stimulus material in full, but the examples indicate that the incongruous sentences in fact contained targets which violated semantic selection restrictions rather than pragmatic expectations (as would be the case in 'The puppy chewed the goldfish'). Initially, congruity was determined by the authors' intuitions, and then verified in a rating study with 52 subjects, who were asked to rate each word in terms of how 'appropriate' it was in the sentence context, using a 4-point scale. The authors do not report whether targets were associatively related to words in the preceding sentence. The overall Cloze value of the targets was .12, with a maximum of .70 (Schuberth, pers. com., quoted from Fischler & Bloom, 1979, p. 13). In the main experiment, the context sentence was presented visually for 1500 ms, which the subject had been instructed to read silently. Immediately after the sentence context disappeared the target on which the subject had to perform a lexical decision was presented for 3 seconds. Lexical decisions were found to be faster for congruous as compared to incongruous targets (there was also a significant main effect of frequency). The overall priming effect was due to both significant facilitation (35 ms for high frequency targets; 14 ms for low frequency targets) and inhibition (22 ms for high frequency targets; 17 ms for low frequency targets), compared to the neutral condition which consisted of a blank white field presented instead of the sentence. The effects of target word frequency and contextual
congruity were additive.

The results were interpreted in terms of the Logogen model (Morton 1969a), which is problematic given that Schuberth & Eimas (1977) used the lexical decision task which is assumed to tap central processes rather than the operation of the input logogens (see eg. Morton, 1982). Furthermore, Schuberth and Eimas (1977) do not specify the semantic information extracted from their context sentences and counted by the logogen. They assume that contextual information, via 'semantic analyses', increases the response strength of 'appropriate' logogens, while leaving the response strength of logogens unchanged if they are not appropriate given the context (rather than lowering their response strength). Since firing of a logogen depends on the response strength for a particular logogen relative to the total response strengths for all logogens, the response strengths of inappropriate logogens are effectively decreased, since response strengths of congruous logogens are raised.

Fischler & Bloom (1979) also found both facilitation and inhibition in a lexical decision task as a function of congruity between target and preceding sentence. Stimulus materials were derived from sentences given in a pretest written Cloze task. Congruous targets had been given as a primary response on average by 53 out of 100 subjects. Associations between words in the sentence and the target were avoided. In the main experiment, a sizeable inhibition effect was found (of around 115 ms). Facilitation, however, was only significant when the target was extremely predictable from the sentence context (the probability of the target as the primary response in the Cloze task was .91). Again there is no full list of stimuli in the paper, and no details are given of how the incongruous anomalous targets were selected; the authors merely state that incongruent targets were chosen to be syntactically, but not semantically appropriate in the sentence context. From the examples it appears that the incongruent targets violated semantic selection restrictions ("She cleaned the dirt from her terms"). The neutral control condition consisted of a string of x’s (another baseline was also used consisting of data from a group of subjects who had only seen anomalous trials. The pattern of results was not affected by the type of baseline condition). Each sentence was presented for 2 seconds on a VDU. After it disappeared, there was a
blank screen for 500 ms, after which the target was displayed until a response was made. The same pattern of results was obtained when context was presented for 3 seconds; neither facilitation nor inhibition increased significantly.

Stanovich & West (1983, experiment 1) found facilitation dominance, and no significant inhibition with their sentence primes, in a task in which the target had to be read out aloud (dubbed the naming task). The neutral condition consisted of the sentence fragment "They said it was the". The sentence contexts, which the subject had been instructed to read silently, were presented on a computer screen for 2500 ms, followed by the target which the subject had to read out aloud (it is not clear whether the sentence context remained on the screen when the target was displayed). The authors also report that the overall contextual priming effect (i.e. the difference between the congruous and incongruous conditions) was greater for difficult targets, i.e. words of low frequency and low Cloze value (11%), as compared to easy targets, i.e. words of high frequency and a Cloze value of 43%. Difficult and easy targets had been designed not to differ in their average relatedness (as established by a rating pretest in which subjects rated the relatedness of targets and subject nouns of the sentence the targets appeared in, and the relatedness of targets and the verbs, using a scale from 1 - 5). Examples of congruous sentences with an easy target and a difficult target are given below:

The skier was buried in the snow
The skier was buried in the avalanche

Incongruous sentences were created by combining a target with an unrelated sentence context. A close inspection of the full list of stimuli (which is given in Stanovich & West, 1981) reveals that semantic violation and pragmatic implausibility are confounded. While in most cases the incongruous target violates semantic selection restrictions of the verb (eg. 'The barber trimmed the gas'), in some cases the incongruous target has the required semantic features but is an implausible continuation (eg. 'The skier was buried in the car'; 'The boy handed his date the drinks').
Tyler and colleagues maintain a strongly interactive view of sentence processing. Lexical-semantic and lexical-syntactic information is argued to be activated early, and rapidly assessed against sentential and discourse contexts. The work is based on the Cohort Model of spoken word recognition. The Cohort Model assumes that auditory word recognition is facilitated by 'contextual factors', i.e. by assessing the semantic and syntactic lexical information activated for each member of the initial cohort against the semantic and syntactic structure of the sentence in which the word occurs. The Cohort Model assumes that sensory input activates a set of word nodes comprising all words that are compatible with the sensory input (even words whose initial phoneme differ in one feature from the input). During this initial access stage lexical and contextual information plays no role. The access stage is completely bottom-up (e.g. Tyler found with her gating experiments that subjects who had to guess which word was presented to them after having heard only the first initial phonemes produce words that are contextually inappropriate). In the following selection phase, where the cohort is reduced to leave one (or few, in later formulations) of the candidates, top-down factors are assumed to operate (leading to word recognition in sentence context to be faster than in isolation). For all words that are part of the cohort, lexical semantic and syntactic information is assumed to be activated in parallel, and the appropriateness of this information relative to the sentential and discourse information built up so far is assessed in parallel for all candidates (e.g. the recognition point for a word in isolation is later in the word than for a word in context: e.g. 'p' in 'trespass' is the recognition point in isolation, i.e. there are no other English words that begin with 'tresp'; however, in context, recognition point might be at the preceding 's', because possible competitors, like trestle, tress, are evaluated to be contextually inappropriate, see Marslen-Wilson, 1987). The assessment of contextual appropriateness occurs while further sensory input is received. Non-sensory information cannot override sensory information, that is contextual information cannot lead to a removal of a word from the initial cohort. However, it can facilitate activation, as is indicated by evidence that word recognition is faster in sentence context than for isolated words.

Marslen-Wilson, Brown & Tyler (1988) showed that the detection of an auditorily presented word in a sentence context is slowed down if the word (a noun) violates the
item-specific information attached to the preceding verb, as compared to the detection of a target noun that is semantically, syntactically and pragmatically appropriate. Response times in a word detection task were not only slowed down for violations of subcategorisation information ('The man slept the guitar') and selectional restrictions ('The man drank the guitar'), but also for pragmatic inferences not specifically stored with the word in the lexicon but computed against the background of the current context ('The man buried the guitar'). The increase in RT (as compared to the correct condition: 'The man carried the guitar' with a mean RT of 241 ms) was largest for the syntactic-violation condition (320 ms mean RT), followed by the semantic-violation condition (291 ms mean RT) and the pragmatic-violation condition (268 ms mean RT). All RTs in the violated conditions were significantly different from the mean RTs for the correct condition, using Min F'. The difference of RTs for the semantic versus pragmatic conditions (291 ms versus 268 ms) was not significant in a Newman-Keuls test when the error term was derived from Min F', but it was significant when error terms were derived from the by-items and the by-subjects analyses of variance. The results indicate that semantic violations interrupt processing slightly more than pragmatic violations (the difference is not statistically significant). The interpretation of this finding is open. Marslen-Wilson et al. (1988) discount the possibility that the difference in response time indicates a difference in processing systems involved. They point out that a semantically-violated sentence (eg. 'John drank the guitar') is, in addition, also pragmatically odd, in fact, they suggest, odder than a pragmatic violation (eg. 'John buried the guitar'). This should cause greater difficulties in creating a pragmatically possible context, leading to an increase in response time in the word detection task. However, the alternative interpretation that the response time difference indicates a difference in processing systems cannot be ruled out by the present data, as Marslen-Wilson et al. (1988) admit. It might be that semantic violations are detected fast because they are processed by a linguistic module, whereas pragmatic violations get noticed only after the first-pass linguistic analysis. If pragmatic inferences start later or take longer (than the detection of a semantic violation), they have less time to affect a detection response. It is possible that semantic violations are effective faster (and therefore delay response times to a greater extent) because they are processed by an encapsulated linguistic system. Such an account would predict that interpretations of
pragmatically violated sentences should be made easier if lead-in sentences were provided, whereas sentences with semantic violations should always slow down processing (as should syntactic violations) notwithstanding any lead-in sentences.

There are some problems with the Marslen-Wilson et al. (1988) study. First of all, no neutral condition was used in which a baseline for speed of word detection could have been established.

Furthermore, the finding that the syntactic-violation condition led to greater slowing down of the detection response is exactly opposite to what Friederici and colleagues (Hahne, Pütz, Friederici & Rösler, 1991) report using a lexical decision task with written material. This might simply be due to the enormous differences in materials, tasks and language studied, and would therefore be of little interest. Friederici draws some quite far-reaching theoretical conclusions from her results, however, which are worth discussing. She found that semantic violations led to longer RT latencies than violations of subcategorisation frames (usually violations of transitive and intransitive requirements, like in Marslen-Wilson et al.'s (1988) syntactic-violation condition). She concludes that semantic and syntactic information are handled by different processing systems, and tries to locate them in different regions of the brain using evidence from event-related electrical brain potentials measured during semantic versus syntactic processing. Another problem in the Hahne et al. (1991) study lies with the fact that in the semantically normal condition, only sentences were used in which verbs and nouns were highly associated. This makes it difficult to interpret the finding that in the violated (and non-associated sentences) lexical decisions were slowed down as compared to normal sentences.

My Experiments 2 and 3 address the questions of whether the semantic requirements of verbs and the semantic categories of subcategorised nouns are activated rapidly by testing the effect of semantic violations on sentence processing. Experiments 4a and 4b also incorporate pragmatic and syntactic violations. These studies were designed to address some of the problems of the Marslen-Wilson et al. (1988) and the Hahne et al. (1991) study. They also extend the classic sentence priming studies by investigating
the semantic influence of sentences on priming lexical decisions when the predictability of target words is extremely low.
CHAPTER 6 : SENSITIVITY TO VIOLATIONS OF SEMANTIC SELECTION RESTRICTIONS: EXPERIMENTS 2, 3, 4 AND 4B

6.1. Introduction

In the previous chapter the notions of basic semantic categories and semantic selection restrictions of verbs were introduced. It was pointed out why these notions are relevant to some of the main issues in psycholinguistics. This chapter presents Experiments 2 and 3, which tested the effect of violating semantic selection restrictions, Experiment 4, which compared the effects of semantic and pragmatic violations, and Experiment 4b, which tested the effects of semantic, pragmatic and syntactic violations.

In these experiments the processing of the meaning of nouns and verbs is investigated in sentence contexts, in contrast to the priming study reported in Chapter 4 in which the semantic relation between two nouns was investigated.

The semantic meaning of nouns is again conceptualized as the basic semantic features discussed by Jackendoff (1983). A few examples are repeated here for illustration. The noun 'horse', for example, has attached to it as part of its lexical entry the feature 'animate Thing', and the noun 'beer' the feature 'inanimate liquid Thing'. The meaning of the word 'earthquake' can be represented by the feature 'Event' (not: Thing), and 'mile' has the feature 'Amount'. Obviously, the full meaning of a noun consists of more information than is conveyed by these features, for example information that differentiates between different inanimate liquid Things (eg. 'milk' versus 'beer') and still more specific information, such as encyclopedic knowledge allowing one to make specific distinctions between, for example, different kinds of beer.

According to Jackendoff's semantic theory (discussed in Chapter 5.3.), the meaning of verbs can be represented as a function-argument structure, eg.
'murder':

[Event CAUSE ([Thing 1], [Event])]

or in more extended format:

[Event CAUSE ([Thing 1, Human],
               [Event GO ([Thing 2, Animate], [Path])])]

What I called Thing 2 above will be coindexed with an NP that has the patient role. The semantic entry of the verb 'murder' requires that the head of this NP maps into the THING category, and, more specifically, into the HUMAN ANIMATE THING category. Thus, the noun 'president' (Thing, Human) would fuse with the verb's argument Thing 2, but 'honey' (Thing, Inanimate) would not, leading to a semantically violated sentence:

'The honey was murdered'

Experiments 2 and 3 study the processing of semantic features of nouns in sentences with violations of semantic selection restrictions. Further examples of verbs which impose semantic selection restrictions are 'to drink' and 'to assassinate':

- to drink something: the subcategorised noun in direct object position has to have the semantic feature 'liquid thing' (* 'He drank the powder' is incorrect).

- to assassinate someone: the noun has to have the semantic features 'animate' and 'human' (*'The table was assassinated' is incorrect)

A sentence such as: 'He drank the powder' is syntactically well-formed, given syntactic-category information and phrase structure rules. Thus, on a purely syntactic account of initial sentence processing (such as Frazier's, 1987a), no disruption of processing should occur when reading this sentence. According to such an account, initially a syntactic analysis is performed which requires access only to lexical information regarding the syntactic class of a word (i.e. is the word a noun, a verb, a determiner, an adjective, a preposition etc.). Once information about the syntactic class of the words in the input is available, syntactic rules specified in a grammar are used
to combine the words into legal phrase structures. A different view has it that	nonsyntactic information is accessed immediately, too (e.g., Altmann et al., 1992; Crain
& Steedman, 1985; Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1987). Thus, as
well as looking up information in the lexical entry of each word in the input regarding
syntactic category information, the basic semantic features might be looked up as well.

Experiments 2 and 3 test whether semantic features of nouns and semantic selection
restrictions of verbs are activated rapidly during sentence processing. If the processing
system is sensitive to such semantic information while it is still engaged in the analysis
of the sentence, it should be possible to demonstrate a disruption of processing. This
disruption is assumed to show up as a delay in performing a secondary task, i.e. the
task of making a lexical decision. In Experiment 4, the effects of semantic (and
pragmatic violations) are tested with a self-paced reading paradigm, and in Experiment
4b, with a self-paced reading paradigm with continuous grammaticality judgements.

Linguistic theories suggest that each word has three types of information stored in its
lexical entry: information specifying its syntactic category (noun, verb, adjective,
preposition etc.), its subcategorisation frame, and its thematic roles. One might simply
assume that a parser, which needs to have access to syntactic category information in
order to build up phrase structures, will automatically also look up all the other types
of linguistic information which is available in the same lexical entry. Weinberg (1987)
holds such an assumption:

"it would be extremely inefficient if the parser did not use such information
[the three types of lexical information specified above] to govern its
construction of well-formed trees, because in this case the parser would have to
construct a representation and then rescan it entirely using information that it
possessed when it constructed the representation in the first place" (Weinberg,
1987, p. 261)

Such an assumption based on efficiency considerations needs to be backed up by
experimental evidence showing whether syntactic and semantic lexical information do
interact in human sentence processing. As a first step Experiments 2, 3 and 4
investigated when semantic information is available in sentence processing.

Experiments 2 and 3 are similar to a study conducted in Germany by Friederici and her colleagues (Hahne, Pütz, Friederici & Rösler, 1991). As reviewed in Chapter 5.4. above, reaction times in a lexical decision task on the final word in a sentence were increased when the sentence contained selection restriction violations as compared to normal sentences. The present studies introduce some important changes designed to allow better interpretation of the effects of semantic selection restrictions. Thus, noun and verb in the non-violated condition will not be highly predictable; a wider range of semantic features, in addition to ANIMACY, will be included, and there will be a neutral baseline condition to check whether semantically correct sentences actually facilitate processing of the target word, and whether semantically violated sentences interfere with processing the target word. Neither Marslen-Wilson et al. (1988) nor Hahne et al. (1991) used a neutral condition.

The present experiments also test whether findings by Marslen-Wilson et al. (1988) for spoken language apply to written language processing as well (see Chapter 5.4. for a review).

6.2. Experiment 2

6.2.1. Introduction

Experiment 2 tests the assumption that when a noun is encountered in a sentential context, its semantic features are looked up immediately and kept active until a verb is encountered which subcategorises this noun. The experiments test the further assumption that upon encountering a verb, not only its subcategorised frame but also its semantic argument structure is looked up. If the semantic argument structure does not fit with the semantic category of the subcategorised noun, processing is expected to be disrupted. A further question that will be addressed by including a neutral baseline condition is whether the activated semantic features of the noun can facilitate access to
the verb when there is semantic fusion. The lexical decision task was chosen as the secondary task since it is sensitive to lexical access (given certain conditions), and to allow a better comparison with the Hahne et al. (1991) study which also used this task.

Thus, the predictions in detail are as follows. Firstly, RTs on the secondary task in the misfusion condition (violated fusion) are greater than those in the correct-fusion condition (positive fusion). This result would indicate that subjects are sensitive to aspects of word meaning as soon as they encounter a word, irrespective of the fact that in both the semantically violated and the semantically correct condition the target is syntactically correct. Secondly, it is predicted that positive fusion facilitates processing of the target verb as compared to processing the verb when no fusion occurs (in the neutral condition). In addition, the violated-fusion condition should yield longer RTs than the neutral condition since the activation of the 'wrong' semantic features (with respect to the verb's requirements) is assumed to be more costly than no pre-activation of semantic features at all.

6.2.2. Pretest

In designing the stimuli for the experimental trials, it was crucial to make sure that the sentences in the Positive-Fusion condition contained nouns that, while fitting the semantic requirements of verbs for their arguments, nevertheless were not highly associated with the following verb. For example, in a sentence like 'The coffee was brewed', the noun is of the semantic category required by the verb (i.e. Liquid), but it is also associated with the verb, as the results of my pretest showed. Given the sentence fragment 'The coffee was ... ' and asked to complete it with a past participle form of a verb, many subjects did answer with 'brewed'. If it were found that a lexical decision to 'brewed' in this sentence was faster than, say to 'brewed' in the sentence 'The writer was brewed', one could not unequivocally attribute this to Positive Fusion and rapid activation of semantic categories. It might have been the case that upon reading 'coffee', all strong associates were activated, and that because of this preactivation, 'brewed' is responded to faster than a non-associate.
As a measure of the association between nouns and verbs, a written Cloze task was run. Subjects were asked to continue sentence fragments which consisted only of the subject NP and an auxiliary (such as "The president was ..."). Subjects were asked to write down the first two verbs in the past participle form that came to their mind as a continuation of the sentence. 84 sentence fragments were presented in a different random order to each subject. There were two sets of sentence fragments, which were given to two different groups of 14 subjects each. The first set contained 57 fragments, and the average number of continuations per fragment was 12.4 (subjects often produced only one continuation). The second set contained 27 fragments, and the average number of responses per fragment was 23.4 (i.e. most subjects gave two responses per fragment). The instructions of the pretest are presented in Appendix G. No verb that was mentioned by any of the subjects in the pretest was included in the experiment in the Positive Condition in order to rule out that high predictability of the verb given the subject noun would lead to faster responses in the lexical decision.

6.2.3. Method

Design

Subjects and Words are treated as random factors. Thus, there are two subdesigns, one by subjects and one by items. In each, Fusion, a fixed factor with three levels (correct, violated, neutral), is a repeated measures factor, that is Subjects and Items are crossed with Fusion. Such a within-subjects and within-words design is more sensitive than the corresponding between-designs (i.e. where the treatment factor is a between-factor), since subjects and words serve as their own controls (see Clark, 1973). The levels of Fusion are illustrated in the following example:

<table>
<thead>
<tr>
<th></th>
<th>The dog was bitten.</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive</td>
<td></td>
</tr>
<tr>
<td>violated</td>
<td>The lake was bitten.</td>
</tr>
<tr>
<td>neutral</td>
<td>The next is bitten.</td>
</tr>
</tbody>
</table>

Three parallel lists of stimuli were formed by randomly assigning each member of a
sentence triplet to a different list (filler and nonword trials were identical in the three lists). List was a between-subjects factor, with subjects assigned at random to one of the lists. Thus, each subject saw only one sentence from each triplet.

The number of trials was the same for each of the three levels of the factor Fusion (positive, violated, neutral). Each subject saw 21 sentences at each level of Fusion, and each sentence of each triplet was seen by 24 subjects at each level of Fusion.

The dependent variables were the response times for making the lexical decision and the errors in making the lexical decision.

Task

The main task for the subject was to read short sentences presented word by word centrally on a microcomputer screen using the RSVP-paradigm. The secondary task was to make a lexical decision on the final word of each sentence, i.e. the target for the lexical decision task was the final word of a passive sentence, which was always a verb in its past participle form (e.g. 'bitten' in: 'The dog was bitten').

Apparatus

Presentation of stimuli and measuring and recording of responses was controlled by BBC microcomputers. Two response boxes were connected to each computer via the userport. Response keys were round buttons mounted on a square plastic box of 7 by 7 by 7 cm. The buttons were labelled with the capital letters Y or N.

Procedure

Subjects were tested individually in sound attenuated rooms in the Department of
Psychology at University College London. The subject was paid £1 at the beginning of the session. They were then seated in front of a BBC-microcomputer and were told that the instructions would be presented on the computer. Appendix D presents the instructions in full. The Experimenter stayed in the room with the subject for the duration of the practice trials. After the subject had read the instructions, the Experimenter answered the questions which a subject might have. The Experimenter repeated to each subject part of the written instruction, i.e. that they should read the sentences, and that the Yes- or No- decision had to be made (by pressing the Yes- or the No-key on the response box) regarding whether the last word was an English word or whether it was a made-up word (a nonword) whose spelling did not correspond to any English word. Subjects were instructed not to judge whether the sentence made sense, but merely whether the target (the final letter string of the sentence) was a word or not.

A BBC master computer controlled the display of instructions and trials, and recorded subject responses (their response time in milliseconds, the key pressed, and whether the decision was correct or false). The trials were presented onto a black screen, with the letters (all lower cases) appearing in white. Trials were presented one word at a time using the RSVP technique. A trial began with the presentation of a white cross (+) warning signal which was displayed for 500 ms in the centre of the screen. Each word was presented centralised on the VDU-screen for 100 ms, with a 100 ms blank screen after the presentation of each word. All words were presented in the same location on the screen. The target remained on the screen until a lexical decision response was made. RT was measured from the onset of the target. After the response, there was a pause of 2000 ms, and then the next trial began. No feedback was given to the subject. After regular intervals (i.e. after 71, 141 and 211 trials), a message appeared on the screen informing the subject that they could take a short break, and press any key on the keyboard when they were ready to continue. The experiment took about 20 minutes per subject, of which about 5 minutes were spent on the practice trials.
Stimuli

Each list consisted of 256 trials: half of those were experimental trials, and half were filler trials (in which adjectives instead of the main verbs were used). 50% of the test and filler trials had nonword targets. All filler trials and the nonword experimental trials were identical in the three lists. The different types of stimuli were intermixed rather than blocked.

The order of trials was determined in a random fashion, with the constraint that not more than 3 trials of the same type (word, nonword, test, filler) occurred in sequence. The order was the same in all three lists.

All targets appeared in sentence final position, and each sentence was four words long, eg. 'Her sausages were boiled'. The sentence structure remained constant, i.e. there was always a determiner (the, his, her), a noun, an auxiliary verb (50% was, 50% were) and a main verb in its past participle form.

A third of the nonword-target trials were designed as neutral 'sentences' (eg. 'the next is foaled'), and two thirds as 'sentences' (eg. 'the robins were acalised').

Each subject saw 21 experimental sentences at each level of the factor Fusion. In addition to these 63 experimental trials, there was one additional sentence, which was in the neutral condition in List 1, in the violated condition in List 2 and in the positive condition in List 3. This resulted from the attempt to create stimuli with the same number of semantic categories at each level of Fusion. This item was not included in the data analyses.

63 sentence triplets were created. In sentences of the misfusion condition, the semantic requirements of the verb did not fit with the semantic category of a preceding NP, coindexed with the verb's argument, as for example in 'The lake was bitten' (underlined word is the target). In the positive fusion condition the target was preceded by a well-fused NP ('The dog was bitten') and in the neutral condition the target was
not preceded by any subcategorised noun ('The next is bitten'). Appendix E gives a full list of the sentences in the positive fusion condition and the misfusion condition of each triplet.

Only unambiguous nouns were used in the sentences. They mapped into a range of semantic categories (Animal, Human, Plant, Body Part, Solid Inanimate Thing, Liquid Inanimate Thing, Abstract Thing, Food, Fruit, Vegetable, Building). Care was taken to ensure that the same number of semantic categories were represented in the violated, positive and neutral conditions, in the word and nonword conditions, and in the filler and experimental trials. Semantic selection restrictions were violated by misfusions of a variety of semantic categories. Appendix E lists the semantic categories of subcategorised nouns in the positive fusion condition and the misfusion condition.

All nouns were of medium to high frequency (Thorndike and Lorge, 1944). Word length varied from one to three syllables. Neither average length nor frequency differed for nouns in correct versus violated sentences, or in word- or nonword-trials. Word frequency and length of the verbs were not controlled for, since each verb occurred at each of the levels of Fusion.

In the following section I describe how the stimuli were created.

1. Experimental Sentences

A pool of verbs was generated which require the subcategorised object noun to be of a particular semantic category. Two sentences were created in the passive form. In one, the subcategorised object noun (which is the subject of the passive sentence) belonged to the semantic category required by the verb, and in the other it was of a different semantic category, leading to misfusion, eg. 'The cat was bitten' and 'The lake was bitten'. All targets had a very low Cloze value (see Pretest below). In table 6.1. I list the most basic semantic categories which the verbs used in Experiment 2 require of their subcategorised object nouns, plus a few exemplary verbs.
### Table 6.1.: Experiment 2: basic semantic categories required by semantic selection restrictions of verbs

<table>
<thead>
<tr>
<th>basic semantic category</th>
<th>example verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>solid, nonhuman</td>
<td>(to skin, to peel)</td>
</tr>
<tr>
<td>body part</td>
<td>(to amputate)</td>
</tr>
<tr>
<td>food</td>
<td>(to marinate, to boil)</td>
</tr>
<tr>
<td>solid (animate or inanimate)</td>
<td>(to bite, to cut)</td>
</tr>
<tr>
<td>plant</td>
<td>(to harvest, to germinate)</td>
</tr>
<tr>
<td>human</td>
<td>(to sack; to interview)</td>
</tr>
<tr>
<td>concrete</td>
<td>(to kiss; to perfume)</td>
</tr>
<tr>
<td>non rigid</td>
<td>(to massage)</td>
</tr>
<tr>
<td>abstract</td>
<td>(to solve)</td>
</tr>
<tr>
<td>animate, non-plant</td>
<td>(to provoke; to tease)</td>
</tr>
<tr>
<td>liquid</td>
<td>(to pour; to sip)</td>
</tr>
<tr>
<td>animate</td>
<td>(to kill; to hurt)</td>
</tr>
<tr>
<td>solid, inanimate</td>
<td>(to sharpen; to knit)</td>
</tr>
</tbody>
</table>

2. Nonword sentences

There were 64 nonword trials designed to be similar to the experimental trials (there were also 64 nonword trials as part of the set of filler items). Half of the nonwords were pseudohomophones (PS for short), which sounded like real words, e.g. 'annalised', 'confurmed', 'feered'. They were derived from real words by changing one or two letters (23 PS differed from real words by 1 letter, 9 by 2 letters) while retaining morphemes such as dis-, re-, mis-, con- and -ed intact. PS-trials were introduced to prevent subjects' complete reliance on a phonological route when reading the targets which would discourage semantic processing. PS trials guarantee that subjects use a lexical route, and thus increase the likelihood of an interaction between lexical processing and context. PS nonwords require a grapheme-phoneme check and undermine a purely phonological route, thus encouraging semantic processing.

11 of the 32 PS were a target in a sentence in which the word that they sounded like would have semantically fitted (eg. 'Her baby was choasen'), 12 PS were derived from a word that would have violated the semantic requirements of the verb (eg. 'The
cucumbers were postponed'), and 9 PS were presented in neutral sentence frames. See Appendix F for a full list of PS.

The other half of the nonwords were nonpseudohomophones (Non-PS for short), which are also given in Appendix F. They were derived from the PS by changing at least one letter, so that all of them differed from real words by at least 2 letters.

Either type of nonword was orthographically and morphologically normal. Thus, a lexical decision could not be made purely on phonological, orthographic or morphological grounds.

3. Filler Sentences

These were introduced in order to distract subjects from the purpose of the experiment. Filler trials differed from the experimental trials in that the target was always an adjective rather than a main verb. The make-up of the 128 filler trials was identical to that of the 128 test trials. There were 64 word trials, and 64 nonword trials. Of the word trials, a third were in positive sentences (eg. 'The mother was sad'), a third in violated sentences ('The factory was sour') and a third in neutral sentences.

4. Practice Sentences

Twenty-four practice sentences were used which consisted of the same proportion of test and filler trials at the different levels of the factor Fusion as the sentences of each list. Furthermore, the range of semantic categories present in the trials of the experiment proper was maintained in the practice sentences, too.
Subjects

In the main experiment, seventy-two subjects were tested. They were randomly assigned to one of the three lists. They were all either students or members of staff of the Department of Psychology, and were paid £1 for their participation.

6.2.4. Results

Filler trials were not analyzed at all, and nonword trials only as indicated. The dependent variables were response times to make lexical decisions and error scores. I will first describe the error analyses.

Error Analysis

Data from 72 subjects and 127 trials (63 word trials and 64 nonword trials) were included in this analysis. There were two types of errors: incorrect lexical decisions (pressing the no-key to a word item or the yes-key to a nonword item) and extremely long response times for correct decisions. Error analyses were carried out separately for the incorrect responses, and for the incorrect and extremely long responses taken together. Extremely long reaction times were defined as falling beyond the cut-off point of 1255 ms which was established after examining the frequency distribution of response times to word items in correct responses. Data above the cut-off clearly fell outside a superimposed normal curve and formed a peak at about 16000 ms, indicating that some different processes were involved in these responses compared to immediate lexical decisions.

Out of the total 9144 (72 x 127) responses, 15.8% were scored as errors: 10.5% were incorrect, and 5.3% were extremely long (the percentage of extremely long responses relative to the total number of correct responses is 5.9%). Table 6.2. shows the distribution of errors for word targets for each level of Fusion, and for nonwords, for
each list. Listed separately are incorrect responses, and incorrect and extreme responses taken together (combined error score).

Table 6.2.: Experiment 2: Error Rates (%) per Stimulus Type and List

<table>
<thead>
<tr>
<th>List 1</th>
<th>incorrect</th>
<th>neutral</th>
<th>violated</th>
<th>nonword</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct</td>
<td>8.7</td>
<td>6.2</td>
<td>15.5</td>
<td>8.9</td>
</tr>
<tr>
<td>combined</td>
<td>11.9</td>
<td>8.1</td>
<td>22.8</td>
<td>14.5</td>
</tr>
<tr>
<td>List 2</td>
<td>incorrect</td>
<td>neutral</td>
<td>violated</td>
<td>nonword</td>
</tr>
<tr>
<td>correct</td>
<td>7.9</td>
<td>17.3</td>
<td>13.9</td>
<td>11.7</td>
</tr>
<tr>
<td>combined</td>
<td>15.9</td>
<td>22</td>
<td>19.1</td>
<td>17.6</td>
</tr>
<tr>
<td>List 3</td>
<td>incorrect</td>
<td>neutral</td>
<td>violated</td>
<td>nonword</td>
</tr>
<tr>
<td>correct</td>
<td>8.1</td>
<td>11.5</td>
<td>13.7</td>
<td>8.5</td>
</tr>
<tr>
<td>combined</td>
<td>10.3</td>
<td>18.1</td>
<td>17.5</td>
<td>14.3</td>
</tr>
</tbody>
</table>

For each item and each subject arcsine transformed proportions of errors in each condition were calculated. These scores were entered into ANOVAS with items and subjects as random variables. Since only word trials were analyzed, I will refer to the incorrect error score as the score of false negatives (i.e. making a No-decision to a target that was in fact a real word). Of the 63 word triplets, 4 had received no false negative decisions at all; and 2 had a combined error score of 0. These items are not included in the by-items analyses of the false negative responses, and the combined error score, respectively. One further item was dropped from the by-items analysis of the combined error score because all 72 responses to this item had been either extremely long or incorrect. In the by-subjects analyses, there was one between factor (List) and one within factor (Fusion), whereas in the by-items analysis, there was only the within factor Fusion.

The ANOVAS on the combined error scores (false negative and correct but extremely long responses) showed a significant main effect for Fusion ($F_{1}[2,138]=13.8$, $p < .0001$; $F_{2}[2,118]=9.16$, $p < .0001$; Min$F^{*}[2,237]=5.51$, $p < .01$). For false negative
responses as the error score, there was also a significant main effect for Fusion 
($F_1[2,138]=13.3, p < .0001; F_2[2,116]=5.34, p = .006; MinF'[2,200]=3.81, p < .025$). In 
addition, in the by-subject analysis on the combined error score, there was a significant 
main effect of List ($F[2,69]=4.8, p < .01$) and a significant interaction List by Fusion 
($F[4,138]=11.78, p < .0001$), but on the error score as number of false negative 
responses, only the interaction List by Fusion, but not the main effect List, was 
significant ($F[4,138]=4.85, p < .001$).

Planned comparisons revealed that the difference between the violated and the positive 
Fusion condition was significant (for the combined error score: $F_1[1,69]=25.78, p < 
.0001; F_2[1,59]=20.67, p < .0001; MinF'[1,124]=11.47, p < .001$; for the error score 
defined as number of false negative responses: $F_1[1,69]=27.53, p < .0001; 
F_2[1,58]=9.99, p = .002; MinF'[1,97]=7.33, p < .01$). Additional paired t-tests showed 
that in the by-item analysis of the combined error score, there was no significant 
difference for the conditions positive versus neutral, but that the conditions violated 
and neutral did significantly differ [$t(59)=2.35, p=.022, 2$-tail]. For the false negative 
responses, the levels positive and neutral differed significantly [$t(58)=2.46, p=.017, 2$- 
tail], but the levels violated and neutral did not.

In the by-subject analysis, paired t-tests showed that both comparisons were 
significant, for both the combined and the false negative error scores (combined error 
score: positive versus neutral: $t(71)=2.26, p=.027, 2$-tail; neutral versus violated: 
$t(71)=2.21, p=.031, 2$-tail; false negative responses: positive versus neutral: $t(71)=2.47, 
p=.016, 2$-tail; neutral versus violated: $t(71)=2.24, p=.028, 2$-tail).

Independent t-tests indicated that the (unexpected) main effect of the between-factor 
List which was found for the combined error score (but not the false negative 
responses) in the by-subjects analysis is due to mean responses in List 2 differing 
significantly from List 1 and List 3 (List 2 versus List 1: $t(142)=3.33, p=.001, 2$-tail); 
List 2 versus List 3: $t(142)=2.74, p=.007, 2$-tail). List 1 and List 3 did not differ 
significantly.
The significant interaction List by Fusion resulted from the following pattern of differences. Paired t-tests on the combined error score revealed that the means for positive and violated differed significantly in List 1 and List 3 (all p's <.016, 2-tail), but not in List 2. The means for the positive and the neutral conditions differed significantly in Lists 2 and 3 (p's <.004, 2-tail), but not in List 1. Finally, the means for the violated and the neutral conditions differed significantly in List 1 (t(23)=8.15, p<.0005, 2-tail), but not in Lists 2 and 3.

The interaction List by Fusion for the proportion of false negative responses was also explored with paired t-tests. The predicted differences between positive and violated were significant in all lists (p's <.038, 2-tail). The means for the positive and the neutral conditions differed significantly in Lists 2 and 3 (p's <.049, 2-tail). The means for the violated and the neutral conditions differed significantly only in List 1 (t(23)=4.78, p<.0005, 2-tail).

Response Times

Data from ten subjects were dropped from the analysis because the subjects had an overall error rate of more than 20% on their 127 experimental trials (i.e. word and nonword trials). Five of the subjects had pressed the wrong button (i.e. 'no' for a real word or 'yes' for a nonword) in more than 20% of their responses. Three subjects produced more than 20% extremely long responses (greater than 1254 ms) out of their total number of correct responses. Only their responses to the word trials (but not the nonword trials) were considered here. Two more subjects were excluded because they had a combined error score of greater than 20%. I also checked whether any subject had a mean response time (for their correct responses below 1255 ms) of more than two standard deviations away from the group mean of the list they had been given. No subject had to be excluded on the basis of this criterion. Out of the remaining 62 subjects, 22 had been given list 1, 18 list 2 and 22 list 3.

The same exclusion criteria were also applied to items. Out of the 63 word trials, 17
had to be excluded. For nine of these, more than 20% of all responses were false negative, two had reached 20% on both criteria (combined error score and false negative responses), one item (the first item of each list) was excluded merely on the basis of having received more than 20% extremely long response times, and five had a combined error score of more than 20%. As was the case for all items, the excluded items, too, had received most errors when they were in the violated condition, less in the neutral and least in the positive condition. Three of the excluded items had been presented one or two trials after a break, and might have yielded better responses if a few buffer trials had been introduced immediately after a break, and at the start of the experiment. There was no order effect for the items that had to be excluded because of false or extremely long responses; however, four out of the five items which had too high a combined error score had occurred in the first third of the stimulus lists. Out of the 64 nonwords, fourteen had to be excluded. Twelve of these were pseudohomophones.

Reaction time analyses were carried out only for correct responses to experimental word trials, but not the nonword trials. Responses from 62 subjects and 46 items were included (for exclusion criteria see paragraph above). As was discussed at the beginning of the Results section, a cut-off was set at 1255 ms, and the few responses above that cut-off (i.e. 1.8% of the correct responses) were dropped. The raw response times were changed using the natural logarithm-transformation. Both the shape of the distributions and homogeneity of variances were thus improved.

Means and midmeans of the log-transformed response times were calculated for each subject and each item for each condition. These scores were entered into repeated-measurements ANOVAs (using the SPSS MANOVA programme) with List as a between factor (in the by-subject design) and Fusion as a within factor (in the by-subject and by-item designs), and either subjects or items as random variables. The midmean score is the mean of the middle 50% of a data set (i.e. the data between the second and third quartiles). The exclusion of the upper and lower 25% of the data protects against the influence of outlying values which can affect the arithmetic mean. Furthermore, the midmean does greater justice to the spread of the data than the
The mean RTs (based on subject means) for making a correct No-decision to nonword targets are presented in Table 6.3.1. (reported are the antilogs).

Table 6.3.1.: Experiment 2: Mean Reaction Times (ms) for Nonword Targets

<table>
<thead>
<tr>
<th>List</th>
<th>pseudohomophones</th>
<th>nonpseudohomophones</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>711</td>
<td>686</td>
</tr>
<tr>
<td>2</td>
<td>723</td>
<td>700</td>
</tr>
<tr>
<td>3</td>
<td>740</td>
<td>706</td>
</tr>
<tr>
<td></td>
<td>725</td>
<td>697</td>
</tr>
</tbody>
</table>

The difference between PS and NW was significant (by-subjects: paired t-tests: $t(61)=4.41$, $p=.0005$, 2-tail; by-items: independent t-tests: $t(51)=1.76$, $p=.085$, 2-tail).
The number of valid responses which made up a subject's mean and midmean at each level of Fusion varied among subjects, depending on how many responses had to be excluded due to a subject having produced a false negative, or an extremely long response. The number of trials per subject per cell ranged from 9 to 18, with a mean of 14 (the means were virtually identical for the three levels of fusion, i.e. 14.58, 14.71, and 14.00, for the positive, neutral and violated conditions, respectively).

In the by-items analysis the number of valid responses that were collapsed to create means and midmeans for each level of fusion varied per item from 11 to 22 with a mean of 19 (again, the means for the three levels of fusion were very similar, i.e. 19.65, 19.83, and 18.87 for the positive, neutral and violated conditions, respectively).

The mean response time for each condition, based on subject means, are given in Table 6.3. below (reported here are the antilogs).

Table 6.3.: Experiment 2: Mean Reaction Times (ms) in each condition

<table>
<thead>
<tr>
<th>List</th>
<th>Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>positive</td>
</tr>
<tr>
<td>1</td>
<td>567.36</td>
</tr>
<tr>
<td>2</td>
<td>578.82</td>
</tr>
<tr>
<td>3</td>
<td>591.70</td>
</tr>
<tr>
<td></td>
<td>578.82</td>
</tr>
</tbody>
</table>

133
Univariate ANOVAs with subject- and item-means indicated that there was a significant main effect of Fusion \( F_1(2,106)=18.96, p<.0005 \), degrees of freedom adjusted according to Greenhouse-Geisser epsilon; \( F_2(2,90)=14.14, p<.0005 \); \( \text{MinF'}(2,186)=8.1, p<.001 \). In the analysis of midmeans as the dependent variable, the nonparametric Friedman test was run on the by-subjects data because the assumptions of normality and homogeneity of variance were violated. This test indicated that the three levels of the factor Fusion differed significantly from each other \( \chi^2=28.000, 2\text{df}, p<.0005 \). Midmeans in the by-items design were entered into the repeated-measures analysis of variance, which indicated a main effect for Fusion \( F_2(2,90)=11.89, p<.0005 \). All multivariate results also showed a significant effect for
Fusion (p<.0001 for all analyses). In the by-subjects analysis based on means, the interaction List by Fusion was significant (again with Greenhouse-Geisser adjusted degrees of freedom: $F(4,106)=2.55$, p<.05), which was in line with only one of the four multivariate criteria used by SPSS (Hotelling's Trace: value=.17293, approximate $F(4,114)=2.46$, p=.049).

In the analyses based on means, planned comparisons indicated a significant difference between the conditions positive and neutral [$F(1,59)=16.20$, p<.0005; $F(2,145)=6.55$, p=.014; MinF'$(1,79)=4.66$, p<.05], and between the violated and the neutral conditions, which was, however, not significant by MinF' [$F(1,59)=7.52$, p=.008; $F(2,145)=7.62$, p=.008; MinF'$(1,102)=3.78$, p<.10]. The overall priming effect, i.e. the difference between the positive and the violated conditions was also significant [$F(1,59)=29.668$, p<.0005; $F(2,145)=27.89$, p<.0005; MinF'$(1,101)=14.38$, p<.001].

The same pattern of results held when midmeans were used as data (positive versus neutral: Wilcoxon Matched-pairs Signed-ranks Tests, by-subjects, across Lists: $Z=4.0931$, p<.0005, 2-tail; also significant in separate analyses per List for Lists 2 and 3 (p's <.0071, 2-tail; paired t-test, by-items: t$(45)=2.35$, p=.023, 2-tail; violated versus neutral: Wilcoxon Matched-pairs Signed-ranks Tests, by-subjects, across Lists: $Z=2.4256$, p=.0153, 2-tail; in separate analyses per List, this difference was only significant for List 1: $Z=2.6416$, p=.0083, 2-tail; paired t-test, by-items: t$(45)=2.80$, p=.007, 2-tail).

The interaction List by Fusion (significant in the analysis based on subject means) was further analyzed with paired t-tests, which indicated that the difference between the positive and the neutral conditions reached significance only for List 3 [t$(21)=4.15$, p<.0005, 2-tail]. The difference between the violated and the neutral conditions was significant for Lists 1 and 2 [t$(21)=3.41$, p=.003, 2-tail; and t$(17)=2.28$, p=.036, 2-tail, respectively].

All 46 items were analyzed individually in order to examine whether there was anything special about those items for which the mean response latencies exactly
followed the prediction of positive < neutral < violated (based on midmeans per item). For 14 out of the 46 items this prediction was true, for 3 the midmeans were ordered in exactly the opposite directions (positive > neutral > violated), and for 29 items parts of the prediction were not true. 12 items did not fulfil the prediction (positive < neutral < violated) because the midmean in the neutral condition was unusually high: while the midmean in the positive condition was smaller than the one in the violated condition (thus supporting the prediction), the midmean in the violated condition was smaller than in the neutral condition (positive < violated < neutral). For 8 items the midmeans were smaller in the positive condition than in the violated condition while the midmean for the neutral condition was smaller or equal to the one in the positive condition (neutral ≤ positive < violated). Items for which the prediction was true did not differ in orthographic regularity from the 32 items for which the prediction was not true. None of the items which did not, partly or fully, support the prediction, had been presented shortly after one of the breaks during stimulus presentation. I also checked the basic semantic categories into which the nouns mapped, and which were required by the verbs. The items which fitted and those which did not fit the prediction covered a range of semantic categories. There is no indication that the prediction is true only of a limited set of semantic categories.

6.2.5. Discussion

The results clearly support the conclusion that the status of fusion affects the responses in a lexical decision task, as was hypothesized. Both latency of response and number of errors increased from sentences with positive fusion, to neutral sentences with no fusion, to sentences in which fusion had been violated. The interaction of the factor Fusion with the between-subjects factor List (significant in the error analyses and the analyses based on means) had not been expected since the lists had been designed in parallel and had been allocated to subjects in a random fashion. I will not try to interpret this finding further since the alpha levels for the univariate and the multivariate tests were quite large (.05 and .49).
Error rates (10.5% incorrect responses, and 5.3% extremely long response times), and
the number of subjects and items that had to be excluded, appear to be quite high
compared to around 5% in lexical decision studies with word pairs rather than
sentences as stimuli. However, other studies using sentence material also yielded quite
a high percentage of errors, eg. up to 9% in Stanovich & West's (1983) study of
reading aloud words preceded by congruous and incongruous sentences. Unfortunately,
the study most similar to the present one (Hahne et al., 1991) does not report error
rates, nor does it present an analysis of the distribution of errors. In the present study,
a few of the items had to be excluded due to a mistake in stimulus presentation: there
were no buffer trials at the beginning of a list, or after the three breaks during the
experiment. These items had an excessive number of extremely long responses. Most
of the excluded items, however, did not follow a break. They show the pattern typical
of all items, i.e. a linear increase of the number of errors from positive to neutral to
violated sentences. In this sense they behave like the other, not-excluded items. The
fact that almost 14% (10 out of 72) subjects had to be excluded may be partly
explained by the fact that the experiment lasted for about forty minutes and might have
been quite strenuous. This, however, is not unusual for experiments of this kind.
Hahne et al. (1991), for example, in a similar study to this one, presented 540 trials, as
compared to 256 in the present study. Furthermore, in the present study subjects had
been given an extensive practice period after which they had been encouraged to
clarify any questions with the experimenter. They had also explicitly been told to make
use of the two breaks one-third and two-thirds through the stimuli, and to press the key
triggering the continuation of the experiment only after they had paused for a while.
However, most subjects were undergraduate students (about a third were academic
staff), who participated in the experiment as part of a tight schedule of lectures and
seminars. Furthermore, some background noise might have disturbed some subjects
even though they were seated in sound-attenuated cubicles. Another possible
explanation is, of course, that having to make lexical decisions on a sizeable number
of trials (256) is difficult and tiring, however ideal the testing conditions might be.
However, these considerations do not coincide with the fact that the number of errors
did not increase as the experiment progressed. In fact, the number of errors (combined
error score) on test trials in the final third of all trials was slightly lower than for the
second and the first third of items (401, 466 and 574, respectively). A mixed AOV was run on error scores for the 63 word items with a post-hoc between-items factor Order (with three levels: 1., 2., and 3. part of the stimulus list), and the within factor Fusion. There was no significant effect for Order, nor a significant interaction of Order with Fusion. In a one-way AOV on the error score of the 64 nonwords, Order was significant \( F(2,61)=8.3, p<.005 \). However, the number of errors did not increase as the experiment progressed (independent t-tests revealed that nonword items in the middle part of the experiment were responded to incorrectly significantly less often than nonword items in the first and third part of the stimulus list, \( p \)'s < .003). Another fact that suggests that the high error rate was not simply due to the experiment being too long, is that only three subjects had to be excluded solely because of long response times, but five solely on the basis of the number of incorrect responses (two more subjects were excluded because of their combined error score). Some subjects reported after the experiment that after some responses they realized they had made a mistake. They had been told in the instructions to quickly press the correct key when they thought they had made a mistake. The purpose of this manipulation was to ensure that subjects did not ruminate over mistakes. However, no record could be kept of this second key press, and it is unclear whether the number of mistakes would be reduced if the second decision was considered, and whether the status of Fusion might affect awareness of having made a mistake.

Apart from the number of errors, which was higher than expected, their distribution across the levels of Fusion was as predicted. For both error measures, in the by-subjects and the by-items analyses, the number of errors steadily increased from the positive to the neutral to the violated condition. This parallels the findings based on response times.

The main result of Experiment 2 is the overall priming effect in lexical decisions if the preceding sentence provides semantic features required by the target word. Performance is faster, and less error-ridden, in sentences with semantic fusion of semantic features of the noun and semantic selection restrictions of the verb as compared to sentences in which semantic fusion is violated. This result is in line with
previous studies of violations of selection restrictions using the lexical decision task
with written language material (Hahne et al., 1991), and a word detection task with
spoken language material (Marslen-Wilson et al., 1988). The results go beyond the
Hahne et al. study, in that predictability of targets was controlled for, and a wider
range of violated semantic features was used. Furthermore, Experiment 2 showed that,
at least for the subjects and items included in the analyses, both the positive and the
violated conditions differed from the neutral condition. Since neither of the two
previous related studies (Hahne et al.’s and Marslen-Wilson et al.’s) included a neutral
condition, no comparison can be made in this respect. However, some of the classic
sentence priming studies (Schuberth & Eimas, 1977; Fischler & Bloom, 1979) also
found facilitation and, under certain conditions, inhibition as well. Before discussing
the differences of the positive and the violated condition from the neutral condition in
the light of the various models that have been suggested to account for the effects of
sentence context on visual word recognition, I will briefly describe Experiment 3
which was designed to replicate the present findings, and to test if the context effect
can be increased by degrading the visual quality of the target word.

6.3. Experiment 3: A follow-up: visual degradation and violations of semantic
selection restrictions

6.3.1. Introduction

Experiment 3 was conducted in an attempt to replicate the context effect found in
Experiment 2. Given that ten subjects and eleven items had to be excluded, the
reliability of the effect could be increased if it could be replicated with different
subjects and different experimental conditions. It was decided to visually degrade the
targets to test whether this would increase the priming effect found in Experiment 2.
Other work has shown that the overall priming effect increases if sensory information
decreases in quality, due to degradation slowing RTs to a greater extent in the
unrelated than in the related condition. This was shown for both words and sentences
as primes, and for naming and lexical decision. (Stanovich & West (1983) used
sentences as primes in a naming task; Becker & Killion (1977) used single associated words as primes in the LDT and the naming task; Schuberth, Spoehr, & Lane (1981, experiment 2) used sentences as primes, but the interaction of relatedness and degradation which they found in the lexical decision task did not reach significance either in the by-subjects analysis with minF').

A more detailed comparison of facilitation and inhibition effects under target degradation (rather than focusing simply on the overall priming effect) has only been done infrequently. The findings are mixed. In studies using the naming task and sentences as primes, reported in Stanovich and West (1983, experiments 6 to 8), the interaction between target quality and context was a reliable effect. Both facilitation and inhibition were increased for degraded as compared to nondegraded targets in their experiment 6. More precisely, in the nondegraded condition, facilitation was significant, and it was increased in the degraded condition. Inhibition was not significant in the nondegraded condition, but was significant in the degraded condition. This effect only held when a block of degraded trials preceded a block of nondegraded trials, and not when the order of blocks was reversed. In the latter case the degradation main effect was very small (37 ms in the neutral condition), and the overall context effect did not differ for the degraded as compared to the nondegraded condition.

However, in Stanovich & West’s study, increased facilitation and inhibition only obtained when degradation was created by reducing the intensity of the target. When the target was degraded by inserting asterisks between each letter of the target word, the interaction between context and target quality was due to facilitation being significantly increased in the degraded as compared to the nondegraded condition, whereas inhibition did not increase (inhibition was not significant in any of the conditions). In this experiment, degraded and nondegraded trials were mixed randomly. In order to establish whether it was the change in the way targets had been degraded, or the mixed presentation of trials, Stanovich & West conducted another experiment (their experiment 8), again with degradation by insertion of asterisks, but blocked presentation of degraded and nondegraded trials as in experiment 6. The interaction of context and target quality was again due to facilitation but not inhibition being
increased in the degraded condition (this was true regardless of the order of blocks). In none of the conditions were inhibition effects significant.

In Stanovich and West’s model (1983) facilitation is explained by automatic spreading activation, and inhibition by attentional, conscious expectancy processes which are assumed to be slower acting. The effects of degradation are discussed in the same terms. Degradation decreases the quality of visual information and slows down lexical access. Increased facilitation is explained by the extra time during which more activation can spread (automatically) from the prime to related words, lowering the thresholds of logogens of related words. Not merely degradation, but other factors such as word difficulty and reading fluency are assumed to be capable of slowing down lexical access in Stanovich & West’s (1983) model. The result is always an automatic compensatory increase of contextual influence via spreading activation from words in the context to related logogens.

The increase of inhibition, on the other hand, was originally (Stanovich & West, 1979) explained by the additional time available to the slow-acting expectancy strategy under target degradation conditions. However, the finding that the type of degradation is important led to a qualification of the interpretation of the inhibition effect. Having extra processing time did not seem sufficient to induce the use of the conscious expectancy strategy, since despite the fact that inserted asterisks did slow down RTs overall, they did not produce inhibition. Stanovich & West (1983) tentatively argue that asterisks do not sufficiently disrupt the uptake of visual information from the target (asterisks leave the shape of individual letters intact). It is only when subjects "are in doubt as to whether purely data-driven mechanisms will be sufficient to specify the stimulus" (Stanovich & West, 1983, p. 19) that they will employ a conscious expectancy strategy, which in turn is assumed to produce inhibition.

The type of target degradation was also found to be important by Durgunoglu (1988, quoted in Neely, 1991). In a naming task with a single word as prime, degradation by means of a mask, which was presented briefly before and after target presentation, increased inhibition (in contrast to Stanovich & West (1983), facilitation was not
affected). When degradation was achieved by inserting asterisks between the letters of the target, inhibition was unchanged, but facilitation increased. Thus, as in Stanovich and West (1983), inhibition was only enhanced by degradation when the letters of the target were degraded either by lowering the contrast (as in Stanovich & West, 1983) or by masking (as in Durgunoglu, 1988, as quoted by Neely, 1991).

6.3.2. Pilot study

In an attempt to optimize the effect of visual degradation, a pilot study was run. Bearing in mind the findings regarding the type of visual degradation, care was taken to ensure that the quality of individual letters was affected. Thus, degradation was created by superimposing over the target a random pattern of white moving dots, which created a 'snowy' effect. The amount of occlusion through the dots was about 40%. The stimulus set was the same as in Experiment 2, but it was reduced by taking out all filler trials (sentences with adjectives). There were 128 trials (plus ten start-up trials), 50% of which had nonword targets and 50% word targets. Again, 30% of the word trials were in the positive fusion condition, 30% in the violated fusion condition and 30% in the neutral condition. Since three parallel lists were constructed, each subject saw a particular target word only once.

Results

Results were obtained from 9 subjects. The overall mean RT in the correct word trials was 894 ms (range=384 to 6464). After a cut-off was set at 1433 ms, the mean RT was 835 ms, which is about 200 ms longer than in Experiment 2. The overall error rate, however, was not higher than in Experiment 2: 7.6% out of all trials had an incorrect response (8.1% out of all word trials).

Mean and median RTs based on individual responses rather than subject- or itemmeans are presented below:
<table>
<thead>
<tr>
<th></th>
<th>pos</th>
<th>neutr</th>
<th>viol</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>764</td>
<td>809</td>
<td>807</td>
</tr>
<tr>
<td>median</td>
<td>732</td>
<td>772</td>
<td>762</td>
</tr>
</tbody>
</table>

Thus, compared to E2, the difference between the positive and the neutral condition was increased (from 20 ms in E2, to 45 ms in E3), but the neutral and violated conditions now did not differ any more. Comparing the median RTs of each individual subject at each level of Fusion, there were only two subjects for whom the predicted sequence (pos < neutr < viol) was true. For three more subjects, the median for positive was smaller than that for violated.

It might be that the mask was too effective, requiring subjects to check the visual input very closely letter by letter. Note that half of the nonwords, i.e. 32, were pseudohomophones. Most of them, 23, differed from real words by only one letter (the other 9 differed by two letters from real words). Even though all 32 nonpseudohomophonic nonwords differed from real words by at least two letters, and thus seemed to be sufficiently distinct from real words, subjects might have adopted a very close spelling checking procedure. One subject reported that he checked every letter of the target before finally making a response. Such a strategy might override the effect of context: if there is detailed visual checking, context cannot exert much of an impact. The finding that the violated condition did not lead to longer responses than the neutral condition might be due to a close visual scanning strategy: if there is no facilitating context, i.e. either a neutral or a violated context, subjects might entirely rely on a close visual scanning strategy, and make a lexical decision based on the output of the visual processor; if that output identifies a target as a word, the fact that it might not fit the context will not be able to overrule the results of the visual analysis.
6.3.3. Method

Design, task, procedure and apparatus were the same as in Experiment 2. The only change consisted in presenting the target in visual noise.

Experiment 3 was run with slightly less masking than had been used in the pilot study. The target was again presented in visual noise (superimposed random dot pattern creating snowy effect), but this time the occlusion was around 30%. There were 42 subjects and 118 trials (10 buffer trials, 54 word trials and 54 nonword trials). A subset of the sentences of Experiment 2 were used, again representing the three levels of the factor Fusion (positive, neutral and violated). Three parallel stimulus lists were created, using a Latin Square, to ensure that each subject saw only one member of each sentence triplet.

6.3.4. Results

The overall error rate (false positives and false negatives) was 10%. The percentage of false negatives on the word trials per condition is listed in table 6.4. below:

<table>
<thead>
<tr>
<th></th>
<th>positive</th>
<th>neutral</th>
<th>violated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.7%</td>
<td>8.7%</td>
<td>11.9%</td>
</tr>
</tbody>
</table>

13.3% of the pseudohomophones and 8.4% of the nonpseudohomophones had false positive responses.

Eleven subjects and eleven items were excluded because more than 20% of the responses (per subject, and per item) were false decisions, or extremely long correct decisions (this was calculated for word trials only). A further three items had to be
dropped because of an error during stimulus presentation. Thus, 11 subjects and 40 items were included in the following analyses. The cut-off for extremely long correct word decisions was set at 1468 ms (i.e. two standard deviations above the mean for all correct word decisions, collapsing over Fusion; one response fell below the lower cut-off of 260 ms). The percentages of extremely long correct decision times to word targets is given for each condition in table 6.5. below:

Table 6.5.: Experiment 3: Percentages of extremely long correct decision times per condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>Positive</th>
<th>Neutral</th>
<th>Violated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.6%</td>
<td>4.7%</td>
<td>4.6%</td>
</tr>
</tbody>
</table>

The analyses reported below include only those subjects and items which did not have to be excluded, and only times for correct decisions to word trials which fell below the cut-off. Because the number of included subjects differed per List, and the number of included items differed per Itemgroup, data were collapsed. Thus, in both the by-subjects and the by-items design there was one repeated measures factor, Fusion, with three levels (positive, neutral, violated). Subject-and item means for each condition were calculated for the logtransformed data and entered into repeated-measures analyses of variance.

The mean RTs for subjects, based on antilogs, are reported below.

Table 6.6.: Experiment 3: Mean response times for correct decisions on word trials per condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>Positive</th>
<th>Neutral</th>
<th>Violated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>735</td>
<td>762</td>
<td>786</td>
</tr>
</tbody>
</table>

There was a significant main effect for Fusion in both the by-subjects and the by-items...
analysis [F1(2,60)=8.60, p=.001; F2(2,78)=8.11, p=.006; MinF'(2,137)=4.17, p<.025].

Planned comparisons indicated that the difference between the positive and the neutral conditions was significant by-subjects and by-items, but not significant according to MinF' [F1(1,30)=4.33, p=.046; F2(1,39)=8.55, p=.006; MinF'(1,60)=2.87, p<.10]. The difference between the violated and the neutral conditions was significant only in the by-subjects analysis [F1(1,30)=4.78, p=.037], but not in the by-items analysis. The overall priming effect, i.e. the difference between the positive and the violated conditions was also significant [F1(1,30)=15.85, p<.0005; F2(1,39)=13.76, p=.001; MinF'(1,69)=7.37, p<.01].

6.3.5. Discussion

Experiment 3 replicated the results of Experiment 2, except that in this experiment, the difference between the violated and the neutral condition was less reliable (significant only in the by-subjects analysis). Visual degradation in this experiment increased the overall correct decision times to word trials by about 200 ms. The number of false negative decisions (i.e. making a No-decision to a target word) is comparable in the two experiments. However, if number of errors and extremely long responses are taken together as a combined error score, more subjects had to be excluded in Experiment 3 than Experiment 2 (26% as compared to 14%). The finding that subjects did not make more false negative decisions in Experiment 3, where the target was visually degraded, indicates that they adopted a very close visual checking strategy. This might also account for the finding that the violated condition differed from the neutral condition only in the by-subjects analysis: as discussed above, violations might be disregarded if close visual analysis led to the identification of a word.

6.4. Discussion of Experiments 2 and 3

In the following discussion I will first briefly sum up my assumptions regarding the representation of different types of information affecting sentence priming, and the
kinds of processing systems involved. I will then discuss existing models of sentence priming vis-a-vis the findings of Experiments 2 and 3, and will finally present my own model to account for the pattern of results found in Experiment 2 and 3.

6.4.1. Assumptions about types of information and processing systems involved in sentence priming

I assume that two processes operate in parallel, the construction of a syntactic and of a semantic representation of the input sentence. An initial visual input analysis leads to the activation of a word node in the lexicon. I assume that each word in the lexicon has attached to it a small set of item-specific information indicating its syntactic category and its basic semantic category. This information becomes activated as soon as the word node is activated via the input analysis. The activated word and the attached information are passed on to the subsystems of the Language Recognition Module (LRM) as symbol strings. The attached semantic information, passed to the semantic subsystem, is assumed to be represented as the function - argument structure suggested by Jackendoff in his competence theory (see Chapter 5.3. and 4.1.).

Semantic fusion was defined as coindexing a verb argument with a noun whose semantic category matches the one required by the argument. The finding that the status of semantic fusion of a sentence affects lexical decision times indicates that subjects analyze the sentence syntactically and semantically despite the fact that the task they were asked to perform, i.e. making a lexical decision on the sentence-final letter string, is independent of sentence context. Also note that exactly the same sentence structures were used for sentences ending in real words and nonwords. Thus, sentence structure did not provide any clues about whether a word or a nonword should be expected in sentence-final position. It seems that subjects cannot avoid undertaking an immediate syntactic and semantic analysis of sentences, even if this is irrelevant to the task they have to perform. Syntactic and semantic processing, at least of these simple sentences, seems to be fast and automatic, and not subject to strategic decisions. It is clear, however, that subjects did not 'forget' their real task to make lexical decisions and instead judged the sense of sentences. They made correct lexical
decisions even if targets were embedded in violated sentences. Furthermore, the time to make an incorrect response was not faster for violated as compared to positive or neutral sentences, which one might have expected if the subject used information regarding violated fusion to make a response in the lexical decision task.

Diagram 6.1. below relates the structures and processes assumed to be necessary in making lexical decisions in sentence contexts.
The lexical system is assumed to consist of a set of words, each with word-specific information regarding the word's visual representation, plus the syntactic category of the word (whether it is a noun, an adjective, a verb etc.) and its basic semantic category, as described above. For present purposes it is not relevant to spell out the relationship between words in the lexical system. From the results of Experiment 1 it is clear that words are not organised according to their basic semantic category, nor is there any reason to assume that they might be organised according to membership in a particular syntactic category (there is no evidence of priming merely on the basis of shared syntactic category). Before discussing the findings of Experiments 2 and 3 with respect to the subsystems depicted in diagram 6.1., I will show how various models proposed in the literature handle sentence priming effects.

6.4.2. Review of models to account for sentence priming effects

The most common interpretation of priming effects (in particular, of facilitation effects) is couched in terms of the spreading activation model originally formulated by Collins & Loftus (1975) (see Neely, 1977; Stanovich & West, 1983). However, this account cannot explain the priming effects in Experiments 2 and 3, since the words constituting the priming context were designed not to be associated with the target, and the Cloze value of the targets was close to zero (see Pretest results). The 'old' studies of sentence priming mentioned above all had used targets that were either highly related to words in the sentence, and/or were highly predictable from the context, as evidenced by Cloze pretests. Interestingly, in the study in which associations between words in the sentence and the target were avoided, facilitation was only found for targets with the highest cloze value of 90%, while inhibition was significant regardless of the predictability of the target (Fischler & Bloom, 1979).

In the spreading activation framework the interpretation of priming effects runs as follows. Facilitation is attributed to activation automatically spreading from the prime to related word nodes, thus preactivating these nodes and lowering their threshold (Neely, 1977; Posner & Snyder, 1975; Stanovich & West, 1979). The automatic nature
of facilitation was demonstrated by Fischler (1977b) for lexical decisions with words as primes. He showed that facilitation (here defined as the reduction in RTs for associated as compared to unassociated pairs) did not increase when previous trials in the experiment contained associated pairs. If stimuli were presented one at a time, and subjects had to make a lexical decision on every stimulus, the explicit instruction that sometimes two adjacent words might be related and that subjects could take advantage of this in making their response, did not lead to greater facilitation as compared to the condition where no such instructions were given. Neely (1977) showed that facilitation in a lexical decision task (with words as primes) occurs at an SOA of 250 ms (even when subjects were instructed to expect a target from a different semantic category than the prime, and even if the proportion of related trials is low), whereas the inhibition effect disappears at this short SOA.

Facilitative context effects are not seen to threaten the lexical autonomy hypothesis since they are based on relations between words inside the lexicon (see Chapter 3.2.1.). Facilitation effects have been obtained with naming and lexical decision tasks for targets that were either semantically or associatively related (eg. Stanovich & West, 1983; Seidenberg, Waters, Sanders & Langer, 1984).

Inhibition is typically only found in lexical decision tasks, but not in naming tasks (but see Becker, below, for different interpretation). In the spreading activation framework, inhibition found in lexical decision tasks is generally interpreted to reflect post-access decision processes, rather than lexical access (Forster, 1979; West & Stanovich, 1982; Stanovich & West, 1983). There is no mechanism, within this framework, for accommodating inhibition as an automatic process affecting lexical access. Seidenberg & Tanenhaus (1986, p. 146) suggest that the cohort model has provided a model for explaining inhibition of lexical access in spoken word recognition. An account along these lines is ruled out for written word recognition. The main argument is the finding that inhibition is typically not found in naming tasks (reading out aloud the target), even when the same set of stimuli is used which produced inhibition in a lexical decision task (Stanovich & West, 1983, experiments 9, 10 and 11; West & Stanovich, 1982). The crucial difference between the two tasks is generally held to be the
postlexical decision component lacking in naming (Stanovich & West, 1983; West & Stanovich, 1982). Several results have been adduced to support the view that inhibition is due to conscious strategic processes. By increasing the SOA in a naming task to several hundred milliseconds, thus encouraging conscious expectancy strategies, an inhibition effect could be created (Stanovich & West, 1979, quoted in Seidenberg & Tanenhaus, 1986, p. 146). Furthermore, lexical decisions take longer than naming the target or making rhyme and category judgements, indicating the operation of slower conscious processes (Forster & Chambers, 1973; Frederiksen & Kroll, 1976, quoted in Seidenberg & Tanenhaus, 1986, p. 147).

Strategies are assumed to affect processes occurring after lexical access has been achieved, at the stage of determining a response. Stanovich & West (1983) suggest that lexical, syntactic and message-level processors send information to a decision-making general problem solver which determines a response in a particular task. In a lexical decision task, a yes- or no-response has to be given. The authors claim that for an unrelated prime, this response is incompatible with information from the message-level processor which might bias the decision-maker to respond 'no'. The decision-maker has to overcome this bias in order to make a correct yes-response in the lexical decision task, which causes an increase in response time (at least for category names as primes and category instances as targets, but not for associatively related prime-target pairs, see discussion of Becker’s model). If SOA’s are shorter than 250 ms, inhibition is diminished, which is taken to indicate that the biasing information from the message-level processor has not had time to affect the decision-maker (Neely, 1977, 1991). In a naming task, no yes/no decision needs to be made, and thus negative information from the message-level processor given an unrelated prime, can not interfere with the required response (similar arguments were advanced by de Groot, Thomassen & Hudson, 1982; Tanenhaus, Flanigan & Seidenberg, 1980; Seidenberg & Tanenhaus, 1986). Stanovich & West (1983) present some studies which indicate that responses in the naming paradigm seem to be immune from strategic influences, even when factors known to induce strategic processing are manipulated. Stanovich & West (1983, experiments 3 to 5, using sentences as primes) blocked related and unrelated trials. They required the subject to make a judgement regarding whether the sentence
made sense after the naming response was made, and they increased the proportion of congruous trials, without inducing inhibition in a naming task.

Stanovich & West (1983) maintain that sentences as primes lead to facilitation in the same way as words as primes, i.e. via spreading activation from words in the sentence to their associates and semantically-related words. It seems that Fischler & Bloom's (1979) study, mentioned above, might be problematic for the spreading activation account, since they found facilitation (and inhibition) even though their targets were not associated with words in the priming sentence. However, the effect only obtained for targets with a Cloze value of 90%. In support of their account that sentences create priming effects via activation spreading from words in the sentence to their associates, Stanovich & West (1983) stress that spreading activation has been shown not to be diminished by words intervening between prime and target, and to persist sufficiently long to be of use in sentences as primes.

The spreading activation account of facilitation does not seem to be able to account for the findings of Experiment 2 and 3 regarding the difference between the positive-fusion and the neutral condition. Targets in the positive condition were designed to be of low Cloze value, and not to be associatively linked to any words in the sentence. Remember that Fischler & Bloom (1979) found facilitation only for targets that were highly predictable in the sentence. In my experiments, the semantic relationship between noun and verb was based on the basic semantic category which had to be shared to yield a well-fused grammatical sentence. Is it possible that the facilitation effect in Experiment 2 resulted from activation spreading across semantic (rather than associative) links between nouns and all verbs with which they could semantically fuse if they were subcategorised by them? There are several reasons why this is a very unlikely interpretation. Firstly, Experiment 1 had found no evidence for semantic priming solely on the basis of abstract semantic features, ruling out that nouns are linked in the lexicon when they share these semantic features. It is somewhat unlikely, then, that these semantic features should be the basis of links between nouns and verbs. A more serious problem for the account of sentence priming as activation spreading from some words in the sentence, has to do with the neglect of the syntactic
Note, however, that in many sentences the analysis of the noun as the subject of the sentence would also have resulted in semantic misfusion. To ascertain whether syntactic structure has been taken into account further experiments are required which explicitly manipulate the need for syntactic analysis.
relationship between words. The syntactic relationship, however, is crucial in the sentences of Experiments 2 and 3 to determine whether the noun is the subject or the object of the verb, and hence the target does or does not semantically fit the sentence. If the syntactic structure of the sentence is not taken into account, the word 'duke', for example, should prime the lexical decision on 'sipped', given that activation spreads along semantic links (the verb 'sip' requires an animate subject noun, and 'duke' has the feature 'animate'), which is unlikely, but not impossible. In the actual sentence, however, 'duke' was the underlying object and not the subject ('The duke was sipped'), and consequently led to inhibition rather than facilitation. This implies that syntactic structure was taken into account. In sum, the spreading activation account of facilitation cannot explain the findings of Experiments 2 and 3: firstly, claiming that activation spreads from words in the sentence to associatively related words fails because my targets had a low Cloze value, and were not associated to any words in the sentence; secondly, invoking semantic links between nouns and verbs fails since the assumed neglect of the syntactic structure of the sentence would incorrectly predict facilitation for my semantically violated sentences.

The spreading activation account of inhibition (and increased inhibition under conditions of stimulus degradation) as the result of having expected a particular target is also unlikely to account for my results. While it might explain the inhibition effect found in Experiment 2 (the difference between the violated and the neutral condition), such an account would have to also incorrectly predict inhibition in the positive fusion condition. Targets in this condition were of low Cloze value, and subjects should have expected a different, highly associated word, which would have led to inhibition when the target was encountered.

Another model, proposing the generation of an expectancy set, also seems to be inadequate to explain the findings of Experiment 2 and 3. This model accounts for priming effects by speeding access to the target word. Becker (1980, 1985) suggested that subjects use the prime to generate an expectancy set, which contains words related to the prime. Then the visual codes of these potential targets are matched to visual information about the actual target stored in iconic memory. If this process fails to find a match, a set of visually similar candidates is compared to information in iconic
memory until a match is found. If there is a neutral prime, or no prime at all, recognition of the target is based entirely on visual information. A set of candidates visually similar to the target is created, which is verified by checking it against visual information in iconic memory. While context in Becker's experiments is typically a prime word rather than a sentence, he explicitly used his model to account for other researchers' findings of priming with short sentences as priming context as well (he discussed Fischler & Bloom, 1979; Schuberth & Eimas, 1977). If sentences rather than a single prime word serve as context, subjects generate a semantic set on the basis of the priming sentence. The size of the generated semantic set is assumed to depend on the strategy used by the subject: Becker proposes a specific prediction strategy, and a general expectancy strategy. The choice of strategy was shown to depend on the kinds of prime-target relations prevalent in the stimulus list. Category names as primes, paired with category members as targets, lead to a fairly large set of expected related targets when they are embedded in a list of stimuli of only such types of pairs. If, however, they are part of a list of pairs which typically allow the prediction of only one, or very few, targets, such as strongly associated prime-target pairs, or antonym pairs (which, given Becker's examples of 'hot - cold', 'smart - dumb', 'dry - wet' also seem to be associated), these same pairs lead to only a small set of predicted targets (Becker, 1980). If semantic set size is relatively small, there will be facilitation dominance: related primes lead to faster recognition compared to a neutral condition, in which the prime does not generate any semantic set; interference from unrelated primes is small because there are only few items in the semantic set which have to be checked. If the semantic set is large, there is interference dominance.

The evaluation of Becker's account vis-a-vis the results of my Experiments 2 and 3 requires a more detailed discussion of the process of generating the semantic set. Firstly, Becker assumes that this process occurs before the set of visually similar candidates has been verified against visual information of the input stored in iconic memory. Generation of the semantic set occurs as soon as the prime word has been recognised. The prime word is analyzed in terms of its semantic features, which are fed into an array of word-detector units. These units count the semantic features that were fed to them and, if their criterion for firing is exceeded, the word corresponding
to a word-detector unit is included in the semantic set. Thus, no input from visual analysis is necessary for a word-detector unit to fire. The next step suggested by Becker is a verification process. The visual representation of each word unit in the semantic set is compared against the visual information of the target word in sensory memory in a serial, randomly ordered, exhaustive way. If there is a match, the word is recognised. This model can account for a pattern of data which shows that, as facilitation increases, interference decreases, and vice versa. Since the semantic set is created as soon as a prime has been recognised, it is available before the target word is presented. Checking the visual representation of candidates in the semantic set against visual information in sensory memory starts before the set of visually-related words is verified against the visual information in sensory memory. In comparison, in a neutral condition where the prime does not lead to the generation of any semantic set, targets can only be recognised by verifying the visually-defined set. Interference effects are accounted for by the further assumption that all candidates in the semantic set have to be checked against visual information in sensory memory before items in the visually-defined set are checked. Compared to a neutral condition, in which no semantic set is created, time is 'wasted' for unrelated targets by checking candidates in the semantic set against the input before verification of the candidates in the visually-defined set can start.

Several issues, I think, are left underspecified by Becker. Firstly, it is not clear whether the generation of a semantic set is a conscious or an automatic process. On the one hand, Becker strongly implies that conscious strategies are involved, with the subject "choosing" (Becker, 1980, p. 509) between two types of semantic process. He refers to the semantic processes as specific prediction strategy and general expectancy strategy (Becker, 1980, p. 509). His basic assumption is that subjects make "predictions about the identity of possible related target words" (Becker, 1980, p. 494). In each of the five experiments reported in 1980, subjects were explicitly told that prime and word targets were likely to be meaningfully related, and they were instructed that upon reading the prime they should try to get its meaning. Obviously, this does not prove that subjects did consciously try to create sets of related words after reading the prime. What these instructions do show, however, is that Becker assumes that explicit instructions can
influence semantic set generation.

On the other hand, Becker also maintains, however, that the mechanism creating the semantic set, i.e. the extraction of semantic features, might equally well be handled by an automatic spreading activation mechanism (Becker, 1980, p. 507). Presumably, such a spreading-activation account entails that words related to the prime get activated when the prime is recognised. However, this would imply that spreading activation is subject to strategic control by the subject: Becker has shown that the number of related words in the semantic set (i.e. the words to which activation spreads) depends on the closeness of semantic relations of the majority of items in a list. In Posner & Snyder’s (1975) theory, expectancy and spreading activation are two different kinds of processes, with only the first, but not the latter, under subjects’ strategic control, and requiring subjects’ attention or awareness. Neely (1991, p. 306) makes a similar point when saying "an ‘unconscious’ expectancy [...] seems to contradict most of the defining features of what an expectancy is".

How could Becker’s model be used to interpret the findings of Experiments 2 and 3? The finding that response times for making a lexical decision on targets in sentences with positive fusion were faster than on targets in neutral sentences, would have to be attributed as indicating that the targets in the positive fusion condition were included in the semantically defined expectancy set. However, it seems unlikely that targets following 'positive' sentences (e.g. 'the student was murdered') would be in this expectancy set. Targets in these sentences had been designed not to be predictable from the preceding sentence. They were not associated to the noun in the priming sentence, nor were they highly likely continuations of the sentence (as possibly 'the student was examined' would have been), as was established by the pretest. All words that had been produced by subjects in the Cloze pretest were excluded from the main experiment. Previous work has shown that predictability would have to be as high as 90% to enable facilitation in a lexical decision task (Fischler & Bloom, 1979).

Thus targets in violated as well as positive sentences would have to be identified by a search through the set of visually similar words. Since this search is assumed to be
determined by word frequency of the candidates in the visually defined set, there should have been no difference in the time it takes for positive and violated targets, since they were matched for frequency.

Furthermore, in contrast to Becker's experiments, subjects in my experiments were not encouraged to build up a semantic representation of sentences and targets.

Another, strongly post-access, account of semantic priming for word pairs is based on the assumption that prime and target together form a cue which is assessed in terms of its episodic familiarity (Ratcliff & McKoon, 1988). However, since it is assumed that familiarity is determined by the associative strength of prime and target, it cannot account for priming data for non-associated stimuli, such as the priming sentences and targets in Experiments 2 and 3.

Thus, it seems likely that the results of Experiments 2 and 3 have to be explained by some other mechanism that operates after the target word has been (at least partly) activated in the lexical system without invoking associative relatedness.

Neely & Keefe's (1989) account of semantic priming for word pairs is based on the assumption that once the target has been accessed, retrospective semantic matching occurs. This mechanism is biased to yield a 'No'-response in the lexical decision task when no semantic relation is detected, because nonword targets are not related to primes. The authors also point out that nonwords are usually derived from real words which are unrelated to the prime, which should increase the bias to link unrelatedness to making a No-decision. Thus, if a real word is encountered as target which is unrelated to the prime, the subject has to suppress the bias to respond 'No' in the lexical decision, which will cost processing time, thus leading to longer RTs for unrelated as compared to related targets. This mechanism, however, cannot easily be invoked for explaining the effects of semantic violation in Experiments 2 and 3. A third of the nonwords (all of which were pseudohomophones) sounded like words that fitted the sentence semantically. They should have undermined the bias to make a No-decision because of the suggested link between unrelatedness and nonword status.
In Norris' (1986) model a conscious expectancy strategy is replaced by a plausibility checking procedure. Norris suggests an initial stage in word recognition in which a set of candidates is activated via visual similarity with the target. In a second stage (which begins while the visual analysis is still ongoing), a plausibility checking procedure operates on this set of word candidates: if a candidate fits with the context established by the preceding word, the recognition threshold of the candidate is lowered; if it does not fit with the context, the recognition threshold is raised. To account for the findings of Experiments 2 and 3, context would have to be the preceding sentence rather than a single word. However, it does not seem to be clear why the recognition of the target would involve the activation of many visually similar word candidates, which all would have to be checked in terms of their contextual plausibility. Furthermore, facilitation and inhibition effects were more reliable in Experiment 2 than 3 even though it was in Experiment 3 that the target was visually degraded and the activation of several visually similar candidates would have been more likely. A further problem with this account is that even in the positive-fusion condition (eg. 'The dog was bitten') there would be visually similar candidates (eg. 'bitter') which would be fed back to the lexicon from the parser and the semantic subsystem as contextually implausible, slowing down recognition of the target.

6.4.3. A new post-access linguistic checking model

Diagram 6.1. is a minimal display of the processing subsystems and their relationships that I suggest are required to interpret the findings from Experiments 2 and 3. I suggest that making lexical decisions involves a decision maker, and does not directly tap any of the subsystems. The decision maker is affected by several, sometimes conflicting types of information. In this model the visual input analysis leads to the activation of the representation of the target in the lexicon, plus information regarding its syntactic and semantic categories (this is similar to Marslen-Wilson's (1987) proposals regarding spoken word recognition). If the target is embedded in a sentence context, the semantic and syntactic information specified in the lexical entry of the word are passed on to the semantic subsystem and the parser, respectively. As part of
the ongoing sentence analysis, the parser uses the syntactic category of the target to integrate the target into the current phrase structure. Sentence analysis, however, also immediately involves the semantic subsystem (otherwise, decision times should not have increased when the target was embedded in semantically violated sentences, such as 'The lake was bitten', which is syntactically well formed). I suggest that the semantic subsystem receives information from the lexicon about the basic semantic features of nouns and the argument structure of verbs, which includes any semantic selection restrictions which the verb might impose. In order to evaluate semantic fusion, the semantic subsystem also requires information from the parser about the syntactic relationship between words (see the above discussion of 'The duke sipped' versus 'The duke was sipped'). Thus, the decision maker receives information from the semantic subsystem, the parser and the lexical system. In the positive-fusion condition, there are Yes-messages from all these subsystems. In the neutral condition, there is no output from the semantic subsystem, since there is no noun which could semantically fuse with the verb. Hence responses are faster in the positive-fusion sentences than the neutral sentences, because there is a further source of positive evidence in the positive-fusion sentences. In the violated-fusion condition, the different outputs (a Yes-message from both the lexical system and the parser, a No-message from the semantic subsystem) interfere with each other: the results show that the decision maker does not simply focus on the output of the lexical system (which would be sufficient to make a lexical decision, since nonwords would not lead to a Yes-message from the lexical system). Interference of outputs from different processing systems, even if they are not relevant for the task to be performed, is a common phenomenon as the STROOP effects demonstrate (Morton, 1969b; Stroop, 1935). The hints in the data that responses for semantically violated sentences were slower than for neutral sentences (which was significant in Experiment 2 in the by-subjects and the by-items analyses, but not by MinF', and significant by-subjects in Experiment 3) can be interpreted as indicating that a No-message from one of the subsystems affects the decision maker to a greater extent than the lack of an output from a subsystem.

Thus, in contrast to proposals of purely syntactic parsers (for example within the Government and Binding theory; Frazier's model), I assume that a semantic subsystem
begins to operate as soon as words in the input sentence have been recognised. The semantic subsystem computes whether there is semantic fusion of coindexed elements, i.e. whether the semantic category of a noun filling an argument position of the verb, and the semantic requirements of the verb, match.

In sum, Experiments 2 and 3 have shown that both syntactic and semantic information is used rapidly during sentence processing. Semantic features of nouns and semantic selection restrictions placed by verbs on their arguments are analysed during sentence processing even when the subject has to perform a task which does not require the analysis of a sentence. Semantic violations produce immediate disruption of sentence processing, and of the secondary task.

The next two experiments (4 and 4b) test the effects of syntactic and pragmatic violations on sentence processing in addition to violations of semantic selection restrictions.

6.5. Experiment 4: The effects of violating semantic and pragmatic information on reading times in a self-paced reading paradigm

6.5.1. Introduction

This experiment was designed to replicate the effects of semantic fusion found in Experiments 2 and 3, using different sentence types and a different task. In Experiment 4, subjects do not have to perform any lexical decisions. A self-paced reading paradigm is used to test whether the time subjects spend reading different types of direct object nouns in active sentences is affected by the semantic status of the object nouns. In addition to the semantically-violated condition, a pragmatically-odd condition was included, in which the direct object noun possesses all the semantic features required by the verb, but is an unusual object given the sentence context. A further change compared to Experiments 2 and 3 consisted in the wider range of sentence
structures used in the present experiment to ensure that subjects did not expect
semantic violations to occur in particular sentence positions.

Example sentences illustrating the semantically normal, the semantically violated and
the pragmatically-odd condition are given in table 6.7. below.

<table>
<thead>
<tr>
<th>Table 6.7.: Example stimulus set for Experiment 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal:</td>
</tr>
<tr>
<td>In the laboratory the chemist thickened the liquid and searched in the cupboard for the right container.</td>
</tr>
<tr>
<td>sem-viol.:</td>
</tr>
<tr>
<td>In the laboratory the chemist thickened the teacher and searched in the cupboard for the right container.</td>
</tr>
<tr>
<td>prag-odd:</td>
</tr>
<tr>
<td>In the laboratory the chemist thickened the gravy and searched in the cupboard for the right container.</td>
</tr>
</tbody>
</table>

The verb 'to thicken' requires its subcategorised direct object noun to possess the
semantic feature LIQUID OBJECT. If during sentence reading, violations of this
requirement are noticed as soon as they occur, reading times on the semantically
violated direct object noun should be longer than in the semantically-normal sentence.
A further question is whether semantically violated direct object nouns cause increased
processing costs simply because they never normally co-occur with the verb. If this
were the case, there should be no difference in the amount of disruption and its
temporal characteristics for the semantically-violated and the pragmatically-odd
condition, since in the particular sentence contexts, the direct object nouns in the
pragmatically-odd condition are also highly unusual. If, however, subjects do not
immediately build up a representation of the whole sentence context, they might not be
slowed down in the pragmatically odd condition because locally the direct object noun
is compatible with the verb. In the semantically-violated condition, on the other hand,
there is no conceivable context in which the verb and object noun combination is possible. This violation should be noticed immediately (whereas the pragmatic violation might only take effect after a few words, if constructing a sentence context is not done immediately), and it should disrupt processing throughout longer parts of the sentence than the pragmatic violation which the subject might resolve by constructing a possible context in which such an unusual action seems plausible.

6.5.2. Pretests

A written Cloze task was run to ensure low Cloze values of the direct object noun in the semantically normal condition. As in Experiments 2 and 3 this was done so that the semantically normal object noun was not highly predictable. Such predictability might encourage short-cuts in processing, which would obviate the need for a semantic analysis. 22 subjects (none of whom took part in the main experiment) were given sheets of paper with a single sentence fragment per page. The order of pages was random across subjects. 20 sentence fragments were given, which ended on the verb before the critical noun phrase (eg. 'In the laboratory the chemist thickened the ....'). Subjects were asked to complete the sentence with the first word that came to their mind, and to write down this word in the space provided after the sentence fragment. The average Cloze value for the nouns finally used in the normal condition of the main experiment was low (8%).

After subjects had finished the written Cloze task, they were shown the same sentences again, this time with a noun provided after the verb (eg. 'In the laboratory the chemist thickened the liquid.'). They were asked to read each sentence again and to judge, on a scale from 1 to 7, the likelihood, in normal discourse, that the sentence would continue with the suggested word which was underlined (1=very likely, 7=very unlikely). Subjects were given a sentence with either the object noun designed for the normal condition of the main experiment (eg. '....liquid.'), or the object noun designed for the pragmatically odd condition (eg. '.... gravy.'). Thus, there were two versions of this pretest, so that one sentence occurred only once per version. There were 10 sentences
with the normal and 10 sentences with the pragmatically odd continuations per version. Each sentence was presented together with the rating scale on a single sheet of paper. Sentences were presented in a different random order per subject, and subjects were assigned at random to one version. Subjects were not timed. The results of the rating study verified that the continuations in the normal sentences were more likely than the ones in the pragmatically odd sentences (mean ratings of 2.6 (SD=1.3) versus 6.25 (SD=0.6), respectively.

6.5.3. Method

Stimuli

15 sentence triplets were created, using the verbs and nouns of Experiment 2 and 3 (see Table 6.7. above for examples). Sentences of different length and structure were used. All contained a main clause with an active sentence. According to the Pretests reported above, the direct object nouns in the normal condition were of low Cloze value. Furthermore, subjects in the Pretest had judged them to be more likely sentence continuations than the direct object noun in the pragmatically odd condition. The sentences in a triplet were identical, except for the direct object nouns. Three parallel stimulus lists were created using a Latin Square so that only one sentence from each triplet occurred per list. Subjects were randomly assigned to a list. Each subject saw 5 sentences per condition, and each sentence triplet was seen by 11 subjects in each condition.

There were 79 filler sentences of various sentence structures, unrelated to the purposes of this experiment, and 13 practice sentences of various sentence structures. There were three practice sentences which were similar to the sentences in the experimental conditions of the present experiment.

For a full list of stimuli, see Appendix H.
Design

In both the by-subjects and by-item analysis there was one repeated-measures factor, Sentence Type, with three levels: normal, semantically violated and pragmatically odd (see Table 6.7. above for examples). In the by-subjects design, List was a between factor, and in the by-items design, Itemgroup was a between factor.

Task

The same self-paced reading paradigm was used as in Experiment 5 (the stimuli of the present experiment were used as filler items in Experiment 5). Sentences were presented one word at a time, whenever the subject pressed a pace button, horizontally across the screen in a cumulative fashion. Subjects were instructed to read and understand the sentences so that they could do a recognition test at the end of the experiment, in which they had to decide whether a sentence had been presented during the main experiment. For further details on the task, the apparatus and the procedure, see Experiment 5.

Subjects

33 subjects, staff and students from University College London, took part in this experiment.

6.5.4. Results

For the purposes of analysis, certain words in the sentence were assigned positions as illustrated below. Position 1 was the critical direct object noun, and positions 2 to 5 were assigned to the subsequent words. In two sentence triplets, position 5 was the final word of the sentence. These triplets were not included in the analyses reported for
position 5.

... the chemist thickened the liquid and searched in the cupboard for the right container.

Table 6.8. below presents the means per level of Sentence Type, separately for each position in the sentence. The means are based on the subject-means of log-transformed data.

<table>
<thead>
<tr>
<th>Position</th>
<th>normal</th>
<th>sem_vio</th>
<th>prag_odd</th>
</tr>
</thead>
<tbody>
<tr>
<td>pos 1</td>
<td>344</td>
<td>372</td>
<td>370</td>
</tr>
<tr>
<td>pos 2</td>
<td>356</td>
<td>432</td>
<td>382</td>
</tr>
<tr>
<td>pos 3</td>
<td>357</td>
<td>465</td>
<td>386</td>
</tr>
<tr>
<td>pos 4</td>
<td>343</td>
<td>411</td>
<td>358</td>
</tr>
<tr>
<td>pos 5</td>
<td>323</td>
<td>396</td>
<td>355</td>
</tr>
</tbody>
</table>

On position 1, there were no significant effects at all (an analysis of variance was conducted on log-transformed data, and a nonparametric Friedman test on non-transformed data, with data collapsed over List and Itemgroup), i.e. there were no immediate effects of the manipulations of the direct object noun.

On position 2, i.e. the word following the direct object noun, both subject-means and item-means of non-transformed and log-transformed data violated the symmetry assumptions. Since the non-log-transformed data also violated the assumption of normality, nonparametric Friedman tests were run on these data. This analysis incorporated only the factor Sentence Type, collapsing over List and Itemgroup. The
difference between the conditions was significant [by-subjects: \( \chi^2 = 15.2727, p < .0005 \); by-items: \( \chi^2 = 10.5333, p = .0052 \)]. Pairwise comparisons using the Wilcoxon Matched-pairs Signed-ranks Test indicated that the normal and the semantically-violated conditions differed significantly [by-subjects: \( Z = 3.4396, p = .0006, 2\text{-tail} \); by-items: \( Z = 2.6126, p = .0090, 2\text{-tail} \)]. None of the other pairwise comparisons reached significance.

Since the log-transformed data in the three conditions were normally distributed, paired t-tests were run. The geometric means per condition in milliseconds (based on subject-means) are presented in Table 6.8. above. The difference between the normal and the semantically-violated conditions was significant [by-subjects: \( t(32) = 3.30, p = .002, 2\text{-tail} \); by-items: \( t(14) = 4.19, p = .001, 2\text{-tail} \)], as was the difference between the semantically-violated and the pragmatically-odd condition [by-subjects: \( t(32) = 2.29, p = .029, 2\text{-tail} \); by-items: \( t(14) = 2.11, p = .054, 2\text{-tail} \)]. The difference between the normal and the pragmatically-odd condition was not significant.

For position 3, for both nontransformed and logtransformed data, the symmetry assumptions were valid only in the by-items analyses. Subject-means of the logtransformed data were, however, normally distributed, and the resulting means per condition are presented in Table 6.8. above. Paired t-tests on the subject-means of the logtransformed data indicated that the sem-vio condition differed significantly from both the normal condition [\( t(32) = 4.15, p < .0005, 2\text{-tail} \)] and the prag-odd condition [\( t(32) = 3.74, p = .001, 2\text{-tail} \)]. The difference between the prag-odd and the normal condition was also significant [\( t(32) = 2.20, p = .035, 2\text{-tail} \)]. (Analyses on the subject-means of the nontransformed data were carried out with the nonparametric Wilcoxon Matched-pairs Signed-ranks Tests since they were not normally distributed. These tests also indicated that the sem-viol condition differed significantly from the normal and the pragmatically-odd conditions (\( p \)'s < .0008, 2-tail), which, however, did not significantly differ from each other).

A mixed AOV on the item-means based on the logtransformed data indicated a significant main effect for Itemgroup [\( F_{2,12} = 4.18, p = .042 \)], but the interaction with
Sentence Type was not significant. There was a significant main effect for Sentence Type \([F(2,24)=15.55, p<.0005]\). Planned comparisons revealed that the sem-vio condition differed significantly from the prag-odd condition \([F(1,12)=18.34, p=.001]\), and the normal condition \([F(1,12)=18.43, p=.001]\). The prag-odd and normal condition did not significantly differ (exactly the same pattern of results was found in a mixed AOV with item-means based on the non-transformed data).

For position 4, the symmetry assumptions were violated in both the by-subjects and the by-items analysis using the subject- and item-means of the nontransformed data. Using log-transformed data, the symmetry assumption was violated for the item-means, but not the subject-means. The cell means, based on the subject-means of the logtransformed data, are given in Table 6.8. A mixed AOV on the subject-means of the logtransformed data indicated a significant main effect for Sentence Type \([F(2,50)=10.26, p<.001, df\text{'s adj. acc. to Greenhouse-Geisser}\]. No other effects were significant. Planned comparisons revealed that the sem-vio condition differed significantly from both the normal condition \([F(1,30)=15.58, p<.0005]\), and the prag-odd condition \([F(1,30)=8.70, p=.006]\). The prag-odd and normal conditions did not significantly differ.

Since the item-means, based on the logtransformed data, were normally distributed, paired t-tests (collapsing over Itemgroup) were run. They showed the same pattern of results as found in the by-subjects analysis, i.e. the sem_vio condition differed from both the normal \([t(14)=6.65, p<.0005]\) and the prag-odd condition \([t(14)=3.96, p=.001]\), which did not significantly differ from each other.

On position 5, two stimulus triplets were taken out since in these sentences the word on position 5 was the last word in the sentence. It was decided to collapse data across Lists and Itemgroups to avoid having to create subject-means over a different number of sentences in the different conditions, and having an unequal number of cases in the different Itemgroups. One-way repeated-measures analyses of variance indicated a significant main effect for Sentence Type both in the by-subjects and the by-items analysis \([F(2,64)=12.70, p<.0005; F(2,24)=13.78, p<.0005; \text{MinF'}(2,72)=6.61, p<.01]\).
Planned comparisons revealed that the sem-vio condition differed significantly both from the prag-odd condition [F1(1,32)=6.40, p=.017; F2(1,12)=5.37, p<.039; but MinF'(1,32)=2.92, p<.10], and the normal condition [F1(1,32)=25.16, p<.0005; F2(1,12)=28.51, p<.0005; MinF'(1,37)=13.37, p<.001]. Furthermore, the difference between the prag-odd and the normal conditions was significant [F1(1,32)=6.39, p=.017; F2(1,12)=10.02, p=.008; but MinF'(1,41)=3.90, p<.10].

6.5.5. Discussion

Somewhat surprisingly, the data showed that on the direct object noun itself (position 1) there was no effect of Sentence Type on reading times. Given that at later positions there were such effects, this does not seem to be due to ineffective manipulations of the direct object noun. Two interpretations for the lack of an effect on position 1 are possible: either subjects do not initially process the semantic features and pragmatic plausibility of nouns when they read sentences; or the present self-paced reading task invites subjects to press the pace button while they are still processing a word; in this case, reading times for a word do not truly reflect the time spent processing that word, but also some processing still occurring on the previous word. Any increase in processing costs for a particular word might only become manifest one word later on. Further experiments using different tasks are needed to decide between these two options. Experiment 4b, reported in the next section, tries to slow down subjects, preventing them from pacing through the sentence by having them perform a grammaticality judgement on each word. Alternatively, a future study measuring eye movements during reading such sentences might be informative, since such a methodology allows to tease apart initial processing from later processing. If first-pass eye movements on the direct object noun also failed to show an effect of Sentence Type, while second pass eye movements (such as regressions) would show an effect of Sentence Type, the conclusion that initially semantic and pragmatic information is ignored would be more plausible.

Thus, while the interpretation of the lack of any effect of the semantic and pragmatic
manipulations on the direct object noun itself is not clear-cut, the results on the following word positions show that these manipulations cause extra processing costs. On positions 2, 3 and 4 there was the consistent finding that the semantically violated sentences required longer reading times than the normal and the pragmatically-odd sentences. The pragmatically-odd sentences did not differ from the normal sentences. On position 5, the semantically violated sentences still took longer than the normal sentences, but now the difference between the semantically violated and the pragmatically odd sentences is less reliable (not significant for MinF’). There is a hint that the pragmatically odd sentences now take longer than the normal sentences (significant by- subjects and by-items, but not for MinF’), while the semantically-violated sentences still clearly take longer than the normal sentences. This pattern of results illustrates that the time course of the effect of semantically violating a direct object noun is different to that of pragmatically violating the direct object noun. The semantic violation leads to an increase in processing time at least from one word after the violation onwards for at least four more words. The pragmatic violation only shows an effect four words after the critical noun. The patterns of results for the semantically violated and the pragmatically odd sentences cannot easily be interpreted if one assumes that the two sentence types only differed in the extent of the unexpectedness of the direct object. If the pragmatically violated direct object noun was merely regarded as an instance of a less severe violation of co-occurrence expectations, one would have expected an increase on processing costs as early as for the semantically violated condition. This increase would, however, be expected to disappear after only a few words, while the semantic violation, being a more severe case of the violation of co-occurrence expectations, would continue to have an effect. This clearly was not evident in the results. The effect of pragmatically odd direct object nouns only manifested itself after three words. It seems more likely that subjects took some time, at least in this type of task, to create a representation of the sentence context which would then make it difficult to integrate the pragmatically odd direct object noun. Note that the pragmatically odd direct object noun is perfectly acceptable if it is processed only in the context of the verb itself; it is only when a representation of a greater part of the sentence has been built up that the oddness of this noun can become apparent. The semantic violation of a direct object noun, however, has an immediate effect (i.e.
on the following word and beyond). Subjects seem to be sensitive to the fact that the
direct object noun does not possess the semantic features that are required from it by
the preceding verb which subcategorizes it. This finding shows that subjects take
nonsyntactic information at least of the local context into account at an early stage
during sentence reading.

The present pattern of results leads to an interesting prediction concerning the time
course of processing local and global nonsyntactic information to be tested in future
research. If nonsyntactic aspects of local context (such as the semantic selection
restrictions of a preceding verb and the semantic features of the subcategorized noun)
are processed quickly, while nonsyntactic aspects of the whole sentence are processed
more slowly, the incompatibility of local nonsyntactic information (as for example in
the semantically violated sentences of this experiment) should cause extra processing
costs, even if a global sentence context could be constructed which allowed a non-
literal metaphorical interpretation of the locally incompatible words, in which the local
incompatibility would be resolved.

As discussed above, further research is also required to establish why subjects did not
take longer to read the semantically violating direct object noun itself. In the next
experiment, subjects were given a self-paced reading task which required them to make
a grammaticality judgement before seeing the next word of the sentence. In such a
task, subjects are less likely to pace through the sentence, and thus it might be more
likely to pick up an effect on the direct object noun itself.
6.6.1. Introduction

Experiment 4b was designed to test the effects of violating semantic selection restrictions and of pragmatic anomalies with a task that would discourage subjects from pacing through the sentence, as might happen in the self-paced paradigm used in Experiment 4. In this task (which was also used in Experiments 6 and 7) subjects have to use one of two buttons: a Yes-button, which they are instructed to press when they judge the sentence, up to the current word, to be grammatical, and a No-button, when they judge the sentence to be ungrammatical. More details of the task and the procedure are presented with Experiment 6.

In this experiment, syntactically violated sentences were used in addition to the normal, semantically violated and pragmatically-odd sentences of Experiment 4. This was done in an attempt to assess the types of information subjects use in making the continuous grammaticality judgements required in this self-paced paradigm. A typical sentence quartet is given in Table 6.9. below.
Table 6.9.: Example stimulus set for Experiment 4b

<table>
<thead>
<tr>
<th>Condition</th>
<th>Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal:</td>
<td>In the laboratory the chemist thickened the liquid and searched in the cupboard for the right container.</td>
</tr>
<tr>
<td>sem-viol.:</td>
<td>In the laboratory the chemist thickened the teacher</td>
</tr>
<tr>
<td>prag-odd:</td>
<td>In the laboratory the chemist thickened the gravy and searched in the cupboard for the right container.</td>
</tr>
<tr>
<td>synt-viol.:</td>
<td>In the laboratory the chemist thickened the too</td>
</tr>
</tbody>
</table>

If making a grammaticality judgement about the sentence presented so far slows down processing (as compared to the self-paced task used in Experiment 4), pragmatic violations should have an effect on the direct object noun itself: the time to make a YES-decision should be longer in the pragmatically odd than the normal condition, even though the object nouns of both conditions are correct syntactically and in terms of the required semantic features. Such a difference would indicate that subjects have build up a representation of the sentence context, and that they cannot disregard pragmatic plausibility even when the task is to judge the grammaticality of a sentence. The number of incorrect No-decisions should not differ for these two conditions.

A comparison of the number of correct NO-decisions in the semantically-violated and the syntactically-violated conditions gives important clues as to how subjects understand the instruction to judge the grammaticality of a sentence. They were told in the instructions that sentences are ungrammatical "if they are incorrect in any conceivable context". Furthermore, the practice sentences provided them with feedback (see Procedure section in Chapter 8.3.). Neither the semantically nor the syntactically-violated sentences continued after the violating word. It was expected that the NO-decisions on this word would take longer in the semantically-violated than the
syntactically-violated condition. In the semantically-violated condition, subjects might try to reinterpret the semantically incorrect object noun in some way which would allow them to ignore the violation of the semantic selection restriction. In the syntactically-violated condition, however, the determiner requires the occurrence of a noun (or adjective or adverb); no word of any other syntactic category can be reinterpreted to fit the sentence grammatically.

6.6.2. Method

Design

Sentence Type is a repeated measures factor in the by-subjects and the by-items design. It has four levels, i.e. normal (the direct object noun fits the verb semantically and pragmatically), semantically violated (the direct object noun violates semantic selection restrictions of the verb), pragmatically odd (the direct object noun fits the semantic selection restrictions but is unlikely as established in the pretest for Experiment 4) and syntactically violated (the word following the determiner violates syntactic phrase structure rules). See Table 6.9. above for example sentences.

Four parallel lists were created using a Latin Square, such that each subject would only see one of the sentences from a sentence quartet. Each subject saw five sentences in each of the four conditions, and each sentence was seen by 11 subjects per condition.

Stimuli

Sentence sets were identical to the ones used in the previous Experiment 4, but an additional 5 sentence sets were created resulting in a total of 20. To each of the sentence triplets used in the previous self-paced experiment, a fourth, syntactically violated, sentence was added. In the syntactically violated condition, the critical noun was replaced by a word of a syntactic category that cannot grammatically follow a
determiner, i.e. a verb, a determiner, a pronoun or a preposition. All sentences included in the present experiment had at least six words following the critical noun. A few further minor stylistic changes were made to the sentences, which are indicated in Appendix H in brackets.

The sentences were derived from the stimulus set used in the selection-restriction experiment (Experiment 2). Pretests were run to ensure low Cloze values of the direct object noun in the normal condition, and to establish the naturalness of the sentences in the normal and pragmatically odd conditions (see previous chapter for details).

In the semantically- and the syntactically violated conditions, the sentence discontinued after the ungrammatical words. In the normal and pragmatically-violated conditions, sentences continued after the critical noun. These sentences became ungrammatical a few words later on.

There were 79 filler sentences of various sentence structures, unrelated to the purposes of this experiment, and 13 practice sentences of various sentence structures. There were three practice sentences which were similar to the sentences in the experimental conditions of the present experiment.

For a full list of stimuli, see Appendix H.

Subjects

44 subjects were recruited from among the staff and students of University College London.
6.6.3. Results

Semantically violated and syntactically violated conditions

Data in the two violated conditions were only analyzed on the word which introduced the violation (i.e. the semantically violated noun, or the syntactically impossible word following the determiner after the verb). Paired t-tests indicated that the average time to make a correct No-decision was longer in the semantically-violated (2394 ms) than the syntactically violated condition (1382 ms) [by-subjects: t(40)=6.61, p=<.0005, 2-tail; by-items: t(19)=4.12, p=.001, 2-tail]. Note that three subjects made only incorrect Yes-decisions in the semantically violated condition. Their values were set to missing in this analysis of the correct No-decisions.

The number of incorrect Yes-decisions differed in the two conditions: 41 % of the responses in the semantically-violated condition were an incorrect Yes-decision on the semantically violated direct object noun, as compared to 15.9% of incorrect Yes-decisions in the syntactically violated condition (significant in the by-subjects and by-items analysis, Wilcoxon Matched-pairs Signed-ranks Tests: p's <.0004, 2-tail). The incorrect Yes-decisions in the semantically-violated condition did not come from only a small set of subjects (on average, 2.07 incorrect Yes-decisions were made per subject; only 4 out of the 44 subjects did not make any incorrect Yes-decisions). Neither were the incorrect Yes-decisions limited to only a small set of sentences (the average number of incorrect Yes-decisions per item was 4.55; no item received no incorrect Yes-decision).

Paired t-tests indicated that the average time to make an incorrect Yes-decision in the semantically violated condition (1468 ms) differed significantly from the time to make a correct Yes-decision in the normal condition (822 ms) [by-subjects: t(39)=2.60, p=.013, 2-tail; by-items: t(19)=2.67, p=.015, 2-tail]. Note again that there were four missing values in the by-subjects analysis: 4 subjects did not make any incorrect Yes-decisions in the semantically violated condition. This difference between the normal and the semantically-violated condition was not significant once extremely long times
for making Yes-decisions were taken out before subject-and item-means were computed (extremely long values were defined as those greater than 1555 ms, which is the cut-off established on the basis of the distribution in the normal condition). Note that the number of responses in the semantically violated condition for incorrect Yes-decisions was quite small once the extremes were taken out: 22 % of the incorrect Yes-responses in the semantically violated condition were extremes (these responses came from 16 different subjects, and were distributed over 13 of the items), as compared to 7% of the correct Yes-responses in the normal condition. Subject-means for this reduced data set in the semantically violated condition were calculated on average over 2.2 responses, and item-means over 3.8 responses (in the normal condition, the corresponding numbers are 4.6, and 10.2, respectively). In the by-subjects analysis there were 11 missing values, i.e. for 11 subjects no data points were left over which to calculate means.

Normal versus pragmatically odd conditions

There were no incorrect No-decisions on the direct object noun in the normal condition, whereas 8% of the responses in the pragmatically odd condition were incorrect No-decisions (they were made by 13 subjects, on 9 different items).

The time to make a correct Yes-decision on the direct object noun was significantly faster in the normal (822 ms) as compared to the pragmatically odd condition (1280 ms), as indicated by paired t-tests [by-subjects: t(43)=7.33, p<.0005, 2-tail; by-items: t(19)=7.12, p<.0005, 2-tail]. The percentage of extremely long decision times (again defined on the basis of the distribution of decisions in the normal condition), i.e. responses greater than 1555 ms, was 7% in the normal condition and 28 % in the pragmatically odd condition. In the normal condition, extremely long responses were distributed over 11 items, and 12 subjects. In the pragmatically odd condition, the extremely long responses were distributed over 31 subjects, and 18 items. When the extremely long responses were removed from the data set before subject-and item-means were computed, the difference between the average decision times for the
normal (733 ms) and the pragmatically odd conditions (882 ms) still differed significantly, as indicated by paired t-tests [by-subjects: $t(43)=4.70$, $p<.0005$, 2-tail; by-items: $t(19)=3.75$, $p=.001$, 2-tail].

6.6.4. Discussion

Experiment 4b showed effects of Sentence Type immediately on the critical word, unlike Experiment 4, in which the effect was manifest only on the following word and beyond. In Experiment 4b subjects were slowed down by the additional task of having to make a grammaticality judgement on every word of the sentence (response times were about twice as long as in the self-paced task used in Experiment 4). This seems to have provided sufficient time for building up a representation of the nonsyntactic representation of the sentence. Subjects were slower to judge a sentence as grammatical when the direct object noun was pragmatically odd, given the sentence context, as compared to the normal condition. This increase in processing time was not due to only a few extremely long responses but to a shift of the distribution of responses in the pragmatically odd condition.

As expected, the time to judge a sentence as ungrammatical was longer in the semantically-violated than the syntactically-violated condition. This probably reflects extra time subjects spent trying to find a possible interpretation of the semantically violated sentences, whereas the syntactic violation signalled more clearly that no grammatical interpretation would be possible.

The results also showed that subjects made more No-decisions in the semantically violated than the pragmatically odd condition (59% versus 8%, respectively). The number of Yes-decisions in the semantically violated condition, however, was quite high, i.e. 41% of all responses in this condition. The time to make such Yes-decisions was longer than the time to make a Yes-decision in the normal condition. The finding that this difference was not significant once extremely long decision times had been taken out should be interpreted with caution: the number of observations over which
subject-and item-means were formed was much smaller in the semantically violated condition (there were 91 Yes-responses, of which 20 were extremely long; for the non-extreme data, subject-means were calculated on average over 2.2 responses, and item-means over 3.8 responses) than the normal condition (of the 220 Yes-responses there were 15 extremes; subject-means were formed on average over 4.6 responses, and item-means over 10.2 responses). Nevertheless, these non-extreme response times for making incorrect Yes-decisions on semantically violated nouns probably reflect those cases where the grammaticality judgement is made without an analysis of semantic selection restrictions. Note that these cases only make up 32% of all (220) responses to semantically violated object nouns.

The few cases of taking an extremely long time to make an incorrect Yes-decision on the semantically violated object noun (22% of all incorrect Yes-decisions in the semantically violated condition, but only 9% of all the responses in this condition) indicate that subjects sometimes noticed the semantic violation but still decided that the sentence was grammatical. This might be because they used an implicit criterion according to which only syntactic violations render a sentence ungrammatical. Alternatively, they might have succeeded in constructing some metaphorical context in which the violation of semantic selection restrictions had been resolved.
CHAPTER 7: THE USE OF NON-SYNTACTIC INFORMATION IN PROCESSING SYNTACTICALLY AMBIGUOUS SENTENCES

7.1. Introduction

Overview

In this chapter I will turn to evidence from the garden path literature to demonstrate if, how and when lexical, semantic and syntactic information interact during sentence processing.

In this introduction I will briefly discuss the various measures of garden path effects, and I will describe the types of information proposed in various models as influencing the processing of syntactically ambiguous sentences. I will then characterize garden path sentences as a special type of the more general class of temporarily syntactically ambiguous sentences. In Section 7.2. I review some of the arguments for and against single-analysis and multiple-analyses models, and discuss two further models of sentence re-analysis. Section 7.3. presents in detail the evidence regarding the interaction of lexical subcategorisation biases with syntactic information in determining the initial decisions of the syntactic processor. Section 7.4. focusses on the role of semantic factors in processing garden path sentences, and Section 7.5. on thematic role assignment and its effects for processing garden path sentences.

Measures of sentence processing

Eye movement data have considerable appeal for psycholinguists who are interested in how readers process syntactic information. Clearly, the most informative data for psycholinguists are those that are sensitive to moment-to-moment changes in processing activity. Eye-movement data "have proved to be very informative in comparison to other types of tasks" (Rayner & Pollatsek, 1987, p. 353) in studying
syntactic processing. As discussed in Chapter 2, a widely held assumption is that eye fixations closely reflect on-going cognitive processes, but this assumption has to be considered in the light of findings that the eyes might skip words, or that fixation on one word might also reflect spill-over effects, i.e. some continuing processing of a previous word. However, a consistency of predictable patterns of specific eye movement measures makes it likely that eye movement data are valid measures of cognitive processes. One particular feature of eye movement data, which is widely exploited in research, is that they can be divided into initial and later fixations on a particular region. This promises a more fine-grain analysis of on-line processes than is possible with other methods, eg. self-paced reading paradigms. It will become clear that other measures, such as various forms of the self-paced reading paradigm, are also necessary. Quite clearly, the measures differ widely, not least in their temporal characteristics: for example, average first fixation durations are around 250 ms, total fixation times around 300 ms, whereas the average time to press a button in a self-paced word-by-word reading paradigm takes about 460 ms. The average total time required to read a whole sentence in the eye movement paradigm is about 1535 ms; the same sentences, when read in a self-paced reading paradigm, takes on average 3388 ms (Ferreira & Henderson, 1990). There are other measures, such as end-of-sentence grammaticality- or acceptability judgements that have an even bigger temporal window (which does not stop some researchers, most commonly linguists, from considering them as the only valid evidence, eg. Pritchett, 1992). The challenge is to understand which processes are tapped by the various tasks and to create a processing framework in which diverse types of evidence can be related to different levels or phases of processing.

Sources of information used in sentence parsing

A most controversial question in the study of sentence processing is which information is used in the initial analysis of sentences. There is no disagreement that the final outcome of sentence comprehension, which is accessible to consciousness, integrates syntactic with semantic and real-world knowledge. Controversy exists, however, about
the early, unconscious processes. In several linguistic theories, words are assumed to have lexical entries which specify three types of information: a) information regarding the word’s syntactic category (i.e. whether it is a noun, a verb, an adjective, a preposition etc.), b) its subcategorisation frame, and c) its thematic roles.

For some linguists, the simplest assumption seems to be that the parser has access to all these three types of information stored in the lexical entry of a word. As discussed in an earlier chapter, Weinberg (1987), for example, proposes that in order to be efficient, the parser would not only look up the syntactic category of a word, which it requires in order to build up phrase structures, but at the same time the other types of lexically stored information.

According to some processing theories, all types of lexical information plus a representation of the current discourse are used immediately in sentence processing. Others argue that during the initial phase in sentence processing in which a syntactic representation of the input is built up (the parsing process) no use is made of context, and only limited use is made of lexical information. Thus it is assumed that only lexical information specifying the syntactic category of a word is employed since this information is required in applying phrase structure rules.

Much of the experimental work in psycholinguistics has focused on sentences which are syntactically ambiguous (they allow, at least temporarily, the application of different phrase structure rules, resulting in different syntactic representations of the input). Syntactic ambiguity allows us to test whether the parser builds up only one single representation at a time, or several representations in parallel, or whether it delays building up any representation for a while and then proceeds to build up a single representation.

If only a single representation is built up, the question is, of course, which particular option is taken up first. There is strong evidence to show that a purely structural (syntactic) bias determines the initial analysis (see the Garden Path model discussed below), but other work has tried to show that non-syntactic information can overrun
the structural preference and guide the initial parse, or that it, at least, affects the
evaluation of a syntactic structure once it has been built up, and guides the reanalysis
of mis-analyzed sentences. In a later section I will discuss in detail the kinds of non-
syntactic information that have been investigated most closely, namely preferences of
verbs for a particular subcategorisation frame, the thematic roles of verbs, and
discourse constraints. It has also been suggested that interpretive information provided
by prior sentence context may determine the initial parse (Crain & Steedman, 1985;
Altmann, Gamham & Dennis, 1992).

Marcus (1980) suggested that, when faced with structurally ambiguous input, the
parser delays syntactic decisions while a look ahead device provides syntactic
information about several words (or phrases) lying ahead of the temporarily
structurally ambiguous phrase. The question about the interaction of different sources
of information within this model (and any other multiple-analyses model) is the
question whether any structural ambiguity leads to a delay to await further input, or
whether non-syntactic information can force the immediate choice of one of the two
possible syntactic analyses, thus removing the need for a delay.

The parallel models that have been suggested differ in their assumptions about how
(and when) the parser decides on one analysis. Obviously, if the parallelism is
maintained only for a very brief period, this model becomes very similar to a single-
analysis model such as the Garden Path model. Altmann (1987) proposes that initially
all possible alternative syntactic structures are built up in parallel, and that at some
later point one is filtered out according to semantic or interpretive criteria.
Alternatively, Garrett (1990) suggests that parallel structures are only maintained for a
short while, after which one of them (the one that follows the principle of minimal
attachment) is adopted as a default. Jackendoff (1992, 1987b) points out that only one
syntactic analysis becomes accessible to consciousness, while multiple analyses might
be temporarily available to the parser. Recently, Macdonald, Just & Carpenter (1992)
proposed a model in which working memory capacity is an important factor
determining how long multiple representations can be maintained for.
Temporary syntactic ambiguity

A variety of sentences with temporarily ambiguous syntactic structures (TASS for short) have been examined over the years. The initial part of such sentences allows two syntactic structures to be built up. Take for example:

'The horse raced past the barn fell.'

(A less familiar example is: 'The cotton shirts are made from comes from Arizona')
The reader might represent the beginning of the sentence as the sequence NP plus main verb plus PP, or alternatively as a noun plus a reduced relative clause (short for 'the horse which was raced past the barn... '). The crucial data is gathered at the point in the sentence when a word is presented which is only compatible with one of the structures. This is called the disambiguating information ('fell' in this example). In fact, the term 'disambiguating information', and the term 'ambiguity' itself, are only appropriate from a linguistic point of view. From a processing perspective, these terms beg the very question that is being investigated. It might well be that for a reader these sentences are not ambiguous (i.e. only one analysis is built up), and the disambiguating information is, to the reader, contradictory information which cannot be integrated into the syntactic structure built up thus far. However, since the term temporal structural ambiguity is generally used (also by proponents of the single-analysis model), I will adopt this terminology here, too, even though it is only appropriate from a linguistic but not necessarily from a psycholinguistic perspective.

Relevant measures of eye movement at the point of disambiguation typically show prolonged processing which indicates that readers originally built up the NP plus VP plus PP analysis which does not in fact fit the sentence given the final part of the sentence. Studying TASS with ambiguous postverbal NPs (such as: 'Sally found out the answer to the difficult physics problem was in the book') Frazier & Rayner (1982) found that disambiguating information (here 'was') led to several eye movement patterns. Some readers fixated the disambiguating word for a very long time, and then read on quite normally; other readers regressed back to the previous ambiguous sentence part after they had fixated the disambiguating word; and yet other readers
continued to read on after the disambiguating material in a disrupted fashion with very long fixations or short saccades, and they re-read the sentence once they had reached the sentence ending. These findings are taken by Frazier & Rayner to demonstrate that readers are generally committed to a single interpretation of the sentence, rather than building up two representations in parallel (but see below for attempts to interpret this finding in a multiple-analysis model).

It is worth stressing that the individual engaged in sentence processing is usually not aware of any ambiguity until contradictory input is encountered (nor are eye fixations longer on the ambiguous word). Furthermore, normally no garden path is experienced at all even when there are several syntactic analyses possible for some part of the input (Garrett, 1990). At the end of a sentence, one is only conscious of one single interpretation, as in this example:

a) The bishop said, "Have the missionaries eaten their breakfast?"

The structural (temporary) ambiguity of a part of this example only becomes available to consciousness when one compares a) with b):

b) The cannibal said, "Have the missionaries eaten for breakfast!"

This example illustrates again the need to test syntactic processes as they unfold; measuring end-products of processing may appear to demonstrate an interaction of strategies based on real-world plausibility and syntactic processes. Without more specific data about on-line processes one cannot rule out the possibility that the interaction merely reflects the output of an independent syntactic processor being affected by knowledge of real-world plausibility.
7.2. Models of processing garden path sentences

7.2.1. Single-analysis: Frazier's garden path model

There is considerable evidence that when confronted with temporarily structurally ambiguous sentences, readers do not await further input but adopt particular parsing strategies which preferentially build up one reading rather than another. In sentences such as

The horse raced past the barn fell.

the main clause reading rather than the reduced relative clause reading is generally preferred. The garden-path model (Frazier, 1987a) proposes that an initial syntactic analysis is assigned on the basis of purely structural information. The preferred structural analysis of the parser incorporates each item from the linguistic input into a constituent structure representation by postulating the minimal number of nodes in the structural tree consistent with the phrase structure rules of the language. From this general characterisation of the parser's preferred strategy, two specific parsing strategies are derived, which Frazier defines as follows:

a) "minimal attachment: attach incoming material into the phrase-marker being constructed using the fewest nodes consistent with the well-formedness rules of the language" (Frazier & Rayner, 1982, p. 180)

and the strategy of late closure which determines the initial analysis when two equally minimal attachments exist:

b) "late closure: if grammatically possible, attach new items into the clause or phrase currently being processed (i.e. the phrase or clause postulated most recently)" (Frazier, 1987a, p. 562)

These two parsing strategies are assumed to be of great generality, subsuming several specific strategies applicable to particular sentence structures. Frazier (1987a and b) argues that they are not accidental but reflect the tendency of the processing system to minimise time and memory costs. Minimal attachment reduces costs since fewer
phrase structure rules need to be accessed (note that this is only true if accessing more rules is assumed to be more costly). Late closure allows earlier structuring since a new item is attached immediately to the existing tree, avoiding the need to hold it in some memory device while further input is received and processed. The strategies are not based on a particular grammatical theory, and, consequently, there has been disagreement about exactly which syntactic analysis of a sentence is the simplest (see Konieczny, Hemforth & Strube, 1991; Pritchett, 1992).

A few examples of minimal attachment are given below, with the temporarily ambiguous constituent underlined:

John hit the girl _with a book_. (Rayner, Carlson & Frazier, 1983, experiment 2)

Following the minimal attachment principle, PP attaches to VP as a sister of V ('hit') and NP ('the girl'). Nonminimal attachment requires an additional more complex NP to be built to which 'the girl' and 'with a book' are attached (here, PP modifies 'the girl').

The horse _raced past the barn_ fell. (Bever, 1970; Rayner et al., 1983, experiment 1; Ferreira & Clifton, 1986)

The minimal attachment principle leads to the main clause analysis rather than the reduced relative subordinate clause analysis, because the latter would require the introduction of another S-node (i.e. for the reduced relative sentence).

Ernie kissed Marcie _and her sister..._. (Frazier, 1987a)

The minimal attachment principle yields the analysis as NP-conjunction since the alternative analysis requires attaching 'her sister' to an additional S-node (as would be correct if the sentence continued '...and her sister laughed').

Late closure is most clearly demonstrated by sentences such as

'Joyce said Tom left _yesterday_'
(Frazier, 1987a; Kimball, 1973),

for which attachment of the adverb to the most recent phrase in the tree (as a modifier of 'left') is preferred over attaching it to the previous verb. Similarly, in
'Jessie put the book Kathy was reading in the library ...' (Frazier & Fodor, 1978)

the PP is attached as a modifier of 'reading' rather than of 'put'.

Evidence that readers apply these strategies is quite considerable. It has mainly come from studies of eye movement (Frazier & Rayner, 1982; Rayner & Frazier, 1987), whereas most of the evidence failing to support the strategies has been found in studies using self-paced reading paradigms. However, this distinction is too simplistic, since there are studies in which the self-paced reading paradigm was used which did support the garden-path model (eg. Ferreira & Clifton 1986; Ferreira & Henderson, 1990).

Recently, studies of event-related brain potentials (ERPs) during sentence reading have also been put forward as support for the single-analysis - minimal attachment model (Osterhout & Holcomb, 1992). Positive-going waves were recorded when reading words which could not easily be fitted into the current syntactic analysis of a sentence (the P600 effect). Sentences contained intransitive and transitive verbs plus clausal complements, such as a) and b):

a) The woman struggled to prepare the meal.
b) The woman persuaded to answer the door.

A P600 effect was found when reading the word 'to' in sentences with transitive verbs (such as b), but not in a). This indicates that in line with the minimal attachment principle subjects reading sentences such as b) analyzed the verb transitively, expecting a direct object NP (and thus taking 'to' as a syntactic violation), rather than interpreting the verb as introducing a reduced relative clause.

However, it is not clear exactly which cognitive (or neural) processes are being measured by P600. The authors assume that P600 is a measure sensitive to syntactic violations. It contrasts with N400, i.e negative-going waves found as a response to reading a word which cannot be integrated into a sentence at a semantic or message level (eg. Kutas & Hillyard, 1980). A detailed account of what P600 is a measure of
is, however, clearly not in sight yet.

Ferreira & Henderson (1990) found evidence in favour of the minimal attachment principle using both eye measurements (their experiment 1) and a non-cumulative word-by-word self-paced reading paradigm (their experiment 2). Their sentences were temporarily syntactically ambiguous between a verb plus direct object reading, and a verb plus reduced complement reading, for example:

He forgot Pam needed a ride with him.

Their results indicate that the minimal attachment principle led the parser to analyze 'Pam’ as the direct object rather than as the subject of a reduced complement clause.

Using both eye movements and a non-cumulative word-by-word self-paced reading paradigm, Ferreira & Clifton (1986) found evidence that the minimal attachment analysis is preferred regardless of plausibility and pragmatic context (see discussion below).

Frazier & Rayner (1982) compared sentences which were parsable on the basis of the minimal attachment strategy (as in a) below) with those for which the minimal attachment strategy creates a garden path (as in b):

a) Sally was relieved when she found out the answer to the difficult physics problem.

b) Sally found out the answer to the difficult physics problem was in the book.

The average total reading time per character was longer in b) as compared to a). Furthermore, in b), total reading times per letter were longer in the disambiguating region ('was in the book’) than either in the ambiguous region ('the answer ...’) or in the region prior to the ambiguity (there was no difference between the ambiguity region and the region prior to it).

Henderson & Ferreira (1990) replicated this result, again studying sentences which are
temporarily ambiguous between an NP reading and a reduced that-complement reading ('She warned Harry brought small gifts'). The NP-reading was preferred by most readers since fixation times on 'brought' were increased, indicating a garden path effect, as compared to fixation times on 'brought' in a sentence which contained 'that' in its surface structure ('She warned that Harry brought small gifts'). Henderson & Ferreira (1990) also showed indirectly that the fixated target ('brought') was difficult if 'that' was absent, because parafoveal preview benefit was reduced in this case.

However, Holmes, Kennedy & Murray (1987) (also reported in Holmes, 1987, experiment 1) showed that processing times increased equally on the disambiguating word ('was' in the example below) in reduced that-complement sentences and complement sentences with an overt 'that' complementizer:

a) A journalist reported the inquests's finding was very wrong.

b) A journalist reported that the inquests's finding was very wrong.

For both sentence types processing times on 'was' were longer than in the post-NP region of the transitive version of the sentence:

c) A journalist reported the inquest's finding in today's paper.

The authors claim that the increased reading times on the post-NP region in a) as compared to c) cannot be interpreted as evidence of garden-pathing resulting from minimal attachment, but instead as reflecting the extra processing cost involved in having to process two clauses in the (reduced and non-reduced) complement sentences instead of just one clause in the transitive sentence. However, it seems difficult to refute Frazier (1987a) who dismisses those results on the grounds that they were based on very long reading times collected in a self-paced reading task. This point gains more credence in the light of Ferreira & Henderson (1990) who directly compared eye movement data with self-paced reading times for the same stimulus materials. In the cumulative self-paced reading paradigm - which is the paradigm used by Holmes et al. (1987) - the effect of 'that'-presence was only very weak (not significant by items, and significant at p < .05 by subjects), whereas both the non-cumulative self-paced
reading paradigm and eye measurements yielded a significant effect. Ferreira & Henderson (1990) suggest that in the cumulative self-paced reading task subjects often adopt a strategy described by Just, Carpenter & Woolley (1982); they press the button to see more words on the screen without immediately processing each word, invalidating the task as an on-line measure of cognitive processes.

Another study which found evidence contrary to the minimal attachment principle is Kennedy, Murray, Jennings & Reid (1989). They, too, found that verb plus sentence-complement sentences were not harder to parse when the complementizer 'that' was absent as opposed to present (the measure here was initial fixation time on the disambiguating phrase 'was an insult' in sentences such as 'The workers considered [that] the last offer from the management was an insult'). They argue that Rayner & Frazier's (1987) results were flawed because their sentences were presented across more than one line, with the line breaks inducing garden paths; however, this argument does not apply to Ferreira & Henderson's (1990) study, which replicated Rayner & Frazier's (1987) results with sentences presented one per line.

7.2.2. Garrett's multiple-analyses model

Garrett (1990) suggests that when faced with TASS, the parser initially builds up in parallel all alternative syntactic structures compatible with the subcat frames. He assumes that this parallelism exists only for a short time, that is for one or two words beyond the head noun of the temporarily structurally ambiguous NP. If by that point there has not been any information in the input forcing one of the analyses, the minimal attachment analysis is adopted as a default (he also suggests that a preference ordering of the subcat frames might determine the ordering of default choices).

Garrett argues in favour of a parallel syntactic processing account in an attempt to integrate two sets of findings which seem at first to be contradictory. Firstly, he reports older findings by Holmes & Forster (1972) and Chodorow (1979) (quoted from Garrett, 1990, p. 153) according to which sentences with purely transitive verbs (see a)
below) were processed faster, and performance was less error-ridden as compared to sentences whose verb allows both a transitive subcat frame and a sentence-complement frame (see b) below). Compare, for example:

a) The helicopter located the wreckage in the ravine.
b) The helicopter discovered the wreckage in the ravine.

Evidence from Holmes et al.'s RSVP study indicated that the accuracy of repeating sentences was lower for the b)- as compared to the a)-sentences. Furthermore, in Chodorow's study, in which sentences were presented auditorily at a faster than normal rate, more errors were made in repeating b)- as compared to a)-sentences. Garrett argues that if the parser had committed itself to an immediate single analysis of b) in line with the minimal attachment principle, it would have analyzed 'the wreckage' as the direct object: but since this is the correct analysis for sentences such as b) the processing differences between a) and b) could not be explained. Within a single-analysis account, the reported differences between a)- and b)-sentences could only be accounted for if one assumed that the parser initially opted for the complement-clause analysis of sentences such as b); but this is against Frazier & Rayner's (1982) evidence which indicates that the complement-analysis is not initially chosen. As an alternative account, Garrett suggests instead that the parser pursues both analyses in parallel, prompted by the two subcat frames for verbs such as 'discover'. He reasons that in sentences like b) "more computational paths must be represented and evaluated" (Garrett, 1990, p. 154), leading to an increased processing load. One 'computational path' corresponds to the parser's decision to attach 'the wreckage' as a sister node of V ('discovered') to make up the VP. The other path involves postulating another S-node (sentence node) and attach 'the wreckage' to it.

However, it is not generally accepted that maintaining multiple analyses does in fact lead to greater processing demands. Typically it is pointed out that the lexical ambiguity research has shown that two alternative word meanings are briefly maintained in parallel (see Chapter 3.3.2.). Some researchers take the multiple, context-insensitive output of a processing system as evidence that the processing
system is an independent module (Seidenberg & Tanenhaus, 1986).

Garrett backs his proposal that multiple parallel syntactic analyses are maintained only briefly by referring to Frazier & Rayner's (1982) finding that no garden path was found for complement clause sentences when the disambiguating information (i.e. the word in the sentence that is evidence for the complement-clause structure) appeared shortly after the ambiguous NP (as in 'Sally found out the answer was in the book'). He does not give any further account of why he assumes that the parallelism is given up after only one or two words.

In assessing the claim of parallelism triggered by the existence of two subcat frames, one would also like to see some direct evidence that processing load is increased as soon as the verb with the two subcat frames is encountered.

7.2.3. The Capacity Constrained Parsing Model

Macdonald et al. (1992) propose a model (the Capacity Constrained Parsing Model) in which they take up an argument often raised in criticising multiple analysis models, i.e. the argument that multiple analyses create too great a short-term memory burden to be viable. They propose that, when faced with a TASS, every individual initially represents two analyses, and that the individual's working memory capacity determines the duration for which both analyses are maintained. Individuals with a high capacity working memory are assumed to maintain both analyses for a longer time than individuals with a low capacity working memory, who will abandon one analysis, the less activated one, more quickly. Thus, subjects with high working memory capacity should perform worse in processing sentences with syntactic ambiguities than subjects with low working memory capacity, because they maintain multiple analyses for longer. They suggest that there is a double cost for maintaining multiple analyses for a long time: firstly, the authors simply assume that maintaining multiple analyses is more costly than maintaining just one analysis; secondly, they propose that maintaining multiple analyses uses up resources required to execute other processes of sentence
comprehension, such as constructing a referential representation of the sentence. Such higher-level processes have to be postponed and time-consuming catching up occurs towards the end of the sentence.

An individual's working memory capacity was defined operationally as her score on the "reading span task" in which subjects read aloud a set of unrelated sentences and had to recall the last word of each sentence of the set. The number of items in a set ranged from 2 up to 6 or more sentences, depending on the subject's performance. Low-span subjects recalled an average of 2.5 or less words and high-span subjects an average of 3.5 or more words. It is not clear what the reading span task measures. The

"rationale behind the test is that the comprehension processes used in reading the sentences should consume less of the working memory resources of high span readers. These readers would thus have more capacity left to hold the final words of the sentences" (Just & Carpenter, 1992, p. 125).

Macdonald et al. (1992) assume that the reading span measure reflects quite different processes than the digit span measure or the ability to recall a list of unrelated words. It seems somewhat surprising, however, that reading span and digit span should be completely unrelated: if a subject had a very low digit span, she would be unlikely to remember many of the sentence-final words in the reading span task, however good her sentence comprehension processes might be. The authors report high positive correlations with other measures of reading comprehension. It is unclear why Reading Span correlates with sentence comprehension measures but Digit Span does not, given that both probably involve the use of the articulatory loop. Such a correlation is also surprising given neuropsychological evidence that short term memory deficits as assessed by the Digit Span measure do not interfere with sentence comprehension (Shallice, 1988).

Macdonald et al. (1992) tested their model with an experiment involving temporarily ambiguous sentences containing a main verb whose past tense form and past participle form are identical thus allowing both a main verb analysis (example a) and a reduced-relative clause analysis (example b):
a) The experienced soldiers / warned about the dangers / before the midnight / raid.

b) The experienced soldiers / warned about the dangers / conducted the midnight / raid.

(Slashes indicate regions used in data analysis). Unambiguous control sentences for a) contained a verb which allowed only the main-clause analysis either because its past tense form is different from its past participle form or because it is strictly intransitive (eg. 'giggled'). Control sentences for b) were created by adding a relative pronoun and an auxiliary. Examples a’) and b’) serve as illustrations:

a’) The experienced soldiers spoke about the dangers before the midnight raid.

b’) The experienced soldiers who were told about the dangers conducted the midnight raid.

Subjects read sentences presented on a computer screen using a word-by-word non-cumulative self-paced reading paradigm (moving window paradigm).

Let us first focus on sentences such as a) above, in which the main clause analysis turns out to be correct. There were no effects in the first analysed region (eg. 'warned about the dangers'). In the next region, there was an ambiguity effect for all subjects, which the authors interpreted as indicating that both subject groups built up multiple analyses in the ambiguous sentences. On the last word (eg. 'raid') high-span subjects took longer than low span subjects in the ambiguous as compared to the unambiguous sentences. The authors conclude from this that low span subjects had given up one of the multiple analyses, whereas high span subjects still maintained both. Furthermore, high-span subjects’ reading times for ambiguous (but not unambiguous) sentences increased throughout the regions and were highest at the last word, which the authors attribute to the subject having to catch up with higher-level processes which were delayed because multiple analyses had been maintained.

Let us now turn to the ambiguous sentences in which the analysis turns out to be
correct (such as b) above). The authors report an interaction of Ambiguity and Region (ambiguous sentences produced longer reading times for regions 2 and 3, but not for region 1). Interestingly, however, there was no interaction of Ambiguity with Reading Span.

Thus, the effect of reading span seems to have been limited to the main clause sentences where a high reading span led to longer reading times on the sentence-final word.

It is somewhat surprising that there is never an increase in reading times in the region which contains the ambiguous verb. If multiple analyses are maintained, then surely they should be maintained immediately after the ambiguous verb occurs, and the processing cost, which the authors claim is caused by maintaining multiple analyses, should be measurable as soon as the ambiguous verb occurs (or at least by the time three more words are encountered, i.e. by the end of the region in which the ambiguous verb occurred). The authors do not address why this burden is not evident earlier, i.e. before the fourth word following the ambiguous verb.

The ambiguity effect in region 2 reported for the ambiguous main verb sentences for both subject groups seems to me to be the piece of evidence potentially most damaging to a single analysis model. A single analysis model predicts that only the (preferred) main clause analysis is maintained, and since there is no disambiguating information in this region, no extra processing should be required. The multiple-analysis model, however, assumes that maintaining two analyses (even if one is activated to a smaller degree than the other) creates a "burden" which should show up as increased reading times.

Macdonald et al. (1992) report a follow up experiment (in which only high span subjects participated) in which only ambiguous main verb sentences (such as a) above) were presented, but no ambiguous relative clause sentences (such as b) above). This experiment was conducted to rule out the possibility that subjects had been induced to generate multiple analyses upon reading the first verb because of the presence of the
ambiguous relative clause sentences in which the unpreferred analysis turns out to be correct. Again, reading times in region 2 (e.g. 'before the midnight', see example a) above) were longer in the ambiguous than the corresponding unambiguous sentences.

However, there seems to be still another alternative interpretation of the increase in reading times in region 2 of the ambiguous as compared to nonambiguous main verb sentences. The increase might reflect comprehension processes which were induced by the questions asked after each sentence. Note that these yes/no questions were designed "to assess which interpretation was assigned to the first verb" (Macdonald et al., 1992, p. 64). The sentence 'The experienced soldier warned about the dangers before the midnight raid', for example, was followed by the question: 'Did someone tell the soldiers about the dangers?' for which (presumably) the correct answer is 'no', since 'warned' was here used as a transitive verb, i.e. the soldiers warned somebody else rather than being warned themselves. Such questions might have alerted the subjects (especially the high span readers who can be assumed to have good verbal skills!) to the possibility that 'warned' could be analysed as a past participle introducing a reduced relative clause. Other studies typically ask simple yes/no questions on about 25 % of the trials mainly to ensure that subjects pay attention to the sentences. These questions do not normally focus on the critical part of the sentence. Thus, it is possible that the multiple analyses in region 2 might have been maintained because of the potentially suggestive comprehension questions after each trial.

One has to note, too, that the effect in region 2 was small (17 ms on average) and that only an analysis by- subjects but not by-items was reported.

The authors dismiss earlier evidence in favour of single analysis models by claiming that these studies must have used a sample of subjects who happened to be medium- or low- span readers: "such readers are more capacity limited and, consequently, will conform more closely to the predictions of a single representation model" (Macdonald et al., 1992, p. 89). The authors do not discuss, however, why their own low span readers’ performance in region 2 of the main-clause sentences supported the multiple analysis model, whereas this was not the case in other researchers’ samples.
In conclusion, it seems safe to accept the authors' summary that "there is a cost incurred in processing ambiguous sentences" (Macdonald et al., 1992, p. 76). There is no clear evidence, however, that reading span determines whether or not an ambiguity effect occurs. Neither can one safely conclude that the increase in reading times for ambiguous sentences indicates extra processing costs ensuing from multiple analyses; these increased reading times might merely reflect experimentally induced strategies of generating multiple analyses in order better to be able to answer the comprehension questions following each trial.

The capacity constrained parsing model has also been tested in a study focussing on the interaction of syntactic and nonsyntactic information which was modelled on Ferreira & Clifton's (1986) experiment. This study will be reported in section 7.4.

7.2.4. Ford's Lexical Preference model

Ford's model of a top down parser (Ford, 1989; Ford, Bresnan & Kaplan, 1982) primarily aims to explain how subjects come to prefer a particular syntactic analysis when processing a syntactically ambiguous sentence. They discuss sentences in which the preferred syntactic analysis does not need to be reanalysed, sentences in which plausibility or syntactic evidence forces cost-free reanalysis and, quite briefly, sentences in which reanalysis leads to garden path effects.

In the following section, I will first discuss Ford et al.'s account of structural preferences and cost-free reanalysis, and will then review how their model handles garden path effects.

Ford et al. point out that the occurrence of closure effects is affected by the lexical items in a particular sentence. For example, the two sentences given below are structurally ambiguous in the same way, but the preferred analysis, indicated by square brackets, differs between them:
The empirical evidence for subjects’ preferred syntactic analyses comes from a paraphrase task given to 20 subjects. Subjects had to read ambiguous sentences which were presented one per page. After having read and interpreted a sentence, they were to turn a page and choose one of two paraphrases of the sentence as the one that corresponded to their interpretation.

In Ford et al.’s model the lexical entry of verbs such as 'want' and 'position' contains two "lexical forms", i.e. two alternative predicate argument structures in which the grammatical functions are specified (Bresnan, 1982, p. 149). The verbs 'want' and 'position', for example, are represented by the following lexical forms, with the preferred lexical form given first:

a) 'want((Subj),(Obj))'
   'want((Subj),(Obj),(PComp))'
   (where PComp stands for prepositional complement)

b) 'position((Subj),(Obj),(PComp))'
   'position((Subj),(Obj))'

Ford et al. (1982) suggest that "the various lexical forms of a given verb have different "strengths" or "saliences", and that the strongest form somehow determines the preferred syntactic analysis" (Ford et al., 1982, p. 745, quotation marks in orig.). What exactly determines the strength of a particular lexical form is not specified, but the authors speculate that it might be affected by frequency of usage. They favour the hypothesis that the strengths are encoded in the lexicon (thus giving rise to uniformity of judgements from different subjects) rather than being computed on-line in the immediate sentence context in which the verb occurs. However, they also give an example where the strength of a lexical form depends on the sentence context. Again using the off-line paraphrase task, they established that subjects preferred the attachment of the PP as a modifier of the NP in the next example:
When he arrived at our doorstep, I could see that Joe carried [a package for Susan].

In a sentence without the introductory subordinate clause, eg:

Joe carried the package for Susan.

the analysis of the PP as modifying the verb (carried [the package] [for Susan]) was preferred, and thus defined to be the stronger lexical form.

The authors assume that in the emerging syntactic representation of a sentence all alternative lexical forms are represented together with the verb, ranked according to their strength. The syntactic analysis of a sentence is affected by the strategy called Lexical Preference: (the authors call this strategy a 'principle', but I will refer to it as a strategy and use the term 'principle' only to refer to grammatical principles such as those formulated in the Government and Binding theory).

"If a set of alternatives has been reached in the expansion of a phrase structure rule, give priority to the alternatives that are coherent with the strongest lexical form of the predicate." (Ford et al., 1982, p. 747).

The notion of re-analysis is first introduced in passim during the authors’ discussion of the sentence:

Although the women talked about John’s killing his very old father on many occasions they repressed how much they disliked the old man.

The initial preference according to the authors is to close the lower VP ('killing') late, i.e. to attach 'on many occasions' as a modifier of the verb 'kill'. The PP 'on many occasions' is, however, "incompatible" as a modifier of the VP 'killing his old father', and "reanalysis will take place" (Ford et al., 1982, p. 760). Given the off-line nature of the paraphrase test of readers' preferences, it is not clear how quickly this incompatibility is noted and to which parts of the sentence subjects direct their attention during reanalysis. Their evidence only shows that in the preferred final analysis of the sentence the lower VP is closed early ([killing his old father]), and the
PP is attached as a modifier of 'talked'. Note that in this example, attachment and reattachment of the ambiguous PP ('on many occasions') is not affected by the strategy of Lexical Preference since the lexical form of 'kill' does not contain a PP as argument. Initial attachment in these sentences results from particular syntactic strategies. The authors discuss other cases of cost-free reanalysis which involve lexical reanalysis, defined as rejecting the strong lexical form of a verb in favour of the weak lexical form to allow building up a grammatical representation of a sentence. In the following example, reanalysis is prompted by syntactic evidence late in the sentence:

The tourists signalled to the guide that they didn’t like.

According to the authors' evidence, the strong lexical form of 'signal' is ((Subj),(toObj),(SComp)). According to the Lexical Preference strategy, this lexical form will determine the initial syntactic analysis as: 'signalled [to the guide] [that they didn’t like]'. When 'like' is encountered (presumably together with the full stop), however, which requires an object, reanalysis has to start during which 'that they didn’t like' is analyzed as a reduced relative clause modifying 'the guide' (which is in line with the weaker lexical form of the verb, i.e. ((Subj),(toObj))).

The authors claim that their theory "predicts" that the above sentence is "relatively easy to analyze, although there would be slight difficulty and a subsequent reanalysis of part of the sentence when the last word is encountered" (Ford et al., 1982, p. 761,762). None of the examples presented so far are assumed to create a garden path effect. The authors maintain that "it seems erroneous to assume that reanalysis of phrasal structure during backup must always cause conscious garden paths" (Ford et al., 1982, p. 763,764). They propose that garden paths occur only under special conditions, one of which is characterized in the following passage:

"once a functional structure for a complete constituent has been recovered, it is very difficult to give a new morphosyntactic analysis of a word within it" (Ford et al., 1982, p. 763).

The authors introduce the term recategorization for this change of morphosyntactic analysis. In a further attempt to distinguish reanalysis without garden paths from
reanalysis with garden paths they propose:

"Reanalysis of phrasal structure without the need to recategorize a word may increase the complexity of local parsing decisions as measured by reaction times, but without the perception of a garden path (...), while reanalysis which requires a new morphosyntactic analysis of a word within the functional structure of a completed constituent may cause a conscious garden path." (Ford et al., 1982, p.764).

Recategorization might involve an actual syntactic category shift, eg. from adjective to noun as in:

The boy got fat melted.

In this sentence (at least in American English), 'fat' is initially incorrectly interpreted as an adjective, and a "complete and coherent functional structure for the simple sentence 'The boy got fat' (i.e. 'The boy became fat') would be recovered" (Ford et al., 1982, p. 763). During reanalysis, prompted by the requirement to integrate 'melted' into the sentence, 'fat' would have to be categorized as a noun. (Note that to test if the two morpho-syntactic forms (eg. of 'fat') are two separate lexical items, or whether they are represented as two entries in one lexical item, one could set up a discourse context which would favour the noun-reading and test whether the garden path effect is avoided in 'The boy got fat melted.' If the two variants are listed under the same lexical item, they should both be primed, and context should not be able to help avoid the garden path effect; alternatively, one could test if the garden path effect is increased if the previous context mentioned a word associated with 'fat' as an adjective).

Ford et al. discuss only a limited range of sentences leading to garden paths (i.e. 'The boy got fat melted' and 'The girl pushed through the door laughed' which is structurally equivalent to the classic example 'The horse raced past the barn fell'). It is unclear how their model could account for other garden path sentences, for example those involving verbs which are optionally intransitive such as

While Ron was sewing the shirt slipped to the ground.
In this sentence, the presumably stronger lexical form ((Subj),(Obj)) has to be replaced by the weaker form ((Subj)). This is an instance of what the authors call Lexical Reanalysis (and, by implication, not an instance of recategorization) and as such it should not cause a garden path.

The problem is that the notion of recategorization is not clearly defined in Ford et al.'s theory. The change from the past tense intransitive active version of 'race' to the past participle of its transitive active version is considered to be an instance of recategorization (as required in 'The horse raced past the barn fell'). These two versions of the verb are assumed to be represented as alternative lexical forms in the lexical entry of the verb 'race'. Note, however, that alternative argument structures and their grammatical functions are also represented by alternative lexical forms. There seems to be no reason why changing from one lexical form to the other should create a garden path in the 'raced'-example but be cost-free in the 'sew'-example.

Ford et al. further suggest that garden paths only occur when the functional structure of a completed constituent has been recovered and needs to be changed. This proposal also fails to correctly predict a garden path in the 'sew' example. To show this, a slightly more detailed analysis of the 'sew' example is required.

Let us assume again that ((Subj),(Obj)) is the stronger lexical form of 'sew'. The syntactic Final Argument strategy would presumably lead to a delay in attaching 'the shirt' into the VP to ensure that the parser can check whether any subsequent input can be incorporated into the NP (the Final Argument strategy was originally invoked to account for late closure in the sentence 'The woman wanted [the dress on that rack]'). Instead of directly attaching 'the shirt' as the direct object of 'sewing', the parser would initially attach it as the head of a complex NP, and would expect, given the grammar rule for complex NP nodes, to find either a PP, a sentence, a VP or an Adjectival Phrase as the next constituent. The Syntactic Preference Strategy would not give any clear preference for any of the options (Ford, 1989), but the next word in the input, ('slipped'), is incompatible with the complex noun attachment. Ford et al.'s model would predict that the necessary reanalysis is cost-free since the initial analysis
did not create a complete constituent (the complex NP with 'the shirt' as head was never completed) which was not yet attached as the direct object.

7.3. The role of lexical subcategorisation biases in parsing

Some single-analysis models maintain that the initial parse is assigned on the basis of subcat preferences of individual verbs, rather than some preference for a particular syntactic structure, as suggested in Frazier’s garden path model. The lexical entry of a verb is assumed to provide information about the preferences with which the verb occurs with its possible subcategorisation frames. The preferential ordering might be due to differential frequency of occurrence of each subcat frame (Ford, 1989; Ford et al., 1982; Holmes, Stowe & Cupples, 1989; Mitchell & Holmes, 1985).

The proposals are quite vague regarding the stage of syntactic analysis during which the subcat preferences are assumed to have an effect on the parser: the effect is supposed to be "at an initial stage of processing during reading" (Holmes, Stowe & Cupples, 1989, p. 686) but "it remains to be determined whether or not they [=lexical expectations] actually guide initial parsing decisions or merely act as a filter after the application of more general parsing strategies" (ibid.).

Holmes (1987, experiment 2, and presented as experiment 1 in Holmes, Stowe & Cupples, 1989) reports an effect of lexical subcategorisation biases for sentences containing verbs that take either a direct object NP or a sentence complement. The verbs allowed both the NP-reading and the complement-sentence reading (i.e. there were none that were strictly intransitive), but they differed in terms of the likelihood with which each of these subcategorisation frames occur. Holmes first conducted a pretest in which 39 subjects had to complete sentence fragments such as 'She believed ...', 'They found ...', 'He noticed ...' (mixed with a number of sentence fragments containing verbs which do not take any clausal complements). She selected 16 verbs that had prompted a direct object continuation "much more often" (Holmes (1987, p. 590) than a clausal continuation (eg. 'see', 'hear', 'read', 'answer') and 16 verbs for
which the pattern was opposite (eg. 'doubt', 'claim', 'know', 'believe'). In the main
experiment, each of these verbs were embedded in reduced and non-reduced clausal
complement sentences (eg. 'The reporter saw [that] her friend was not succeeding').
These sentences, and filler sentences of different syntactic structure, were presented in
a word-by-word self-paced cumulative reading task, in which the subject pressed a
button after each word when she thought that the sentence could continue
grammatically (that is, the subject made a grammaticality judgement after each word).
(Half of the trials were filler trials which contained syntactic violations or were
syntactically incomplete). The results were clear-cut: when sentences containing verbs
with object-NP bias were encountered, decision times on the auxiliary in reduced-
complement sentences were significantly longer than decision times on auxiliaries in
sentences containing that-complements (the average difference was 403 ms; the
average decision time for aux in the reduced sentences was around 1100 ms, and in the
sentences containing 'that' around 700 ms). The decision times for the next word
('not' in the above example) did not differ.

The results were quite different for the sentences containing verbs with a preference
for clausal-complement continuations (such as 'The candidate doubted [that] his
sincerity would be appreciated'). Firstly, decision times for the determiner and the
noun of the ambiguous NP were significantly longer in the reduced as compared to the
non-reduced sentences. Holmes interprets this as indicating that subjects reading the
reduced sentences expected a 'that' as an introduction of a sentential complement after
having read a verb with a subcat preference for sentential complements. Secondly,
there was a garden-path effect, but it was much smaller than in the sentences with NP-
bias verbs. Here, the difference in decision times at the critical auxiliary (the
disambiguating word) was increased in the ambiguous reduced complement sentences
by (on average) only 110 ms (which was significant by items and by subjects). The
average decision for aux in the non-reduced sentences was about 680 ms, and in the
reduced sentences about 800 ms. The author argues that subjects initially assumed a
sentential-complement analysis (suggested by the verbs' subcat preferences), but that
they gave up this hypothesis when no 'that' occurred after the verb, which would have
been the most likely introduction of a clausal complement (whether 'that' really is the
most likely beginning of a sentential complement has not been empirically ascertained; in fact, Frazier (1987a, p. 566) claims that "sentential complements usually begin with subject noun phrases", i.e. not with 'that'). Instead of 'that' there is an NP, which is analyzed as a direct object, so that upon encountering the aux, a garden path occurs. The pronounced drop in the garden path effect (for verbs which prefer a complement-clause continuation as compared to verbs which prefer an object-NP continuation) is interpreted as indicating that subjects reading the sentences with clausal-bias verbs "re-establish[ed] a structural option that has already been tested", i.e. the clausal-complement analysis that they initially expected after reading the verb with clausal-complement bias. Subjects reading sentences with NP-bias verbs had to "chang[e] to an entirely new option" (Holmes, 1987, p. 593) when encountering aux, which turned out to be more time consuming.

Holmes et al. (1989) ran a follow-up experiment (experiment 2) which differed from the one just reported only on the task the subjects were asked to do. Rather than making a grammaticality judgement on each word, they only had to press a pace button in this version of a cumulative self-paced reading paradigm, and they were told that they should try to remember the sentence so they could repeat it after it had been presented (which they were asked to do on every 4th trial). The results now showed no garden path effect at all for the reduced sentences with clausal-bias verbs (and no differences at the determiner- and noun- position either). The reduced sentences with NP-bias verbs did lead to a garden path, which was evident not on the disambiguating aux, but on the following word (average reading time on this word was around 580 ms in the reduced sentences, and around 470 ms in the non-reduced sentences (note: since the means are not reported they have to be estimated from the figures)).

The effect of subcat preferences on sentence processing shown by Holmes et al. (1989) was further refined in their third experiment which varied the length of the ambiguous noun phrase (in this experiment subjects had to answer a yes/no question after each trial; the display of words was non-cumulative, and a self-paced paradigm without grammaticality judgements was employed). Compare a) and b) below which show the short and the long version of an ambiguous sentence with an NP-bias verb:
a) The reporter saw the woman was not very calm.

b) The reporter saw the woman who had arrived was not very calm.

Length of the ambiguous NP did not make any difference in sentences with NP-bias verbs: in long and short sentences there was garden pathing as evidenced by longer response times on the disambiguating word (the aux) in the ambiguous sentences as compared to nonambiguous sentences (which contained the complement 'that').

With regard to garden pathing in sentences with clausal-bias verbs, however, length was important. Examples a) and b) below show the short and the long version of an ambiguous sentence with a clausal-bias verb:

a) The passenger claimed the luggage should be given to him.

b) The passenger claimed the luggage which we took should be given to him.

When the ambiguous NP was short, there was no garden-pathing (replicating the results Holmes et al.'s (1989) experiment 2 in which only short sentences were used); when the ambiguous NP was long, however, garden-pathing did occur (reading times on aux which were on average 47 ms longer in the reduced than in the non-reduced sentences). How can this length effect be accounted for?

Holmes et al. (1989) suggest that garden-pathing in sentences with clausal-bias verbs and long ambiguous NPs is caused by a switch from an initial clausal analysis consistent with the verb's subcat bias, to the direct object analysis. They suggest two tentative explanations. Firstly, they propose that if an NP remains unattached for too long (awaiting the verb of the putative new clause) memory storage problems might arise. This would then lead "the reader to question the current structural hypothesis" (Holmes et al., 1989, p. 683). A re-analysis would be done guided by the verb's second (unpreferred) subcat frame, i.e. the transitive frame. This reanalysis will turn out to be incorrect later in the sentence. The problem with this account is that the NP need not remain "unattached", waiting for the verb which confirms its analysis as the subject noun of a complement-clause. If the verb has a preference for the clausal-bias
subcat frame, the NP should initially be attached as a daughter to a newly postulated sentence node even when further supporting syntactic evidence is not immediately received. Being attached in this way should not cause any greater memory overload as an NP being attached as a direct object.

As an alternative interpretation the authors propose that subjects might expect that NP-subjects of embedded sentences do not extend beyond two or three words, due to English being a predominantly right-branching language in which embedded constituents are more likely to occur towards the end of sentences rather than in subject position. This expectation might lead to the rejection of the initial attachment of the long ambiguous NP as the subject of a sentential complement in the long sentences with clausal-bias verbs (whereas a long direct-object NP is to be expected in a right-branching language). This is an interesting interpretation. It is important, however, to specify how a general expectation about the length of subjects in a right-branching language is represented in a way which affects the operation of the parser.

As an alternative to Holmes et al.'s proposals I would suggest that the strength of the bias for one of the two subcat frames might be of importance in determining the case with which the initial analysis is given up. In Holmes et al.'s pretest, the NP-bias verbs co-occurred with direct-object continuations in 70% of the responses, whereas for the clausal-bias verbs, sentential clausal complements occurred in only 55% of the responses. If the parser, looking up the subcat information of a verb, does not find a clear preference for one of two (or more) subcat frames of the verb, it might do two things to fit a temporarily ambiguous NP into its analysis: either it applies some default syntactic preference, such as minimal attachment: in Holmes et al.'s experiment this did not happen since there was no garden path for sentences such as 'The passenger claimed the luggage should be given to him'. Alternatively, in the absence of a strong subcat preference, the parser could make use of a look-ahead facility: it delays the attachment of the ambiguous NP and looks ahead for some words (only about three, according to Marcus, 1980); in Holmes et al.'s short condition, a verb (or aux) is found within this window (eg. 'should') which unambiguously forces the sentence-complement analysis. If the NP is long, however, (as in 'the luggage which
we took'), no information is found that forces the complement-analysis, and the parser opts for the only other syntactic analysis possible, given phrase structure rules: it attaches the ambiguous NP as a direct object.

A question that Holmes et al.'s data do not conclusively answer is whether the length of the ambiguous NP or its structure underlies the length effect. The ambiguous NP was always lengthened by adding a subsequent qualifying phrase (eg. '... claimed the luggage which we took ...'). One could, however, instead add a preceding qualifier, such as in 'The observer forgot the extremely narrow outcome would be crucially important'. In the first case, the beginning of the NP gives no clue as to its length, whereas in the second case it is clear immediately after encountering the determiner that the NP is going to be long.

Mitchell & Holmes (1985) also report an effect of lexical preference for subcategorization frames on garden pathing. They looked at a variety of syntactic ambiguities, also employing the self-paced reading technique (presenting phrases rather than words at a time). However, self-paced reading, especially with phrases as units of presentation, might not be sensitive enough a task to tap into the initial syntactic processes; it may reflect initial parsing and subsequent reanalysis, and does not make possible a decision about the exact point in sentence processing at which lexical effects took place.

Mitchell (1987) reports findings which he interprets as demonstrating the existence of two submechanisms, both of which he includes in the parser. One, the assembler, blindly proposes syntactic structures, the other, the monitor, filters and revises them. The monitor is assumed to have access to detailed lexical information as well as semantic and pragmatic information. Mitchell presents an experiment intended to show that only the monitor has access to subcategorisation information.

An example of the type of sentence he studied is the following:
a) After the child had visited the doctor prescribed a course of injections.

The verb 'visit' can take a transitive or an intransitive frame. If, as the principle of late closure would imply, the following NP is attached as a direct object, garden pathing occurs since then 'prescribed' cannot be fitted into the phrase structure. Sentences like a) were compared to sentences like b) which contained a strictly intransitive verb:

b) After the child had sneezed the doctor prescribed a course of injections.

The data come from a phrase-by-phrase non-cumulative self-paced reading task. Sentences were split into two parts: the first part ended after the ambiguous NP (i.e. after 'the doctor') (Mitchell included another condition in which the first part ended after the verb, but this is not relevant for the present discussion). Each trial started with a prompt. When the subject pressed the space bar, the first part of the sentence was presented left-justified on the screen. The subject was instructed to press a button after reading it, which triggered the first display to be replaced by the second part, which had a full-stop included at the end. There were 24 experimental sentences and 36 filler sentences made up of a variety of syntactic structures. Yes/no comprehension questions were asked after 12 of the filler trials (none after the experimental trials).

There are two relevant results. Firstly, reading times for the second part of the sentence were significantly longer if the sentence contained verbs which can be both transitive and intransitive as opposed strictly intransitive verbs (the means are 3346 ms versus 2354 ms). This Mitchell interprets as indicating a garden path effect for the transitive sentences but no garden path for the sentence contained intransitive verbs. Mitchell further reports that the reading times for the first part of the sentence were significantly longer when the verb was strictly intransitive than when it allowed a transitive and an intransitive analysis (3449 ms versus 2740 ms on average). He takes this finding as reflecting time-consuming monitoring- and reanalysis processes. He assumes that the assembler churned out the direct object analysis (in line with the principle of late closure), blind to lexical subcat information, and that this output was then checked by the monitor against lexical information. The assembler would then have to generate another proposal. However, while this interpretation might be
perfectly adequate, longer reading times per se do not help to distinguish between early (blind syntactic) and later (monitoring) processes. Longer reading times might reflect that it takes longer to create a new node in the syntactic tree (a sentence node to which the ambiguous noun is attached) than to attach an NP as a direct object to an already existing VP node. The data do not help to decide between these two interpretations.

Ferreira & Henderson (1990) report that in their studies lexical subcat information (either verb plus NP or verb plus complement clause) did not affect garden path effects. They directly compared data from eye measurements and two word-by-word self paced reading studies, one with non-cumulative, the other with cumulative displays. The data of the latter were suspect and will not be discussed. The stimulus materials (80 experimental sentences and 72 fillers of a variety of sentence types) were identical in these three experiments. Comprehension questions were asked after 20% of the trials throughout the experiments. The syntactic ambiguity was due to two readings for the second NP given that the verbs had two subcat frames. The NP (‘Pam’ in the following example) could be analysed either as an object NP, or as a subject NP of a sentence-complement:

He forgot Pam needed a ride with him.

The experiments compared sentences of the form subject plus verb plus sentence complement which either did or did not contain the complementizer 'that'. The crucial data come from measurements at the disambiguating region (here 'needed').

In order to test whether subcat information affects parsing, two types of main verbs were used: type-a verbs which rarely take a direct object (as 'wish') (some of these verbs were strictly intransitive) and type-b verbs which allow both the direct-object and sentence-complement readings (as 'forgot'). The classification of verbs into these two types was done on the basis of the experimenters’ intuitions and of norms collected by Connine, Ferreira, Jones, Clifton & Frazier (1985). If verb information is used, garden paths should only arise for type-b verbs.
For data from the eye movement study (first fixation duration and total reading time, i.e. the sum of first fixations and refixations) on the disambiguating word, there was a significant main effect of complementizer presence, with more time spent when the complementizer was absent as compared to present. Crucially, there was no interaction of complementizer presence (absent versus present) and verb type (type-a versus type-b). There were no other effect for first fixation duration. For total reading time there was a hint of an increase on the ambiguous NP (e.g. 'Pam') when the complementizer was absent as compared to when it was present (significant by items, but only p < .07 by subjects). There was no interaction with verb type. Since this effect was not significant for the first fixation measure, it must be due to subjects' refixating the ambiguous word after they were garden-pathed on the disambiguating word.

The results of the non-cumulative word-by-word self-paced reading task can be summarized as follows:

1. there were no effects on the ambiguous word
2. on the disambiguating word, there was a significant complementizer effect (shorter reading times if 'that' was present); again there was no interaction with verb type
3. on the first word after the disambiguating word there was a continued complementizer effect and a significant interaction: reading times were highest when 'that' was absent and the verb was of type b (i.e. allowing both readings)
4. The mean total reading time for a whole sentence was 3388 ms. Complementizer presence was a significant effect, and so again was the interaction.

Thus the second experiment confirmed the findings of the first, indicating that verb bias for subcat frames cannot block the garden-path effect. The slower measure obtained in experiment 2 indicates that verb bias is used in some later processes, presumably of reanalysis, such that re-analysis is faster when the verb is biased towards the sentence-complement subcategorisation frame. The evidence of verb type affecting reanalysis was stronger in experiment 2 than in experiment 1 (in E1 there was only a marginal effect of verb type on total reading times of the post-disambiguating word, and some hints that more regressions occurred back to the disambiguating word if the main verb was of type-b, whereas in E2 the interaction
verb type and complementizer presence was significant on the word following the
disambiguating information, and for the mean reading times for the whole sentence).
This is plausible, suggest the authors, if one assumes that verb bias is relied upon more
if re-analysis is more difficult, as in the non-cumulative self-paced reading task, in
which subjects cannot look back at previous words of the sentence.

Thus, this evidence quite convincingly showed that the minimal attachment principle
determined (blindly) which syntactic structure is initially built up, regardless of
subcategorisation biases of individual verbs. Garden-pathing was found if the
complementizer 'that' was absent, irrespective of the subcategorisation biases of the
main verb. Re-analysis of a sentence, after garden-pathing occurred, was aided by the
verbs' biases since reading times in later sections of the sentence, and total sentence
reading times, were shorter if the verb had a bias not to take object-NP continuations.
The authors make a strong case for using on-line measures since otherwise only the
final interpretation of a sentence is tested which, as most would agree, takes into
account non-syntactic factors. The controversial issue remains the timing of the use of
different sources of information during sentence processing.

Clifton et al. (1984, experiment 1) found evidence that subcat preferences affect
processing difficulty as measured by response latencies on a secondary task (lexical
decision). In a pretest the preferences of optionally transitive verbs to occur either with
a transitive or an intransitive subcat frame were established. In the main experiment
each verb was presented in both a transitive and an intransitive sentence frame, eg. 'to
read', which is a predominantly transitive verb, occurred in:

The babysitter read the story to the sick child.
The babysitter read to the sick child.

Sentences were presented one word at a time. Each word was presented centrally on
the screen for 300 ms, with a blank of 50 ms between each word. The letter string on
which the lexical decision had to be made was presented at the far left of the video
screen, surrounded by asterisks. The target was always an unrelated real word when
the test stimuli were presented; nonwords were only presented for some of the filler sentences. The target for the lexical decision was presented immediately following the first word after the verb: this word unambiguously signals either a transitive subcat frame (e.g. a determiner or a possessive pronoun signals an NP) or an intransitive subcat frame (i.e. a preposition). Response times on the LDT were significantly faster when the actual sentence frame coincided with the subcat preference of the verb than when it did not (mean response times a) for transitive sentences: transitive subcat preference: 908 ms, intransitive subcat preference: 1008 ms; b) for intransitive sentences: intransitive subcat preference: 877 ms, transitive subcat preference: 1000 ms). If lexical decisions are taken to reflect immediate processing (see e.g. Tyler, 1989), this result would provide clear evidence that lexical subcat information is used to affect initial parsing decisions: the parser builds up the syntactic structure that fits the lexical preference, and has to make a reanalysis when further input proves the initial structure to be incorrect. Clifton et al. (1984), however, conclude that the data "do not, for instance, constitute evidence for a lexically driven parser" (ibid., p. 706). Alternative mechanisms they suggest would make use of the preferred subcat frames only for checking the analysis that the parser has independently produced; or there might be a mechanism which labels syntactically predicted nodes with lexically determined probabilities.

7.4. Plausibility and semantic constraints in parsing

Holmes et al. (1989, experiment 1 and 2) addressed the question whether a structurally ambiguous NP is more likely to be analyzed as a direct object rather than as a subject of a clause-complement when its plausibility as a direct object is very high, even if the verb predominantly subcategorises complement-clauses (see previous section for details of the subcategorisation biases). Plausibility was established in a pretest in which subjects rated 32 declarative sentences on a 5-point scale. Half of the declarative sentences were designed to have implausible objects. The pretest produced clearly distinct ratings for high versus low plausibility object NPs (results were reported separately for NP-bias and clausal-bias verbs: average ratings for plausible NPs was
4.5 and 4.3, and for implausible NPs 2.2 and 2.0, respectively). Apart from these operational definitions, the notion of plausibility receives no further explanations. An inspection of the stimuli shows that while some NPs are implausible as direct objects according to one's experience with the world, others are violations of semantic selection restrictions of the main verb. These two types of NPs, however, might be processed in quite different ways, a possibility which will be addressed in my Experiments 4 and 4b.

As described in detail above, Holmes et al. used a cumulative word by word self-paced reading task with continuous grammaticality judgements. Words were presented across the screen starting from the left-hand side. Subjects had to press a yes-button when they thought the sentence could continue grammatically. This would trigger the display of the next word. Otherwise, they would press a no-button. After false negatives, stimulus presentation continued but response times for the remaining words were excluded from the data analyses.

Subcat bias had a clear effect on garden pathing in ambiguous sentences such as 'The reporter saw her friend was not succeeding' (see previous section): reading times were longer on the disambiguating word (eg. 'was') when the main verb 'was' was biased towards a direct object analysis, but not when the main verb (eg. 'claim') was biased towards a complement-clause analysis. This effect did not interact with the plausibility of the ambiguous NP as a direct object. In particular, for clausal-bias verbs, plausibility did not increase reading times for the disambiguating word: garden pathing was avoided (or rather significantly reduced compared to NP-bias verbs) for sentences with clausal-bias verbs, and high plausibility of the ambiguous noun as a direct object did not 'lure' the parser into giving up the preference based on the verb's subcat bias to analyze the NP as the subject of the subordinate clause. The following example shows a clausal-bias verb with a plausible (a) and implausible (b) ambiguous NP:

a) 'The candidate doubted his sincerity would be appreciated.'

b) 'The candidate doubted his champagne would be appreciated.'
Equally important is the finding that, for NP-biased verbs, the garden path effect was not decreased when the ambiguous noun was implausible as a direct object.

In Holmes et al.’s (1989) experiment 2, in which the grammaticality judgement at each word was replaced with merely pressing a pace button to receive the next word, plausibility again had no effect on the size of garden-path effects.

Thus, Holmes et al.’s (1989) study provides only very weak evidence for plausibility affecting syntactic analysis. The size of the garden path effect was unaffected by the plausibility of the ambiguous noun as direct object.

There are some concerns which I think justify some caution before reaching any firm conclusions. As pointed out above, plausibility manipulations in this study involved a mixture of violations of real-world expectations and violations of semantic selection restrictions. Furthermore, 4 of the 16 sentences with NP-bias verbs had direct object nouns designed to be implausible which appear to have possible continuations turning them into compound nouns which would in fact be plausible direct objects (1. 'The secretary read the fashion ...' would be plausible if 'fashion magazine' was expected; 2. 'The doctor found the fever ...' - fever curve; 3. 'The salesman wrote the market ...' - market review; 4. 'The scientist showed the travel ...' - travel guide). Thus, plausibility might not have properly been manipulated.

In other studies, too, it is quite difficult to define clearly what is meant by semantic and pragmatic information.

A number of studies focused on the influence of semantic/pragmatic information in sentences that are temporarily ambiguous between a main clause analysis and a reduced relative clause analysis, such as:

The defendant examined by the lawyer turned out to be unreliable.

There is a preference, based on minimal attachment, to interpret the first verb as a main verb (this structure requires less syntactic nodes than the alternative reduced
relative analysis). When the disambiguating phrase 'by the lawyer' is read, which signals that the verb has to be analyzed as the verb of a reduced relative clause, reading times are typically longer as compared to reading times for the PP in the unambiguous, non-reduced version 'The defendant that was examined by the lawyer ...'.

An early study which focused on plausibility and garden pathing for such sentences was conducted by Rayner, Carlson & Frazier (1983). They explicitly did not want to commit themselves to statements about the exact nature of the plausibility factor, except to say that most likely both semantic and pragmatic information was involved. In their first experiment, they compared sentences such as a) and b):

a) The florist sent the flowers was very pleased.

b) The performer sent the flowers was very pleased.

Since performers are assumed to be more likely the recipients rather than the senders of flowers, the relative-clause analysis, which is the non-preferred analysis according to the minimal attachment principle, would be quite plausible for sentence b): thus, there should be no garden-path effect for sentences such as b). Eye movements were measured during sentence reading. To ensure subjects paid attention to the sentences, they were asked to paraphrase every fifth sentence after they read it. Subjects were not told in advance whether a sentence had to be paraphrased. While the data are quite clear, again the interpretation is more complicated. Reading times per character were longer on the disambiguating regions ('was very') of reduced relative clauses as compared to nonambiguous sentences for both plausible and implausible sentences. The average fixation durations on the word following the disambiguating region were also longer in the ambiguous relative sentences than the control sentences, again for plausible and implausible sentences, and the number of regressions from within or after the disambiguating region was greater for plausible and implausible reduced relative sentences as compared to the control sentences. The paraphrases were also analyzed. Correct paraphrases were given more often (88%) for plausible reduced relative sentences than for implausible ones (64%), which the authors interpret as
showing that the final interpretation of the sentences was affected by their plausibility.

Rayner et al.'s (1983) second experiment tested sentences which were structurally ambiguous, but in which both structural analyses were correct on purely syntactic grounds and both were compatible with later parts of the sentence. These sentences contained a pragmatic rather than a syntactic error message, as in: 'The spy saw the cop with the revolver ...'. The preferred syntactic analysis (according to the minimal attachment principle) attaches the PP ('with the revolver') to the VP, as further information specifying the act of seeing. Since pragmatically, 'revolver' is not a good candidate as an instrument used in the act of seeing, 'revolver' provides a pragmatic signal indicating that the initial syntactic analysis cannot be correct. The question is whether for this type of sentences the pragmatic information is used by the parser before it provides its initial syntactic analysis, or whether the pragmatic information can only be used after the first analysis, performed purely on the grounds of syntactic preferences, has been completed. This question was addressed by comparing sentences in which pragmatic plausibility coincided with the initial syntactic analysis preferred by minimal attachment versus sentences in which pragmatic plausibility coincided with the revised syntactic analysis, such as a) and b) respectively:

a) The spy saw the cop with binoculars ...
b) The spy saw the cop with a revolver ...

If pragmatic plausibility determines the initial analysis, reading times should not differ for the two types of sentences, because no re-analysis would be required in b) even though the attachment is the syntactically non-preferred one to the NP. The results did not confirm this: total reading times per character and first fixations per character were longer for sentences such as b) as compared to a), and first fixations were significantly longer on the noun in the PP than on preceding words for sentences of type b) but not a).

Thus, Rayner et al.'s (1983) experiment 2 indicates that the parser can be triggered to reanalyse a sentence even if there is no syntactic error but instead a pragmatic signal indicating an incorrect analysis. Both experiments further suggest that the parser does
not, however, use pragmatic information before it has computed its preferred structure. The authors suggest that there exists a syntactic and a thematic processor, which operate largely independently from each other. The syntactic processor computes only one analysis (the one that requires the least nodes in a syntactic tree); the thematic processor examines all thematic structures possible for a given verb and selects one by comparing their plausibility given the nouns in the sentence. Exactly how pragmatic plausibility (eg. of 'revolver' as an instrument of seeing) is evaluated is not specified. The authors also do not spell out how the outputs of the thematic processor and the parser are interfaced. In particular, is it because the two outputs clash that the syntactic processor re-analyzes the sentence, again guided by purely syntactic information? (Once the minimal attachment structure is rejected, the parser can only build up one other structure, given the phrase structure rules of the language). Or does the output of the thematic processor actually include information about the alternative syntactic structure that the parser can simply accept, without going through the input again?

The evaluation of Rayner, Carlson & Frazier's (1983) study is surprisingly varied, and obviously reflects the evaluator's own theoretical position. While, for example, Garrett (1990) takes the general outcome of this research to show that "minimal attachment applies even for cases in which semantic constraint militates against it" (ibid., p. 147), McClelland (1987) criticises the experiments and, after presenting some data of a study conducted by himself and Taraban (Taraban & McClelland, 1988), comes to the opposite (albeit cautiously stated) conclusion

"that the findings of Rayner et al. need not be interpreted as favouring any version of autonomous syntax hypothesis. Though syntactic cues are sometimes so strong that they overshadow semantic constraints, we find that under other conditions semantic constraints do appear to exert relatively immediate effects" (McClelland, 1987, p. 28).

McClelland (1987) criticises Rayner et al.'s (1983) first experiment on the grounds that the plausibility constraints introduced there were very weak. The implausibility of, for example, performers sending flowers to somebody else is actually not that great.

Rayner et al.'s results from their experiment 2 are put into a different light by
McClelland's claim that the materials contained "a consistent semantic bias in favour of the minimal completion" (McClelland, 1987, p. 24, 25). His evidence comes from a rating study which he conducted using Rayner et al.'s sentences. Subjects were presented with the sentences up to (and including) the preposition (e.g. 'The spy saw the cop with -') and were asked to generate an expectation for the next words. They were then presented with either of the completions used by Rayner et al. for this sentence and had to judge how well it matched their expectation. Completions that allowed the VP attachment (e.g. 'binoculars') were judged to be closer to their expectations than nouns requiring NP attachment (3.62 versus 2.9 on a 5-point scale).

In a new study, Rayner et al.'s sentences plus 20 new ones were presented. The new sentences were designed to produce expectations of NP attachment, as in 'I read the article in the magazine'. All sentences were presented in a non-cumulative word-by-word self paced reading task. After each sentence a question had to be answered (no details are given about the type of question). Rayner et al. (1983) had required their subjects to paraphrase the sentence on 25% of the trials. For the sentences adopted from Rayner et al., reading times on the last word of the sentences replicated the effects found by Rayner et al.: reading times were longer for PP's which fitted the NP attachment (e.g. 'He saw the cop with the revolver') as compared to those that required the VP attachment (e.g. 'He saw the cop with the binoculars'). However, the new set of stimuli exhibited exactly the reverse pattern of reading times. The same pattern of results was found in a follow-up experiment in which the sentences were extended beyond the PP and reading times were summed over the noun in the PP and the next three words. These findings provide evidence against the minimal attachment principle since the preference for a particular structural analysis was shown to vary depending on the contents of sentences with the same surface form.

Taraban & McClelland (1988, experiment 2) report a further study that tried to specify which aspects of subjects' expectations were violated by the less expected noun of the PP (the critical noun for short). It might have been the switch of attachment (from NP to VP or vice versa), but other possibilities exist, which are best described with an example.
The janitor cleaned the storage area with the

a) broom
b) solvent
c) manager
d) odour

because of many complaints.

Written cloze pretests were used to identify sentences in which subjects expected the PP to modify the verb (as in a) above), or the NP (as in: 'The hospital admitted the patient with cancer'). For each sentence, the noun in the PP was varied as is illustrated above. Noun a) is the most expected continuation, whereas d) requires a different attachment (to the NP rather than the VP) because it cannot be an instrument of cleaning; the switch of attachment also involves a switch from the thematic role Instrument to what the authors call the role of Modifier of the direct object noun; b) is consistent with the attachment expectation and the Instrument role (i.e. a solvent is a good Instrument) but it is less expected as an instrument in the given sentence context than 'broom'; c) is consistent with the VP attachment but it does not play the Instrument role. Rating studies established that nouns b), c) and d) were equally (un-)expected, and the events described in the sentences equally (un-)realistic. In the main experiment, sentences were again presented in a non-cumulative word-by-word self-paced reading task. Average reading times for the noun of the PP showed no significant effects. On the following two words, however, there was an increase in reading times for d) and c) as compared to a); crucially, d) and c) did not differ from each other. Thus it seems that violations of the expected thematic role of the PP produces the greatest disruption, and violating the expected syntactic attachment does not produce any additional difficulty over and above that associated with violation of thematic role expectations.

I would like to return now to Rayner et al.'s (1983) experiment (1) showing that preference for the main clause analysis (in sentences such as 'The florist sent the flowers was very pleased') is not abolished by reducing the plausibility of the subject NP (eg. 'performer' instead of 'florist'). McClelland (1987, see above) criticised the plausibility manipulation as very weak. Apart from this example, an inspection of the materials reveals that for 8 out of the total 12 sentences the supposed implausibility of
of the noun as subject in the main clause analysis appears questionable (the tourist
wired the money; the customers mailed the information; the reporter passed the caviar;
the teenager sold the car; the visitors expected to arrive; the clerk paid the money; the
child read the story). The authors did not try to validate their intuitions regarding
plausibility or otherwise by any kind of pretest. There is one piece of evidence,
however, that the sentences did differ due to their plausibility: as mentioned above,
paraphrases were accurate 88% of the time for reduced plausible sentences, but only
64% of the time for reduced implausible sentences. However, only half of the
experimental sentences had to be paraphrased (the authors wanted to prevent subjects
from adopting special strategies for dealing with the experimental sentences).
Furthermore, no absolute numbers of paraphrases for reduced plausible and implausible
sentences are reported. It is stated only that subjects had to paraphrase at least one
reduced plausible and one reduced implausible sentence. Thus, this evidence appears
too weak to support the conclusion that plausibility was properly manipulated.

A study that did find clear plausibility effects is reported in Crain & Steedman (1985,
experiment 1). They, too, focused on sentences analyzable as either a reduced relative
clause or a main clause, and manipulated the plausibility of the main clause analysis
by the choice of the subject noun. Compare, for example:

a) The postman delivered the junk mail threw it in the trash.

b) The tenant delivered the junk mail threw it in the trash.

In b) the reduced relative clause analysis is more plausible than in a), and hence a
garden path should be avoided in b). Note that in reduced relative clauses of this type
the disambiguating information (i.e. 'threw') occurs quite late in the sentence (in
Ferreira & Clifton's (1986, see below) it followed almost always immediately after the
first verb). Crain & Steedman directly compared the plausible and implausible reduced
sentences. Unfortunately, Crain & Steedman’s experimental task can only indicate the
final interpretation of the sentence. Subjects were asked to make a grammaticality
judgement after they had read a sentence. The sentences were presented across the
screen, one word at a time, at a presentation rate of 300 ms per word. The words, once
presented, remained on the screen. At the end of the sentence, 3 asterisks were displayed (for about 300 ms), after which the sentence was removed from the screen. Subjects were then asked to make the grammaticality judgement, which was not timed. The number of No-judgements was significantly higher for implausible than plausible sentences (no mean numbers are reported), but obviously, given the task and presentation conditions, no conclusions can be reached about the effect of plausibility during sentence analysis.

The study by Ferreira & Clifton (1986) tried to remedy two shortcomings of the previous work: a) they used a measure that taps processing as it unfolds (rather than Crain & Steedman’s end-of-sentence task), and b) they imposed what they claimed to be more constraining plausibility manipulations. In contrast to Crain & Steedman’s sentences, the disambiguating information occurred earlier in their sentences, usually immediately after the first verb.

Experiment 1 in Ferreira & Clifton (1986) tested whether the main verb reading might be blocked, and a garden path avoided, if the subject noun violated thematic role requirements of the verb. The verb 'examine', for example, requires an agent as subject. The authors assume that only nouns with the semantic feature + animate are possible agents, and thus nouns that lack this feature should prompt the alternative reduced-relative analysis, as for example in:

'The evidence examined by the lawyer turned out to be unreliable.'

Unambiguous non-reduced control sentences contained an overt relative pronoun, as in:

'The evidence that was examined by the lawyer turned out to be unreliable.'

Eye movements were measured and analyzed for the different regions of the sentence (reading times were divided by number of character for each region to control for effects of region length). The data did not support the hypothesis that an inanimate subject noun would block the syntactic preference to initially analyze the ambiguous
sentence as a main clause rather than a reduced relative clause. As expected, first-pass reading times (first fixations) on the disambiguating region ('by the lawyer') indicated a garden path effect (by subjects and by items) when the first noun was animate (e.g. 'The defendant examined by the lawyer ...'). However, when the first noun was inanimate, the garden path effect was also significant, albeit only in the by-items analysis. Furthermore, first pass reading times on the first verb (e.g. 'examined') were longer in the reduced than in the unreduced sentences when the subject noun was inanimate (significant at .05 by items), possibly indicating that readers were sensitive to the fact that the preferred (main clause) analysis for reduced sentences resulted in anomaly when the noun was inanimate. Time spent re-reading parts of the sentence (excluding regressions from the disambiguating region to the ambiguous region, or to an earlier point in the disambiguating region) was longer in the reduced than in the unreduced sentences, and there was no interaction with animacy. Thus, the findings show that whilst first-pass reading times on the verb were increased when the subject noun did not semantically fit, this anomaly did not appear to guide syntactic analysis, and was not used to avoid the preferred main verb analysis which resulted in the garden path.

However, the reliability of this study might be questioned, since only sixteen sentences were used, and 15% of the data were lost due to technical malfunctions.

Furthermore, the increased fixation duration on the disambiguating region with inanimate subject nouns might not genuinely reflect a garden path effect. The increased latencies on the verb in this condition might indicate that the parser begins to compute the reduced relative clause reading, forced by both the semantic feature of the noun (inanimate) and the selection requirements of the verb (requires an animate subject noun). Computing the reduced relative clause reading might take longer than computing a main clause because a new sentence node has to be postulated (representing the relative sentence). If this were the case, processing might spill over to the fixation of the next word (i.e. the disambiguating region). Thus, while in the case of the ambiguous sentences with animate subject nouns longer fixation times on the disambiguating region might genuinely reflect that the parser has received a syntactic
error message vis-a-vis its main clause analysis, these same increased fixation times might, for the ambiguous sentences with inanimate subjects, be spill-over effects resulting from the build up of a more complex syntactic structure which was started immediately upon encountering the ambiguous verb. A possible test for this account would require sentences in which the word after the verb was not the crucial disambiguating phrase, but another word which would provide extra time for the reduced relative clause analysis to be built up, for example:

The evidence examined and rejected by the lawyer ...

In such sentences, no garden path should occur on the disambiguating by-phrase. Alternatively, a different on-line measure could be used which would allow more time to be spent on each word.

The claim that a main clause structure was initially built up regardless of the animacy of the subject could be more directly tested. If a main clause analysis had been built up, then a direct object immediately following the verb should be perfectly acceptable (again regardless of the animacy or otherwise of the subject noun). Thus reading times on the direct object should be the same for a) and b) below:

a) The defendant examined the newspaper and ...

b) The evidence examined the newspaper and ...

Altmann & Steedman (1988) maintain that the preference for the main clause analysis in Ferreira & Clifton’s (1986) study did not result from the minimal attachment principle but from the influence of pragmatic presuppositions. They suggest that sentences presented in isolation (i.e. without a preceding paragraph setting the scene) support presuppositions that work against a reduced relative reading. Restrictive relative clauses presuppose that there is a set of entities of the kind denoted by the noun which are specified further by the modifier. Thus, 'The defendant examined...' should be analyzed as a reduced relative only if the preceding text had made it clear that there were several defendants. The absence of a preceding context favours the
main-clause analysis (since there is no reason to assume that the NP is one of a set that needs to be specified). In the case of inanimate nouns, this analysis is implausible (and is noted to be so as evidenced by the increased fixation times on the verb), but because it is merely implausible rather than completely semantically impossible (given the inanimate nouns chosen by Ferreira & Clifton, 1986), the anomaly does not override the pragmatic preference to assume a main clause analysis. Altmann & Steedman consider the anomaly (of a main clause reading of eg. 'the evidence examined ...') as implausible rather than completely semantically impossible because, they argue, a fairly plausible main clause interpretation is still possible for most of the sentences used in the experiment (the same observation is made by Carlson & Tanenhaus, 1988, p. 278). Note that Ferreira & Clifton did not run a pretest to ensure the implausibility of the main clause analysis when the subject noun is inanimate. Altmann & Steedman only give one example of a plausible main clause analysis for a sentence with an inanimate subject noun: 'The meal brought ...' could continue as 'The meal brought back memories of ...'. I think other sentences used by Ferreira & Clifton might plausibly continue like this:

'The skin felt soft to the touch'
'The trash smelled horribly'
'The car towed the caravan'
'The stories told of heroic deeds'
'The ship sighted land'

Another criticism of the Ferreira & Clifton study is that the verbs were not pretested for any differences in subcat preferences (Tyler, 1989). Unfortunately, no such analysis of the stimuli is given post hoc, so that it is not clear whether the results were in fact confounded by this uncontrolled factor.

Just & Carpenter (1992) conducted an experiment which was very similar to Ferreira & Clifton's (1986, experiment 1). The only differences consisted in eliminating a few sentences whose inanimate subject noun could be an instrument ('the car towed ...'), and in introducing a true/false comprehension question after each trial (Ferreira & Clifton 1986 asked a true/false question only after every fourth trial). The main purpose of Just & Carpenter's experiment was to show that reading span determines
whether nonsyntactic and syntactic information interact during sentence processing. Just & Carpenter argued within the capacity constrained parsing model discussed in section 7.2.3. above. They propose that

"individuals with a large working memory capacity may be more able to keep both syntactic and nonsyntactic information activated, and hence their syntactic processing would be more likely to be influenced by the nonsyntactic information" (Just & Carpenter, 1992, p. 126).

First-pass reading times of high-span readers but not of low-span readers were significantly shorter on the disambiguating by-phrase when the first noun was inanimate as opposed to animate. This was the case in both the unambiguous and the ambiguous sentences. However, both subject groups produced longer first-pass reading times on the by-phrase in the ambiguous as opposed to the nonambiguous sentences, even when the subject noun was inanimate. Thus, while high-span readers take In/animacy into account, they do not use it to avoid the garden path. Just & Carpenter misleadingly, in my view, interpret their findings as showing that the syntactic processor in high span subjects is less informationally encapsulated than the syntactic processor in low-span subjects. They claim that high-span subjects "take the nonsyntactic information into account in initially interpreting the syntactic ambiguity" (Just & Carpenter, 1992, p. 126). This is misleading since high span subjects did show a garden path effect in the inanimate condition.

Britt, Perfetti, Garrod & Rayner (1992) showed that discourse context (a preceding text) did not help subjects to avoid being garden pathed by sentences with main verb/reduced relative clause ambiguities. In their self-paced reading paradigm sentences were presented one word at a time using the moving window presentation. Even after having read a context sentence, subjects took longer to read the region after 'the rug' in sentence a) than in the unambiguous sentence b):

a) The coffee spilled on the rug was difficult to conceal.

b) The coffee spilled on the rug and was difficult to conceal.

However, they did find that discourse context eliminated the ambiguity of PP-
attachment sentences: in context, nonminimal attachment sentences such as a) did not take any longer than minimal attachment sentences such as b):

a) He examined the child with the doll.

b) He examined the child with the needle.

Trueswell & Tanenhaus (1991) tried to shed new light on the apparent impenetrability of sentence processing in the kind of ambiguous sentences tested by Ferreira & Clifton (1986). They investigated whether discourse context affects the initial processing of ambiguities such as 'the defendant examined ... ' by providing a temporal context in which the preferred main verb analysis would not be felicitous. In a sentence completion task, subjects were more likely to continue an ambiguous fragment (eg. 'The student spotted ... ') as a reduced relative clause sentence (eg. 'The student spotted by the proctor was ... ') rather than as a past tense main clause sentence ('The student spotted the proctor ... ') when the preceding text was written in future tense as compared to past tense. In this example, the preceding texts were as follows:

Several students were sitting [will be sitting] together taking an exam yesterday [tomorrow]. A proctor came up [will come up] and noticed [notice] one of the students cheating.

The authors also report the results of a self-paced reading study in which reduced and unreduced relative clause sentences were presented two words at a time with the moving window technique (eg. The student [who] was spotted by the proctor received [will receive] a warning.). Reading times on 'the proctor' showed an increase in reading time in the ambiguous as compared to the unambiguous sentences only when the preceding context was written in the past tense, but not when it was written in the future tense. However, reading times on 'spotted by' were longer in the ambiguous than the unambiguous sentences regardless of the tense of the preceding text, which seems to indicate that the preceding paragraph did not help to avoid the garden path altogether.

Stowe (1989) reports a study which also conflicts with Ferreira & Clifton's results (1986, experiment 1). This experiment is reviewed in detail in the next section, after
which the experiments are reported which I conducted as a further test of Stowe’s model of the semantic factors influencing syntactic decisions.

7.5. Thematic roles and processing garden path sentences

7.5.1. Experimental evidence

Stowe (1989) found that the semantic feature Animacy did influence syntactic decisions, thereby directly contradicting Ferreira & Clifton’s (1986, experiment 1, F & C for short) conclusions. Her tasks and materials differed strongly from F&C’s and will be described in detail in this section.

In her first experiment, Stowe compared sentences such as:

a) Even before the police stopped the driver was getting nervous.

b) Even before the truck stopped the driver was getting nervous.

Verbs like 'stop' have two subcat frames, and, associated with them, two thematic frames. One is the intransitive frame ('The truck stopped'), and the thematic role of the subject NP is Theme. The other is the transitive frame (somebody stopped something): here, the subject NP has the thematic role of Agent, and the object that is stopped is the Theme. In sentence b) above, the noun in the subject NP-position has the semantic feature +inanimate. It is often been assumed that inanimate nouns cannot take on the Agent role (at least not in a non-metaphorical reading) since Agents are obligatorily intentional (Fillmore, 1968). Thus, in the sentence with the inanimate subject noun, the transitive subcat frame which requires the subject noun to be an Agent does not fit, whereas the intransitive frame does since it requires the subject noun to be a Theme and the Theme role can be taken up by inanimate nouns. An animate subject noun is compatible with both the Agent and the Theme role, but Stowe assumes that the Agent role is initially preferred because of a strong default principle of thematic assignment.
Stowe tested whether sentences such as b) would lead to the same garden path as sentences such as a). A garden path on encountering 'was getting' follows from the principle of late closure. The principle of late closure would lead the parser to assume that the NP ('the driver') following the verb is the direct object, rather than closing off the subordinate clause after 'stopped', and analyzing the following NP as the subject of the following main clause.

A self-paced reading paradigm with continuous grammaticality judgements was used. The crucial evidence comes from decision times on the disambiguating region ('was already') in the ambiguous sentences as compared to unambiguous sentences. The latter were created by inserting an adverbial phrase after the first verb which syntactically signalled the end of the subordinate clause (eg. 'Before the police [truck] stopped at the restaurant the driver was already getting nervous').

There was a garden path effect for ambiguous sentences with an animate subject noun but not for those with an inanimate one: reading times were 613 ms longer and false negative grammaticality judgements more frequent when the disambiguating region was encountered than in unambiguous sentences. It appears that the inanimacy of the subject noun in ambiguous sentences such as b) did effectively prevent the parser from producing an incorrect sentence analysis.

In her second experiment the plausibility of the second noun was manipulated, eg. compare:

a) When the police [truck] stopped the driver became very frightened.

b) When the police [truck] stopped the silence became very frightening.

If the parser takes into account semantic information, then 'the silence' should not be analysed as a direct object and, therefore, there should be no garden pathing.

Using the same experimental procedure as in her first experiment, Stowe found a main effect of plausibility for reading times on the second noun ('driver' or 'silence'): the mean response time for implausible nouns was 899 ms and for plausible nouns 753
ms. However, there was no interaction with Animacy of the first subject noun, which is surprising since the postverbal noun should not have been analysed as a direct object in the inanimate condition, and hence its plausibility or otherwise as a direct object should not have affected decision times.

On the disambiguating word ('became') the interaction of Animacy and plausibility was significant: decision times were longer in the animate than the inanimate condition (replicating the results from experiment 1) but only when the second noun was plausible.

Implausibility is, unfortunately, not clearly defined by Stowe. It covers pragmatic violations (i.e. the noun is possible but very unlikely as a direct object of a particular verb), and semantic violations (i.e. the noun violates semantic selection restrictions of the verb, eg. 'he shifted his health': violates the requirement for a direct object noun with the semantic feature + concrete). In 14 out of the 20 test sentences implausibility was caused by such semantic violations.

It also seems important to ensure that the violated nouns are not intrinsically more difficult to process than the non-violated ones, for example because they are more abstract. In Stowe's materials, "about half of the implausible nouns are more abstract than the corresponding plausible noun" (Stowe, 1989, p. 341). Stowe points out herself that longer decision times on the implausible nouns in the animate condition cannot unequivocally be interpreted as reflecting the recognition that they are implausible as direct objects.

It seems necessary to include an unambiguous control condition (as Stowe did in her first experiment, but not her second), for example by introducing an adverbial phrase after the ambiguous noun, which would signal the end of the subordinate clause. Only then could one assess whether the garden path in the animate condition is affected by the plausibility of the postverbal noun as a potential direct object.

In summary, Stowe's experiments show that inanimacy appears to affect thematic role
assignment which in turn influences syntactic analyses (some have even concluded that her results show that inanimacy affects the initial syntactic interpretation of the sentence, Tyler, 1989). These findings directly contrast with the results Ferreira & Clifton (1986, experiment 1) reported. How can this apparent contradiction be accounted for?

The first point to note is methodological. Stowe's paradigm required subjects to make a grammaticality judgement at each word about the sentence read so far. Thus, the response times (which are on average about 700 ms) do not only reflect reading times but a conscious metalinguistic decision. The task might also have encouraged subjects to anticipate how the sentence could continue grammatically, and decision times might have reflected a disruption if the anticipated word did not come up. Furthermore, because subjects received feedback about the decisions, they might have been encouraged to think carefully before making a decision.

If Stowe's results were merely due to the greater amount of time that subjects were allowed to spend on each word, then there really is no contrast with F&C. F&C argue that "it is at the stage of reanalysis, not initial analysis, that nonsyntactic information sources are consulted" (Ferreira & Clifton, 1986, p. 366). They adopt Frazier's proposal (Frazier, 1985; Rayner et al., 1983) of a thematic processor which operates only on the output of the syntactic processor, in particular on the verbs and their arguments. Their first experiment, they argue, is compatible with the suggestion that reanalysis is only started when a syntactic error message is encountered (eg. the 'by-phrase' in sentences such as 'The defendant examined by the lawyer ...'). They also point out, however, that in other types of sentences there is no syntactic error signal, but yet there is evidence of reanalysis (longer reading times), eg. in sentences such as 'The spy saw the cop with the revolver': here, the only error message is the implausibility of seeing with the help of a revolver (which the thematic processor is assumed to have noted), which then causes the reanalysis of the preferred VP-attachment analysis (but see Taraban & McClelland (1988), discussed above, who claim that VP attachment is not necessarily preferred). To incorporate this finding they suggest that a semantic error message can cause the parser to give up its initial
analysis. They emphasize that initially the parser builds up a single analysis according to its very own structural preferences, which is reanalysed either if further input cannot be attached in the current structure, or if the thematic processor fails to assign thematic roles to the constituents in the suggested structure, due to semantic or plausibility factors. In their first experiment, however, a thematic anomaly did not prompt the parser to start a reanalysis: "the readers apparently did not resolve this anomaly on a semantic basis, but instead waited for syntactic information" (Ferreira & Clifton, 1986, p. 355). From this one might conclude that in their model a semantic error message is only taken up by the parser at the end of a sentence; this would explain why a semantic error message is used (to prompt a reanalysis) in sentences such as 'The spy saw the cop with the revolver', but is not used in 'The evidence examined ...'. However, in Stowe's experiment 1, semantic anomaly (i.e. the inanimacy of the first noun) had an effect before the syntactic error message was encountered, and certainly before the end of the sentence was reached. This rules out the proposal that the parser accepts semantic error messages only at the end of sentences.

Given the length of time that was available for processing each word in Stowe's paradigm, it is difficult to decide whether thematic information determined the initial syntactic analysis in the inanimate condition (i.e. early closure of the subordinate clause after the verb) or whether, as a default, the direct object analysis was initially preferred, in line with the principle of late closure, and then revised (all of which might be reflected in her dependent variable). If the latter option were correct, one would expect longer decision times on the verb (eg. 'stopped'), in line with Ferreira & Clifton's (1986) findings of longer reading times on verbs like 'examined'. In Stowe's data the decision times on the verbs are the same in all sentence conditions, i.e. around 700 ms. Even though these response times are very long, one could argue that if extra processing did occur on the verb in the inanimate condition, due to a reanalysis of the default transitive analysis, this should have shown up as an increase of decision times on top of the extra time required for making a grammaticality judgement. Still, use of a different on-line measure with a smaller temporal window would allow one to test whether the output of the thematic processor determined the initial analysis of the verb or whether it prompted a reanalysis of the initial analysis which would have been built
up independently by the syntactic processor. Stowe (1993, pers. com.) replicated her findings of experiment 1 using a self-paced reading task which did not require grammaticality judgements: there was an interaction of Ambiguity and Animacy (and significant main effects for both factors) one word after the disambiguating word. Again, there were no effects on the first verb (e.g. 'stopped'), indicating that no reanalysis occurred.

One difference in the materials used by Ferreira & Clifton (1986) and Stowe (1989) is that the critical verbs in F&C (like 'examine') only allow one subcat frame and thus one thematic frame (a subject NP and an object NP, with the roles of Agent and Theme, respectively), whereas Stowe’s verbs allowed two subcat and thematic frames (the intransitive subcat frame, in which the subject NP is Theme, and the transitive subcat frame in which the subject NP is Agent, and the object NP theme). The impact of these differences is difficult to assess given that the details of how the thematic processor operates are not specified by Stowe. It is not clear, for example, how it 'knows' that a given noun is the subject: this requires syntactic information. Stowe assumes that the thematic processor passes on to the syntactic processor the information that the subject is Theme, and that the corresponding subcat frame is the intransitive one. The syntactic processor uses this information to close the clause, rather than expecting a direct object. In F & C’s sentences in which the verb only had one thematic and subcat frame, no alternative thematic role can be assigned to an inanimate noun when the Agent role is inappropriate. The output of the thematic processor in this case might either be a message that no role can be assigned, or the assignment of Agent as a default (most subjects, after all, are Agents). Since the thematic information does not provide a clear alternative analysis for the parser to pursue, the parser might maintain its default analysis (as a main clause) until further input can be used. Note that an account along these lines assumes that semantic information (re. animacy) can only be used via the thematic processor as one determinant of the assignment of thematic roles, and that it is the assignment of thematic roles, but not semantic information itself, which affects the decisions of the parser. This fits F&C’s data (as well as Stowe’s) because in F&C, semantic information could only block the assignment of the Agent role, but the thematic
processor could not propose an alternative. There is some evidence, however, showing that inanimate nouns in sentences such as F&C's can affect syntactic decisions (and avoid the garden path); Tanenhaus, Carlson & Trueswell (1989) report unpublished results of an eye movement experiment they modeled on F&C. They ensured that sentences with inanimate subject nouns did not allow plausible past tense completions (see the discussion in section 7.4. above). With these materials, they showed that there was no garden path when the subject noun was inanimate (i.e. there was no difference in fixation times on the 'by'-phrase for ambiguous and nonambiguous sentences such as: 'The van recognized by the spy ...' and 'The van that was recognized by the spy ...'). From this it would seem that the impossibility of assigning the Agent role is sufficient to affect syntactic processing.

7.5.2. Models of the interaction of thematic and syntactic processing

Rayner, Carlson & Frazier (1983)

Rayner, Carlson & Frazier (1983)'s model assumes two independent processors which are characterised by different properties, i.e. a thematic processor and a syntactic processor. The thematic processor considers more than one analysis (of which sort is to be described below) at a time, and it uses real-world knowledge and knowledge of the preceding discourse. The syntactic processor does not have access to these knowledge sources, and only considers one analysis at a time following the principles of minimal attachment and late closure. According to this model, the syntactic processor outputs one syntactic analysis of the input string and attempts a revision if it is incompatible with the output of the thematic processor. The input to the thematic processor are the phrases into which an initial syntactic analysis has parsed the input string, eg. 'John', 'the lunar eclipse' and 'with a telescope' and 'saw', plus some information stored in the lexical entry of the verb, i.e. the thematic structures associated with that verb. The verb 'see' has two thematic structures:

a) Experiencer Theme
b) Experiencer Theme Instrument
The authors do not specify how the phrases that are input to the thematic processor are assigned to the elements of the thematic structures. Using real world knowledge, and knowledge about the preceding discourse, the thematic processor examines both thematic structures "to determine the relative plausibility" (Rayner et al., 1983, p. 37, italics in the original) of each, and selects and outputs the most plausible thematic structure. The authors do not discuss which aspects of the meaning of the phrases has to be retrieved, or constructed, in order to decide how likely a thematic frame is. Furthermore, the authors do not specify the format of the output of the thematic processor, and how it is compared with the output from the syntactic processor. It also remains unspecified what happens if the syntactic processor has to start a reanalysis. In the eye movement study with sentences such as 'The spy saw the cop with the revolver ...', no regressions of the eyes are reported, but from other studies it is clear that, often, the eyes would go back to the ambiguous part of the sentence: this would imply that the syntactic processor has some kind of short term memory facility in which maybe part of the parse tree built up so far is stored, or a record of the last choice point is kept.

Stowe (1989)

In Stowe’s (1989) and Tanenhaus & Carlson’s (1989) model the thematic processor is assumed to determine the choices of the syntactic processor. Stowe’s model was designed to account for her finding (in an experiment using a self-paced reading paradigm in which a grammaticality judgement had to be made after each word) that garden pathing, typically found in sentences such as

    a) Before the police stopped the driver was ...

was avoided when the first noun was inanimate, and therefore not a plausible agent.

As was the case for Rayner et al.’s (1983) model, several aspects of Stowe’s model, too, are underspecified. Firstly, it is not clear what exactly constitutes the input to the thematic processor. Stowe suggests that the thematic processor evaluates how well a
given noun fits with the thematic roles specified in the thematic frame of the verb. It is unclear how the thematic processor 'knows' which nouns are to be taken as arguments in subject position, and which in object position. Furthermore, subject- or object-nouns might not be single nouns but a noun phrase consisting of relative clauses ('the actor who had only recently been employed...') or coordinate structures ('the actor and the cameraman ... '). The point is that the thematic processor requires some information about the structure of the sentence to be able to perform its task of evaluating how well the head of a particular phrase fulfils a thematic role. What remains to be specified is at which stage in processing the syntactic processor passes on this type of structural information to the thematic processor. An intrinsic logical problem of such a model is that the parser is assumed to output only one structure at a time, and yet the thematic processor, which requires some syntactic input from the parser, is assumed to evaluate several thematic frames in parallel.

The second area of underspecification is the order in which the thematic processor evaluates multiple thematic frames of a verb. Take for example the verb 'stop' which has two subcat frames, and, associated with them, two thematic frames. A first noun that is animate fits both frames (note: the first NP is the subject):

   a) [NP (+animate) NP]  
    [Agent        Theme]  
    eg.: He stopped the car.

   b) [NP]  
    [Theme]  
    eg.: He stopped.

When the subject noun is inanimate, only the second frame is possible since here no particular semantic features are required. When the subject noun is animate, the thematic processor preferentially assigns it the Agent-role, according to Stowe. Ferreira & Clifton (1986), too, consider Animacy as a decisive feature in whether or not a subject noun can take the role of Agent. Stowe maintains that only in exceptional cases need the semantic features not match the semantic requirements of the thematic role. These cases exist when the context allows a metaphoric interpretation. As an example, she mentions the sentence:
The rock murdered the millionaire.

In this sentence, the Agent role is assigned to the inanimate noun 'rock', which is subsequently interpreted as having the semantic features "necessary for the role" (Stowe, 1989, p. 325), eg. intentionality (which might be plausible in the context of a science fiction story). Stowe stresses that assignments of thematic roles resulting in an anomalous literal interpretation are rare, and are avoided if an alternative more plausible thematic assignment is available (as in the causative/ergative verbs for which the subject noun can be either Agent or Theme). Stowe does not take into account that the plausibility of the whole transitive sentence, rather than the semantic features of the subject noun, might be relevant in thematic assignments. Instead, she stresses that

"by manipulating the semantic feature animacy [of the subject noun], the possibilities for the thematic structure of the rest of the sentence can be altered" (Stowe, 1989, p. 328)

and that

"when the subordinate clause contains an inanimate subject, the agent role is not suitable to be assigned to the subject noun phrase" (Stowe, 1989, p. 329).

However, her experiments do not unequivocally allow such a conclusion because the animate subjects were always plausible Agents, in the context of the verb and the object noun, whereas the inanimate subjects were always very implausible as agents in the actions described by the verb and the object noun.

In Schlesinger (1992) the Agent role is defined not by any necessary or sufficient characteristics (such as Animacy, which Stowe seems to consider to be necessary) but by a set of features which are typically found in Agents. A similar idea has been put forward by McClelland & Kawamoto (1986) who suggest that thematic roles can be represented as sets of "role descriptors", some but not all of which apply to a thematic role in a particular sentence. In Schlesinger’s proposal, these features describe the nature of the entity that an NP (in subject position) is referring to. As agentive features, Schlesinger (1992) lists the following: cause, characteristic responsible, motion, change of state, control, intention and responsibility. Prototypical Agents are
assumed to have "most" of these features.

My Experiments 5 and 6 were designed to test whether inanimate subject nouns can take up the Agent role in sentences where the resulting transitive action is a plausible one. If this were to be supported by the data, one of the agentive features usually seen as most typical of Agent-hood, i.e. Intentionality, would seem to be not a necessary feature at all.

An alternative way of investigating the assumed default assignment of the Agent role to animate subject nouns is described by Tanenhaus & Carlson (1989). They mentioned that they were currently testing whether the Theme- rather than the Agent role is preferentially assigned to an animate subject noun when a preceding sentence context makes this plausible, as in the following example:

Although dinner was ready, the man was in the middle of varnishing his desk.

By the time the man finished the meal was cold.

Note that in Stowe's experiment 1, the Theme role (and its associated subcat frame) might have been activated when the subject noun was animate, given that animate nouns can be a Theme. However, this thematic frame might have been deactivated when the second noun, which was always a very plausible direct object, was encountered. To test more directly if only the <Agent, Theme>-frame is activated when the first noun is animate one could use sentences in which there is syntactic information signalling that an intransitive structure follows immediately after the verb (eg. 'Before the police stopped he was ....'); if a transitive structure had been expected, extra processing cost might be involved to switch to the intransitive analysis, as compared to sentences in which only the intransitive Theme frame is possible (see discussion at the end of this section). Stowe had included sentences of this kind as control sentences (eg. the unambiguous control sentence 'before the police/truck stopped at the restaurant the driver was ...'). She did not, however, report response times for the disambiguating region ('at the restaurant').
Carlson & Tanenhaus (1988)

Carlson & Tanenhaus' (1988) model of the role of thematic processing during sentence comprehension is very similar to Stowe's. I will only point out some of their assumptions about the processing of thematic roles as far as they are relevant for sentences with causative/ergative verbs (such as 'stop').

The authors propose that thematic roles are assigned to arguments of the verb as soon as possible, at least by the time the verb has been encountered, but possibly even earlier. This early assignment sometimes needs to be changed, i.e. when further parts of the sentence are incompatible. Carlson & Tanenhaus (1988) assume that arguments occurring before the verb (e.g. the subject) might be assigned a thematic role before the verb is processed, based on some default patterns of thematic assignment, e.g. to assume that an animate subject is Agent, that objects are Themes, or that inanimate subjects are Themes. Default assignments would be particularly useful in SOV languages in which the verb occurs at the end of the sentence (see Frazier, 1987c).

Assignment of thematic roles is assumed to be based on semantic and pragmatic factors, but it is not discussed in any detail:

"It seems plausible that the meaning of an argument, the core meaning of the verb, as well as general world knowledge are taken into account in making provisional thematic assignments" (Carlson & Tanenhaus, 1988, p. 276).

As an example of how world knowledge affects the (provisional) assignment of thematic roles, the authors point out that in sentence a)

a) John packed the suitcase.

'the suitcase' is preferentially assigned the thematic role Location, in the sense that John packed something into the suitcase. Sentence a) might however continue as in b)

b) John packed the suitcase onto the ship.

In this case, Location has to be reassigned to 'onto the ship', and 'the suitcase' has to
be assigned the thematic role of Theme. The authors argue that it is general knowledge about suitcases and the action referred to by the verb (rather than any grammatical properties of the NP or verb) which determines the initial provisional thematic assignment, as in sentence a). The initial assignment would be different for an NP referring to a different thing, as in c)

c) John packed his clothes.

Here, 'his clothes' would initially be assigned the thematic role Theme (based on general knowledge about clothes and what to do with them), an assignment which would be confirmed if the sentence continued with, for example, 'into the bags', but disconfirmed if the sentence continued with 'with the pebbles he had found'.

Carlson & Tanenhaus interpret Stowe's data as support for their views that semantic features influence thematic assignment: "animacy affects thematic assignment", and that an inanimate subject noun "is more likely to be a Theme than an Agent" (Carlson & Tanenhaus, 1988, p. 277). However, they do not discuss the status of semantic features versus the other factors they assume to affect thematic assignment, i.e. general knowledge and the core meaning of the verb.

They further assume that the parser has to check the (provisional) output of the thematic processor each time before it attaches a possible argument of the verb to determine whether there is a thematic role available for it. Thus, thematic assignment guides the operations of the parser rather than filtering the output produced independently by the parser. In fact, syntactic processing is seen as a filter for interpretations proposed by the thematic processor. Tanenhaus & Carlson (1989) speak of argument structure information as guiding the parse by allowing the parser to project structure yet to be encountered. This model accounts for Stowe's findings in the following way. The assignment of the Agent role to the animate subject noun activates the <Agent, Theme> thematic frame and prompts the expectation that an NP in direct object position will follow to take on the Theme role. As a consequence, the parser analyzes the postverbal NP as a direct object, even though syntactic rules would have allowed an alternative analysis. Syntactic input later on in the sentence rules out
the transitive analysis which had been prompted by the assignment of the Agent role to the subject.

The next point to be explained in this account is why there is no fast and cost-free reassignment of thematic roles in Stowe's sentences. Carlson & Tanenhaus (1988) distinguish two types of verbs. There are verbs such as 'pack', introduced above, which allow cost free re-assignment of thematic roles (e.g.: 'the suitcase' is initially assigned the role Location: 'He packed his suitcase', but when later parts of the sentence force re-assignment, no garden path effect occurs, as demonstrated by 'He packed his suitcase onto the ship'). The authors assume that the alternative thematic roles are represented in the same thematic frame, and that they remain active and available apparently irrespective of the time between initial assignment and re-assignment.

Other verbs, however, such as Stowe's causative/ergative verbs, are assumed to have two different thematic frames, i.e. <Agent, Theme> and <Theme>. The authors propose that lexical access makes available all thematic frames associated with the verb but that the thematic frame which is not initially chosen becomes inactive, and hence unavailable for cost-free re-analysis. Thus, the garden path effect found for sentences with animate subject nouns demonstrates that once the <Agent, Theme> frame was assigned, no alternative thematic roles were available in the same frame, and, furthermore, that the alternative <Theme> frame has become unavailable. Unfortunately, nothing is said about how fast the alternative thematic frame is assumed to become inactive, whether it deactivates because it has not been assigned to an argument or because a role of the other frame has been assigned to an argument, and whether this might be affected by strategies.

In addition to Carlson & Tanenhaus' point that the alternative grid becomes increasingly deactivated, there might be another reason why there was no cost-free reassignment of thematic roles in Stowe's experiment. Note that upon encountering disambiguating information in Stowe's sentences, not merely a re-assignment of a thematic role is required, but also the undoing of a syntactic attachment (i.e. of 'the
driver' as direct object) which might exert extra processing costs. No such syntactic reanalysis is necessary in sentences such as 'He packed his suitcase/He packed his suitcase onto the ship'. Following this interpretation one would predict that a reanalysis which only requires thematic reassignment but no undoing of syntactic analyses would require less time than a reanalysis that requires thematic and syntactic reanalyses (in fact, under the strong claim that the alternative thematic frame remains active for a while, no extra processing costs should be predicted if thematic reassignment without syntactic reanalysis was required). A situation of thematic reassignment without the need for syntactic reanalysis exists in Stowe's control sentences. Unfortunately, she does not report any reading times for the crucial words. Take for example:

a) Before the police stopped at the restaurant the driver was ...

Here, 'stop' should lead to an initial assignment of the Agent role to 'the police' and the ensuing syntactic analysis as a transitive sentence. This turns out to be a garden path when the phrase 'at the restaurant' is found, since this phrase, of course, forces the intransitive analysis. The re-analysis only involves assigning the subject noun a new thematic role, i.e. the Theme role of the other active thematic frame, but no undoing of syntactic analyses. With the sentence 'Before the police stopped the driver was ...' there will be a much larger garden path effect because re-analysis not only involves assigning the subject noun a new thematic role but also a change in syntactic analysis.

7.5.3. Empirical problems in Stowe's (1989) experiments

In addition to the above criticisms of Stowe's model of thematic assignments and their role in sentence processing, there are several aspects of her experimental materials which undermine her interpretations.

Just to re-iterate, the main features of her first experiment were as follows. The sentences were of the type exemplified by a) and b) below:
a) Even before the police stopped the driver was ...
control: Even before the police stopped at the restaurant the driver was ...

b) Even before the truck stopped the driver was ...
control: Even before the truck stopped at the restaurant the driver was ...

Her main finding is that reading times on 'was' is longer in a) than in its control, whereas in b), reading time on 'was' is the same as in its control. Reading times on 'was' in the control sentence of a), in b) and in its control sentence do not differ.

Problem 1: the selection of the first subject noun

It is not clear whether the semantic feature of Animacy of the head of the subject NP (eg. 'police) is crucial for assigning the thematic role of Agent, which prompts the expectation of a transitive structure. Stowe’s sentences in the animate condition were all highly plausible transitive sentences (eg. the police stopped the driver; the actor finished the last scene; the lecturer began the talk; the children rolled the ball; the owner opened the shop).

The syntactic processor might not have been influenced only by lexical semantic information, but also by pragmatic factors and general knowledge which would be necessary to evaluate plausibility.

A test is required to determine whether the garden path effect (i.e. the transitive analysis) in the animate condition would hold when the first noun is animate but also pragmatically neutral in the context of the verb and the postverbal noun. Alternatively, sentences in the inanimate condition should be constructed such that an inanimate subject noun could plausibly be the subject in a transitive action involving the verb and the postverbal noun.

In her second experiment, Stowe intended to make the transitive sentence in the
animate condition less plausible by changing the second noun (e.g. the police stopped the silence). As noted above, most second nouns were actually semantically impossible rather than implausible. Stowe found that when the second noun was semantically impossible as a direct object in the animate condition, no garden pathing (increased reading times on the disambiguating verb) was found. The interpretation of this finding is difficult, however, as Stowe points out herself. Decision times on the semantically impossible nouns were higher than on the possible nouns not only when the subject noun was animate, but also when it was inanimate (where, according to Stowe’s model, no transitive analysis should have been attempted). These increases were probably not due to the intended anomaly of the second noun as direct object (which should only have shown up in the animate condition), but due to some lexical difference between the impossible and the possible second nouns. In fact, the impossible second nouns were more abstract than the possible second nouns. Thus, one cannot unequivocally argue that the semantic impossibility of the second noun as a direct object caused longer reading times and prompted the re-analysis of the transitive structure into an intransitive one in the animate condition.

Problem 2: the selection of verbs: the implausibility of the intransitive frame in the animate condition

The verbs were chosen by Stowe because they can occur with two subcat frames. The two frames, however, are not equally 'natural': sentences were constructed such that in the animate condition, the intransitive frame seems odd for most verbs, and in the inanimate condition, the transitive frame seems odd.

In the animate condition, the intransitive frame often seems possible only in the nonambiguous condition. For example, compare 'When the dancer shifted' versus 'When the dancer shifted into position': 'into position' does not seem to be exchangeable with any other adjunct, such as 'in the afternoon', which, however, should be the case if the only function of the disambiguating information was to signal the end of the VP. Note that in one of her animate sentences (While his friend was
cooking the meat ...) there is no ergative intransitive analysis: 'his friend' is never the Theme, not even in Stowe's unambiguous sentence (While his friend was cooking in the kitchen ...). The unambiguous sentence is still transitive, with the direct object understood rather than overtly expressed. A truly intransitive reading of 'While his friend was cooking ...' (similar to the analysis of the sentence 'While the soup was cooking in the oven...') is only possible in a cannibalistic society!

In Stowe's experiment 2, however, the intransitive analysis seems to have been accepted in the animate condition when the transitive frame was blocked by the implausibility of the second noun as the direct object. This result does not invalidate my criticism of the artificiality of the intransitive frame in the animate condition in Experiment 1. The set of sentences used in Experiment 2 was not identical to the one used in Experiment 1. In E2, 20 sentences were used, 7 of which did not occur in Experiment 1. Thus it is not certain that all sentences in the animate condition of the first experiment allowed the intransitive frame in a natural way.

Problem 3: testing the effect of an implausible direct object noun

As discussed above, in Stowe's second experiment the 'implausible' postverbal noun received longer response times than the plausible postverbal noun not only in the animate but also in the inanimate condition (the interaction between animacy and plausibility was not significant). Thus, the increased response times on the implausible postverbal noun might have been due to some confounded difficulty. They are not clear evidence for costly processes of thematic re-assignment in the animate condition. Contrary to Stowe's assumptions of default thematic assignments of the agent role to animate subject nouns, one cannot rule out that both thematic frames were initially activated on the first verb.
7.5.4. Summary

This chapter has outlined in detail the empirical problems of Stowe’s study which was intended to support a strong thematic model with a default assignment of thematic roles to nouns on the basis of semantic features, which determines the decisions of the syntactic parser when faced with temporarily syntactically ambiguous sentences. It has also highlighted that the models put forward to account for the interaction of thematic and syntactic information, and for the factors determining the assignment of thematic roles, are still underdetermined empirically. In the following chapter, three experiments are reported that were designed to shed some new light on the role of semantic information in assigning thematic roles, and on the impact of the thematic processor during syntactic analysis.
CHAPTER 8: THE USE OF SEMANTIC INFORMATION IN PROCESSING GARDEN PATH SENTENCES: EXPERIMENTS 5, 6 AND 7

8.1.: Experiment 5: Animacy/inanimacy of the subject noun and the garden path effect: a self-paced reading study

8.1.1. Introduction

Experiment 5 was designed to test whether inanimate nouns can be assigned the Agent role when the particular action is highly plausible - i.e. to test whether there is thematic animacy without semantic animacy. Note that what is under test is not whether inanimate nouns can take an Agent role in a metaphorical context, but in a literal sense (as in 'the lights stopped the traffic'). Causative/ergative verbs can assign one of two thematic roles to the subject noun. The underlying hypothesis of the experiment is that the thematic processor can only discard one of these roles when it has evaluated not only the semantic features of the nouns, but the plausibility of the sentence. This hypothesis contrasts with Carlson & Tanenhaus' (1988) proposal that there is an early provisional thematic assignment, and also with their and Stowe's claim that Animacy is of primary importance in assigning thematic roles.

I will test whether garden pathing occurs for sets of sentences which differ only in the subject noun, as demonstrated below:

a) After the boy turned the wheel came loose ...
b) After the motor turned the wheel came loose ...

If pragmatic plausibility can overrun semantic information (i.e. the feature inanimacy), a garden path should occur for sentence b) as much as for sentence a), because in b), too, the transitive analysis should be chosen.

To establish whether garden pathing occurred, the same type of unambiguous control
sentences were chosen as in Stowe's experiment. Reading times for the ambiguous sentences, especially on the verb of the main clause ('came loose'), will be compared to reading times for unambiguous control sentences in which an additional syntactic constituent makes the transitive analysis impossible. In these sentences, 'the wheel' is unambiguously analyzed as the subject of the main clause rather than the direct object of the subordinate clause, and thus the second verb should fit perfectly with this analysis, as in

a') After the boy turned abruptly the wheel came loose ...

b') After the motor turned abruptly the wheel came loose ...

The statistical analysis will focus on certain words at particular positions in the sentence. Table 8.1. gives a full stimulus set, indicating the critical positions.
Table 8.1.: Experiment 5: example stimulus set

<table>
<thead>
<tr>
<th>animate-ambiguous (a_amb):</th>
<th>animate-nonambiguous (a_namb):</th>
<th>inanimate-ambiguous (i_amb):</th>
<th>inanimate-nonambiguous (i_namb):</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 I 2 3 4 5</td>
<td>1 I 2 adv1 adv2 3 4 5</td>
<td>1 I 2 3 4 5</td>
<td>1 I adv2 2 adv1 3 4 5 6</td>
</tr>
<tr>
<td>Although the cheerful clowns had changed the mood remained the same for a while.</td>
<td>Although the cheerful clowns had changed beyond recognition the mood remained the same for a while.</td>
<td>Although the cheerful campaign had changed the mood remained the same for a while.</td>
<td>Although the cheerful campaign had changed beyond recognition the mood remained the same for a while.</td>
</tr>
</tbody>
</table>

8.1.2. Hypotheses

According to the strong thematic hypothesis, semantic features of a noun are decisive in assigning thematic roles. If Stowe's model is correct, the animacy of the subject noun will lead to the assignment of the Agent role, and to the activation of the <Agent, Theme> thematic frame, which is associated with the transitive subcat frame. Inanimacy of the subject noun, however, should rule out the assignment of the Agent role and thus the transitive analysis. Thus, on this model, a garden path effect is expected to occur if the subject noun is animate, but not when it is inanimate, as...
measured by longer reading times on the second verb and beyond (positions 5 and 6).

Furthermore, if animacy leads to the activation of the transitive frame, there should be a strong expectation of an NP after the verb has been read; thus, in the animate-nonambiguous condition, there should be longer reading times on the adverbials (which close off the verb phrase after the verb) as compared to the inanimate-nonambiguous condition where there should not be a strong expectation for a direct object to occur.

The alternative hypothesis is to doubt the strong thematic hypothesis: the semantic feature of the first noun is not decisive in assigning thematic roles. Instead, it is the pragmatic plausibility of the transitive analysis that is important (i.e. how likely is it that an action is carried out by the first NP affecting the second NP). Thus, even when the head of the first NP is inanimate, transitive analysis occurs (with the ensuing garden path effect on position 5) when the transitive analysis is pragmatically plausible whether the first subject noun is animate or inanimate.

Furthermore, there should be no difference in processing time on the adverbials, since the semantic features of the subject noun alone do not force the activation of the <Agent, Theme>-thematic frame, and the associated transitive subcat frame. Thus, in the inanimate-nonambiguous condition, early closure, i.e. an adverbial following the verb, is no more likely than as in the animate condition.

**8.1.3. Pretest**

First, a pretest was conducted to establish that the transitive analysis is in fact equally likely for sentences with animate and inanimate first nouns (such as a) and b) above).

The first part of the ambiguous sentences were presented as simple active declarative sentences, i.e. without the conjunction and the disambiguating information. Subjects read pairs of sentences which only differed in the animacy of the first noun, eg.
a) The boy turned the wheel.
b) The motor turned the wheel.

Each sentence pair was presented on a separate sheet of paper, with a response box next to each member of the pair. In a forced-choice paradigm, subjects had to decide which of the two sentences of a pair they found more plausible by making a tick in the appropriate box.

Procedure

Ten subjects were given a booklet, consisting of 36 pages. On each page there was a sentence pair which only differed in the subject noun. Subjects were asked to read both sentences and to indicate in a box next to each sentence which sentence of the pair 'sounded more natural' to them. They were asked to make a decision even when some sentences sounded almost 'equally OK' to them. They were encouraged to work rapidly through the pages, and not to ponder over their decisions.

Materials

36 sentence pairs were presented, 16 of which used causative/ergative verbs from Stowe's sentence set, and 20 using new causative/ergative verbs. Sentences with animate subject nouns were the first sentence of a pair as often as sentences with inanimate subject nouns. Each sentence pair was presented on a separate sheet of paper. The pages were arranged in a different random order for each subject.

Results

A sentence pair was defined as biased if eight or more subjects preferred one of the sentences of the pair.
For ten pairs there was a bias to prefer the sentence with the animate subject noun. These ten sentence pairs were not used in the main experiment. Six pairs had a bias towards the sentence with the inanimate subject noun. These pairs were included in the main experiment because such a bias is conservative with respect to the strong thematic hypothesis.

8.1.4. Method

Subjects

Forty-four subjects took part in this experiment. They were students at the Psychology Department of University College London and received course credit for their participation.

Apparatus

Presentation of instructions and stimuli, and recordings of subjects' key presses were controlled by Elonex PC's, using a program written with MEL-software (Schneider, 1988, 1990). PC's were situated in sound-attenuated experimental booths.

Stimuli

Twenty-four sentence sets such as the one presented in Table 8.1. above were designed. They were pooled from the set of sentences used in the pretest, with the exclusion of the sentence pairs for which there had been a clear preference towards the sentence with the animate subject noun. Experimental sentences were intermixed with filler sentences which were constructed to resemble the experimental sentences to various degrees. Sentences were presented in a different random order to each subject. The filler sentences were of four different types. There were 15 active transitive
sentences, five with a noun in direct object position which was semantically violated, and five with a pragmatically inappropriate direct object noun. 22 filler sentences were short complement-clause sentences in which the complement clause was either in subject- or object position (eg. "It is lucky that you remembered your umbrella"; "That he will come is certain"). Another 22 filler sentences consisted of short transitive and intransitive declarative sentences in which the subject noun was neither an Agent nor a Theme (but, for example an Instrument: "The fence keeps the cattle out", or a Source: "The story hurt the famous actress"). Finally, there were 10 filler sentences with the same sentence structure as the experimental sentences (conjunction + subordinate clause + main clause). Half of these sentences had an animate first subject noun, and half an inanimate first subject noun (eg. "When the taxi took the tourists to the airport it got caught in an accident"). Thus, each subject saw 93 sentences altogether. The filler items were identical across the four lists. Furthermore, all subjects saw 13 practice sentences which represented the different types of sentences used in the main experiment. Appendix I presents the experimental trials of Experiment 5.

Design

The design was a fully-factorial repeated-measures design, with Animacy (animate versus inanimate) and Ambiguity (ambiguous versus nonambiguous) as the repeated-measures factors. To ensure that subjects saw only one sentence from each experimental stimulus set, four parallel lists of sentences were created using a Latin Square to assign sentences to the four experimental conditions. Subjects were assigned to one of the four lists at random. Thus, each subject saw six experimental sentences per condition and never saw more than one sentence from each sentence quartet. List was included as a between-group factor in the by-subjects analyses reported below. In the by-items analyses, the corresponding between-factor, dubbed Itemgroup, represented the four subgroups of sentence quartets which were presented in the four conditions of a list. Each sentence of each quartet was seen by 11 subjects per condition.
Task

A self-paced reading paradigm was used in the cumulative moving-window format similar to Stowe’s (1989) paradigm. No grammaticality judgements were required. Subjects had to press a key on the keyboard to see the next word of a sentence. Sentences were presented a third of the screen down, on one or two lines. Line breaks never occurred until after at least two words following the crucial disambiguating region.

Procedure

Subjects were tested individually in sound-attenuated experimental cubicles each fitted with an Elonex PC. Subjects were told by the Experimenter that they were to read and understand sentences to be presented on the computer screen, and that they had to fill in a paper-and-pencil recognition test at the end of the experiment to indicate which sentences they recognised as having read during the experiment. Instructions and all trials were presented on the PC. During the presentation of the instructions and the practice trials the Experimenter stayed in the cubicle. After the practice trials there was a short break and the subject was encouraged to clarify any questions they might have. During the experiment, the subject was alone in the cubicle, and immediately upon leaving, was handed the recognition test to be filled in straight away.

Practice trials followed a screen of instructions and were initialized by pressing the plus-key at the right of the key board. This key was also the button subjects had to press to see the next word during sentence presentation. It was chosen to be the key for seeing the practice trials to ensure that subjects had familiarized themselves with the location of the key before the real trials began.

Each trial was initiated by the message "Press the space bar to begin the next trial" presented at the top of the screen. After the subject pressed the space bar, the message at the top of the screen was replaced with the message: "Press the + key to see the
next word", which stayed there throughout the trial.

Sentences were presented cumulatively one word at a time, at the rate of the subject's key presses. They were presented in white letters on black background, on one or two horizontal lines about a third down the screen. Line breaks occurred only on position 7 or later in the sentence. No punctuation marks except for the full stop at the end of each sentence were given.

Scoring regions

As indicated in table 8.1. above, certain words in the experimental sentences are identified by particular positions for purposes of analysis. Data analyses focuses on positions 4, 5 and 6 and the adverbial (adv1 and adv2).

8.1.5. Results

Recognition Test

Subjects did not have any difficulty in discriminating the sentences of the recognition test which had not been part of the experiment from those that they had read during the experiment. 81% of the subjects correctly identified the foils; the average number of mistakes per subject was less than 1.

Self-paced reading experiment

Overview of analyses

The variability in subjects' response times was quite considerable. For positions 5 and 6 the data violated the assumption that the variance-covariance matrices are equal for
all levels of the between-factor. For position 6, the data based on the means of logtransformed responses times did not violate this assumption, and the results of the analyses of variance will be reported below. Furthermore, response times for positions 5 and 6 in a sentence were added, and analyses carried out on these summed response times. Given that subjects differ in their overall speed of responding, some subjects might get garden-pathed on position 5 whereas others might show the garden-path effect only on position 6. Such a late garden path effect is particularly likely with subjects who tend to press the pace-key very fast. They might not be processing the word that comes up when they press the button, but a word one or two words previously. Thus, the response times summed over positions 5 and 6 might be a better indicator of a garden path effect than responses at either position alone.

In an attempt to make use of the variability in the data, further analyses were carried out on the extremely long response times to establish whether they occurred more frequently in the ambiguous than the nonambiguous conditions in positions 5 and 6.

Finally, responses on the adverbials were compared for the two non-ambiguous conditions to test whether adverbials cause greater processing complexity in the animate condition. According to the strong thematic hypothesis, in sentences with an animate subject noun, there should be a strong expectation of a direct object noun following the verb. If instead the adverbial is presented, this should lead to increased processing costs in the animate, but not the inanimate condition.

Position 6

After logtransforming response times, subject- and item-means were calculated and entered into analyses of variance. Table 8.2. below gives the geometric means, based on the subject-means, per condition:
Table 8.2: cell means of Experiment 5, position 6

<table>
<thead>
<tr>
<th></th>
<th>animate</th>
<th>inanimate</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ambig.</td>
<td>n_amb</td>
<td>ambig.</td>
<td>n_amb</td>
</tr>
<tr>
<td>List 1</td>
<td>386</td>
<td>390</td>
<td>389</td>
<td>378</td>
</tr>
<tr>
<td>List 2</td>
<td>431</td>
<td>370</td>
<td>375</td>
<td>393</td>
</tr>
<tr>
<td>List 3</td>
<td>443</td>
<td>406</td>
<td>368</td>
<td>350</td>
</tr>
<tr>
<td>List 4</td>
<td>471</td>
<td>414</td>
<td>345</td>
<td>448</td>
</tr>
<tr>
<td></td>
<td>431</td>
<td>395</td>
<td>368</td>
<td>391</td>
</tr>
</tbody>
</table>

There was a significant main effect for Animacy $[F(1,40)=10.12, p=.003]$;
F2(1,20)=6.42, p=.020; but MinF'(1,44)=3.93, p<.10]. Furthermore, the interaction Animacy x Ambiguity was significant [F1(1,40)=5.37, p=.026; F2(1,20)=9.09, p=.007; but MinF'(1,60)=3.38, p<.10]. No other effects were significant. Post-hoc paired t-tests indicated that responses in a_amb were significantly longer than in a_namb, both across Lists and across Itemgroup (p's < .0275, 1-tail).

Responses summed over positions 5 and 6

For each subject, the response time at position 5 and the response time at position 6 of the same sentence were added, and the sums logtransformed. Subject-means were then calculated over the sums from the six sentences of each condition. Similarly, for each sentence, the two responses from a particular subject at position 5 and 6 in the sentence were added, the sums logtransformed, and item-means calculated over the sums from the 11 subjects of each condition. Table 8.3. shows the geometric means, based on the subject-means:

Table 8.3.: cell means of Experiment 5 for responses summed over positions 5 and 6

<table>
<thead>
<tr>
<th></th>
<th>animate</th>
<th>inanimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ambig.</td>
<td>n_amb</td>
</tr>
<tr>
<td>a_amb</td>
<td>781</td>
<td>803</td>
</tr>
<tr>
<td>a_namb</td>
<td>824</td>
<td>751</td>
</tr>
<tr>
<td>i_amb</td>
<td>920</td>
<td>816</td>
</tr>
<tr>
<td>i_namb</td>
<td>978</td>
<td>845</td>
</tr>
<tr>
<td></td>
<td>872</td>
<td>803</td>
</tr>
</tbody>
</table>

There was a main effect for Animacy [F1(1,40)=14.04, p=.001; F2(1,20)=11.02, p=.003; MinF'(1,49)=6.17, p<.025]. The interaction Animacy x Ambiguity was also significant [F1(1,40)=4.42, p=.042; F2(1,20)=7.67, p=.012; but MinF'(1,60)=2.80, p<.10]. In the by-items analysis, the interaction Itemgroup x Ambiguity was marginally significant [F2(3,20)=3.25, p=.044]. No other effects were significant. Post hoc paired
t-tests indicated that responses in the a_amb condition were significantly longer than in the a_namb condition (by-subjects: t(43)=1.78, p < .0415, 1-tail; by-items: t(23)=2.55, p < .009, 1-tail).

Given that the garden path effect in the animate condition (the comparison of the a_amb versus the a_namb condition) was only just significant in the by-subjects analysis, I analyzed the garden path effect in relation to the speed of a subject. It is possible that slower subjects, who do not pace through the sentence, do show a garden path effect whereas fast subjects do not. Plotting the difference in mean response times in the ambiguous versus the nonambiguous condition of each subject against the speed of making a response in the nonambiguous condition, however, did not show a clear relationship. Slower subjects, as defined here by their mean response time in the animate-nonambiguous condition, were not more likely to show a garden path effect (a_amb - a_namb) than fast subjects.

**Extremely long responses**

Further evidence for a garden path effect might be found by considering the distribution of extremely long responses on positions 5 and 6, and to test whether they occur more frequently in the ambiguous than the nonambiguous conditions.

Responses were defined as extreme when they were greater than the cut-off point established by collapsing responses from all four sentence types at a particular position. As cut-off I chose the value that lies 1.5 times the interquartile range above the 75th percentile of the distribution of decision times collapsed over conditions (this definition is sensitive to the spread of the middle 50% of the data, without being affected by the highest values of the distribution which are likely to inflate the definition of a cut-off based on standard deviations).

The cut-off for position 5 was 715 ms, and a total of 8% of all responses at position 5 fell beyond it. Table 8.4. shows the percentage of extremely long values per condition.
Table 8.4.: Experiment 5: percentage of extremely long responses per condition at
position 5

<table>
<thead>
<tr>
<th></th>
<th>a_amb</th>
<th>a_namb</th>
<th>i_amb</th>
<th>i_namb</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.5%</td>
<td>9%</td>
<td>5%</td>
<td>6%</td>
<td></td>
</tr>
</tbody>
</table>

The cut-off at position 6 was 806 ms, and a total of 7% of all responses were extremes. Table 8.5. shows the distribution of extremes per condition:

Table 8.5.: Experiment 5: percentage of extremely long responses per condition at
position 6

<table>
<thead>
<tr>
<th></th>
<th>a_amb</th>
<th>a_namb</th>
<th>i_amb</th>
<th>i_namb</th>
</tr>
</thead>
<tbody>
<tr>
<td>11%</td>
<td>6%</td>
<td>4%</td>
<td>6%</td>
<td></td>
</tr>
</tbody>
</table>

Thus, there were only few extremely long responses, even in the a_amb condition. In fact, many subjects did not have any extremely long responses in a particular condition (i.e. looking just at the a_amb condition, 26 subjects had no extremes on position 5, and 26 subjects had no extremes on position 6).

In a further step, I analyzed the number of extremes separately for a fast and a slow group of subjects. The two subject groups were established on the basis of their overall mean response time collapsed over all conditions, and over positions 1 to 6 of all sentences. The 22 subjects with the highest overall mean response time were designated the slow group, and the 22 other subjects, the fast group. Of the 22 fast subjects, 19 had no extremely long responses in a_amb on position 5, and 20 had no extremely long responses in a_amb on position 6 (for the i_amb condition, both figures are 19). However, of the 22 subjects in the slow group, only 7 did not have any extremes in a_amb on position 5, and only 6 had no extremes in a_amb on position 6 (for i_amb, the figures are 14 and 15, respectively). Table 8.6. below gives the percentage of extremely long responses out of all responses per condition for positions 5 and 6 for the slow group only (the total number of observations per cell here is 132,
i.e. 22 subjects times 6 sentences).

Table 8.6.: Experiment 5: percentage of extremely long responses per condition at positions 5 and 6 for the group of slow subjects

<table>
<thead>
<tr>
<th></th>
<th>a_amb</th>
<th>a_namb</th>
<th>i_amb</th>
<th>i_namb</th>
</tr>
</thead>
<tbody>
<tr>
<td>pos. 5</td>
<td>23%</td>
<td>14%</td>
<td>8%</td>
<td>11%</td>
</tr>
<tr>
<td>pos. 6</td>
<td>21%</td>
<td>9%</td>
<td>7%</td>
<td>12%</td>
</tr>
</tbody>
</table>

Wilcoxon Matched-pairs Signed-ranks Tests performed for the slow group only indicated that on position 5 the number of extremes differed significantly for the a_amb and the i_amb condition (p=.0157, 2-tail); however, there was no significant difference in the number of extremely long responses of the a_amb versus the a_namb condition. On position 6, however, the number of extremely long responses was significantly higher in a_amb than a_namb (p<.012, 1-tail). The number of extremes in the two inanimate conditions did not differ on either position.

Early positions

An analysis of extremely long responses was also carried out for the positions early in the sentence to ensure that the differences in number of extremes for the a_amb versus the a_namb condition on position 6 did reflect the effect of greater processing complexity due to garden pathing, rather than some difference in the difficulty of the sentences in the different conditions. Table 8.7. gives the percentages of extremely long responses for positions 1, 2, 3 and 4 and the cut-offs for each position. For each position, extremes were defined on the basis of the distribution of all responses at that position, collapsed over condition. Responses of all subjects were included. According to both the strong thematic hypothesis and the alternative hypothesis, the NP in positions 3 and 4 should show fast reading times in the ambiguous condition where subjects are expected to falsely analyse the NP as the direct object.
Table 8.7.: Experiment 5: cut-offs and percentages of extremely long responses per condition at the early positions

<table>
<thead>
<tr>
<th>pos.</th>
<th>cut-off</th>
<th>a_amb</th>
<th>a_namb</th>
<th>i_amb</th>
<th>i_namb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>542</td>
<td>6%</td>
<td>4%</td>
<td>6%</td>
<td>6%</td>
</tr>
<tr>
<td>2</td>
<td>632</td>
<td>7%</td>
<td>5%</td>
<td>6%</td>
<td>7%</td>
</tr>
<tr>
<td>3</td>
<td>672</td>
<td>3%</td>
<td>9%</td>
<td>4%</td>
<td>9%</td>
</tr>
<tr>
<td>4</td>
<td>683</td>
<td>9%</td>
<td>9%</td>
<td>6%</td>
<td>7%</td>
</tr>
</tbody>
</table>

As can be seen from this table, the extremely long responses are distributed quite evenly across the different conditions. It is only at position 3 that there are more extremes in both nonambiguous conditions as compared to the ambiguous conditions. These extremes were not confined to particular items (they occurred for 13 items in each condition), but the majority of them came from subjects of the slow group (the extremes in the a_namb condition were due to 17 subjects, 15 of whom are part of the slow group; the extremes in the i_namb condition came from 18 subjects, 15 of whom were in the slow group).

The whole range of responses at the early positions, not just the extremes, were also analyzed in analyses of variance. There were no significant effects at positions 1, 2 or 4. At position 3, there was a significant main effect of Ambiguity, with longer mean responses in the nonambiguous (geometric mean: 375 ms) as compared to the ambiguous conditions (329.5 ms) [F1(1,40)=40.25, p<.0005; F2(1,20)=14.65, p=.001; MinF'(1,35)=10.74, p<.01].

Adverbial

Responses in the two non-ambiguous conditions were compared to establish whether an animate subject noun leads to the expectation of a transitive analysis, which, if ruled out by the adverbial following immediately after the verb, leads to increased processing complexity as compared to the inanimate condition, where, according to the strong thematic hypothesis, no such expectation exists.
In some sentences, the adverbial phrase consisted of more than one word. Initially, only responses to the first word of the adverbial were analyzed. Paired t-tests in both the by-subjects and the by-items analyses failed to indicate any significant difference between the two nonambiguous conditions. Furthermore, the number of extremely long responses in the two conditions (with the cut-off established on the basis of the joint distribution of responses at a_namb and i_namb) was exactly 8% in both conditions.

Finally, the average response time per letter across all the words making up the adverbial was computed. Again, paired t-tests failed to yield a significance level small enough to justify the rejection of the null hypothesis that there was no difference between the a_namb and the i_namb condition on the adverbial.

8.1.6. Discussion

The results of Experiment 5 quite clearly show that there is no garden path effect in the inanimate-ambiguous condition. Responses in this condition were faster than in the animate-ambiguous condition, and they did not differ from responses in the inanimate-nonambiguous condition. Furthermore, the number of extremely long responses did not differ from the inanimate-nonambiguous condition, not even for the sub-group of the slower subjects. For the sentences with an animate subject noun, however, there was evidence for a garden path effect, as indicated by longer mean response times on position 5, and position 5 and 6 taken together in the animate-ambiguous as compared to the animate-nonambiguous condition. The number of extremely long responses in these two conditions differed only on position 6, and only for subjects whose overall response speed was slower than the average for all subjects tested in this experiment. This pattern of results is entirely consistent with Stowe’s findings (1989, Experiment 1) and the strong thematic hypothesis according to which the semantic feature INANIMATE of a subject noun leads to the assignment of the Theme role, rather than the Agent role. Only one of the thematic frames of the causative/ergative verbs is compatible with the Theme role of the subject. The intransitive subcat frame associated with the <Theme> frame leads to the early closure of the verb phrase after the verb,
and the analysis of the ambiguous noun phrase as the beginning of a new clause.

There are some aspects of the findings, however, which do not seem to fit easily with such an interpretation.

Firstly, there is the Ambiguity main effect at position 3. While response times (and number of extremes) did not differ in the early positions in the sentence (positions 1, 2 and 4), there was a difference on position 3, where average responses were longer, and there were more extremely long responses, in the two non-ambiguous conditions as compared to the two ambiguous conditions. In the context of the strong thematic hypothesis, it is somewhat problematic to account for the finding of an Ambiguity main effect rather than an interaction with Animacy. In the animate condition, this difference between ambiguous and nonambiguous sentences could straightforwardly be interpreted as reflecting the fact that, in the nonambiguous sentences, the word at position 3 marks the beginning of a new clause (the main clause), whereas in the ambiguous sentences, it marks the beginning of a new phrase. It seems likely that the beginning of a new clause requires more processing time than the beginning of a new phrase. The problem is, of course, that in the inanimate conditions, the word at position 3 should be analyzed as the beginning of a new clause in both the nonambiguous and the ambiguous sentences. The lack of a garden path effect on positions 5 or 6 for the inanimate-ambiguous sentences invites the interpretation that the ambiguous noun phrase (positions 3 and 4) was analyzed as the beginning of a new clause; that is, the word in position 3 in the inanimate-ambiguous condition signals the beginning of a new clause in the same way as it does in the inanimate-nonambiguous conditions. But how should one account for the longer response times on position 3 in the inanimate-nonambiguous as compared to the inanimate-ambiguous condition?

One could argue, of course, that the word in position 3 in the i_amb condition was not initially analyzed as the beginning of a new clause in the inanimate-ambiguous condition, but as the beginning of a direct-object NP (which would also be in line with the finding that on positions 3 and 4 responses in the i_amb condition did not differ
from those in the a_amb condition); one would then have to assume that this analysis was given up rapidly and in time before the words on position 5 and 6 appeared.

Another aspect of the findings is also difficult to interpret if the strong thematic hypothesis is true. In the animate condition, subjects should strongly expect a direct object after the verb. Thus, when they see an adverbial instead, which forces the intransitive analysis, they should take longer to process it than when they did not strongly expect a direct object, as in the inanimate condition. The results of Experiment 5 clearly did not show any effect of Animacy on the adverbial, however.

Thus, the results of Experiment 5 present a somewhat mixed picture. On the one hand, the evidence is quite strong showing a garden path effect for the animate, but not the inanimate condition. On the other hand, other predictions derived from the strong thematic hypothesis are not supported by the evidence. Firstly, there was no difference on the adverbial in the two nonambiguous conditions. Secondly, the interpretation of the increase in processing time in the two non-ambiguous conditions on position 3 as reflecting extra processing costs incurred in analyzing the beginning of a new clause, does not fit with the lack of such an increase on position 3 in the i_amb condition, which should also be analyzed as the beginning of the new clause if the strong thematic hypothesis is true.
Chapter 8.2.: Experiment 6: Animacy/inanimacy of the subject noun and the garden path effect: self-paced reading with continuous grammaticality judgements

8.2.1. Introduction

Like Experiment 5, this experiment tests the influence of semantic information on processing syntactically ambiguous sentences. Again, the sentences are temporarily ambiguous because the postverbal noun phrase can be analyzed either as a direct object, or as the subject noun phrase of the following main clause, as in:

Although the cheerful clowns had changed the mood remained the same for a while.

Like Experiment 5, Experiment 6 tests the strong thematic hypothesis favoured by Stowe (1989), according to which a garden path only occurs if the first subject noun is animate, as in the example above. According to this hypothesis, Animacy leads to the default assignment of the Agent role, which in turn activates the <Agent, Theme> frame of the verb, and the associated transitive subcat frame. If the first subject noun is inanimate, Stowe predicts that it is assigned the Theme role, which activates the other thematic frame of the verb, <Theme>, which is linked to the intransitive subcat frame of the verb. The postverbal noun phrase can thus not be attached in the verb phrase as direct object, because there are no free slots left in the subcat frame, and no thematic roles to assign by the verb. Thus, the postverbal noun phrase has to be analyzed as the subject of a new clause, and no garden path effect should occur on the following verb.

However, in Stowe's experiment semantic features of the subject noun was confounded with the plausibility of the transitive action expressed in the subordinate clause. Experiment 6, like Experiment 5, investigates the impact of semantic features of subject nouns in sentences where the plausibility of the transitive action does not differ between the conditions.

Experiment 6 uses a task in which the subject has to make continuous grammaticality
judgements while reading a sentence. This is the task used by Stowe (1989). It is likely to provide longer response times than the self-paced reading paradigm used in Experiment 5, thus possibly allowing to tap the sentence processing system at later stages than in Experiment 5. The additional task requirement of having to decide at each word whether the sentence is grammatical might also discourage subjects from pacing rapidly through the sentence, a strategy that at least some subjects seemed to have adopted during the self-paced paradigm of Experiment 5.

8.2.2. Hypotheses

As in Experiment 5, it was predicted that in the animate ambiguous sentences processing complexity would increase, indicating a garden path effect, on or after the disambiguating word. If the plausibility of a transitive sentence rather than the semantic feature animacy affects syntactic analysis, there should also be a garden path effect in the inanimate ambiguous condition. Increased processing complexity should manifest itself in longer response times, and more false negative responses on the disambiguating word/s in the ambiguous as opposed to the unambiguous condition.

8.2.3. Method

Subjects

There were 44 subjects in this experiment, who were recruited from staff and students of University College London.

Task

The self-paced reading paradigm of Experiment 5 was adjusted to include repeated grammaticality judgements from the subject. The same moving-window paradigm was
used, i.e. words were presented one word at a time in a cumulative fashion. Again, sentences were presented on one or two lines a third of the way down the screen. In this experiment, subjects did not press a button to indicate that they were ready to see the next word as in the previous self-paced experiment. Instead, they pressed a key to indicate whether the sentence was grammatical at the word they were reading. They were instructed to press a key labelled Y (for YES) to indicate their decision that the sentence was grammatical up to this point, and to press a key labelled N (for NO) to indicate their decision that the word they were currently reading did not fit into the sentence grammatically. Pressing the YES-key (and pressing the NO-key when the sentence was in fact grammatical) prompted the presentation of the next word. Subjects were given feedback, which is described in the Procedure section below.

Design

The design is identical to the one used in the previous self-paced experiment: in both the by-subjects and the by-items designs there are two crossed repeated-measures factors with two levels each, i.e. Ambiguity (ambiguous and non-ambiguous) and Animacy (animate and inanimate). Furthermore, in the by-subjects analysis, there was the between-group factor List, and in the by-items analysis the between-group factor Itemgroup. The example stimulus set presented in table 8.1. above is reproduced here as table 8.8. below to illustrate the different conditions. Again, the word positions are indicated on which the statistical analyses were carried out. I occasionally refer to positions 5 and 6 as the disambiguating information.
Table 8.8. Experiment 6: example stimulus set

<table>
<thead>
<tr>
<th>animate-ambiguous (a_amb):</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6</td>
</tr>
<tr>
<td>Although the cheerful clowns had changed the mood remained the same for a while.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>animate-nonambiguous (a_namb):</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 adv1 adv2 3 4 5 6</td>
</tr>
<tr>
<td>Although the cheerful clowns had changed beyond recognition the mood remained the same for a while.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>inanimate-ambiguous (i_amb):</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>Although the cheerful campaign had changed the mood remained the same for a while.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>inanimate-nonambiguous (i_namb):</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 adv1 adv2 3 4 5</td>
</tr>
<tr>
<td>Although the cheerful campaign had changed beyond recognition the mood remained the same for a while.</td>
</tr>
</tbody>
</table>

Dependent variables

In addition to logging decision times (time in milliseconds from display of a word until a response is made) as the dependent variable, as in the first self-paced experiment, I recorded which key the subject pressed and whether the decision was correct or false given the input sentence.
Stimuli

There were 98 sentences, 50 of which were ungrammatical. 24 sentences were experimental trials, the rest were filler trials. Ungrammatical sentences ended with a word which either belonged to an incorrect syntactic category (e.g. 'That the within'; '... receive their his') or contained a morphological mistake (e.g. '... prepared an vegetable'; '... which he solds') or an agreement error ('the new assistant are'). Ungrammatical words occurred in various positions in the sentences. In 30 sentences the ungrammatical word occurred more than 9 words into the sentence, and in 20 sentences the ungrammatical word was the 9th or earlier word in the sentence.

The sentences were the same as those used in Experiment 5, with a few modifications. In some sentences slight stylistic changes were made, or sentences were made longer to allow more positions in the sentence in which an effect of processing complexity might become apparent.

For the experimental trials, the same 24 sentence quartets were used as in the previous self-paced experiment. A few minor changes were made. The complete set is given in Appendix I, with any changes from the set used in the Experiment 5 indicated by brackets. Sentence sets 3, 4, 11 and 24 were made longer to ensure that in all sentences at least three words followed the disambiguating word. Sentence quartets 15 and 23 were changed stylistically (in 15, the end of the subordinate clause in the non-ambiguous condition now read 'sank beneath the waves' rather than 'sank in the sea'; in sentence set 23, the conjunction was changed from 'before' to 'when'). Four sentence quartets (sets 2, 15, 16 and 21) were made ungrammatical late in the sentence, i.e. on a word at least four words after the disambiguating information on which a garden path might occur.

Again, to ensure subjects saw only one sentence from each quartet, four parallel lists of sentences were created using a Latin Square to assign sentences to the four experimental conditions. Each subject saw six sentences in each condition, with one sentence per condition containing an ungrammaticality late in the sentence.
The same filler sentences were used as in Experiment 5. All fillers were unrelated to the present experiment and were chosen to represent a variety of sentence structures. Fillers were identical in the four lists.

There were 14 practice trials, which were identical in the four lists. 8 practice sentences were ungrammatical. The practice sentences represented the different types of stimuli and ungrammaticalities used in the main experiment.

**Procedure and Apparatus**

Subjects were tested individually in small, sound-attenuated experimental cubicles. Presentation of instructions and stimuli, and recording of responses was again controlled by Elonex PC's (using a program written with MEL-software). Two keys of the Elonex keyboard were re-labelled: the plus-key to the right of the numeric keypad was designated as the yes-key, and covered with a white sticker with a black Y printed on it; the minus-key just above it was covered with a white sticker with a black N.

After the subject was seated in front of the computer, s/he first read the Instructions which were presented on the screen, and then did the practice trials. During this time the Experimenter was present. In the Instructions the subjects were told that "half of the sentences are ungrammatical, that is they are incorrect in any context one could think of". After the practice trials there was a short break and the subject was encouraged to ask any questions about the procedure they might have. The experimenter then told the subject that after s/he had seen all trials, a message would appear on the screen informing her/him that the experiment was over. The experimenter encouraged the subject to respond as fast and as accurately as possible and left the cubicle.

Each trial consisted of the following events. First, the message "Press the 'YES-key' to see the first word and then the 'YES-' or 'NO-key' " was presented at the top of the screen, and remained there till the end of the trial. Sentences were presented one word
at a time (following a correct Yes-decision, or an incorrect No-decision) in a
cumulative fashion. Sentences were presented on one or two horizontal lines a third
down the screen, in white letters on black background. After the subject had made a
response to the last word of a sentence (which was either a grammatical word plus a
full stop, or an ungrammatical word) the sentence disappeared and a feedback message
("Well Done!" or "Wrong!") flashed up for 1000 ms in the centre of the screen. Then
the message "Press the space bar to begin the next trial" was presented at the top of
the screen and remained there until the subject initialized the next trial.
Following Stowe’s procedure (to enable a close comparison with her results) the
following kind of feedback was given during sentence reading. In order not to interrupt
reading, explicit feedback was given only at the end of sentences. Since a sentence
always stopped after the ungrammatical word, there was always feedback after an
ungrammatical sentence: the message "Well done!" was displayed if subjects had
pressed the no-key, and "Wrong!", if they had pressed the yes-key. If the sentence
ended with a grammatical word (presented with a full-stop) and the subject pressed the
yes-key, the message "Well done!" was displayed. If the subject pressed the no-key
after a grammatical sentence-final word, the message "Wrong!" appeared.

No explicit feedback was given after a decision on a grammatical word which was not
sentence-final: after either a correct yes-decision or a false negative response, the next
word was presented without feedback (subjects had been told that a sentence would
only continue if the last word they had seen was grammatical).

8.2.3. Results

Overview of analyses

Error data and response times for correct decisions were analyzed. The errors analyzed
here are the instances when a subject made a No-decision when the sentence was in
fact grammatical. This is of particular interest for positions 5 and 6 in the ambiguous
sentences, where the subject sees a word which is not compatible with the transitive
The analysis of response times for correct decisions at the crucial positions was done in several ways. Initially, for each subject the mean response time over the six sentences that were presented per condition were calculated (and item-means were formed over the responses from the 11 subjects per condition), and entered into analyses of variance. Secondly, the distribution of extremely long responses across the different conditions was analyzed. Thirdly, all extremely long responses were removed, and analyses of variance were run with subject-and item-means that had been calculated over the remaining responses. This allows us to test whether the distributions of the different cells did not only differ in the number of extremes, but also in their means.

**Error Data**

The percentages and absolute number of errors, i.e. of false negative responses, per sentence condition and word position in the sentence (Pos) are given in Table 8.9. below. The total number of observations per cell was 264 (44 subjects by 6 sentences).

<table>
<thead>
<tr>
<th></th>
<th>a_amb</th>
<th>a_namb</th>
<th>i_amb</th>
<th>i_namb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos 1</td>
<td>1 (.4)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pos 2</td>
<td>1 (.4)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pos 3</td>
<td>1 (.4)</td>
<td>2 (.8)</td>
<td>0</td>
<td>1 (.4)</td>
</tr>
<tr>
<td>Pos 4</td>
<td>2 (.8)</td>
<td>8 (3)</td>
<td>1 (.4)</td>
<td>2 (.8)</td>
</tr>
<tr>
<td>Pos 5</td>
<td>46 (17)</td>
<td>2 (.8)</td>
<td>18 (7)</td>
<td>2 (.8)</td>
</tr>
<tr>
<td>Pos 6</td>
<td>15 (6)</td>
<td>2 (.8)</td>
<td>5 (2)</td>
<td>1 (.4)</td>
</tr>
<tr>
<td>Pos 7</td>
<td>7 (3)</td>
<td>3 (1)</td>
<td>5 (2)</td>
<td>5 (2)</td>
</tr>
</tbody>
</table>

Table 8.9. above clearly shows that very few incorrect decisions were made at the
early positions, before the disambiguating word occurred (i.e. before Position 5). The only errors of any magnitude occurred at Positions 5 and 6. At these positions, the distribution of errors across sent_type followed the pattern of the response time data described below.

**Position 5**

At Position 5, 12 subjects made no false decisions at all (they were evenly distributed across the four Lists), and there were two items which received no false decisions at all. In the two non-ambiguous conditions, only two subjects made a false decision. In the animate-ambiguous condition, 27 subjects made one or more false decisions, as did 14 subjects in the inanimate-ambiguous condition. Of the total 24 items, 20 received at least one false decision in the animate-ambiguous condition, and 11 items received at least one false decision in the inanimate-ambiguous condition. Wilcoxon Matched-pairs Signed-ranks Tests confirmed that the number of No-decisions in the animate-ambiguous and the inanimate-ambiguous conditions differed significantly (p's < .003, 2-tail). The differences between i_amb and i_namb, and a_amb and a_namb were also significant (p's < .0099, 2-tail, and < .0001, 2-tail, respectively).

**Position 6**

There were only 13 instances in which a subject made a false decision at Position 6 in the same sentence in which they had made a false decision at Position 5 (the converse is that of the 23 incorrect decisions on Position 6, only 10 were in a sentence where the subject had made a Yes-decision on Position 5). Breaking this down by sentence type, the following pattern emerges: in a_amb, in 10 out of the 46 cases in which the subject made a false negative response at Position 5, they also pressed the No-key at Position 6; in i_amb, they did so in two of the 18 cases that were false at Position 5, and in i_namb, in one of the two false responses at Position 5. Especially with respect to the two ambiguous conditions, it is surprising that the majority of subjects who
made a No-decision at position 5 went on to make a Yes-decision at position 6, given that the word at position 6 does not give any additional syntactic information that was not already available at position 5 (it merely adds more information in line with the syntactic analysis that the word at position 5 had enforced). One possibility is that when the word at position 6 appeared, subjects realized that they had made a mistake at position 5 (remember that a sentence never continued after a word that truly made it ungrammatical), and quickly pressed the Yes-key because they wanted to see more input, rather than because they had resolved the garden path. A fast Yes-decision at position 6 might also have resulted from a judgement of the compatibility of the words at positions 5 and 6, rather than from a judgement of the grammaticality of the whole sentence up to the word at position 6. However, the data clearly indicate that the time to make a Yes-decision on position 6 after having made a No-decision on position 5 in the same sentence is very high (median: 2286 ms in the a_amb condition, 1992 ms in the i_amb condition), whereas the time to make a Yes-decision at position 6 after having also made a Yes-decision at position 5 in the same sentence is much lower (median: 643 ms for a_amb, 656 ms for i_amb). Thus, it seems likely that subjects did not press the Yes-key at position 6 after having pressed the No-key on position 5 merely to get some more input, or because of the local compatibility of the words in positions 5 and 6. Instead, the long Yes-decision times on position 6 probably indicate that subjects are trying to resolve the garden path, trying to find some re-analysis of the sentence that was clearly not available to them at position 5 where they had rejected the sentence as ungrammatical.

**Decision Times**

In all analyses reported below, decision times were analyzed only for correct decisions. Times for incorrect responses were dropped and subject- and item-means per condition were calculated over the response times for the remaining correct decisions. These means were entered into analyses of variance with Animacy and Ambiguity as within factors, and either List (in the by-subjects analysis) or Itemgroup (in the by-items analysis) as between factors.
Decision Times at Position 5

Figure 8.2 shows the means per sentence type (based on subject-means).

The cell means are presented separately per List in table 8.10. below.
Table 8.10.: Experiment 6, cell means (in ms) for position 5

<table>
<thead>
<tr>
<th></th>
<th>animate</th>
<th></th>
<th>inanimate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a_amb</td>
<td>a_namb</td>
<td>i_amb</td>
<td>i_namb</td>
</tr>
<tr>
<td>List 1</td>
<td>1335</td>
<td>724</td>
<td>1127</td>
<td>688</td>
</tr>
<tr>
<td>List 2</td>
<td>1498</td>
<td>658</td>
<td>1116</td>
<td>599</td>
</tr>
<tr>
<td>List 3</td>
<td>1913</td>
<td>1019</td>
<td>938</td>
<td>750</td>
</tr>
<tr>
<td>List 4</td>
<td>2248</td>
<td>845</td>
<td>901</td>
<td>785</td>
</tr>
<tr>
<td></td>
<td>1748</td>
<td>811</td>
<td>1020</td>
<td>705</td>
</tr>
</tbody>
</table>

It is noticeable that the means within each sentence type vary quite considerably, especially in the a_amb condition. It turns out that three of the six items (items 5, 9 and 21) which subjects of List 1 saw in the a_amb condition were among the five items with the shortest mean decision times of all 24 items in this condition. The individual responses (rather than the mean responses averaged over subjects) to those three items in condition a_amb were consistently fast across the eleven subjects of List 1.

I also checked the distribution of sentences across the different lists to identify sentences which lead to particularly strong garden paths as indicated by extremely long decision times. Extremely long decision times were defined as scores falling beyond the cut-off point of 1986 ms, which is the value that lies 1.5 times the interquartile range above the 75th percentile of the distribution of decision times collapsed over conditions (this definition was chosen because it is sensitive to the spread of the middle 50% of the data, without being affected by the highest values of the distribution which are likely to inflate the definition of a cut-off based on standard deviations).

The maximum number of extremely long responses for a sentence in condition a_amb was 6 (remember that each sentence was seen by 11 subjects per condition). (The maxima for the other conditions were 2 for a_namb, 4 for i_amb and 1 for i_namb) In condition a_amb, there were three sentences (itemid 2, 7 and 11) with the maximum of
six extremely long responses. Two of the sentences were among the sentences in the a_amb condition seen by subjects of List 3, and the other sentence was among the sentences in the a_amb condition seen by subjects assigned to List 4. These three sentences did not lead to extremely long decision times in the animate-nonambiguous condition, i.e. it is unlikely that the extremely long responses to these sentences in the animate-ambiguous condition were due to some difficulty or oddness in the sentence, which would have made processing difficult even in the nonambiguous version of the sentence. Thus it seems likely that the three sentences lead to particularly strong garden path effects. By chance, they were all in the stimulus sets which subjects of List 3 and 4 saw in the a_amb condition.

The decision time data, whose means were presented in Table 8.10 above, were entered into repeated-measure analyses of variance. Note that exactly the same pattern of results as reported below was obtained when the data were first log-transformed before computing subject- and item-means. The cell means based on the antilogs of the subject data are presented in Table 8.11 below.

<table>
<thead>
<tr>
<th>Table 8.11: Experiment 6, cell means (geometric means) in ms for position 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>animate</td>
</tr>
<tr>
<td>ambig.</td>
</tr>
<tr>
<td>a_amb</td>
</tr>
<tr>
<td>List 1</td>
</tr>
<tr>
<td>List 2</td>
</tr>
<tr>
<td>List 3</td>
</tr>
<tr>
<td>List 4</td>
</tr>
<tr>
<td>1260</td>
</tr>
</tbody>
</table>

The analyses of variance based on the non-transformed decision time data indicated that there was no main effect for either List or Itemgroup. There was a significant main effect for Ambiguity [F1(1,40)=67.86, p<.0005; F2(1,20)=60.50, p<.0005; MinF'(1,51)=31.98, p<.001] and for Animacy [F1(1,40)=26.22, p<.0005;
In the by-subjects analysis, Animacy interacted with List [F(3,40)=3.14, p=0.036], with values of the transformed variable testing the Animacy main effect differing significantly between List 1 and List 4 [t(40)=-2.09, p=0.043, 2-tail]. The interaction Ambiguity x Animacy was significant [F(1,40)=18.07, p<0.0005; F(2,120)=18.23, p<0.0005; MinF'(1,53)=9.07, p<0.01]. Post hoc paired t-tests revealed that responses in the a_amb condition were significantly longer than in any other condition, in the by-subjects and the by-items analyses (all p's < 0.0005). In the by-subjects analysis, this interaction was marginally affected by List (whereas in the by-items analysis, the interaction Itemgroup x Ambiguity x Animacy was not significant). The interaction List x Animacy x Ambiguity was significant at p=0.047, with F(3,40)=2.89. Post hoc paired t-tests indicated that the responses in the a_amb condition were on average longer than in the i_amb condition (significant for Lists 3 and 4 (p's <0.003, 1-tail, only marginally significant for List 2 [t(10)=1.5, p=0.08, 1-tail] and not significant for List 1). Decision times in the a_amb condition were significantly longer than those in the a_namb condition for all Lists (all p's < 0.009, 1-tail). Times in the i_amb condition were significantly longer than in the i_namb condition for all Lists except List 4 (p's < 0.04, 1-tail).

These patterns of results are true for most subjects and for most items. The a_amb condition received longer responses than the i_amb condition for 21 out of all 24 items, and by 32 out of all 44 subjects. Responses were higher in a_amb than a_namb for 23 items and 38 subjects. Finally, responses for i_amb were higher than those for i_namb for 20 items and 36 subjects.

These results are based on correct decisions only at position 5. However, included are responses to sentences in which the subject might have made a false decision at an earlier position in the sentence. I checked the pattern of false decisions at position 4 in order to track subjects' subsequent responses at position 5. There were 13 false decisions at Position 4, 2 in a_amb, 8 at a_namb, 1 at i_amb and 2 at i_namb. Why a_namb received more false responses than the other conditions is not clear. It might be that subjects tried to attach N2 as the direct object, but obviously could not do so in the unambiguous condition. They would have then judged the sentence to be
ungrammatical at Position 4 because it apparently violated English word order rules (eg. 'Because the boy was burning fiercely the pots ...'). For 12 out of the 13 responses that were false at Position 4, the response at Position 5 was correct. Decision times for these 12 correct responses at Position 5 were extremely long only in two cases (one in a_namb, the other in i_namb) (an extreme response was defined here as lying more than two standard deviations away from the subject's median for that cell). In 9 out of the 12 cases, decision times were the maximum of that subject in that cell. Thus, response times for a_namb at Position 5 might have been slightly inflated, which, if anything, might have made the interaction between Animacy and Ambiguity more difficult to detect.

**Extremely long decision times at position 5**

The above finding that decision times in the animate-ambiguous condition were significantly higher than in the inanimate-ambiguous condition (while there was a garden path effect in both ambiguous conditions) prompted further data analyses for position 5 for the following reasons. These results might indicate that the garden path effect is less strong in the inanimate than the animate condition. The results might, however, also result from a mixed pattern of performance, according to which garden paths occur less frequently in the inanimate than the animate ambiguous conditions; when a garden path occurs in the inanimate condition, it might be as pronounced as in the animate condition, but if it occurs less often, the mean decision times will be lower than in the animate condition. Thus, we need an index for a pronounced garden path effect which is based on individual rather than averaged decision times. In the following analysis, the cut-off of 1987 ms for defining extremely long decision times introduced above was used again. Decision times greater than 1987 ms were considered as potential indicators of garden pathing (of course, extreme values might also result from erratic responses, which, however, should occur randomly in all conditions). A total of 11% of all correct decisions at position 5 were extremely long. Table 8.12. shows the percentage of extremely long decision times per condition.
For each subject (and separately for each item) the number of extremely long decision times was calculated. Wilcoxon Matched-pairs Signed-ranks Tests indicated that the animate-ambiguous and the inanimate-ambiguous condition differed significantly, as did the animate-ambiguous as compared to the animate-nonambiguous, and the inanimate-ambiguous versus the inanimate-nonambiguous conditions (all p's < .005, 2-tail). As expected, the two nonambiguous conditions did not significantly differ.

Thus, there is evidence that the number of extremely long responses was in fact greater in the animate ambiguous than the inanimate ambiguous condition, which might have resulted in a more pronounced garden path effect in the animate ambiguous condition. The two distributions might however differ not only in the number of extremes, but also in their central tendency. In a final analysis of responses at position 5 only decision times smaller than 1987 ms were included, and subject-and item means were entered into analyses of variance (there was one missing value in the by-subjects analysis: one subject of List 4 had no responses left in the animate-ambiguous condition: three of her responses had been extremely long, the other three had been a No-decision). Table 8.13. below gives the cell means for position 5 (based on subject means) after all responses greater than 1986 ms had been removed.
Table 8.13: Experiment 6, cell means in ms on position 5, for responses smaller than 1987 ms

<table>
<thead>
<tr>
<th></th>
<th>animate</th>
<th></th>
<th>inanimate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ambig.</td>
<td>n_amb</td>
<td>ambig.</td>
<td>n_amb</td>
</tr>
<tr>
<td>List 1</td>
<td>742</td>
<td>653</td>
<td>813</td>
<td>688</td>
</tr>
<tr>
<td>List 2</td>
<td>924</td>
<td>597</td>
<td>759</td>
<td>599</td>
</tr>
<tr>
<td>List 3</td>
<td>1062</td>
<td>819</td>
<td>786</td>
<td>684</td>
</tr>
<tr>
<td>List 4</td>
<td>1312</td>
<td>782</td>
<td>728</td>
<td>707</td>
</tr>
<tr>
<td></td>
<td>1003</td>
<td>711</td>
<td>772</td>
<td>669</td>
</tr>
</tbody>
</table>

Figure 8.3. displays the means per sentence type:
There were significant main effects for Animacy [F1(1,39)=26.55, p<.0005; F2(1,20)=8.44, p=.009; MinF'(1,33)=6.40, p<.025] and Ambiguity [F1(1,39)=37.64, p<.0005; F2(1,20)=28.32, p<.0005; MinF'(1,48)=16.16, p<.001]. Furthermore, the interaction Animacy by Ambiguity was also significant [F1(1,39)=12.71, p=.001; F2(1,20)=7.47, p=.013; MinF'(1,43)=4.70, p<.05]. In the by-subjects analysis, there was a significant main effect for List [F1(3,39)=4.07, p=.013]. List interacted with Animacy [F1(3,39)=8.84, p<.0005]. The interaction List x Animacy x Ambiguity was also significant [F1(3,39)=4.18, p=.012]. In the by-items analysis, Itemgroup interacted with Animacy [F2(3,20)=5.77, p=.005]. The interaction Itemgroup x Animacy x Ambiguity was also significant [F2(3,20)=7.02, p=.002]. Post-hoc paired t-tests indicated that the difference between the animate-ambiguous and the inanimate-ambiguous conditions was significant both across Lists and Itemgroups (p's < .020, 2-tail), and also separately for List 2 and Itemgroup 4 (p's <.002, 2-tail). Mean response times in a_amb were longer than a_namb for Lists 2 and 4 (p's <.0001, 1-tail), and longer for i_amb than i_namb for Lists 2 and 3 (p's <.014).

**Decision Times at Position 4**

Response times were logtransformed before the means were computed for subjects and items for each cell, to ensure that the variance-covariance matrices for the within variables were equal for all levels of the between factor. The antilogs of the means per cell (based on the subject means) are presented in table 8.14. below.
Table 8.14.: Experiment 6, cell means in ms (geometric means) for position 4

<table>
<thead>
<tr>
<th></th>
<th>animate</th>
<th></th>
<th>inanimate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ambig.</td>
<td>n_amb</td>
<td>ambig.</td>
<td>n_amb</td>
</tr>
<tr>
<td>List 1</td>
<td>a_amb</td>
<td>722</td>
<td>i_amb</td>
<td>887</td>
</tr>
<tr>
<td></td>
<td>a_namb</td>
<td>739</td>
<td>i_namb</td>
<td>647</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>749</td>
</tr>
<tr>
<td>List 2</td>
<td>a_amb</td>
<td>709</td>
<td>i_amb</td>
<td>635</td>
</tr>
<tr>
<td></td>
<td>a_namb</td>
<td>652</td>
<td>i_namb</td>
<td>639</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>659</td>
</tr>
<tr>
<td>List 3</td>
<td>a_amb</td>
<td>824</td>
<td>i_amb</td>
<td>777</td>
</tr>
<tr>
<td></td>
<td>a_namb</td>
<td>854</td>
<td>i_namb</td>
<td>669</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>781</td>
</tr>
<tr>
<td>List 4</td>
<td>a_amb</td>
<td>718</td>
<td>i_amb</td>
<td>648</td>
</tr>
<tr>
<td></td>
<td>a_namb</td>
<td>957</td>
<td>i_namb</td>
<td>681</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>751</td>
</tr>
<tr>
<td></td>
<td></td>
<td>742</td>
<td>792</td>
<td>729</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>658</td>
</tr>
</tbody>
</table>

There was a significant main effect for Animacy [F1(1,40)=20.54, p<.0005; F2(1,20)=8.77, p=.008; MinF'(1,37)=6.15, p<.025]. List x Animacy was significant [F(3,40)=6.31, p=.001], as was the interaction List x Ambiguity [F(3,40)=4.86, p=.006]. The interaction Animacy x Ambiguity was significant (F1(1,40)=11.55, p=.002; F2(1,20)=7.89, p=.011; MinF'(1,46)=4.69, p<.05), and interacted further with List [F(3,40)=3.47, p=.025] and Itemgroup [F(3,20)=3.43, p=.037]. Post hoc paired t-tests indicated that in the ambiguous condition, collapsing data across Lists, and across Itemgroups, there was no significant difference between animate and inanimate (except for List 1: t(10)=3.97, p=.003, 2-tail). In the non-ambiguous condition, however, response times were higher in the animate than the inanimate condition across Lists and Itemgroups, and separately for all Lists except List 2 (all p’s <.014, 2-tail), and for Itemgroups 3 and 4 (p’s <.016, 2-tail).

There was a main effect for Itemgroup [F(3,20)=5.16, p=.008], with response times in Itemgroup 3 differing significantly from the average in the other Itemgroups [t(20)=3.89, p=.001; mean for Itemgroup 3: 829 ms, IG 1: 708, IG 2: 692, IG 4: 735 ms].

Why should it be the case that at Position 4, at least for some items and most subjects, the animate-nonambiguous condition should lead to longer response times than the inanimate-nonambiguous condition? Also remember that the number of errors was highest for the a-namb condition. It might be useful to repeat some example sentences.
to illustrate the two conditions. The second noun ('concert') is at position 4:

animate-nonambiguous (a_namb):
After Dame Kiri concluded triumphantly the concert ...

inanimate-nonambiguous (i_namb):
After the duet concluded triumphantly the concert ....

It might be that subjects find it more difficult to integrate N2 into the sentence in the a_namb condition because the transitive subcat frame is more strongly activated than in the i_namb condition, and is not completely discarded even though syntactic evidence against it has been received (the ADV which closes off the subordinate clause after the verb). Note, however, that at Position 4, there was no difference between a_amb and i_amb: in the ambiguous condition, N2 is integrated into the ongoing sentence analysis at the same pace, regardless of whether the subject noun was animate or inanimate (It might of course be the case that the decision times at Position 4 reflect different processes/different syntactic analyses in the a_amb as compared to the i_amb condition, which happen to take the same time. It is likely, though, that in both a_amb and i_amb N2 (at Position 4) is analyzed as the direct object, since at Position 5, at the disambiguating word, there is a garden path effect in both conditions, albeit of greater size in a_amb than in i_amb).

It will be informative to look at decision times on the ADV itself: times should be longer in the a_namb than the i_namb condition if in the animate condition the transitive frame is strongly active (which implies a strong expectation of a direct object noun) and in the inanimate condition, the transitive frame is less strongly activated, and thus syntactic evidence incompatible with a transitive analysis is accommodated more easily.

Disambiguating Adverbial

The disambiguating adverbial are the words (sometimes only one word) in the non-ambiguous sentences which indicate that the verb phrase of the subordinate clause has
to be closed off early as an intransitive clause. I will first focus on the first adverbial (advl for short), i.e. the first word of the disambiguating adverbial phrase. For example, advl is 'in' in the sentence: 'After the guide had ended in the afternoon ...'.

If the transitive frame is more highly activated in the animate as compared to the inanimate condition, then decision times might be expected to be longer on advl in the animate as compared to the inanimate condition, because anything but a word that is part of an NP (which would then be attached as the direct object) might be unexpected in the animate condition.

In this analysis, only two conditions (a_namb and i_namb) of the original design are used. The between factors (List and Itemgroup) were not incorporated into the analysis since, using just the two conditions, individual Lists did not contain items from all itemgroups. Again, only times for correct decisions were included. There were on average 6 observations per cell in the by-subjects design, and 11 observations per cell in the by-items design. Of the total 528 responses, 13 (2.5%) were No-decisions; in a_namb, 10 (3.8%) of all 264 responses were No-decisions, and in i_namb, 3 (1.1%) (all three were made to different items). No subject made more than one false decision per condition. Item 17 received more errors than any other item (5 in the a_namb condition, and 1 in i_namb).

Taking the subject- and item means of the times for making correct Yes-decisions per condition as data, paired t-tests failed to indicate any significant differences between a_namb and i_namb at the position advl (means per condition based on subject-means: 858 ms for a_namb, and 793 ms for i_namb, standard deviations 378 ms and 360 ms, respectively).

Further paired t-tests were carried out on subject- and item means for a_namb versus i_namb after calculating the data in a number of different ways (taking the means of subject- and item-medians, logtransforming the decision times before taking means, adjusting the decision time for advl by the length of the word, and calculating the mean decision time per character for all words making up the adverbial phrase), all of which failed to reach significance.
I also checked whether there was any difference in the number of extremely long Yes-decisions on adv1 in the a_namb and i_namb conditions. I defined as extremely long all those Yes-decisions that were greater than 1321 ms (this is the value that lies 1.5 times the interquartile range above the 75th percentile of the distribution of decision times collapsed over conditions a_namb and i_namb). In the a_namb condition, 33 Yes-decisions were extremely long, and in the i_namb condition, 27. Wilcoxon Matched-pairs Signed-ranks Tests on the number of extremely long Yes-decisions for each subject, and each item, in conditions a_namb and i_namb failed to show any significant differences.

Thus, there are no grounds for rejecting the null hypothesis that a_namb and i_namb do not differ: subjects did not take more time to incorporate the disambiguating adverbial phrase into the VP in the animate as compared to the inanimate condition.

**Position 6**

The reason for analyzing decision times at position 6 in the sentence was to check if time-consuming reanalysis processes still continue in the ambiguous as compared to the non-ambiguous conditions.

As was done for the other positions, only response times for correct decisions were analyzed. Additionally, responses were removed if the subject had made a false decision in the same sentence at position 5 (55 responses in total) or position 4 (a further 11 responses). Removing these data was motivated by the following considerations. First of all, any further display of words after the subject decided the sentence was ungrammatical serves as feedback for the subject that they had made a mistake, and incorporating this feedback into the sentence analysis is likely to be time-consuming. Furthermore, we know from the above analyses that a Yes-decision on position 6 which follows a No-decision on position 5 takes a long time. However, these times do not so much show the continuation of a reanalysis process originally started at position 5, but rather a new process of reanalysis initiated after the subject
had terminated the initial reanalysis process with a No-decision on position 5. The earlier analyses also showed that Yes-decisions at position 5 in the non-ambiguous conditions tended to be higher when they were preceded by a No-decision on position 4 than when they were preceded by a Yes-decision on position 4.

The exclusion criteria described above led to removing 20% of the observations from condition a_amb; 4% from a_namb; 8% from i_amb and 1% from i_namb. Thus, the number of observations over which subject- and item-means were formed was smallest in condition a_amb (here, the number of observations over which subject- means were formed was 1 to 6, with a mean of 4.8; item-means were formed over a range of 4 to 11 observations, with a mean of 8.8 observations). Thus, the results of the analyses reported below should be interpreted with caution.

The data were log-transformed, and subject- and item-means for each sentence-type were computed and entered into analyses of variance as before.

Table 8.15. below gives the means for each sentence type per List (reported here are the antilogs), based on the subject-means.

<table>
<thead>
<tr>
<th>List</th>
<th>animate ambig.</th>
<th>animate a_amb</th>
<th>inanimate ambig.</th>
<th>inanimate n_amb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>708</td>
<td>728</td>
<td>725</td>
<td>707</td>
</tr>
<tr>
<td>2</td>
<td>668</td>
<td>643</td>
<td>658</td>
<td>676</td>
</tr>
<tr>
<td>3</td>
<td>672</td>
<td>790</td>
<td>705</td>
<td>781</td>
</tr>
<tr>
<td>4</td>
<td>881</td>
<td>927</td>
<td>666</td>
<td>768</td>
</tr>
<tr>
<td></td>
<td>728</td>
<td>765</td>
<td>688</td>
<td>731</td>
</tr>
</tbody>
</table>

In the by-subjects analysis, the only significant effect was for the interaction List x Animacy \[F(3,40)=4.26, p=.011\]. An inspection of the means in the table above indicates that for subjects of List 4, the average decision times in the two animate
conditions were much higher than in the two inanimate conditions. In the by-items
analysis, the main effect for Animacy was marginally significant \( F_2(1, 20) = 4.52, 
\ p = 0.046 \). The interaction Itemgroup \times Animacy \times Ambiguity was also significant
\( F_2(3, 20) = 3.86, \ p = 0.025 \). There were no other significant effects.

8.2.5. Discussion

Let me first repeat that a garden path effect was defined in several ways in the
analyses above: firstly, as the increased decision times on position 5 in the ambiguous
versus the non-ambiguous condition, eg. on the word 'was':

ambig.:
After the guide [downpour] had ended the walk was ...

non_ambig:
After the guide [downpour] had ended in the afternoon the walk was ...

and secondly, as extremely long response times. Additionally, the number of false
negative decisions also gives clues about a garden path effect.

In Experiment 6 there was clear evidence from decision times, extremely long
responses and from the error data for a garden path effect both in the animate and the
inanimate condition. However, decision times at position 5 were longer (and there
were more extremely long responses and more errors) in the animate-ambiguous than
the inanimate-ambiguous condition, indicating that the garden path effect was more
severe for animate sentences. This interaction of Animacy with Ambiguity was also
significant once extremely long response times had been excluded. On position 6, there
was no evidence for any effect of Ambiguity any more, indicating that subjects had
resolved the garden path and found an alternative syntactic analysis. This result also
suggests that decision times reflected processing complexity at the word currently
presented, and that subjects did not tend to press the Yes-key to see the next word
while they were still focusing on the previous word, as some subjects seemed to have
done in the previous Experiment 5.
Let me summarize the other findings at the various theoretically interesting positions in the sentence.

At the first verb on position 2 (e.g. 'After the guide [downpour] had ended ...'), two different syntactic analyses, a transitive and an intransitive, could be chosen, given that the verb has two subcat and two thematic frames. Is the choice of syntactic analysis affected by the animacy of the subject noun? My data from Experiment 6 clearly show that this is not the case. Firstly, there was no difference between the animate and inanimate conditions in the decision times (and error rates) on position 4 for the postverbal NP in the ambiguous condition. If the postverbal NP in the inanimate-ambiguous condition had been analyzed as the beginning of a new clause rather than as part of the VP, one might have expected increased processing costs.

Secondly, there were no decision time differences between the animate and inanimate conditions on the adverbial following the verb in the nonambiguous intransitive condition (e.g. 'After the guide [downpour] had ended in the afternoon ...'). Thus, the bias after the verb for the transitive and the intransitive analysis was the same in the animate and the inanimate conditions.

The garden path findings at position 5 indicate that when the postverbal NP is being processed, subjects show a preference for the transitive analysis and attach this noun as a direct object despite the fact that the intransitive subcat frame of the verb also allows closing off the VP early, immediately after the verb, and analyzing the postverbal noun as the subject of a new clause. This preference is not blocked by the inanimacy of the first subject noun.

There was one slightly surprising finding regarding decision times at position 4 in the non-ambiguous sentences, e.g. on 'walk':

After the guide [downpour] had ended in the afternoon the walk ...

Decision times on 'walk' were longer in the animate than the inanimate condition, which is unexpected since in both conditions the syntactic evidence, i.e. the adverbial,
clearly forces the intransitive analysis, leaving the attachment of 'walk' as the subject of the main clause as the only grammatical analysis. One could speculate that in the animate condition the transitive analysis is more strongly activated than in the inanimate condition, even in the face of syntactic evidence (the postverbal adverbial) to the contrary. This might lead some subjects to assume initially that there is a word order mistake (eg. assuming that 'After the guide had ended in the afternoon the walk ...' was an incorrect version of the sentence 'After the guide had ended the walk in the afternoon ...').

Alternatively, this finding might indicate that subjects more strongly expected a particular NP after the adverbial in the animate than the inanimate condition. There seems to be an expectation that the main clause should continue to focus on the 'guide' (eg. 'After the guide had ended the walk he ....'), whereas the corresponding expectation for the inanimate sentence seems to be weaker, and a range of continuations seems to be acceptable.

Let us focus again on the main findings of this experiment: the garden path effect (at position 5) was bigger in the animate than the inanimate ambiguous conditions in spite of the result that the animate and inanimate conditions did not differ in decision times on the postverbal noun itself. Given that decision times were the same in the animate and inanimate conditions on the postverbal noun, the garden path effects should have been equal, but they clearly were not. This data pattern implicates a two-stage theory. During the first stage, when the postverbal noun is processed, there is initially an equal tendency to attach this as a direct object in the animate and inanimate conditions, as was said before. Then there must be a second stage, ending with the response to the disambiguating word at position 5 ('was'), which forces a re-analysis, during which several processes might be operative: the initially preferred transitive analysis might be strengthened by semantic biases, i.e. by the animacy of the first subject noun: this would cause the transitive analysis to be more difficult to abandon at position 5 in the animate than the inanimate condition, or it might cause the intransitive analysis to be more difficult to construct; alternatively, after attaching the postverbal NP as direct object, the intransitive analysis in the inanimate condition might not deactivate as
much as in the animate condition, leaving it more available to help in re-analysis when the initial transitive analysis has to be given up.

How can one choose between these alternatives?

In the next section I will first discuss a range of existing models of sentence processing, and I will show how each fails to account for the present set of data. I will then present a new model which can accommodate my garden path findings.

Models

I will discuss each model only briefly, presenting the processing predictions it would make at each point of my test sentences, and evaluating these predictions against the set of findings from Experiment 6. As a reference point I repeat here an example sentence from the experiment:

animate ambiguous:
After the guide had ended the walk was ....

inanimate ambiguous:
After the downpour had ended the walk was ...

1. Frazier’s single analysis model

Frazier proposes a serial parser which computes only a single analysis and, when several analyses are possible, makes an immediate commitment to a single analysis by following the minimal attachment (MA) and late closure (LC) strategies. These strategies are justified in terms of reducing STM load and processing complexity. The possibility that the parser might process alternative syntactic analyses in parallel is completely discarded.

The parser accepts as input from the lexicon the syntactic category of words (but
ignores the subcat information of verbs) and produces as output a phrasal tree to represent the syntactic structure of a sentence. In its first-pass analysis, it does not use subcat information, or any semantic or pragmatic information. This model was used to account eg. for the classical gp-effects for the Bever-type sentences, eg.

The horse raced past the barn fell.

According to this model, the reduced relative clause analysis is more complex (and thus is not initially computed) because a more complex NP has to be projected, i.e. one that has as its daughters the simpler NP 'the horse', plus the reduced relative sentence 'raced past the barn'. In the main clause analysis only a VP node has to be created.

With regard to my sentences, Frazier's model, in particular the principle of Late closure, makes the right prediction in so far as it predicts that N2 is attached as a direct object, and that V2 presents syntactic evidence that this initial analysis is incorrect. It fails to account for the fact that the garden path effect is bigger in the animate than the inanimate condition, since the parser is assumed to use only syntactic information. Syntactically, however, sentences in the animate and the inanimate condition are of course identical, and therefore should not differ in the size of the garden path effect which they produce.

Frazier makes two suggestions regarding processes of re-analysis, but they, too, fail to account for the differential effects at position 5 in the ambiguous sentences of the present experiment. One of the suggestions (see discussion in Section 7.5.2.) applies to re-analysis in sentences where there is no syntactic error signal, eg.

He saw the girl with the red hair.

There is no syntactic error signal forcing a change to the initial attachment of the PP to the verb, in line with the minimal attachment strategy. In this case, as Rayner et al. (1983) propose, the error signal results from conflicting outputs from the parser and the thematic subsystem (note that there is no such conflict in the sentence 'He saw the
The output of the thematic subsystem is the thematic frame <Experiencer,Theme>, plus the associated subcat frame (NP, NP) (note: the first NP is the subject NP, which is not strictly subcategorised by the verb; it is listed here in the subcat frame merely to match the Experiencer role in the thematic frame), whereas the phrase structure which the parser outputs assigns two complements to the verb, an NP and a PP. This conflict can act as an error signal and initiate re-analysis (Frazier, 1987a, p. 576). This re-analysis model obviously has no predictions to make regarding the re-analysis in the ambiguous sentences of the present experiment since these sentences clearly contain a syntactic error signal (the word at position 5) which might trigger re-analysis.

The other proposal for re-analysis is made by Frazier & Rayner (1982; also discussed in Frazier, 1987a). It also fails to account for the differential garden path effect in my experiment. According to this strategy, in certain cases re-analysis is predicted to be quite effortless, but the predictions make no appeal to semantic factors. Take the sentence:

The city argued the mayor's position was incorrect.

The ultimately correct analysis of this sentence violates MA: during the initial minimal analysis, 'the mayor's position' is attached as a direct object; during re-analysis, this NP has to be attached as the subject of the reduced complement clause. Re-analysis of this sentence is cost-free because, so the authors argue, the ambiguous postverbal NP is "available" to become the subject of the following clause: the parser determines that 'was' misses its subject, and also knows (in some unspecified way) that the preceding NP is a legitimate subject for this VP (Frazier & Rayner, 1982, p. 183). In contrast, 'the horse' in 'The horse raced past the barn fell' is deemed not to be available (and thus re-analysis is predicted to be costly) not only because it is further removed from the site of the error signal but also because the nature of the original error is more difficult to identify.

According to this re-analysis principle, there should be no costly garden path for the sentences of my experiment because 'the walk' is "available" to be re-attached as
subject noun, and the location and the nature of the error is the same for the two types of sentences. Furthermore, this model attempts to explain why in some cases re-analysis is cost-free; what needs to be explained in the context of the present experiment, however, is the higher cost involved for some sentences; but cost there is for all ambiguous sentences.

Frazier and Rayner (1982) also speculate that the costs of revising a syntactic error might be affected by semantic interpretation. Semantic interpretation is assumed to lag behind syntactic analysis; however, if an ambiguous phrase is long, semantic interpretation might have had time to be completed before the syntactic error signal is received. A further assumption is that more costs are involved in re-analyzing a phrase that has been semantically interpreted than one that has not yet been semantically interpreted. This proposal, however, also fails to account for the findings of the present experiment. Here the ambiguous phrases were of the same length in the two conditions and thus there should not have been any reason why re-analysis costs varied. One might extend the authors' suggestions by assuming that the semantic interpretation of the direct object attachment might somehow be processed faster in the animate than the inanimate condition. However, given the authors' assumption that semantic interpretation is not immediate, it is unlikely that semantic interpretation of an NP and its attachment could be complete (even under the most beneficial circumstances, whatever they might be) as soon as the next word was processed.

2. Ferreira & Clifton (1986)

Ferreira & Clifton (1986), adopting Frazier’s model, stress that the parser does not use nonsyntactic information (eg. plausibility of a sentence) in its initial analysis, but only for re-analysis, which is assumed to start only when the initial analysis is complete (eg. the PP in 'I saw the girl with red hair' has been attached to the VP). They also explicitly speculate that thematic information is not only used in re-analyzing syntactically correct sentences (as in 'I saw the girl with red hair.') but that it is also used by the parser when there was a syntactic error signal:
"It does not seem plausible that the parser would use high level information only to reanalyze fully ambiguous sentences and ignore the information when it has available a syntactic error signal. Instead, the parser uses the error signal as an additional clue to aid in the construction of a different analysis." (Ferreira & Clifton, 1986, p. 366).

No details are given about exactly how the re-analysis process might be affected by nonsyntactic information, nor do the authors comment on their finding in experiment 1 that second pass reading times for the disambiguating phrase were in fact longer in the inanimate than the animate condition, whereas they should have been shorter in the inanimate condition if plausibility factors somehow aided re-analysis.

If the aid-to-reanalysis proposal were true, one might speculate that for my sentences the thematic subsystem preferred the intransitive frame to the transitive frame in the inanimate condition, and that this information somehow aided re-analysis once the syntactic error signal indicated that a re-analysis has to take place. The problem is that at position 4 (the postverbal noun) response times were not longer in the inanimate ambiguous sentences than the animate ambiguous sentences, even though, according to the above interpretation, the attachment of the noun as direct object is implausible in the inanimate condition (note that in my Experiment 4 response times were longer when the ambiguous noun semantically violated selection restrictions of the verb). Thus, since there is no evidence that the implausibility of the ambiguous noun as direct object was noted when the noun occurred, it is unlikely that it could aid in re-analysis which was initiated only one word later on.

Ferreira and Henderson (1991) present a more detailed model of the interaction between parser and thematic subsystem. They, too, propose that the thematic subsystem assigns thematic roles to the head of the phrases output by the parser (this assignment is proposed to occur as soon as possible, i.e. even before the verb is encountered, by using default assignments). As in Frazier’s model, the parser is assumed to use only syntactic category information, and phrase structure rules, plus the syntactic principles of MA and LC. The thematic subsystem, on the other hand, in addition to receiving phrases from the parser, also receives, from the lexicon, all thematic and subcat frames of a verb. Thus, a verb’s preference for a particular frame
(eg. mainly intransitive verbs) cannot affect the initial parsing decision. The parser, according to the principle of Late Closure, will always attach a postverbal noun as a direct object (as in 'While the girls race the boys ...'); this NP is then input to the thematic subsystem which will assign to it the Theme role, selecting the transitive frame. Reanalysis, which is assumed to be only initiated when a syntactic error signal is received, will depend on the availability of the alternative frame. The alternative frame is assumed to decay as soon as the other frame is initially chosen. This decay is assumed to be greater for longer than for shorter ambiguous NPs because the thematic subsystem is committed to the alternative frame for longer. The amount of decay (and ease of reanalysis) is also assumed to depend on the initial activation of the alternative frame. For example if a verb occurs more frequently in an intransitive rather than a transitive frame, the intransitive frame is assumed to be highly activated, and (if the alternative transitive frame is selected) it will have decayed to a lesser level when reanalysis begins.

It might be possible to extend this model to account for my results by assuming additionally that the thematic subsystem uses semantic properties of the first noun in activating the thematic frames. This would lead to a greater level of activation for the intransitive frame in the inanimate than the animate condition. Consequently, it would decay to a lesser level when the alternative transitive frame was chosen, and would thus be more easily available when re-analysis had to occur. However, given this account, response times on the postverbal adverbial in the non-ambiguous condition should have been faster in the inanimate as compared to the animate condition, which was, however, not the case.

3. Pritchett’s single analysis model

With regard to my sentences, Pritchett’s model (1992) correctly predicts a garden path, but for quite different reasons than those given by the single-analysis models discussed so far. He assumes that the parser has access to the verb’s subcat and thematic frames, and that it attempts to assign all thematic roles of a verb as early as possible. He
assumes that a parser will have to re-assign thematic roles very frequently, but that this reanalysis is in most cases cost-free (i.e. it does not lead to a conscious garden path at the end of a sentence; note that conscious garden paths are the only type of evidence he considers relevant for evaluating parsing models). Reanalysis will incur costs (in fact it will require conscious re-processing) if the ambiguous phrase has to be re-assigned to an argument position which is not governed or dominated by the phrase's original position.

With respect to my sentences, the model correctly predicts that the ambiguous noun is first assigned as an argument of the first verb (the Theme) in order to discharge as many thematic roles as possible. The garden path effect is correctly predicted because during re-analysis the ambiguous noun has to be re-assigned as the external (subject) argument of the second verb, i.e. it has to be attached in an argument position which is not governed by the initial argument position. Since his model only refers to phrases and their positions in a phrasal tree, it fails, however, to account for the finding of a greater garden path effect in the animate than the inanimate condition.

4. Multiple-analyses models

When faced with ambiguous input, models of this type propose that the parser computes all alternatives in parallel, on the basis of grammatical rules. There are various accounts of how the final selection of one analysis is made, eg. on the basis of plausibility given the previous context (Altmann & Steedman, 1988; Steedman & Altmann, 1989), or the plausibility within a sentence (Jackendoff, 1992), or the greater frequency of one analysis (eg. Jackendoff, 1987b).

With respect to my sentences, one could assume that the parser computes both the early closure (intransitive) analysis, and the late closure (transitive) analysis in parallel; the garden path effect at position 5 indicates that in both conditions, the transitive frame was preferred. The problem is that both Jackendoff's and Altmann & Steedman's model predict that one analysis is selected and the other is rejected.
(according to various criteria); my findings, however, indicate that the selection (and rejection) is not all or none: in the inanimate condition, the transitive analysis must have been preferred (otherwise there would not have been a garden path effect), but the intransitive analysis seems to have deactivated to a lesser extent than in the animate condition, because the garden path effect was smaller in the inanimate than the animate condition.

The two main features of Marcus' (1980) multiple-analyses model which are of concern here are firstly the parser's ability to compute several alternative syntactic analyses in parallel, and secondly its use of a look-ahead buffer with a size of three constituents. With respect to my sentences (and other garden path constructions) this model fails to correctly predict the occurrence of garden paths. After having attached the first noun in my sentences in the only grammatical way, the buffer should contain the elements [had ended], [the walk] and [was]. On the basis of this information (esp. given that 'was' is in the buffer) the parser should always initially choose the correct analysis, and no garden path should occur.

5. Ford et al.'s (1982) lexical preference model

As was discussed in Chapter 7.2., Ford et al.'s model focusses on syntactic reanalysis processes which are cost-free and only briefly discuss sentences in which reanalysis leads to garden path effects. The central claim in their model is that lexical entries of verbs contain several lexical forms which correspond to alternative syntactic analyses, and that the various lexical forms have different strengths or saliences. To account for the smaller garden path effects in the inanimate than the animate sentences of Experiment 6, one would have to assume that the strengths of lexical forms is not only determined by the lexical entry of the verb, but is also affected by the sentence context. One would have to assume that the intransitive form of the verb was preferred more often in the inanimate than in the animate condition. The observed (smaller) garden path effect in the inanimate condition could then be interpreted as being due to only a few sentences (in which the transitive form was preferred). However, my data
showed that there was a garden path effect for 20 out of all 24 sentences in the inanimate condition, implying that for the majority of sentences in the inanimate condition, the intransitive form was not strongly preferred. One might, however, argue that the inanimacy of the first subject noun did selectively activate the intransitive form, but that the occurrence of an NP in direct object position led to abandoning that form and choosing the alternative transitive form. But then one would have to predict a garden path effect on the NP in direct object position in the inanimate condition (because it is here that the alternative form was chosen), which is clearly not borne out by the data.

6. Strong thematic models

In this class of models, semantic features of the first subject noun (at least animacy, inanimacy) are retrieved and thematic roles assigned to the noun immediately (i.e. before the verb is encountered) using some default strategies such as 'assign subject the agent role' (Ferreira & Henderson, 1991), or 'assign an animate subject the agent role', and 'assign an inanimate subject the theme role' (Stowe, 1989). According to the strong thematic model by Stowe, this default assignment of thematic roles to the first subject noun determines which subcat frame of the verb is activated (in fact it is not entirely clear whether she assumes that both subcat frames are initially activated, with the one compatible with the default theta assignment to the subject receiving greater activation, or whether the default assignment actually blocks the activation of the alternative theta frame). When the ambiguous postverbal noun is input, there is no blind syntactic strategy (e.g. late closure); instead, the analysis by the parser is determined by which subcat frame is active: the noun is attached as direct object if the <Agent, Theme> frame is active, and it is attached as a subject of the following main clause if the <Theme> frame is active. Carlson & Tanenhaus (1988) propose a very similar model, in which the parser has to check the output of the thematic subsystem each time before it attaches a possible argument of the verb: no attachment is made if there is no thematic role available for it (e.g. because the only theta role has already been assigned to the subject).
Obviously, this model fails to predict the garden path in the inanimate condition which was found in Experiment 6, whereas it is in line with Stowe's (1989, experiment 1) results. In this study, which also used the self-paced reading paradigm with continuous grammaticality judgements, Stowe did find a garden path only in the animate but not the inanimate condition. The main difference to my study is that the postverbal (syntactically ambiguous) noun was a plausible object in her animate condition, but it was not at all plausible in her inanimate condition (e.g. 'Even after the music ended the song ....'; 'When the yacht was about to sail the boat ....'; 'After the report returned the report ...'(sic!); see Chapter 7.5. for a detailed discussion of Stowe's experiment). Thus, her sentences do not allow a fair test of the role of in/animacy in determining syntactic processing. My sentences in the inanimate condition indicate that if the transitive action is plausible, inanimacy of the subject noun does not block the transitive analysis (and the subsequent garden path).

7. My alternative parallel model

Having discussed how various models fail to account for my data, I will now sketch a model which can accommodate the findings. In this model, the parser has access not only to syntactic category information in the lexicon, but also to the subcat frames of verbs. I will describe the various subsystems assumed in the model and specify which information is available to them at the crucial words, and how parsing decisions are made.

When the parser processes a verb which allows two subcat frames, it computes two analyses in parallel, i.e. in my sentences, in one analysis the verb phrase is closed early, after the verb, and in the other analysis the verb phrase is not closed off. Thus in this model, in contrast to Stowe's and Frazier's, both analyses are initially computed regardless of the semantic properties of the subject noun, and the syntactic preferences of the parser. Even if the thematic subsystem might prefer one particular analysis, my data indicate that further input following the verb, whether it syntactically supports one or the other of the two analyses, is accepted to the same extent in the animate and the
inanimate condition (there were no response time differences between the two conditions on the adverbial in the non-ambiguous early closure sentences, nor on the postverbal ambiguous noun in the late closure sentences).

Once both analyses have been computed and are passed on to the temporary linguistic store, their further level of activation depends both on the evaluation by the thematic subsystem and on further syntactic decisions.

When the next word (the ambiguous noun) is processed by the parser, the late closure analysis receives compatible syntactic evidence (which is also plausible, unlike in Stowe’s sentences), and is thus activated further. The early closure analysis does not receive any syntactic evidence in its support, and its activation level decreases. Unlike Steedman & Altmann and Jackendoff (see above), I do not assume that the selection between the alternative analyses is absolute; in particular, I assume that deactivation of the intransitive analysis, which begins on position 4 due to the lack of syntactic support, is affected by the output of the thematic subsystem. I propose that the thematic subsystem receives from the lexical system information about the thematic frames of a verb (eg. for 'stop': <Agent, Theme> and <Theme>). Its task is to evaluate the plausibility of each frame. To accomplish this task, it checks the information in the temporary linguistic store: it accesses all phrase structures which involve the verb whose thematic frames it is currently evaluating. The phrase-structure trees in the temporary store are assumed to be enriched by semantic information output from the semantic subsystem which specifies the semantic features of the nouns, and the semantic fit between verbs and the nouns in its argument positions. The thematic subsystem computes the plausibility of the alternative thematic frames by a) checking the semantic features of the nouns in the argument positions of the verb, and b) by passing on information to the central system to check the plausibility of the frames against the knowledge of likely actions and things in the world (eg. to determine how plausible it is that a motor can turn a wheel). In my sentences, the transitive actions expressed in the sentences were designed to be equally plausible in the animate and the inanimate condition. Thus, real-world plausibility did not differentially affect the evaluation of the alternative thematic frames by the thematic subsystem. However, the
first criterion, i.e. checking the semantic features of the nouns, leads to a preference of the <Agent, Theme> frame in the animate condition due to a preference to assign an animate subject noun the Agent role, and an inanimate subject the Theme role. These preferences are fed back to the temporary linguistic store and the level of activation of the thematically preferred analysis is raised. Thus, when the intransitive frames begin to deactivate on position 4 because of lack of syntactic support, their level of activation will remain higher in the inanimate condition than in the animate condition, because the evaluation by the thematic subsystem has increased their activation. Syntactic support is assumed to have a greater effect on the activation levels of (alternative) analyses than thematic support. Thus, the attachment of the postverbal noun (position 4) as direct object increases the activation level of the direct object late-closure analysis in the temporary store, and decreases the activation of the alternative intransitive analysis to the extent that, on the following word (position 5), the alternative intransitive frame is not sufficiently activated in either the animate or the inanimate condition, creating a garden path effect for both.

The finding that the garden path effect is smaller in the inanimate than the animate condition can be accounted for by the level of activation which the intransitive frames reach after they begin to deactivate on position 4: since the intransitive frames receive support from the thematic subsystem in the inanimate but not the animate condition, they will maintain a higher level of activation in the inanimate condition. When on position 5 syntactic evidence forces re-analysis, the parser checks the temporary store for any alternative analysis that it had originally built up. The alternative intransitive analysis is more highly activated in the inanimate than the animate condition, and thus the parser will be able to adopt this analysis more quickly and compute the new attachment of the postverbal noun as the subject of the next clause.

From this account it follows that deactivation should be affected by the length of the ambiguous postverbal noun phrase: the longer it is, the longer the parser will be committed to the late closure transitive analysis, and the more the alternative analysis should deactivate. (Ferreira & Henderson (1991) found evidence of such length effects in end-of-sentence grammaticality judgements using sentences with optionally
transitive verbs). It would be interesting to establish whether the alternative intransitive analysis always deactivates less in the inanimate than the animate condition (because of its support from the thematic subsystem). It might be the case that, given a sufficiently long period in which the initially chosen transitive analysis is maintained, the level of activation of any alternative analysis, thematically supported or not, decreases below a lower bound, so that it is not available any more to help in re-analysis when the syntactic error signal forces the transitive analysis to be given up.

At present, this model does not yet specify whether the output of the thematic subsystem can only increase the activation of an analysis in the temporary store, or whether it can also decrease the activation. In the present set of sentences, it was syntactic information that prompted deactivation: first, the postverbal NP (in the ambiguous sentences) supported the transitive but not the intransitive analysis, creating higher activation for the former, and deactivation for the latter; and then the verb at position 5 (in the ambiguous sentences) ruled out the transitive analysis, forcing its deactivation.

In Experiment 6 the influence of semantic information on the parsing process was tested indirectly by demonstrating that it is an important factor which determined the preference for a particular thematic frame by the thematic subsystem. This preference affected the level of activation of the corresponding analysis in the temporary linguistic store.

The following Experiment 7 was designed to test the influence of semantic information on the parsing process by investigating whether evidence from the semantic subsystem against the direct object analysis would prevent a garden path, or affect the size of a garden path effect.
Chapter 8.3.: Experiment 7: Semantic properties of the direct object noun and the garden path effect: self-paced reading with continuous grammaticality judgements

8.3.1. Introduction

Experiment 7 focuses on the semantic features of the postverbal noun in bringing about a change in the initial syntactic analysis. Take for example the sentence:

a) Wherever Alice walks her funny old dog attracts a lot of attention.
b) Wherever Alice walks her funny old hat attracts a lot of attention.

Experiments 5 and 6 had shown garden path effects when the postverbal noun was a plausible and semantically correct direct object (as in a) above). Experiment 6 also showed that the garden path effect was smaller depending on semantic properties of the first subject noun. This finding was interpreted as evidence for the influence of semantic information on the evaluation of alternative thematic frames by the thematic subsystem. It was suggested that the thematic subsystem evaluated the <Theme>-frame, corresponding to the intransitive analysis, as highly plausible when the subject noun was inanimate, and less plausible when it was animate, thus increasing the level of activation of the <Theme> frame and its associated syntactic frame in the temporary linguistic store. When syntactic evidence forced the parser to give up the initially preferred transitive analysis, reanalysis was easier in the inanimate than the animate sentences, because in the former the intransitive analysis had a higher level of activation in the temporary linguistic store.

It is still an open question whether and when the semantic features of the syntactically ambiguous postverbal noun affect syntactic analysis. Is the initial preference for a transitive analysis maintained even if the potential direct object noun violates the semantic selection restrictions of the verb?

Stowe (1989, experiment 2) tried to establish if a second noun which is implausible as a direct object would block the preferred transitive analysis, thereby avoiding the
garden path effect later in the sentence. Her experiment was inconclusive, as was discussed in detail in Chapter 7.5. above.

My experiments on violations of semantic selection restrictions (Chapter 6) indicated that semantic misfusions of verb and subcategorized direct object noun are registered immediately during reading, and that they disrupt sentence processing. The question under investigation here is whether semantic misfusion (such as in b) above) prompts deactivation of the initially chosen transitive syntactic analysis, and forces reanalysis (or, alternatively, whether semantic misfusion blocks the initial direct-object attachment of the postverbal noun). If this were the case, no garden path should occur on 'attracts' in example b) above because the reader has already switched to the intransitive analysis which turns out to be the only correct one.

8.3.2. Hypotheses

A garden path effect is predicted for the ambiguous semantically normal sentences, i.e. significantly longer decision times, and more false negative decisions at position 5 and later words in the ambiguous as opposed to the non-ambiguous condition.

For the sem-viol sentences longer decision times are predicted on the postverbal noun itself (position 4) in the ambiguous as compared to the non-ambiguous condition (and also as compared to the postverbal noun in the normal ambiguous condition) because in this condition, the preference to attach the postverbal noun as direct object clashes with the semantic incompatibility of the noun as direct object. Furthermore, if semantic violation can influence the decisions of the parser, no garden path effect is expected in the sem-viol condition: decision times and number of false negatives on position 5 are predicted not to differ between the ambiguous and non-ambiguous conditions.

8.3.3. Method

The task, procedure and dependent variables were the same as in Experiment 6, i.e. a
self-paced cumulative moving window paradigm with grammaticality judgements was used.

Design

Two factors were varied in this experiment. One was Ambiguity (with the levels ambiguous and non-ambiguous), and the other was Semantic Status, with the levels normal and semantically violated (sem-viol). The two factors were crossed. They were repeated measures factors in both the by-items and the by-subjects designs. Again, List and Itemgroup (resulting from constructing 4 stimulus lists) were between-group factors in the by-subjects and the by-items design, respectively.

The following table presents sample sentences to illustrate the four conditions of this experiment:
Table 8.16. Experiment 7: example stimulus set

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
</table>

Table 8.16. Experiment 7: example stimulus set

normal, amb.:

Wherever Alice walks her funny old dog attracts a lot of attention.

sem-viol, amb:

Wherever Alice walks her funny old hat attracts a lot of attention.

normal, non-amb.:

Wherever Alice walks in the city her funny old dog attracts a lot of attention.

sem-viol, non-amb:

Wherever Alice walks in the city her funny old hat attracts a lot of attention.

Ambiguous sentences were again of the structure: Conj Det NP1 Vb1 NP2 Aux (or Vb2) plus various other phrases. Non-ambiguous sentences differed from the ambiguous ones by an additional adverbial phrase inserted between Vb1 and NP2 which closed the subordinate clause. The positions used to identify particular words in the sentence are indicated in the table above.

In the sem-viol condition the head of the postverbal noun phrase (N2) violated the semantic selection restrictions of the verb in its transitive reading, whereas in the normal condition it fitted these semantic restrictions, i.e. it was a semantically possible direct object. Note that in this experiment, there is no Animacy factor: the first noun is always animate. It was the animate condition which had shown the clearest preference for an initial transitive analysis, with the biggest ensuing garden path effect, in the previous experiments. It is in this condition in particular where a semantically violated postverbal noun is likely to be of greatest impact.
Stimuli

A total of 48 sentences were used, 24 of which were fillers. 50% of the sentences were made ungrammatical by introducing errors of the same kind as was done in Experiment 6. In 10 of the ungrammatical sentences, the error occurred within the first 9 words of the sentence, and in 14 sentences, the error occurred more than 9 words into the sentence.

Experimental sentences

24 sentence quartets were created, with each member representing one condition, as was illustrated in table 8.16. above. 16 of the causative/ergative verbs used in the previous ambiguity experiments were used again, and 8 new verbs were included (to ride, to walk, to drive, to break up, to fly, to develop, to return, to finish). Sentences making up a quartet were created according to several criteria: the postverbal noun in the normal condition was a semantically possible direct object; the postverbal noun in the sem-viol condition lacked a semantic feature which the verb required its direct object to have; the nouns in the N2 position were matched for length (number of letters) and frequency; both normal and sem-viol sentences were equally natural in the non-ambiguous conditions (according to the judgements of three native speakers).

Four sentence quartets (sets 1, 10, 15 and 20) were made ungrammatical late in the sentence, i.e. at least 6 words after the second verb. See Appendix J for a listing of all stimuli.

Four lists of stimuli were created, each containing only one sentence from each sentence quartet. Subjects were assigned to a list at random to ensure that a subject would only see one member of each sentence quartet. A Latin Square was used to assign sentences to the four experimental conditions, such that in each list there was an equal number of sentences (i.e. 6) in each condition. In each list, one sentence per
condition contained an ungrammaticality late in the sentence.

**Filler sentences**

The 24 filler sentences were identical in the four lists. They were unrelated to the purpose of this experiment, and were designed to reflect a variety of sentence structures. Furthermore, there were ten practice sentences which were identical in the four lists. They were made up to reflect the range of sentence types used in the experiment. Half of the practice sentences were ungrammatical.

**Subjects**

There were 40 subjects in this experiment, all of which were students at the Department of Psychology, University College London.

**8.3.4. Results**

**Overview of analyses**

Responses at positions 4, 5 and 6 were analyzed with respect to the time it takes to make a correct decision, and the distribution of false negative decisions and extremely long responses for correct decisions. Furthermore, times for correct decisions at position 5 were also analyzed after extremely long responses had been removed.

**Position 4**

There were only 9 false No-decisions altogether. They were made by a range of different subjects on several different items. No-decisions were not included in the
response times analyses reported below, and were not analyzed further.

Subject- and item-means of the logtransformed (correct) decision times were entered into analyses of variance. Table 8.17. below gives the cell means (reported here are the antilogs) based on subject-means.

Table 8.17. Experiment 7: cell means (in ms) of correct decision times for position 4

<table>
<thead>
<tr>
<th></th>
<th>ambiguous</th>
<th></th>
<th>non-ambiguous</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>normal</td>
<td>sem-viol.</td>
<td>normal</td>
<td>sem-viol.</td>
</tr>
<tr>
<td>List 1</td>
<td>648</td>
<td>985</td>
<td>650</td>
<td>745</td>
</tr>
<tr>
<td>List 2</td>
<td>714</td>
<td>729</td>
<td>633</td>
<td>685</td>
</tr>
<tr>
<td>List 3</td>
<td>799</td>
<td>950</td>
<td>778</td>
<td>725</td>
</tr>
<tr>
<td>List 4</td>
<td>588</td>
<td>895</td>
<td>668</td>
<td>726</td>
</tr>
<tr>
<td></td>
<td>683</td>
<td>884</td>
<td>680</td>
<td>720</td>
</tr>
</tbody>
</table>
Figure 8.4 presents the means collapsed over list.

There were no significant main effects for either List or Itemgroup. There were significant main effects for both Ambiguity and Semantic Status [Ambiguity: $F(1,36)=10.40, p=.003$; $F(1,20)=11.40, p=.003$; $\text{MinF'}(1,53)=5.44, p<.025$; Semantic Status: $F(1,36)=29.90, p<.0005$; $F(1,20)=17.67, p<.0005$; $\text{MinF'}(1,42)=11.11, p<.01$]. The interaction List by Semantic Status was significant [$F(3,36)=4.62, p=.008$], as was the interaction Itemgroup by Semantic Status [$F(3,20)=3.26, p=.043$], and Itemgroup by Ambiguity [$F(3,20)=4.69, p=.012$]. The interaction Ambiguity by Semantic Status was also significant [$F(1,36)=5.99, p=.019$; $F(1,20)=10.78, p=.004$; but $\text{MinF'}(1,56)=3.85, p<.10$].

Post-hoc paired t-tests indicated that in the ambiguous condition, average response
times were longer in the semantically violated as compared to the normal sentences (significant across Lists and Itemgroups, p’s < .00025, 1-tail; and also separately for List 1 and List 4 (p’s <.0035, 1-tail) and Itemgroups 1, 2 and 4 (all p’s <.052, 1-tail). In the non-ambiguous conditions, the differences between semantically violated and normal sentences were not significant across Lists or Itemgroups. Furthermore, the semantically violated ambiguous sentences yielded significantly longer decision times than the semantically violated nonambiguous sentences (across Lists and Itemgroups: all p’s <.002, 1-tail; also separately for List 1 and 3 (all p’s < .052, 1-tail), and for Itemgroups 2, 3 and 4 (all p’s <.04, 1-tail)). As expected, response times did not differ for normal sentences in the ambiguous versus the nonambiguous condition. Response times were higher in the violated-ambiguous than the normal-ambiguous condition for 28 out of 40 subjects, and for 20 out of 24 items; and response times were higher in the semantically-violated ambiguous than in the semantically-violated nonambiguous condition for 27 of the 40 subjects, and 20 of the 24 items.

**Position 5**

The analysis of decision times included only those responses that were correct at position 5 and for which a subject had not made an error in the same sentence at position 4. As a result, 27 responses (i.e. 2.8% of the data) had to be excluded.

The distribution of errors (false negatives) per condition at position 5 is given in Table 8.18. below. Note that there were 240 observations per cell (40 subjects saw 6 items per condition).

**Table 8.18.** Experiment 7: Number and (in brackets) percentages of No-decisions per condition at position 5

<table>
<thead>
<tr>
<th>Normal Amb</th>
<th>Semi-Viol Amb</th>
<th>Normal Nam</th>
<th>Semi-Viol Nam</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 (6.7)</td>
<td>4 (1.7)</td>
<td>1 (0.4)</td>
<td>1 (0.4)</td>
</tr>
</tbody>
</table>

In the semantically-violated ambiguous condition, two of the four errors occurred in
cases where a subject had already made a false decision for the same sentence one word earlier, at position 4. Remember that at position 4, there were six errors in the semantically-violated ambiguous condition. This means that in 4 cases, subjects changed their minds when reading a sentence: they pressed the No-key at position 4 but the Yes-key at position 5 of the same sentence. All of the 16 errors listed in the table above for the normal ambiguous condition are the first error by a particular subject in a particular sentence. Each of the errors in the last two conditions were made in a sentence where the subject had already pressed the No-key at position 4.

For each subject and each item the means of the logtransformed times for correct decisions were computed for each condition. The (antilogs of the) cell means, based on the subject means, are presented in Table 8.19. below.

**Table 8.19** Experiment 7: cell means (in ms) of correct decision times for position 5

<table>
<thead>
<tr>
<th></th>
<th>ambiguous</th>
<th></th>
<th>non-ambiguous</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>normal</td>
<td>sem-viol.</td>
<td>normal</td>
<td>sem-viol.</td>
</tr>
<tr>
<td>List 1</td>
<td>1009</td>
<td>1130</td>
<td>645</td>
<td>664</td>
</tr>
<tr>
<td>List 2</td>
<td>934</td>
<td>781</td>
<td>676</td>
<td>649</td>
</tr>
<tr>
<td>List 3</td>
<td>1233</td>
<td>738</td>
<td>763</td>
<td>727</td>
</tr>
<tr>
<td>List 4</td>
<td>1171</td>
<td>755</td>
<td>666</td>
<td>648</td>
</tr>
<tr>
<td></td>
<td>1080</td>
<td>838</td>
<td>686</td>
<td>671</td>
</tr>
</tbody>
</table>
Figure 8.5. graphically presents the means collapsed over List.

![Bar graph showing reaction times for correct decisions for position 5.](image)

There were no significant main effects for either List or Itemgroup. There were significant main effects for Ambiguity \( [F_1(1,36)=92.51, p<.0005; F_2(1,20)=35.77, p<.0005; \text{Min}F'(1,35)=25.80, p<.001] \) and Semantic Status \( [F_1(1,36)=22.16, p<.0005; F_2(1,20)=8.79, p=.008; \text{Min}F'(1,36)=6.29, p<.025] \). In the by-subjects analysis, Semantic Status interacted with List \( [F_1(3,36)=7.11, p=.001] \). The interaction Ambiguity by Semantic Status was significant \( [F_1(1,36)=11.27, p=.002; F_2(1,20)=13.86, p=.001; \text{Min}F'(1,54)=6.22, p<.025] \). In the by-subjects analysis, the interaction List by Ambiguity by Semantic Status was also significant \( [F_1(3,36)=3.41, p=.028] \).

Post-hoc paired t-tests revealed that response times were significantly higher in the
semantically normal than the semantically violated sentences in the ambiguous condition (across Lists: \(t(39)=3.69, p=.0005, 1\text{-tail}\), and also separately for all Lists except List 1, all \(p\)'s < .009, 1-tail; across Itemgroups: \(t(23)=3.25, p=.002, 1\text{-tail}\), and separately for Itemgroup 3 and 4, \(p\)'s < .047, 1-tail), but not in the non-ambiguous condition. For the semantically normal sentences, response times were significantly higher in the ambiguous than the nonambiguous condition across all Lists, and for each List separately (all \(p\)'s < .005, 1-tail) and across Itemgroups and for all Itemgroups separately, except Itemgroup 1 (all \(p\)'s < .0036, 1-tail). This Ambiguity effect also held for the semantically-violated condition, across Lists \(t(39)=4.77, p<.00025, 1\text{-tail}\), and each List separately, except List 3 (all \(p\)'s < .02, 1-tail), and across Itemgroups \(t(23)=3.38, p=.0015, 1\text{-tail}\), and for Itemgroup 2 separately \(t(5)=3.57, p=.008, 1\text{-tail}\). These patterns of results hold quite consistently for individual subjects and items (decision times were greater in normal-ambiguous than normal-nonambiguous for 38 out of the 40 subjects, and 23 of the 24 items; greater for semantically-violated ambiguous than semantically-violated nonambiguous for 30 subjects and 18 items; and greater for normal-ambiguous than semantically-violated ambiguous for 25 subjects and 18 items).

**Further analyses at position 5**

The combined findings that response times at position 5 were higher in the ambiguous normal than the ambiguous semantically-violated condition, while there was a garden path effect in both conditions, requires further analysis. Is it the case that subjects are garden-pathed to a lesser extent in the semantically-violated than the semantically-normal condition? Or are some subjects not garden-pathed at all in the semantically-violated condition, whereas others are garden-pathed? Note that the overall garden path effect in the semantically-violated condition could be smaller in the above analyses simply because subjects are "lumped together": data from subjects who might not have been garden pathed at all are combined with data from subjects who might have experienced a garden path of the same magnitude as in the semantically-normal condition. Furthermore, there is a second type of 'lumping together': in the response
time analyses reported above, the mean response time of each subject was computed over six items per condition; it might be that subjects in the semantically-violated ambiguous condition are garden pathed only on some of the sentences, but that the garden path-effect, for those sentences where it occurs, is as big as in the normal ambiguous condition.

To test these possibilities, responses at position 5 were analyzed further in terms of the number of extremely long responses as the index of a garden path effect. Additionally, the effect of the different conditions on making a correct decision was analyzed again, this time after the extremely long responses had been removed.

Table 8.20. lists the percentage of extremely long Yes-decisions per condition on position 5. Responses were defined as extreme if they were greater than 1620 milliseconds (this cut-off was established in the same way as in Experiment 5 and 6).

<table>
<thead>
<tr>
<th></th>
<th>ambiguous</th>
<th>non-ambiguous</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>32%</td>
<td>3%</td>
</tr>
<tr>
<td>sem-viol.</td>
<td>15%</td>
<td>3%</td>
</tr>
<tr>
<td>normal</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>sem-viol.</td>
<td>3%</td>
<td>3%</td>
</tr>
</tbody>
</table>

For each subject (and, separately, for each item) the number of extreme responses per condition was calculated. Sign- and Wilcoxon Matched-pairs Signed-ranks Tests indicated that the number of extreme responses was higher in the normal-ambiguous than the semantically-violated ambiguous condition (all p's < .004, 1-tail), and higher in the normal-ambiguous than the normal-nonambiguous condition (all p's < .0001, 1-tail), and in the semantically-violated ambiguous than the semantically-violated nonambiguous condition (all p's < .02, 1-tail). As expected, there was no difference in the number of extreme responses in the two nonambiguous conditions.
In the final analysis for position 5, the effect of the different conditions on response times was analyzed for those responses which were not extremely long. Subject- and item-means of response times for correct decisions which were smaller than 1621 ms were calculated per condition and entered into AOV’s. There was a significant main effect for Ambiguity \( F_1(1,35)=15.78, p<.0001; F_2(1,20)=7.74, p=.011; \) \( \text{Min}F'(1,39)=5.19, p<.05 \), but no other effects were significant, i.e. in particular, the interaction Ambiguity by Semantic Status was not significant.

**Position 6**

Data analysis for this position only included response times for correct decisions if the subject had not made a No-decision on any earlier word in the same sentence. The hypothesis under investigation here is that no garden path effect is expected for this position in the semantically-violated ambiguous condition, but that for the normal-ambiguous condition, a garden path effect is expected. On position 5 the garden path was bigger in the normal-ambiguous than the semantically-violated ambiguous condition, and thus reanalysis processes in the normal-ambiguous condition might still be occurring on position 6, but are less likely to persist for semantically-violated ambiguous sentences.

Unfortunately, subject- and item-means of both the untransformed and logtransformed data were not normally distributed, and the variances across groups were not equal. Analyzing the effect of the two repeated-measures factors with the non-parametric Friedman test (collapsing over Lists and Itemgroups) showed that the conditions did not differ significantly for the non-transformed data (neither in the by-subjects nor the by-items analysis). For the logtransformed data, the analysis by items, but not the analysis by subjects indicated that the conditions differed significantly \( p<.0345 \).

In a further step, the number of extremely long responses in the different conditions on position 6 was analyzed. Extremes were defined as responses greater than 1212 ms.
The percentage of extremely long responses per condition is given in the next table.

<table>
<thead>
<tr>
<th></th>
<th>ambiguous</th>
<th>non-ambiguous</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>normal</td>
<td>sem-viol.</td>
</tr>
<tr>
<td></td>
<td>14%</td>
<td>7%</td>
</tr>
</tbody>
</table>

Table 8.21. Experiment 7: percentage of extremely long Yes-decision times per condition on position 6

For each subject (and for each item) the number of extremes per condition was calculated. Wilcoxon Matched-pairs Signed-ranks Tests indicated that the number of extreme responses was significantly different in the normal-ambiguous as compared to the semantically-violated ambiguous condition (p’s < .024, 1-tail) and higher in the normal-ambiguous than the normal-nonambiguous condition (p’s < .046, 1-tail). The number of extremes did not significantly differ in the semantically-violated ambiguous as compared to the semantically-violated nonambiguous condition, and in the two nonambiguous conditions (the normal-nonambiguous versus the violated-nonambiguous conditions).

8.3.5. Discussion

The results confirm the prediction that subjects notice the semantic violation of the postverbal noun (position 4) in the ambiguous sentences, eg. of 'hat' in the following example:

Wherever Alice walks her funny old hat ....

Decision times on position 4 were higher in the semantically-violated ambiguous condition than in the normal-ambiguous condition (where the postverbal noun was 'dog', in the example above), and also higher than in the semantically-violated nonambiguous condition in which the semantically violated noun was not in direct-
object position, but followed the adverbial (as in 'Wherever Alice walks in the city her funny old hat ...'). This increase in decision times indicates that a noun which directly follows a verb is, at least temporarily, analyzed as the direct object; if (only) the intransitive analysis had been chosen, this noun would have been analyzed as the subject of a new clause, and as such would not have violated the semantic selection restrictions of the verb, causing extra processing costs. The increase probably indicates that subjects noticed the semantic violation and tried to find a possible interpretation for the semantically violated transitive analysis. The question is what happens during this extra processing time on position 4 in the ambiguous semantically-violated condition.

Did subjects eventually reject the semantically violated direct object analysis and adopt the alternative (intransitive) syntactic analysis instead? Such a result would suggest that the error signal from the semantic subsystem forces the parser to abandon the transitive analysis and to take up the intransitive analysis, which the semantic subsystem accepts as correct. If this were the case, there should be no garden path effect on position 5 (eg. on 'attracts' in: 'Wherever Alice walks her funny old hat attracts ...').

On the other hand, the transitive analysis might simply deactivate during this extra processing time (spent on 'hat') which prevents further syntactic evidence from coming in. The output of the semantic subsystem indicating that the transitive analysis is semantically violated might add to its deactivation. However, if the parser does not use the output of the semantic subsystem in making its initial decisions, it will not give up its preferred transitive analysis, and a garden path should occur on position 5, even if the semantically violated transitive analysis led to extra processing costs on position 4.

Before turning to the results on position 5, another aspect of the findings for position 4 has to be clarified. As reported above, the semantically-violated ambiguous sentences received more No-decisions than sentences in any other condition. Note, however, that the total number of No-decisions was quite low even in this condition, i.e. 2.5% (or six No-decisions in absolute terms). Does this mean that in the majority of cases a sentence is judged grammatical even when it contains a semantic violation?
To answer this question one has to establish first of all if the subjects who made a No-decision on position 4 were in fact garden pathed on position 5 (if they were not, the postverbal noun could not have been attached as direct object, and the No-decision at position 4 would not reflect the semantic status of the postverbal noun as direct object). Additionally, it has to be ascertained that the Yes-decisions on position 4 were followed by a garden path on position 5 (the Yes-decisions on position 4 cannot be taken as evidence that subjects accepted as grammatical a sentence with a semantically violated direct object noun if there is no further evidence that they in fact did attach the postverbal noun as the direct object; subjects might have overwhelmingly made a Yes-decision on 4 because they did not so attach the noun).

Let us first focus on the responses on position 5 following a No-decision on position 4. Taking as evidence for a garden path effect at position 5 either a No-decision, or an extremely long response time for a correct decision, it becomes apparent that most No-decisions on 4 were followed by a garden path on 5. In two instances, there was a No-decision on position 5, and in three of the four remaining cases the time to press the Yes-key on position 5 was extremely long (greater than 2345 ms). Thus, five of the six No-decisions on position 4 were followed by a garden path on position 5. These cases, then, are the few instances in which subjects rejected a sentence as ungrammatical because the direct object noun was semantically violated.

Let us now turn to the Yes-responses on position 4 and determine whether they were followed by a garden path on position 5. As argued above, if there is no clear evidence for a garden path effect on position 5 after the subject had made a Yes-decision on position 4, the semantic status of the noun as direct object might have been irrelevant for making the long Yes-decision on position 4. The lack of a garden path would indicate that the noun was not attached as direct object, and pressing the Yes-key on position 4 would not unequivocally support the interpretation that the subject accepted a sentence as grammatical which contained a semantically violated direct object noun.

The results of the initial analyses of variance clearly indicated a garden path effect at position 5 in the semantically-violated condition. There should not have been an
It is possible that this increase in response time does not reflect garden pathing but spill-over from processing on position 4 in the semantically-violated ambiguous sentences. Further research is required to distinguish between these possibilities for example by introducing a modifier after the postverbal noun in the semantically-violated ambiguous condition to absorb any spill-over effects.
increase in response time (relative to the violated-nonambiguous condition) if subjects had given up the transitive analysis on position 4 and adopted instead the intransitive analysis. Note, however, that in this analysis data from each subject were collapsed over six sentences to create a score for each condition: it is possible that a subject was garden pathed only on a few sentences, which would have resulted in extremely long responses, but that on most sentences, s/he was not garden pathed at all. The analysis of the number of extremely long Yes-decisions on position 5 as a further indicator of a garden path effect (Table 8.20. above) showed in fact that only 15% of all responses on position 5 in the violated-ambiguous condition were extremely long.

Let us sum up the implications of these further analyses for the above finding that the majority of subjects seem to judge a sentence as grammatical on position 4 even in the violated-ambiguous condition.

We first established that in the semantically-violated ambiguous condition there are only 5 cases in which a No-decision at position 4 was followed by a garden path on position 5. Furthermore, Yes-decisions on position 4 were followed by a garden path on position 5 (here defined as either an extremely long response or a No-decision) in only 36 cases (2 No-decisions, and 34 extremely long Yes-decisions). Thus, while No-decisions in the ambiguous-violated condition on position 4 were extremely rare, one cannot safely conclude that the opposite was also true, i.e. that the majority of subjects judged a sentence as grammatical even when it contained a semantically-violated direct-object noun. The application of a strict definition of a garden path effect (number of extremely long responses on position 5) forces us to make the more careful conclusion that 15.4% of all Yes-decisions at position 4 were followed by a clear garden path on position 5. Thus, only these cases can be unequivocally interpreted as evidence that subjects accepted as grammatical a sentence on the position where a semantically violated direct-object noun occurred.

It would be instructive to know if the length of time a subject took to make a Yes-decision on position 4 in the semantically-violated ambiguous condition affected the likelihood of a garden path on position 5 in the same sentence. An initial test showed...
that there was no significant correlation between times for making a Yes-decision on position 4 and times for making a Yes-decision on position 5. An inspection of a scatterplot indicated, however, that there were only few extremely long responses on position 5 when the response on position 4 had been extremely long; and there were many extremely long responses on position 5 when the responses on position 4 were not extremely long.

This impression was tested in the following way. Responses on position 4 and 5 were first divided into two groups, extremes and non-extremes. Cut-off points were again defined as the value that lies 1.5 times the interquartile range above the 75th percentile of the relevant control distribution of correct decision times. The cut-off point for responses at position 4 in the violated-ambiguous condition was established on the basis of the distribution of the relevant control condition, i.e. the normal-ambiguous condition at position 4. This cut-off was 1280 ms. The cut-off for responses at position 5 for the violated-ambiguous condition was taken from the distribution for the violated-nonambiguous condition at position 5. Its value was 1173 ms. The data for positions 4 and 5 in the violated-ambiguous condition were then labelled as extremes if they fell beyond the relevant cut-off points and non-extremes if they did not.

Subjects produced more extremely long response times on Position 5 if their response time on Position 4 in the same sentence had NOT been extremely long, as compared to when it had been extremely long on Position 4 (Sign test: p=.0003, 2-tail).

Let us now consider the findings for position 5. The initial results for position 5 indicated, as predicted, a garden path effect in the normal sentences. In addition, there was clear evidence for a garden path effect in the semantically-violated sentences, too. However, the increase in decision time was greater in the semantically-normal than the semantically-violated ambiguous sentences.
A potential problem in interpreting the latter finding is that data for each subject per condition were collapsed over (maximally) six items (and for each item over 10 subjects), and furthermore that data from all subjects (and all items) were collapsed to form the mean for each condition. As was mentioned in the Results section, one has to check whether in the violated-ambiguous condition some subjects were not garden pathed at all (or only garden pathed on some sentences), while others were garden pathed to the same extent as in the semantically-normal condition. Such a mixed pattern would have led to a smaller overall effect, as found in the initial analysis. Note that if this were the case, the difference in size of the garden path effects in the normal-ambiguous and the violated-ambiguous conditions cannot straightforwardly be attributed to semantic information facilitating reanalysis.

The analysis of the number of extremely long Yes-decisions on position 5 repeated the pattern of results found in the initial analyses of variance: there were more extremes in both ambiguous conditions, with the normal-ambiguous condition having more extremes than the violated-ambiguous condition.

Is the initial finding of a more pronounced garden path effect in the normal-ambiguous as compared to the semantically-violated ambiguous condition due to the greater number of extremely long responses in the normal-ambiguous condition? Equally importantly, is the garden path effect in the semantically-violated ambiguous condition only a result of the greater number of extremes in the semantically-violated ambiguous as compared with the semantically-violated nonambiguous condition?

The final analysis on position 5 addressed these questions. Included here were only responses that were not extremely long. This analysis indicated that responses on position 5 in the ambiguous conditions, when extreme values had been removed, did not differ significantly from each other, but were higher than in the non-ambiguous conditions.

Thus, in both ambiguous conditions there seems to be a garden path effect which is not due only to a few extremely long responses. The finding that there are significantly
more extremely long Yes-decisions (and more No-decisions) on position 5 in the
normal-ambiguous than the violated-ambiguous condition suggests that recovery from
the garden path turns out to be extremely difficult more often in the normal-ambiguous
than the violated-ambiguous condition.

In conclusion, then, it seems that while subjects noticed that the postverbal noun
violated semantic selection restrictions of the verb in the semantically-violated
ambiguous sentences, they did not generate the alternative syntactic analysis which
would have prevented them from being led up the garden path on position 5. The
semantic violation, however, somehow helped them to recover more easily from the
garden-path effect.

What are the implications of these results for the model developed above (see
discussion section of Experiment 6)?

In this model it is assumed that two syntactic analyses are computed in parallel by the
parser, prompted by the two subcat frames of the verb. The two analyses are assumed
to be stored in the temporary linguistic store, and their level of activation is assumed
to be a function of further syntactic input as well as of the evaluation of their
plausibility by the thematic subsystem. When an analysis fails to receive further
syntactic support (eg. the intransitive analysis does not receive any compelling
syntactic support immediately after the verb) this analysis begins to deactivate. The
level of activation that a particular analysis reaches depends on the output of the
thematic subsystem: an analysis that is judged to be very plausible (given semantic
features of the subject noun and canonical theta role assignments) deactivates less than
one that is judged less plausible. The level of activation of an analysis in the
temporary linguistic store affects reanalysis processes because the parser, when forced
to re-analyse, checks the alternatives in the temporary linguistic store. An alternative
analysis with a higher level of activation (due to its support from the thematic
processor) is more readily available and can be taken up by the parser more easily than
an alternative with a lower level of activation.
The present set of results indicates that semantic misfusion at position 4 in the ambiguous condition (i.e. on the postverbal noun in 'Wherever Alice walks her funny old hat ...') causes an increase of processing complexity, which I suggest reflects an error message from the semantic subsystem. It is clear, however, that during this extra processing time subjects do not adopt the alternative syntactic analysis (i.e. they do not undo the direct object analysis to reattach the NP as the subject of a new clause): when subjects processed the verb or auxiliary in position 5, which provided syntactic evidence that the postverbal NP had to be attached as the subject of a new clause, they were garden pathed even in the semantically-violated condition. Thus, the error signal from the semantic subsystem at position 4 in the semantically-violated ambiguous condition does not force the complete deactivation of the initial syntactic analysis. This apparently contradicts Rayner et al.'s (1983) findings regarding PP-attachment-ambiguities, such as in sentences a) and b) below:

a) He saw the girl with the red hair.

b) He saw the girl with the binoculars.

In sentence a) the analysis which Rayner et al. assume is preferred on syntactic grounds (i.e. attaching PP as a modifier of the verb 'seeing') is not the one that subjects finally come up with. In spite of the fact that there is no syntactic error signal, subjects seem to give up the (postulated) preferred analysis to arrive at the correct analysis, i.e. attaching the PP as a modifier of the NP. This has been taken as an example where a pragmatic error ('hair' is not an instrument of seeing) forces a syntactic reanalysis. The contrast to my findings might be due to the fact that in my sentences, a semantic error signal occurred before the whole sentence had been processed. It might be that a nonsyntactic error signal can force a syntactic reanalysis only at the end of a sentence when the subject has time to come up with an alternative.

The finding that the garden path effect on position 5 was smaller in the semantically-violated than the normal ambiguous condition, plus the fact that there were fewer extremely long decision times in the semantically-violated than the normal ambiguous condition, suggests that the output of the semantic subsystem must have helped
reanalysis. I suggest that this was due to the different levels of activation of the alternative analyses in the temporary linguistic store. The semantic subsystem computed the semantic misfusion and sent an error message to the temporary linguistic store, which initiated a decrease in activation of the transitive analysis. This initial deactivation of the transitive analysis occurred as soon as the postverbal noun was processed. Deactivation was, however, not sufficient. The transitive analysis remained the most strongly activated analysis in the temporary linguistic store. However, when the syntactic error message (on position 5) forced deactivation of the transitive analysis, this was accomplished faster in the semantically violated sentences because the transitive analysis had already begun to deactivate as a result of the earlier semantic error message.

In conclusion, the present model attributes greater influence to the output of the parser than to the output of the semantic subsystem in determining the level of activation of alternative syntactic analyses held simultaneously in the temporary store. A word or phrase which syntactically supports one analysis leads to an increase in activation of that analysis, and to a decrease of the alternative. A word or phrase which is syntactically incompatible with an analysis forces its deactivation. This syntactically caused deactivation can be slowed down, however, by the output of the thematic processor (see results of Experiment 6): the positive evaluation of an analysis by the thematic processor increases the activation level of the analysis. A syntactically caused deactivation can be sped up if the prior negative evaluation of an analysis by the semantic subsystem (based on the evaluation of semantic fusion) has reduced its activation level.

In sum, the findings from Experiment 6 reported in the previous section allowed us to rule out a strong thematic model in which the semantic feature in/animacy forces a default assignment of thematic roles which could raise the level of activation of the syntactically non-preferred analysis to the extent of avoiding a garden path effect. The semantic features determined the evaluation of thematic frames by the thematic subsystem, and increased the activation of the syntactically nonpreferred intransitive analysis, which helped later reanalysis processes. In Experiment 7, I varied the
semantic status of the postverbal noun, but only included animate nouns as the subjects of the subordinate clause. This experiment showed that semantic misfusion increases processing complexity and speeds up the reanalysis process but that it does not force a syntactic reanalysis.
CHAPTER 9: CONCLUSIONS

9.1. Introduction

In this chapter I will briefly sum up the main findings of the research reported in this thesis and discuss them within the framework introduced in Chapter 1.

The main goal of this thesis was to specify which aspects of word meaning subjects analyze when they process words, and how this semantic information affects the processes of syntactic analysis. I tried to address this goal in three ways. In the first part of the thesis the focus was on the processing of single words rather than words in sentence contexts. I used the semantic priming effect as a test of a theoretically derived definition of semantic relatedness between nouns. In a second line of investigation, I examined word meaning in the context of sentence processing by testing how sensitive subjects are to semantic violations in syntactically well-formed sentences. The third way of addressing the main goal of the thesis consisted in taking the issue of the influence of semantic information during sentence processing a step further, by studying whether semantic information can actually determine syntactic choices.

In addressing each of these three main questions of the thesis, my experiments were designed to provide new evidence about particular controversial issues which I have identified in the relevant literature. But the experiments were also designed to shed further light on two of the general problems in the study of language processing, i.e. the levels problem (the relationship between linguistic theories of grammar and theories and evidence from language processing) and the methodological problem to elucidate the time course of linguistic processes rather than their end product by employing on-line methods.
9.2. Processing the meaning of single words

The first experiment addressed the question of whether semantic features of nouns are activated rapidly when they are processed without a sentence context. The semantic priming effect was used as a technique to test a particular proposal about the semantic relations between two nouns (derived from Jackendoff, 1983). In many semantic priming studies, the notion of semantic relatedness is used interchangeably with associative relatedness. Experiment 1 was designed to provide new evidence to decide whether word pairs that are semantically but not associatively related, and which do not share a typical functional relationship, produce the semantic priming effect.

I used Jackendoff's semantic theory (Jackendoff, 1983), developed in linguistics, as a basis to define whether or not two nouns were semantically related. In Experiment 1 I focused on nouns which mapped into the THING-hierarchy. I compared noun pairs which were associated (according to standard association norms), for example 'doctor - nurse', with noun pairs which were not associated but which were semantically related by virtue of sharing the same mapping function, for example 'son - nurse', which both map into HUMAN THING. Priming was measured as the reduction of the time required to make a lexical decision on targets in the related pairs as compared to unrelated control pairs, in which the target was preceded by a noun which mapped into a different basic concept (eg. 'baptism - nurse').

The results of Experiment 1 indicated that associatively related primes facilitated making a lexical decision on targets (as compared to the unrelated control condition), but that a semantically related prime did not lead to reliable facilitation. Furthermore, responses in the associatively-related condition were faster than in the semantically-related condition. The priming effect found for associatively related pairs indicates a) that the experimental paradigm was sensitive to priming effects, and b) that 'semantic priming' is not the proper term for this effect: it was not a semantic relationship, but the fact that the targets were high associates of the primes, which was crucial in producing priming.
Priming effects in the lexical decision task can be interpreted in different ways. Following the classic spreading activation model, priming is the result of activation spreading across links between word nodes in a network. The lexical decision task is seen as a measure of the ease of lexical access to the target word. When activation automatically spreads from the prime word to the target, the level of activation of the target is raised, and both lexical access and making a lexical decision are assumed to occur faster than without such preactivation. The links between nodes along which activation spreads are designed to represent the associative relationship between words. To account for priming for semantically related words which are not associated, one could simply postulate further links to connect words that map into the same basic semantic concept; these additional links would allow activation to spread automatically to words which share the same basic concept, and would thus speed up lexical access and performance in a lexical decision task assumed to be sensitive to lexical access.

However, research carried out in the context of the Logogen model has shown that the view of the lexical decision task as directly reflecting lexical access cannot be correct. Firstly, this view does not lead to correct predictions with respect to the interaction of contextual and visual information. Morton (1982) refers to evidence (Becker & Killion, 1977; Stanners, Jastrzembki & Westbrook, 1975) showing that word frequency and visual degradation of the target have an additive effect in the lexical decision task. If the lexical decision task measured lexical access, then frequency of words should interact with the visual quality of the word. Furthermore, Becker (1976) showed that word frequency and difficulty of a secondary task interacted in the lexical decision task, indicating that making a lexical decision is affected by attentional demands. An alternative interpretation of priming effects is to consider lexical decisions as reflecting a post-access decision process, which uses as evidence not only the activation of a word in the lexicon, but any other information that could be useful in distinguishing real words from nonwords.

How would semantic priming in the lexical decision task function according to the alternative account? Suppose that words in the lexicon are not connected on the basis of shared semantic features. How would the decision maker get information which
would determine its choice? Diagram 9.1. presents the processing systems that are potentially relevant in making lexical decisions on a target preceded by a prime word.

A minimal characterisation of the lexical system (see Chapter 1) states that it contains abstract visual representations of words in a format which allows output from the initial visual analysis process to make contact with word representations in the lexical system. It is further assumed that each word has a tag, specifying its basic semantic and syntactic features. Upon activation of a word in the lexicon, the semantic tag is assumed to be passed on to the semantic subsystem. If two words are input in close temporal proximity, the semantic subsystem might register whether the two words share semantic features. The decision maker could make use of the output from the semantic subsystem (specifying that the target shares semantic features with an
immediately preceding word) to come to a faster decision.

The lack of reliable priming for semantically-but-not-associatively related word pairs in Experiment 1 does not support either the spreading activation- or the alternative account. Thus, on the one hand, this result indicates that there are no links in the lexicon which are based on shared semantic features of the type investigated here, along which activation could spread automatically. On the other hand, this finding also indicates that in this task the semantic subsystem does not produce any output to the decision maker which would signal the relationship between two nouns in terms of shared semantic features.

Let us now turn to the other finding of Experiment 1, the priming effect for associatively related words. Priming in this experiment clearly did not depend on the semantic relationship between two words, but was determined by the existence of an associative relationship between them. There is no entirely satisfactory account of associative priming in the literature. As discussed above, the classic spreading activation account can be criticized for regarding lexical decisions as direct measures of lexical access rather than the outcome of a post-access decision process. Which sources of information are available to the decision maker? To account for the processing of associatively related pairs which also share basic semantic features, the alternative mechanism proposed above for semantically related words could be invoked again: upon lexical access, the semantic subsystem registers that prime and target share semantic features, and this match leads to a faster response from the decision maker. However, this cannot be the full story. Why would the semantic subsystem operate in this way when associatively related words are processed, but not when semantically related words are processed? Experiment 1 showed that there was no facilitation for semantically-related pairs, and that lexical decisions on targets in the associative condition were faster than in the semantic condition. Thus, sharing semantic features is not a sufficient condition for creating priming. Furthermore, shared semantic features do not seem to be necessary for priming, since associatively related words do not necessarily share basic semantic features (in fact, 8 out of the 36 associatively related pairs of Experiment 1 did not share basic semantic features). Associatively related
words might instead, for example, share some functional relation, as in priest - church; time - clock; breast - woman; igloo - eskimo; bee - honey, etc. While in Experiment 1 the type of semantic relations in associatively related pairs was not systematically manipulated, other research (eg. by Moss et al., 1992) found priming for associatively related words which do not share basic semantic features but a functional relationship. One might want to propose that such functional relations, too, are represented in the lexicon as semantic tags which get passed on to the semantic subsystem; however, one still faces the problem of why there was priming for associatively related pairs but not for semantically related pairs. In the present model, there is only one further processing system, in addition to the lexical system and the semantic subsystem already discussed, which might provide information that is used by the decision maker, and this is the central system. Activation of a word in the lexicon might trigger the activation of prototypical knowledge about the referent of the word in the central system. The prototypical knowledge would include highly likely scenes or scripts in which the referent participates, and the typical entities that the referent would co-occur with. Activating the typical knowledge about the referent of a target word is assumed to be faster if it overlaps with the knowledge representation activated after accessing the prime word. Since nonwords do not have any referents about which there could be any prototypical knowledge, the decision maker could use the output of the central system, indicating that the referent of the target is part of the prototypical knowledge about the referent of the prime, to come to a decision about the lexical status of the target string.

In sum, the pattern of results of Experiment 1 suggests that the semantic subsystem is not involved in making lexical decisions when processing pairs of words. It was argued that the priming effect found for associatively related pairs cannot be primarily caused by their sharing various types of semantic relationships. Instead, the associative priming effect was interpreted as reflecting the operation of a post-access decision process which takes advantage of the overlap of prototypical knowledge about the referents of the prime and target word in making lexical decisions.

The next experiments addressed the question whether the semantic subsystem,
operating on specific semantic features, might be active when words are presented in a linguistic context rather than in the context of one preceding word.

9.3. Processing word meaning in sentence contexts

The following experiments addressed the role of basic semantic features of words during sentence processing. Even though priming was only found for associatively related word pairs in the previous experiment, it might be the case that basic semantic features are required in sentence processing.

In Experiments 2, 3, 4 and 4ba, I focused on particular semantic features which have syntactic consequences, namely semantic selection restrictions of verbs (eg. Chomsky, 1965). I tested the psychological plausibility of certain linguistic theories which propose that the lexical entry of a word contains, inter alia, a specification of semantic selection restrictions. These are semantic requirements which a verb places upon the nouns which it subcategorizes (the verb 'to drink', for example, requires its object noun to be a Liquid). In terms of processing, semantic selection restrictions of a verb can only be of any effect if the lexical entries of nouns contain information about at least those semantic features that are required by verbs. Interestingly, the kind of semantic information required by semantic selection restrictions can be described as abstract semantic features (such as Liquid; or Animate: the verb 'to kill', for example, requires the direct object noun to be animate), not unlike the ones captured by the mappings into basic ontological concepts in Jackendoff's theory. Are these basic semantic features accessed automatically during sentence processing?

The experiments on semantic selection restrictions are also relevant to the relationship between different types of linguistic information during sentence processing. Controversy reigns as to whether there is an initial phase in which syntactic information is processed independently of semantic information.

Experiments 2 and 3 addressed this issue by presenting syntactically well-formed
sentences, a subset of which, however, violated semantic selection restrictions (eg. 'The lake was bitten'). Subjects read those sentences, and their speed in performing a secondary task (i.e. a lexical decision on the sentence-final word) was measured as an index of the difficulty in processing a sentence. If initial sentence processing merely involves syntactic analysis which is independent of semantic information, violations of semantic requirements should not cause any increase in processing difficulty.

The results indicate that syntactically relevant semantic features of nouns are processed immediately during sentence processing, and that semantic violations disrupt processing. In particular, my data show that semantic violations slow down performance on the secondary lexical decision task, relative to both the neutral baseline condition, and the semantically normal condition. Furthermore, semantically normal sentences speed up lexical decisions as compared to both the semantically violated and the neutral baseline condition. This pattern of results was discussed with respect to a number of psychological models of context effects on word recognition.

The experiments go beyond previous work in several respects. First of all, the verbs included in my materials required a range of different semantic features rather than merely the Animacy/Inanimacy feature which has been dominant in previous work. Furthermore, a neutral baseline condition was introduced, in addition to the semantically violated and the non-violated conditions, to allow an assessment of the relative strength of the facilitating versus the inhibiting influence of semantic information. My materials also improved on another shortcoming in previous studies, in that sentences were designed to contain verbs and subcategorised nouns which were low in predictability in the semantically normal condition. Previously, semantic status had been confounded with predictability, with semantically normal sentences also displaying high predictability between noun and verb in a sentence. Thus, my materials more validly allow the conclusion that subjects are sensitive to violations of semantic selection restrictions at a very early stage during sentence processing.

The findings from Experiments 2 and 3 allow us to specify further the post-access model of lexical decisions developed to account for the findings of Experiment 1.
Again the decision maker has to decide whether the target is a word or not, and again I assume that it uses all kinds of information available to it. In Experiments 2 and 3 sentences were introduced (rather than word pairs as in Experiment 1) to involve a further processing subsystem, the parser, in addition to the lexical system and the semantic subsystem. Note that the central system is not included in this account. It was invoked in the discussion of Experiment 1 to account for the associative priming effect. In Experiments 2 and 3, however, there were no associative relations between words in the priming sentence and the target. Diagram 9.2. presents the version of the model relevant for discussing the results of Experiments 2 and 3.

Note that the use of sentences as input do not force us to invoke an additional parser; in principle, it is possible for the semantic subsystem to operate on single words, without 'knowing' about their syntactic relationship. However, since the semantic
subsystem is assumed to operate only on basic semantic features, it would, given words such as 'dog' and 'bitten', register that the former maps into THING and the latter into ACTION, and would decide that there is no semantic relationship between them (i.e. that they do not share basic semantic features). This account, however, does not fit with the finding that a lexical decision on, for example, 'bitten' was facilitated when it was preceded by the sentence: 'The dog was bitten', as compared to the sentence 'The lake was bitten'.

Alternatively, one might want to claim that the semantic subsystem checks semantic Fusion for isolated verbs and nouns before their syntactic structure has been analyzed by the parser, thus treating 'dog' 'bitten' as related. This cannot be correct, however, because words like 'duke', 'sipped', should create facilitation even when they occurred in the sentence 'The duke was sipped'. If the semantic subsystem had operated on the two isolated words, it would have checked the verb's semantic selection restrictions against the semantic features of the available noun, and would have found that the noun has the semantic features which the verb requires of its subject noun. However, since there was no facilitation but instead inhibition for violated sentences such as 'The duke was sipped', it seems correct to assume that the semantic subsystem evaluates semantic Fusion only after the noun and the verb have been structured syntactically. Incidentally, for the majority of violated sentences in Experiments 2 and 3, checking whether an isolated noun and a verb fuse would have been uninformative in any event: most sentences were violated in two ways, since the noun did not fulfil the semantic selection restrictions of the verb either as subject noun or as direct object noun.

Note that I do not propose that the semantic subsystem only operates on words pre-packaged syntactically by the parser. It is well-established (eg. Slobin, 1966) that non-

^But see footnote 1, p. 152 b
reversible passives, such as a) are understood more easily than reversible passives such as b):

a) The engine was repaired by the mechanic
b) The boy was pushed by the girl

This phenomenon has received various interpretations. One account (Wales & Marshall, 1966) assumes that b) is more difficult because of an incomplete analysis of the input. The analysis is initially incomplete because 'The boy ... pushed ... the girl' is semantically well-formed and consequently 'was' and 'by' are not properly analyzed. The initial analysis 'The engine ... repaired ... the mechanic', however, is semantically deviant and forces a full analysis of 'was' and 'by'. Interestingly, Morton (1966) characterises the semantic status of the two sentences quite differently. He assumes that the string 'engine, repaired, mechanic' is not ambiguous or deviant. The logical structure of the sentence is implicit in the string and thus no further analyses are required. The string 'boy, hit, girl', however, is semantically ambiguous and requires further processing to obtain a unique interpretation.

According to my proposal that the semantic subsystem might operate on words which have not been analyzed syntactically, non-reversible passives are easier to understand because information from the parser is not required to determine which noun is object and which is subject. By checking the semantic features of the nouns and the semantic selection restrictions of the verb, the semantic subsystem can determine that only the noun 'mechanic' has the semantic feature ANIMATE required by the verb 'repaired' from the underlying subject noun, and that it violates the semantic requirements for being the object noun of 'repaired'. In reversible passives, however, checking the semantic selection restrictions cannot determine which noun is subject and which is object, and evidence from the parser is required.

Let us now get back to the discussion of the findings of Experiment 2 and 3, in the context of the model depicted in diagram 9.2. I suggest that the parser, assumed to use phrase structure rules to build up a syntactic representation of a sentence, receives
from the lexical system information specifying the syntactic category of the words in
the input (e.g. noun, verb, adjective etc.). This is the information required for the
application of phrase structure rules combining words of different syntactic categories
to phrases. Note that syntactically, the violated sentences of Experiments 2 and 3 are
well-formed, eg. 'The lake was bitten' should be accepted by the parser, and a decision
which only considered output from the lexical system and the parser, should be made
without any delay. To account for the finding that there is a delay in making a lexical
decision in the violated condition compared to both the neutral baseline condition and
the semantically correct condition, I suggest a specific type of interaction between the
parser and the semantic subsystem. I assume that the semantic subsystem receives
from the lexical system information about the basic semantic features of nouns and the
semantic selection restrictions of verbs. It also receives input from the parser
specifying the relationship between words in the input, i.e. in the sentences of
Experiments 2 and 3 the analysis of the noun (e.g. 'dog') as part of the direct object of
the verb 'bite'. On the basis of these two types of information, the semantic subsystem
can compute whether or not the semantic features of the noun can fuse with the
semantic requirements of the verbs. If there is no fit, i.e. if the semantic selection
restrictions of the verb are violated, the decision maker has to decide among three
types of evidence: a No-message from the semantic subsystem, a Yes-message from
the parser, and a Yes-message from the lexical system. I assume that the three types of
information interfere with each other, and that the decision maker cannot simply focus
on the output of the lexical system (which, in itself, would be sufficient for making a
lexical decision since nonwords are not represented in the lexical system). This is the
same kind of interference as the one found in STROOP effects (see Morton, 1969b;
Stroop, 1935): conflicting information slows down processing even if some of the
information is not relevant for performing the task at hand.

In Experiments 2 and 3, the semantically violated sentences lead to a No from the
semantic subsystem, and a Yes both from the parser and the lexical system, whereas in
semantically normal sentences, the output is a Yes-message from all systems. In the
neutral baseline condition, there is no output from the semantic subsystem at all: there
is no noun whose semantic features have to fuse with the semantic selection
restrictions of the verb: thus, sentences in the semantically normal condition lead to facilitation when compared to the neutral condition, because, in the former, there is additional positive evidence. The slowed-down decisions in the semantically violated condition as compared to the neutral condition indicate that a No-message from one of the processing systems involved affects the decision maker to a greater extent than the lack of information from one of the processing systems.

Experiment 4 provided evidence that violations of semantic selection restrictions are processed differently to violations of pragmatic plausibility. In this study, active sentences were used in which the direct object noun either violated the semantic selection restrictions of the verb (eg. '.... the kids sharpened the bird ...'), or in which the noun was implausible as a direct object given the sentence context (eg. 'In the park the kids sharpened the pencil ...'). Reading times (measured in a cumulative self-paced reading paradigm) were longer on the first word following the semantically violated direct object noun (and also longer on every further following word, up to the fourth word, which was the last one analyzed); reading times in the pragmatically-odd condition, however, were longer than in the normal condition only on the fourth word following the direct object noun, but not on the words following it more immediately. These data suggest that semantic fusion and plausibility are evaluated by different subsystems. As before, semantic fusion is evaluated by the semantic subsystem which is assumed to have access to information about semantic features and semantic selection restrictions. It cannot know the plausibility of a direct object noun in a particular sentence context. Thus I suggest that pragmatic plausibility is assessed by the central system on the basis of knowledge about likely actions and events in the real world. To perform such assessment, the central system requires information about the linguistic structure of the sentence, incorporating the syntactic and semantic representation of the sentence. This information is assumed to be input from the temporary linguistic store into which both the semantic subsystem and the parser feed their current representations. Note that the temporary linguistic store can also function as the link between the parser and the semantic subsystem. In Diagram 9.2, the syntactic information specifying which noun is the direct object (required by the semantic subsystem to evaluate semantic fusion between verb and direct object noun)
was transmitted directly from the parser to the semantic subsystem. However, it is more parsimonious to consider the temporary linguistic store as keeping a record of all currently available sentence representations, with the semantic subsystem picking up from there the syntactic information it requires. Diagram 9.3. shows this additional store and the connection to the central system.

The findings from Experiment 4 show that the decision to press the pace-key to see the next word is not immediately affected by the implausibility of a word. The present model allows various options for why this is the case. Firstly, the decision maker might delay using the output from the central system about pragmatic implausibility; secondly, the central system might require extra time to assess the plausibility of a word in a given sentence context, and thirdly, the current linguistic representation of
the sentence is updated and available to the central system only after some delay, but
not immediately after each new word has been semantically and syntactically analyzed.
The finding that there was disruption of processing in the pragmatically inappropriate
condition (albeit several words after the direct object noun occurred) indicates that the
decision maker was eventually affected by the No-message from the central system.

The arrow in Diagram 9.3. from the central system to the decision maker, which was
introduced in the discussion of Experiment 1 to account for associative priming as the
result of the activation of prototypical knowledge about the referents of words, is now
interpreted to represent a No-message indicating that the word about which some
decision needs to be made is pragmatically implausible in the current sentence.

Experiment 4b, which required subjects to make grammaticality judgements on each
word of the sentence, showed that making a correct No-decision on a semantically-
violeated direct object noun takes longer than making a No-decision on a syntactically
violeated word in the position where syntactically a direct object noun has to occur ('... they sharpened the birds' versus '... they sharpened the under'). This finding can
easily be interpreted in the current model as indicating that in the syntactically violated
condition the decision maker receives negative information from both the parser
('under' is not of the required syntactic category), and the semantic subsystem ('under'
does not fit the sentence semantically). In the semantically-violated condition,
however, the parser gives a Yes-message ('birds' is a noun and as such a correct
continuation of the NP starting with a DET), but the semantic subsystem gives a No-
message.

Interestingly, in 41% of the responses, a Yes-decision was made on the semantically-
violeated noun. In half of these cases, the time to make the Yes-decision was longer
than in the normal condition, indicating that subjects did not simply ignore semantic
information. Note, however, that subjects did eventually make a Yes-decision, after
having taken extra time on the direct object noun: in these cases the output from the
parser has greater impact on the decision maker than the output from the semantic
subsystem.

343
In summary, then, the results of Experiments 2, 3, 4 and 4b do not only provide evidence for a model of context effects on word processing, they also indicate that the linguistic notions both of semantic selection restrictions of verbs and of basic semantic features of nouns correspond to information specified in the lexical entries of words in the mental lexicon, and, furthermore, that such semantic information is activated rapidly during sentence processing.

9.4. Word meaning and the choice of syntactic options

The previous set of experiments showed that semantic information (i.e. semantic features of nouns and semantic selection restrictions of verbs) is processed immediately, while the syntactic analysis of a sentence is still on-going. A further question about the interaction between semantic and syntactic information needs to be addressed: is semantic information used early on during sentence processing to determine the outcome of syntactic choices?

Previous research has produced conflicting results regarding the influence of semantic features on syntactic choices. The evidence comes from studies focusing on a range of types of garden path sentences and employed empirical measures which tap sentence processing at different phases. To do justice to an area of research such as this, in which the major conflict is not about whether semantic and syntactic information interact during sentence processing, but rather when they interact, I reviewed the findings in the literature with particular concern for the temporal requirements of different empirical methods.

The influence of semantic and other nonsyntactic information on syntactic choices can best be studied in situations where the effect of adopting a particular syntactic analysis can be measured clearly in the course of the analysis of a sentence. Garden path sentences provide such a set-up. They contain words that can be analyzed syntactically in (at least) two alternative ways, with only one analysis being compatible with later parts of the sentence. Garden-path sentences are the subset of temporarily ambiguous
sentences in which processing is severely disrupted. The disruption results from a
preference, displayed by the majority of subjects, to opt for the syntactic analysis
which turns out to be incompatible with later parts of the sentence. A number of fairly
clearly demarcated models exist in the literature, specifying the various factors
assumed to determine the initial preferences. I discussed the problematic consequences
which the direct application of linguistic notions to processing models can have with
two very different models of syntactic preferences, i.e. the models developed by by
Ford et al., and Frazier. Frazier claims that the parser, when faced with a structural
ambiguity, follows certain purely syntactic strategies, assumed to be independent of
non-syntactic information. These strategies are justified in terms of their minimising
processing complexity and STM requirements, and are assumed to lead to the simplest
analysis. The problem is that Frazier’s analyses rely heavily on specific linguistic
descriptions of sentences. Thus, what counts as the simpler, less complex analysis, is
only true, and can only be used as a justification of the parser’s preference for a
particular analysis, if the linguistic description is accepted in the first place. However,
for Frazier’s analysis this is not universally the case (eg. Konieczny, Hemforth &

I also showed that evidence on garden path sentences can be better interpreted if
assumptions about architecture are made explicit. As was the case with respect to the
classic Derivational Theory of Complexity, the implicit default assumption is still often
that of a serial architecture in which a syntactic parser is assumed to consider only one
analysis at a time.

Experiments 5 and 6 tested the influence of semantic features in determining the
preferred syntactic analysis by concentrating on the features of Animacy and
Inanimacy, and on a particular type of garden path sentence in which two syntactic
analyses of a postverbal noun are temporarily possible. Example sentence (1) illustrates
that this noun can be analyzed either as a direct object or as the subject of a new
clause since the verb which it follows can be used both transitively and intransitively:

(1) After the boy turned the wheel came loose ...
The verb 'to turn' has two subcat frames, i.e. it occurs either without a direct object (the intransitive frame) or with a direct object (the transitive frame). Note that 'the wheel', in the absence of any punctuation marks (or any intonational cues), is temporarily syntactically ambiguous: it could be analyzed as the direct object noun, or it could be postulated to be the subject noun of a following main clause (which is in fact the analysis which is confirmed by the phrase 'came loose' later on in the sentence).

The two subcat frames, which specify the syntactic constituents the verb subcategorizes (if any), are linked to two thematic frames, which specify the thematic roles played by the constituents. The thematic frame corresponding to the intransitive subcat frame assigns the role <Theme> to the subject NP. The thematic frame corresponding to the transitive subcat frame is <Agent, Theme>: the role of Agent is assigned to the subject NP, and the role of Theme to the direct object NP.

Sentences such as (1) above pose several questions with respect to my model. Firstly, if a verb has two subcat frames, are both of them stored in the lexical entry of the verb as a tag, and are they both passed on to the parser? Does the parser use both subcat frames to build up two analyses, or is there a default syntactic strategy which leads to one preferred analysis? What is the impact of semantic features of nouns on the parsing process? Stowe (1989) reported experiments investigating semantic features of the first subject noun, and the plausibility of the postverbal noun as a direct object. She proposes a thematic processor which is assumed to assign thematic roles to nouns on the basis of semantic features: animate nouns (e.g. 'boy') are, by default, assigned the Agent role, and inanimate nouns, the Theme role. Stowe reports that sentences such as (1) above led to garden paths, but sentences such as (2) below did not, because, she argues, in (2) the default thematic assignment of <Theme> to the first subject noun is only compatible with the <Theme> frame of the verb, which corresponds to the verb's intransitive subcat frame.

(2) Before the play finished the last scene became quite emotional.
Thus, these data would suggest that the parser receives input from the thematic processor about the preferred thematic frame for a verb in a current sentence. Since thematic frames are linked to subcat frames, the parser can use this information about the preferred subcat frame to decide which syntactic analysis to pursue. However, Stowe confounded the manipulation of semantic features with the plausibility of the subordinate sentence. Thus, one cannot safely conclude that the semantic features Animacy and Inanimacy determined the assignment of thematic roles by the thematic processor, and hence the decision of the parser. There are clearly sentences in which an inanimate noun can be assigned the Agent role rather than the Theme role, as in (3):

(3) After the motor turned the wheel....

In sentences such as (3) the inanimate subject noun is a plausible Agent in the particular transitive action described in the sentence. Experiments 5 and 6 addressed the question whether semantic features force a default assignment of thematic roles.

My data show a clear garden path effect in both the animate and the inanimate conditions, thus contradicting Stowe’s results, which showed that the garden path was avoided in the inanimate-subject condition. Furthermore, I found that the garden path effect was smaller in the inanimate than the animate condition, thus demonstrating that semantic manipulations had an effect on the size of the garden path effect. My data point to the shortcomings of a range of existing models, such as certain versions of multiple-analyses models, Ford et al.’s subcat preference model, Pritchett’s theta-assignment model etc. In particular, the present set of results cannot correctly be predicted from one of the most popular single-analysis model, namely Frazier’s strong syntactic preference model. Her model does not predict a difference in the size of the garden path effect. My data also go against Stowe’s strong thematic model, which suggests that semantic features such as Inanimacy force the assignment of a particular thematic role and the activation of a particular thematic and associated subcat frame, which in turn affects the syntactic analysis of the structurally ambiguous noun. Her model incorrectly predicts the absence of a garden path in the inanimate condition. My
data indicate that semantic features alone do not determine the assignment of thematic roles. Clearly, in my sentences, the transitive analysis was preferred, even in the inanimate condition. Thus, the Agent role must have been assigned to the inanimate subject noun, contrary to the predictions of the default thematic assignment model. My results indicate that thematic assignment is affected by the plausibility of the transitive action expressed by the noun - verb - noun sequence, which I had designed to be equal in the animate and the inanimate conditions (unlike Stowe). Thematic assignment does not seem to be fixed by some default principle which would only rely on the semantic features of a noun.

On the basis of my data I proposed a parallel analyses model in which the level of activation of alternative syntactic analyses kept in the temporary linguistic store is predominantly determined by syntactic preferences of the parser, such as Late Closure. Thematic evaluation of the alternatives, and output from the semantic subsystem about semantic fusion of the constituents in the alternative analyses, affect the level of activation of the analyses, but, apparently, not to the same extent as the parser. They cannot force the complete deactivation of a syntactic analysis if there is syntactic evidence for it, or sufficiently raise the level of activation of an analysis that does not have support from the parser.

The results of Experiment 6, however, showed that the output of the thematic subsystem ensures that the intransitive alternative analysis is more highly activated in the inanimate than the animate condition. Thus, when syntactic evidence forces the parser into reanalysis, the alternative analysis is more easily available in the inanimate than the animate condition, and the garden path effect is less strong. Diagram 9.4. shows the addition of the thematic subsystem to the model and indicates its inputs and outputs.
The thematic subsystem receives as input
a) the thematic frames of a verb, input from the lexical system;
b) the current syntactic analyses of the sentence enriched with the semantic features of
   the nouns, input from the temporary linguistic store, and
c) information about the plausibility of certain transitive actions, input from the central
   system.

There is output from the thematic subsystem to the central system, submitting a
particular thematic assignment for an evaluation of its plausibility. This link is
necessary to reconcile Stowe's (1989) finding that there was no garden path effect for
sentences with inanimate subject nouns. As discussed before, in her sentences the
postverbal noun was a highly implausible direct object in the inanimate-subject sentences, but not in the animate-subject sentences, thus confounding the effect of semantic features with the plausibility of the transitive action described by the sentence. In my sentences of Experiment 6, in which the plausibility of the postverbal noun as direct object was held constant in the animate and the inanimate condition, the output from the thematic subsystem did not prevent the garden path in the inanimate condition. The thematic subsystem also outputs information to the temporary linguistic store which affects the level of activation of the analyses kept there: the activation for an analysis corresponding to the thematic frame that the thematic subsystem evaluated as plausible will increase. The decisive factor determining the evaluation by the thematic subsystem in Experiment 6 were the semantic features of the subject nouns: animate nouns were preferentially assigned the Agent role, and inanimate ones, the Theme role.

Experiment 7 demonstrated the effect of semantic information on the parser in another way. Here, the noun in direct object position was designed to violate the semantic selection restrictions of the verb (eg. 'Wherever Alice walks her funny old hat ...'). Again, a garden path effect was not only found in the semantically normal condition (eg. 'Wherever Alice walks her funny old dog attracts a lot of attention'), but also in the semantically violated condition. Interestingly, processing times on the direct object noun itself were longer than in the control condition, suggesting that subjects noticed the semantic violation. However, the semantic evidence which disfavoured the direct object analysis was not strong enough to deactivate the syntactically preferred transitive analysis sufficiently to avoid the garden path. The semantic evidence, however, did enable faster reanalysis once syntactic evidence against the preferred syntactic analysis had been received. Faster reanalysis was interpreted as resulting from the greater speed with which the transitive analysis was given up due to deactivation having already started on the ambiguous noun itself.

This multiple constraint satisfaction model needs to be tested against further types of garden path sentences. It also allows us to make some specific predictions with respect to the garden path structure investigated in the present set of experiments, for example
concerning length effects: if the postverbal NP is long (either by adding adjectives, which would move the head of the NP further away from the verb, or by adding relative clauses after the head of the NP), the assumed higher level of activation of the non-preferred syntactic analysis (resulting from thematic and semantic support) should disappear, such that garden path effects (i.e. difficulty of reanalysis) should become identical in the animate as compared to the inanimate condition.

The series of experiments reported in this thesis have specified several aspects of the framework introduced in Chapter 1. This research demonstrates that the framework is of great use in discussing current findings in the literature and that it allows the generation of testable predictions about the processing of word meaning.
BIBLIOGRAPHY


Bresnan, J. (1978). A realistic transformational grammar. In M. Halle, J. Bresnan, &


facilitation of word recognition as measured from a neutral prime. *Memory & Cognition, 10*, 358-370.


comprehension: syntactic structures and task demands. Brain and Language, 42(1), 1-29.


56-98.


McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention and Performance XII: The psychology of reading* (pp. 3-36). Hove: LEA.


McClelland, J. L., St. John, M., & Taraban, R. (1989). Sentence comprehension; A


papers: The proceedings of the 1966 Edinburgh conference (pp. 10-17).
Edinburgh: Edinburgh University Press.


Swinney, D. A., Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*, 645-660.

Tabossi, P. (1988a). Accessing lexical ambiguity in different types of sentential


APPENDIX A: INSTRUCTIONS FOR EXPERIMENT 1

To mark the passages of text that appeared on the screen at one time, I inserted ‘/’ in the following text.

"Please read these instructions carefully. On each trial, you will first see a cross for a short moment. Then you will see a string of letters immediately followed by a second string of letters. Some of these strings will be English words and some of them will not be correct English words. Look carefully at both strings. Your task is to decide as fast and as accurately as possible whether or not the second string is a correct English word. The first string may or may not help you in this decision. There are two response keys. If the second string is a real English word, press the Yes-key as fast as you can. If it is not a real English word, press the No-key as fast as you can. You can use one or both hands for pressing the buttons. If you accidentally pressed the wrong button, just press the correct button immediately afterwards. After you have pressed a key, there will be a blank screen for a few moments, and then you’ll see the cross again, followed by more letter strings. Please respond as fast as possible. Here we go!"
APPENDIX B: WORD-WORD PAIRS USED IN EXPERIMENT 1

Note: The first line in each triplet shows an A-pair (a relation of High Association between Target and Prime), the second line an S-pair (a semantic relation) and the third line a U-pair (unrelated Control relation).

The word in square brackets at the end of the second line gives the basic ontological concept into which the two semantically related words map. The word in brackets in the third line indicates the ontological concept into which the control prime maps.

<table>
<thead>
<tr>
<th>PRIMES</th>
<th>TARGETS</th>
</tr>
</thead>
<tbody>
<tr>
<td>bug</td>
<td>insect</td>
</tr>
<tr>
<td>fish</td>
<td>insect [ANIMAL]</td>
</tr>
<tr>
<td>honesty [PROPERTY]</td>
<td>insect</td>
</tr>
<tr>
<td>chair</td>
<td>table</td>
</tr>
<tr>
<td>broom</td>
<td>table [INDOOR THING]</td>
</tr>
<tr>
<td>brilliance [PROPERTY]</td>
<td>table</td>
</tr>
<tr>
<td>music</td>
<td>sound</td>
</tr>
<tr>
<td>sight</td>
<td>sound [PERCEPTION]</td>
</tr>
<tr>
<td>outing [EVENT]</td>
<td>sound</td>
</tr>
<tr>
<td>breast</td>
<td>woman</td>
</tr>
<tr>
<td>surgeon</td>
<td>woman [HUMAN]</td>
</tr>
<tr>
<td>length [AMOUNT]</td>
<td>woman</td>
</tr>
<tr>
<td>buoy</td>
<td>water</td>
</tr>
<tr>
<td>copper</td>
<td>water [NATURAL THING]</td>
</tr>
<tr>
<td>fight [ACTION]</td>
<td>water</td>
</tr>
<tr>
<td>stem</td>
<td>flower</td>
</tr>
<tr>
<td>fruit</td>
<td>flower [PLANT]</td>
</tr>
<tr>
<td>pound [AMOUNT]</td>
<td>flower</td>
</tr>
<tr>
<td>justice</td>
<td>law</td>
</tr>
<tr>
<td>faith</td>
<td>law [ABSTRACT THING]</td>
</tr>
<tr>
<td>inch [AMOUNT]</td>
<td>law</td>
</tr>
<tr>
<td>priest</td>
<td>church</td>
</tr>
<tr>
<td>train</td>
<td>church [OUTDOOR THING]</td>
</tr>
<tr>
<td>party [EVENT]</td>
<td>church</td>
</tr>
<tr>
<td>ocean</td>
<td>sea</td>
</tr>
<tr>
<td>air</td>
<td>sea [NATURAL THING]</td>
</tr>
</tbody>
</table>
game [ACTION]

stove
hurricane
meeting [EVENT]

street
mosque
degree [AMOUNT]

moon
metal
colour [PROPERTY]

collar
towel
truce [EVENT]

floor
armchair
invitation [ACTION]

wrath
notion
extent [AMOUNT]

breeze
ozone
kilo [AMOUNT]

ass
tROUT
shape [PROPERTY]

foot
shampoo
wealth [PROPERTY]

dagger
toaster
warmth [PROPERTY]

digit
hour
creation [ACTION]

rat

sea

fire
fire [NATURAL THING]
fire

road
road [OUTDOOR THING]
road

star
star [NATURAL THING]
star

shirt
shirt [INDOOR THING]
shirt

ceiling
ceiling [INDOOR THING]
ceiling

anger
anger [ABSTRACT THING]
anger

wind
wind [NATURAL THING]
wind

donkey
donkey [ANIMAL]
donkey

shoe
shoe [INDOOR THING]
shoe

knife
knife [TOOL]
knife

number
number [MEASURE]
number

mouse
butterfly
calorie [AMOUNT]
time
bathtub
disease [EVENT]
chisel
comb
tuition [EVENT]
capsule
juice
kindness [PROPERTY]
barley
coffee
style [PROPERTY]
beach
sun
lesson [EVENT]
cud
salmon
nightmare [EVENT]
bee
gin
misery [PROPERTY]
carpet
doctor
politics [ACTION]
cheek
beak
litre [AMOUNT]
silver
marble
performance [EVENT]
nostril
snout
aid [ACTION]
hatred

mouse [ANIMAL]
mouse
clock
clock [INDOOR THING]
clock
hammer
hammer [TOOL]
hammer
pill
pill [EDIBLE THING]
pill
oats
oats [FOOD]
oats
sand
sand [OUTDOOR THING]
sand
cow
cow [ANIMAL]
cow
honey
honey [FOOD]
honey
rug
rug [INDOOR THING]
rug
face
face [BODY PART]
face
gold
gold [PRECIOUS THING]
gold
nose
nose [BODY PART]
nose
love
intellect
quantity [AMOUNT]

igloo
dancer
minimum [AMOUNT]

image
perfume
poverty [PROPERTY]

mattress
attic
amount [AMOUNT]

love [ABSTRACT THING]
love
eskimo
eskimo [HUMAN]
eskimo

picture
picture [PERCEPTUAL OBJECT]
picture

bed
bed [INDOOR THING]
bed
APPENDIX C: WORD-NONWORD PAIRS USED IN EXPERIMENT 1

1. PSEUDOHOMOPHONE TRIALS

<table>
<thead>
<tr>
<th>Primes</th>
<th>Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td>nail</td>
<td>cote (coat)</td>
</tr>
<tr>
<td>coin</td>
<td>rouf (roof)</td>
</tr>
<tr>
<td>throat</td>
<td>kea (key)</td>
</tr>
<tr>
<td>creed</td>
<td>offise (office)</td>
</tr>
<tr>
<td>tooth</td>
<td>cley (clay)</td>
</tr>
<tr>
<td>research</td>
<td>laryer (lawyer)</td>
</tr>
<tr>
<td>humour</td>
<td>figger (figure)</td>
</tr>
<tr>
<td>theft</td>
<td>floot (flute)</td>
</tr>
<tr>
<td>comfort</td>
<td>hoatel (hotel)</td>
</tr>
<tr>
<td>health</td>
<td>lettice (lettuce)</td>
</tr>
<tr>
<td>item</td>
<td>musium (museum)</td>
</tr>
<tr>
<td>maximum</td>
<td>chicken (chicken)</td>
</tr>
<tr>
<td>session</td>
<td>bentch (bench)</td>
</tr>
<tr>
<td>universe</td>
<td>wawl (wall)</td>
</tr>
<tr>
<td>festival</td>
<td>joory (jury)</td>
</tr>
<tr>
<td>skill</td>
<td>bair (bear)</td>
</tr>
<tr>
<td>contest</td>
<td>lighbrary (library)</td>
</tr>
<tr>
<td>crisis</td>
<td>stoan (stone)</td>
</tr>
</tbody>
</table>

2. NONPSEUDOHOMOPHONE TRIALS

<table>
<thead>
<tr>
<th>Primes</th>
<th>Targets</th>
</tr>
</thead>
</table>
shoulders parents (parents)
product stex (sex)
worker shir (ship)
glory fungut (fungus)
mind liggage (luggage)
rice murd (mud)
inquiry lond (pond)
badge ralt (malt)
response custlomer (customer)
defense tice (ice)
weakness poncil (pencil)
volume virlin (virgin)
agreement oxmygen (oxygen)
weather ven (van)
election horest (forest)
hue iteal (ideal)
speed arfist (artist)
chest gratt (grass)
APPENDIX D: Instructions for Experiment 2

To mark the passages of text that appeared on the screen at one time, I inserted '/' in the following text.

"Please read these instructions carefully.

On each trial, you will first see a cross for a short moment. Then a short sentence of four words is presented. Each word flashes up only very briefly, except for the last word which remains on the screen.

Sometimes the last word is a correct English word, but sometimes it is a word that does not exist in English.

Your task is to read the sentence and to decide as fast and as accurately as possible whether or not the last word is a word that exists in English. /

To make this decision, please press either the Y-key or the N-key on the response boxes which you find next to the keyboard.

Please use one hand for the Y-box, and one for the N-box.

If you accidentally pressed the wrong button, just press the other one immediately afterwards.

After you have pressed a key, there will be a blank screen for a few moments, and then you'll see the cross again, followed by more sentences. /

First there will be some practise trials so that you can get used to the task.

If you have any questions, please check with the Experimenter before you carry on.

Please respond as fast as possible. Here we go!"
APPENDIX E: Stimuli for Experiment 2

Note: first line of each pair: positive fusion
second line of each pair: misfusion

the semantic categories of nouns are given in block capitals

Her curtains were ironed. SOLID THING
Her chicken were ironed. ANIMATE THING

His leg was amputated. BODY PART
His lotion was amputated. LIQUID THING

The rabbit was marinated. FOOD
The actor was marinated. HUMAN

The cat was bitten. ANIMATE THING
The lake was bitten. LIQUID THING

Her sausages were boiled. FOOD
Her pianos were boiled. SOLID THING

The bananas were harvested. FRUIT
The assistants were harvested. HUMAN

The opinion was suppressed. ABSTRACT THING
The onion was suppressed. VEGETABLE

His coats were soaked. SOLID THING
His ideas were soaked. ABSTRACT THING

The judges were sacked. HUMAN
The windows were sacked. INANIMATE THING

Her hand was kissed. BODY PART
Her thought was kissed. ABSTRACT THING

The pope was ordained. HUMAN
The pea was ordained. VEGETABLE

His passengers were fined. HUMAN
His muscles were fined. BODY PART

The beans were germinated. VEGETABLE
The speakers were germinated. HUMAN
His neck was massaged.  BODY PART
His street was massaged.  INANIMATE THING

Her poems were intoned.  ABSTRACT THING
Her brothers were intoned.  HUMAN

Her hair was perfumed.  BODY PART
Her advice was perfumed.  ABSTRACT THING

His bed was upholstered.  SOLID THING
His visitor was upholstered.  HUMAN

The sticks were sharpened.  SOLID THING
The birds were sharpened.  ANIMAL

The princess was interviewed.  HUMAN
The finger was interviewed.  BODY PART

The skirt was knitted.  SOLID THING
The prince was knitted.  HUMAN

The tiger was startled.  ANIMAL
The butter was startled.  FOOD

The bankers were sued.  HUMAN
The mangos were sued.  FRUIT

Her rooms were rented.  INANIMATE THING
Her grandparents were rented.  HUMAN

The mutton was cut.  SOLID THING
The water was cut.  LIQUID THING

The soup was sipped.  LIQUID THING
The duke was sipped.  HUMAN

The seeds were husked.  FRUIT
The girls were husked.  HUMAN

The liquid was thickened.  LIQUID THING
The teacher was thickened.  HUMAN

The brandy was swigged.  LIQUID THING
The camel was swigged.  ANIMAL

The peaches were skinned.  FRUIT
The butlers were skinned.  HUMAN
Her husband was strangled. HUMAN
Her garden was strangled. INANIMATE THING

The vegetables were sown. PLANT
The sailors were sown. HUMAN

The gate was oiled. SOLID THING
The milk was oiled. LIQUID THING

The leaders were betrayed. HUMAN
The apples were betrayed. FRUIT

The coffee was spilled. LIQUID THING
The writer was spilled. HUMAN

Her children were teased. HUMAN
Her potatoes were teased. VEGETABLE

The turkey was gutted. ANIMAL
The juice was gutted. LIQUID THING

The woman was baptised. HUMAN
The orange was baptised. FRUIT

The priest was interrogated. HUMAN
The car was interrogated. INANIMATE THING

The ducks were slaughtered. ANIMAL
The bricks were slaughtered. INANIMATE THING

The town was besieged. INANIMATE THING
The fish was besieged. ANIMAL

The singer was arrested. HUMAN
The stomach was arrested. BODY PART

Her doctor was provoked. HUMAN
Her basket was provoked. INANIMATE THING

The soldier was scolded. HUMAN
The nose was scolded. BODY PART

The carpets were woven. INANIMATE THING
The dolphins were woven. ANIMAL

The player was tortured. HUMAN
The kitchen was tortured. INANIMATE THING
Her ink was poured.
Her knee was poured.
The worker was punished.
The snow was punished.
His fields were fertilised.
His shoes were fertilised.
His questions were solved.
His tables were solved.
The thigh was injected.
The moon was injected.
The friend was imprisoned.
The road was imprisoned.
The lemon was peeled.
The mill was peeled.
Her tea was gulped.
Her sister was gulped.
His shops were looted.
His eyes were looted.
The student was murdered.
The honey was murdered.
The sheet was knotted.
The boy was knotted.
The building was thatched.
The sparrow was thatched.
The guest was raped.
The salt was raped.
His floors were marbled.
His neighbours were marbled.
His horses were killed.
His books were killed.
His wrist was hurt.
His house was hurt.
The elephants were wounded. ANIMATE THING
The diamonds were wounded. INANIMATE THING

The egg was hatched. FOOD
The queen was hatched. HUMAN
APPENDIX F: NONWORD TRIALS USED IN EXPERIMENT 2

1. Pseudohomophones

His pigeons were faught.
His radios were woshed.
The cucumbers were postpoaned.
Her bread was annalised.
Her mother was cheeted.
His dogs were convurted.
The oil was bilt.
The mountains were condowned.
The heroes were desined.
The colonel was gleigned.
The grapes were waighed.
The sun was repeated.
Her wine was birnt.
Her baby was choasen.
His foot was acheeved.
The people were compeared.
Her toys were poasted.
The hospitals were conseled.
His enemies were denighed.
Her desk was rooled.
Her cake was beleaved.
His skin was is beeten.
The pig was confurmed.
The next is foalded.
The next is chaised.
The next is hird.
The next is feered.
The next is coambed.
The next is sean.
The next is seezed.
The next is diagnoased.
The next is treigned.

2. Non-pseudohomophones

The professors were enbloid.
The robins were acalised.
The strangers were heemed.
The president was disrised.
His answers were instroid.
His uncles were gierved.
The government was knooded.
The captains were crimbed.
The sea was wershed.
Her bottles were taigned.
The air was rifered.
His pictures were snotched.
Her arm was repheamed.
His castle was liffited.
The hills were pleered.
His cherries were preburred.
Her bags were snoaken.
His heart was remoulved.
His wife was clilt.
The cabbage was presarred.
The next is hutten.
The next is misvurted.
The next is pean.
The next is leaten.
The next is fimured.
The next is blean.
The next is strooned.
The next is admoored.
The next is threen.
The next is smollowed.
The next is woot.
The next is snoolen.
APPENDIX G: INSTRUCTIONS FOR WRITTEN CLOZE TASK USED AS A PRETEST FOR EXPERIMENT 2

Dear Subject,

On these pages there is a list of beginnings of sentences. Please think of verbs in the passive to complete each sentence. Each verb should be a reasonable completion of the sentence.

For example, if the beginning of the sentence was: "The parcel was .....", completions might be 'mailed', or 'weighed', or 'sent'.

Or, if the beginning was "The mouse was ...." , completions might be 'caught', or 'killed', or 'chased'.

Work quickly. We want the first verbs that you can think of that fit well into the sentence. Please write down at least two completions for each sentence.

Remember that you have to write down verbs in the passive but no adjectives (for example, don't write: "The mouse was small").

Don’t think too hard about your responses, just write down what comes to your mind. If you have any questions about the task, please ask the experimenter for clarification.

Thank you!
APPENDIX H: STIMULI USED IN EXPERIMENTS 4 AND 4B

Stimuli of Experiment 4 are given first, followed, in square brackets, by the changes (if there were any) that were made to create the stimuli for Experiment 4b

Note: in Experiment 4b, an X preceding a word indicates that the sentence becomes ungrammatical at that point; the subject does NOT see the X; the sentence does not continue after the X-word

1.
In the cooking programme on TV the cook boiled some sausages [boiled some rice] and later prepared a vegetable stew [prepared an Xvegetable]

In the cooking programme on TV the cook boiled some pianos and later prepared a vegetable stew [boiled some Xpianos]

In the cooking programme on TV the cook boiled some beer and later prepared a vegetable stew [prepared an Xvegetable]

[In the cooking programme on TV the cook boiled some Xthese]

2.
The farm workers had harvested the bananas but they did not receive their share of the profits from the sales [they did not receive their Xhis]

The farm workers had harvested the assistants but they did not receive their share of the profits from the sales [had harvested the Xassistants]

The farm workers had harvested the cannabis but they did not receive their share of the profits from the sales [they did not receive their Xhis]

[The farm workers had harvested the Xwhich]

3.
In his workshop the craftsman upholstered old beds [old sofas] and repaired broken chairs [which he Xsolds]

In his workshop the craftsman upholstered old visitors and repaired broken chairs [upholstered old Xvisitors]

In his workshop the craftsman upholstered old shelves and repaired broken chairs [which he Xsolds]

[In his workshop the craftsman upholstered old Xour]
4. In the park the young kids sharpened the sticks and poked holes into the ground
poked holes into Xand

In the park the young kids sharpened the birds and poked holes into the ground
[sharpened the Xbirds]

In the park the young kids sharpened the pencil and poked holes into the ground
[poked holes into Xand]

[In the park the young kids sharpened the
Xunder]

5. Following an invitation from Buckingham Palace the senior journalist interviewed the
princess about her future plans [Xouting]

Following an invitation from Buckingham Palace the senior journalist interviewed the
finger about the future plans [interviewed the Xfinger]

Following an invitation from Buckingham Palace the senior journalist interviewed the
cleaners about their future plans [Xouting]

[Following an invitation from Buckingham Palace the senior journalist interviewed the
Xthose]

6. During the winter evenings the old women knitted many skirts [knitted beautiful
sweaters] and told a lot of stories [and told a lot of Xinto]

During the winter evenings the old women knitted many princes and told a lot of
stories [knitted beautiful Xprinces]

During the winter evenings the old women knitted many [knitted beautiful] bikinis and
told a lot of stories [and told a lot of Xinto]

[During the winter evenings the old women knitted beautiful Xsome]

7. The visitors to the zoo startled the tigers and disturbed many of the rare animals
[disturbed many of the other Xeats]

The visitors to the zoo startled the butter and disturbed many of the rare animals
[startled the Xbutter]

The visitors to the zoo startled the goldfish and disturbed many of the rare animals
[and disturbed many of the other Xeats]

[The visitors to the zoo startled the Xeats]

8. At the State dinner the waiters cut [sliced] the mutton and served all courses in a very unobtrusive way [served all courses in a very Xinto]

At the State dinner the waiters cut the water and served all courses in a very unobtrusive way [sliced the Xwater]

At the State dinner the waiters cut [sliced] the hamburgers and served all courses in a very unobtrusive way [served all courses in a very Xinto]

[At the State dinner the waiters sliced the Xwhere]

9. In the laboratory the chemist thickened the liquid and searched in the cupboard for the right container [searched in the cupboard Xpicture]

In the laboratory the chemist thickened the teacher and searched in the cupboard for the right container [thickened the Xteacher]

In the laboratory the chemist thickened the gravy and searched in the cupboard for the right container [searched in the cupboard Xpicture]

[In the laboratory the chemist thickened the Xtoo]

10. After the revolution the soldiers betrayed the leaders and put up a lot of resistance [and put up a lot Xthese]

After the revolution the soldiers betrayed the apples and put up a lot of resistance [betrayed the Xapples]

After the revolution the soldiers betrayed the child and put up a lot of resistance [and put up a lot Xthese]

[After the revolution the soldiers betrayed the Xbecame]

11. In the shaking aeroplane the stewardess spilled some coffee and had trouble keeping her balance [keeping her Xthey]
In the shaking aeroplane the stewardess spilled the writer and had trouble keeping her balance [spilled the Xwriter]

In the shaking aeroplane the stewardess spilled some petrol and had trouble keeping her balance [keeping her Xthey]

[In the shaking aeroplane the stewardess spilled some Xwithout]

12. Before the festival the farmers slaughtered many pigs and the bakers made lots of delicious cakes [made lots of Xfrom]

Before the festival the farmers slaughtered many bricks and the bakers made lots of delicious cakes [slaughtered many Xbricks]

Before the festival the farmers slaughtered many pets and the bakers made lots of delicious cakes [made lots of Xfrom]

[Before the festival the farmers slaughtered many Xhaving]

13. British farmers fertilize their fields with many chemicals which might be bad for the environment [which might be bad Xthe]

British farmers fertilize their shoes with many chemicals which might be bad for the environment [fertilize their Xshoes]

British farmers fertilize their mangoes with many chemicals which might be bad for the environment [which might be bad Xthe]

[British farmers fertilize their Xwere]

14. The architect decided to thatch the building in keeping with the character of the street [in keeping with the character Xstreet]

The architect decided to thatch the sparrow in keeping with the character of the street [decided to thatch the Xsparrow]

The architect decided to thatch the mosque in keeping with the character of the street [in keeping with the character Xstreet]

[The architect decided to thatch the Xgoes]
15. The schoolboy reported that the angry dog bit the cat and tried to chase the chicken [and tried to Xinto]  
The schoolboy reported that the angry dog bit the lake and tried to chase the chicken [bit the Xlake]  
The schoolboy reported that the hungry [angry] dog bit the ant and tried to chase the chicken [and tried to Xinto]  
[The schoolboy reported that the angry dog bit the Xbetween]

Sentence sets 16 to 20 were only used in Experiment 4b:

16. Immediately after the corruption scandal the authorities sacked the judges and reorganized the whole legal system which Xwhy  
Immediately after the corruption scandal the authorities sacked the Xwindows  
Immediately after the corruption scandal the authorities sacked the tramps and reorganized the whole legal system which Xwhy  
Immediately after the corruption scandal the authorities sacked the Xbetween

17. In the final scene of the film the hero kissed her hand and promised that he would never Xsang  
In the final scene of the film the hero kissed her Xhope  
In the final scene of the film the hero kissed her bag and promised that he would never Xsang  
In the final scene of the film the hero kissed her Xtheir

18. When helping his mother in the kitchen the boy broke the teapot and started to cry because Xstupid  
When helping his mother in the kitchen the boy broke the Xpuddle  
When helping his mother in the kitchen the boy broke the bicycle and started to cry because Xstupid
When helping his mother in the kitchen the boy broke the Xwhose

19. Nobody knew why the child had strangled the cat even though a number of theories were Xwould

Nobody knew why the child had strangled the Xgarden

Nobody knew why the child had strangled the cow even though a number of theories were Xwould

Nobody knew why the child had strangled the Xwhy

20. The persistent rain had soaked his coat and spoilt the walk which he Xwhich

The persistent rain had soaked his Xideas

The persistent rain had soaked his ear and spoilt the walk which he Xwhich

The persistent rain had soaked his Xalso
APPENDIX I: STIMULI FOR EXPERIMENTS 5 AND 6

Stimuli used in Experiment 5 are given first, followed, in square brackets, by the changes (if there were any) that were made to create stimuli of Experiment 6

Note: in Experiment 6, an X preceding a word indicates that the sentence becomes ungrammatical at that point; the subject did NOT see the X. The sentence does not continue after the X-word.

1. After the guide had ended the walk was welcomed by everybody.
   After the guide had ended in the afternoon the walk was welcomed by everybody.
   After the downpour had ended the walk was welcomed by everybody.
   After the downpour had ended in the afternoon the walk was welcomed by everybody.

2. Before the butcher opened the door was jammed by the great number of people who tried to get in [by the great number of X where]
   Before the butcher opened in the evening the door was jammed by the great number of people who tried to get in [by the great number of X where]
   Before the theatre opened its doors were jammed by the great number of people who tried to get in [by the great number of X where]
   Before the theatre opened in the evening its doors were jammed by the great number of people who tried to get in [by the great number of X where]

3. When the toddler shifted the bricks were scattered everywhere [and they were in everybody’s way].
   When the toddler shifted along the path the bricks were scattered everywhere [and they were in everybody’s way].
   When the van shifted the bricks were scattered everywhere [and they were in everybody’s way].
   When the van shifted along the path the bricks were scattered everywhere [and they were in everybody’s way].
4. While the skipper was drying the windscreen was being cleaned [thoroughly].

While the skipper was drying in the sun the windscreen was being cleaned [thoroughly].

While the wipers were drying the windscreen was being cleaned [thoroughly].

While the wipers were drying in the sun the windscreen was being cleaned [thoroughly].

5. Although the partners split up the property remained in joint ownership.

Although the partners split up after a quarrel the property remained in joint ownership.

Although the company split up the property remained in joint ownership.

Although the company split up after a quarrel the property remained in joint ownership.

6. While his friend was warming up the meat was defrosting on the shelf.

While his friend was warming up in the kitchen the meat was defrosting on the shelf.

While the oven was warming up the meat was defrosting on the shelf.

While the oven was warming up in the kitchen the meat was defrosting on the shelf.

7. Before the child turned the wheel came loose and frightened everybody.

Before the child turned very rapidly the wheel came loose and frightened everybody.

Before the motor turned the wheel came loose and frightened everybody.

Before the motor turned very rapidly the wheel came loose and frightened everybody.

8. After Dame Kiri concluded the concert was praised by everybody.

After Dame Kiri concluded triumphantly the concert was praised by everybody.
After the duet concluded the concert was praised by everybody.

After the duet concluded triumphantly the concert was praised by everybody.

9.
While the lad was swinging the tyre fell to the ground.

While the lad was swinging in the air the tyre fell to the ground.

While the crane was swinging the tyre fell to the ground.

While the crane was swinging in the air the tyre fell to the ground.

10.
Although the cheerful clowns had changed the mood remained the same for a while.

Although the cheerful clowns had changed beyond recognition the mood remained the same for a while.

Although the cheerful campaign had changed the mood remained the same for a while.

Although the cheerful campaign had changed beyond recognition the mood remained the same for a while.

11.
Before the captain sailed the Atlantic ocean seemed very calm [and the weather was good].

Before the captain sailed to America the Atlantic ocean seemed very calm [and the weather was good].

Before the boat sailed the Atlantic ocean seemed very calm [and the weather was good].

Before the boat sailed to America the Atlantic ocean seemed very calm [and the weather was good].

12.
When the tiger started the panic spread quickly around the town.

When the tiger started so violently the panic spread quickly around the town.

When the noise started the panic spread quickly around the town.
When the noise started so violently the panic spread quickly around the town.

13. Although the giant was shaking the wall was not damaged at all.

As the man moved the cup fell to the floor and broke.

As the man moved carelessly the cup fell to the floor and broke.

As his hand moved the cup fell to the floor and broke.

As his hand moved carelessly the cup fell to the floor and broke.

15. As the woman sank the lifeboat was on its way back to the harbour [to the Xwithin]

As the woman sank in the sea [sank beneath the waves] the lifeboat was on its way back to the harbour [to the Xwithin]

As the rock sank the lifeboat was on its way back to the harbour [to the Xwithin]

As the rock sank in the sea [sank beneath the waves] the lifeboat was on its way back to the harbour [the Xwithin]

16. Because the boy was burning the pots were quickly filled to extinguish the fire [the Xthese]

Because the boy was burning fiercely the pots were quickly filled to extinguish the fire [the Xthese]

Because the kiln was burning the pots were quickly filled to extinguish the fire [the Xthese]

Because the kiln was burning fiercely the pots were quickly filled to extinguish the fire [the Xthese]
17.
Because the girl was growing lots of fruit was consumed in the house.
Because the girl was growing that summer lots of fruit was consumed in the house.
Because the tree was growing lots of fruit was consumed in the house.
Because the tree was growing that summer lots of fruit was consumed in the house.

18.
After the clumsy maid had circulated the dust took a long time to settle.
After the clumsy maid had circulated through the room the dust took a long time to settle.
After the continuous draught had circulated the dust took a long time to settle.
After the continuous draught had circulated through the room the dust took a long time to settle.

19.
Although the delivery boy swirled the newspapers did not fly away.
Although the delivery boy swirled in the streets the newspapers did not fly away.
Although the fierce tempest swirled the newspapers did not fly away.
Although the fierce tempest swirled in the streets the newspapers did not fly away.

20.
Although the farmer was running the mill burnt to the ground.
Although the farmer was running fast the mill burnt to the ground.
Although the water was running the mill burnt to the ground.
Although the water was running fast the mill burnt to the ground.

21.
When the flock decreased the fertile land became more valuable than before [than before X than]
When the flock decreased in size the fertile land became more valuable than before [than before X than]
When the landslides decreased the fertile land became more valuable than before [than
before X than]

When the landslides decreased in size the fertile land became more valuable than
before [than before X than]

22.
Although the driver bounced the precious goods reached their destination on time.

Although the driver bounced on the road the precious goods reached their destination
on time.

Although the lorry bounced the precious goods reached their destination on time.

Although the lorry bounced on the road the precious goods reached their destination on
time.

23.
Before the technicians [When the technicians] had rotated the propellers were cleaned
thoroughly in the factory.

Before the technicians [When the technicians] had rotated according to plan the
propellers were cleaned thoroughly in the factory.

Before the engine [When the engine] had rotated the propellers were cleaned
thoroughly in the factory.

Before the engine [When the engine] had rotated according to plan the propellers were
cleaned thoroughly in the factory.

24.
Although the tourist was boiling the eggs were very cold [and needed reheating].

Although the tourist was boiling in the heat the eggs were very cold [and needed
reheating].

Although the water was boiling the eggs were very cold [and needed reheating].

Although the water was boiling in the heat the eggs were very cold [and needed
reheating].
APPENDIX J: STIMULI FOR EXPERIMENT 7

Note: an X preceding a word indicates that the sentence becomes ungrammatical at that point. Subjects did NOT see the X. In each sentence quartet, the first sentence is of the normal ambiguous condition, the second of the semantically-violated ambiguous condition, the third of the normal nonambiguous condition, and the fourth of the semantically-violated nonambiguous condition.

1. After the teacher had finished the adventurous walk was welcomed by everybody in the school which Xwhy

   After the teacher had finished the adventurous boy was welcomed by everybody in the school which Xwhy

   After the teacher had finished at lunchtime the adventurous walk was welcomed by everybody in the school which Xwhy

   After the teacher had finished at lunchtime the adventurous boy was welcomed by everybody in the school which Xwhy

2. Before the butcher opened the new shop had been thoroughly cleaned by his assistant.

   Before the butcher opened the new path had been thoroughly cleaned by his assistant.

   Before the butcher opened in the morning the new shop had been thoroughly cleaned by his assistant.

   Before the butcher opened in the morning the new path had been thoroughly cleaned by his assistant.

3. Before we had dried our damp clothes had really begun to get on our nerves.

   Before we had dried the strong rain had really begun get on our nerves.

   Before we had dried in front of the fire our damp clothes had really begun to get on our nerves.

   Before we had dried in front of the fire the strong rain had really begun to get on our nerves.

4.
Although the partners split up their shared property was enough reason to help them come back together again.

Although the partners split up their shared history was enough reason to help them come back together again.

Although the partners split up after a quarrel their shared property was enough reason to help them come back together again.

Although the partners split up after a quarrel their shared history was enough reason to help them come back together again.

5.
When his friend had warmed up the meal was a real success as had been expected.

When his friend had warmed up the night was a real success as had been expected.

When his friend had warmed up after a while the meal was a real success as had been expected.

When his friend had warmed up after a while the night was a real success as had been expected.

6.
As the man turned the enormous boat took everybody completely by surprise.

As the man turned the enormous noise took everybody completely by surprise.

As the man turned into the harbour the enormous boat took everybody completely by surprise.

As the man turned into the harbour the enormous noise took everybody completely by surprise.

7.
After Dame Kiri had ended the fantastic concert was highly praised by the organisers.

After Dame Kiri had ended the fantastic audience was highly praised by the organisers.

After Dame Kiri had ended triumphantly the fantastic concert was highly praised by the organisers.

After Dame Kiri had ended triumphantly the fantastic audience was highly praised by the organisers.
8. While the girl was swinging her arms were all over the place.
   While the girl was swinging her thoughts were all over the place.
   While the girl was swinging in the playground her arms were all over the place.
   While the girl was swinging in the playground her thoughts were all over the place.

9. Although the cheerful clowns had changed the gloomy mood remained the same for the rest of the day.
   Although the cheerful clowns had changed the gloomy sky remained the same for the rest of the day.
   Although the cheerful clowns had changed beyond recognition the gloomy mood remained the same for the rest of the day.
   Although the cheerful clowns had changed beyond recognition the gloomy sky remained the same for the rest of the day.

10. Before the captain sailed the Atlantic ocean seemed very pleasant to him and the
    Before the captain sailed the Atlantic climate seemed very pleasant to him and the
    Before the captain sailed to America the Atlantic ocean seemed very pleasant to him and the
    Before the captain sailed to America the Atlantic climate seemed very pleasant to him and the

11. When the workmen started the very frightening noise forced everybody to seek refuge.
    When the workmen started the very frightening storm forced everybody to seek refuge.
    When the workmen started early in the morning the very frightening noise forced everybody to seek refuge.
When the workmen started early in the morning the very frightening storm forced everybody to seek refuge.

12. Although the furious elephant was shaking the strong trees hindered each one of his aggressive attacks.

Although the furious elephant was shaking the strong wind hindered each one of his aggressive attacks.

Although the furious elephant was shaking with rage the strong trees hindered each one of his aggressive attacks.

Although the furious elephant was shaking with rage the strong wind hindered each one of his aggressive attacks.

13. As the children moved their new toys got very badly damaged by accident.

As the children moved the new lawn got very badly damaged by accident.

As the children moved carelessly their new toys got very badly damaged by accident.

As the children moved carelessly the new lawn got very badly damaged by accident.

14. After the pirates sank the armoured vessel moved on fast and soon reached the mainland.

After the pirates sank the stormy weather moved on fast and soon reached the mainland.

After the pirates sank without a struggle the armoured vessel moved on fast and soon reached the mainland.

After the pirates sank without a struggle the stormy weather moved on fast and soon reached the mainland.

15. Although the tourist was boiling the water remained surprisingly cool for a long time Xhours

Although the tourist was boiling the house remained surprisingly cool for a long time Xhours
Although the tourist was boiling in the summer heat the water remained surprisingly cool for a long time X hours

Although the tourist was boiling in the summer heat the house remained surprisingly cool for a long time X hours

16.
Before the king rides his beautiful big horse is always looked after by his staff.

Before the king rides his beautiful big park is always looked after by his staff.

Before the king rides in the morning his beautiful big horse is always looked after by his staff.

Before the king rides in the morning his beautiful big park is always looked after by his staff.

17.
Wherever Alice walks her funny old dog attracts a lot of attention.

Wherever Alice walks her funny old hat attracts a lot of attention.

Wherever Alice walks in the city her funny old dog attracts a lot of attention.

Wherever Alice walks in the city her funny old hat attracts a lot of attention.

18.
When the salesman was driving the brand new car struck an old lady and hurt her slightly.

When the salesman was driving the brand new tyre struck an old lady and hurt her slightly.

When the salesman was driving down the road the brand new car struck an old lady and hurt her slightly.

When the salesman was driving down the road the brand new tyre struck an old lady and hurt her slightly.

19.
While the fire burnt a huge amount of petrol escaped from several big tanks.

While the fire burnt a huge amount of water escaped from several big tanks.
While the fire burnt ferociously a huge amount of petrol escaped from several big tanks.

While the fire burnt ferociously a huge amount of water escaped from several big tanks.

20.
As the girl had grown her beautiful black hair suited her better than ever before
As the girl had grown her beautiful black dress suited her better than ever before
As the girl had grown fast in the past few months her beautiful black hair suited her better than ever before
As the girl had grown fast in the past few months her beautiful black dress suited her better than ever before

21.
Although the couple had broken up their family continued to be their joint responsibility.
Although the couple had broken up their children continued to be their joint responsibility.
Although the couple had broken up after several years their family continued to be their joint responsibility.
Although the couple had broken up after several years their children continued to be their joint responsibility.

22.
As the pilot flew the shaky unreliable aeroplane caused great anxiety amongst everybody.
As the pilot flew the shaky unreliable stewardess caused great anxiety amongst everybody.
As the pilot flew during the storm the shaky unreliable aeroplane caused great anxiety amongst everybody.
As the pilot flew during the storm the shaky unreliable stewardess caused great anxiety amongst everybody.
23. Although the baby was developing an awful chronic disease caused him great discomfort and suffering.

Although the baby was developing an awful chronic famine caused him great discomfort and suffering.

Although the baby was developing quite normally an awful chronic disease caused him great discomfort and suffering.

Although the baby was developing quite normally an awful chronic famine caused him great discomfort and suffering.

24. After the head-teacher had returned the first important report was widely publicized in the school magazine.

After the head-teacher had returned the first important meeting was widely publicized in the school magazine.

After the head-teacher had returned from holiday the first important report was widely publicized in the school magazine.

After the head-teacher had returned from holiday the first important meeting was widely publicized in the school magazine.