An Evaluation of a Cognitive Architecture of Sentence Processing: Computational and Empirical Assessment

Umesh Patil

Doctoral dissertation submitted to the Faculty of Human Sciences at the
University of Potsdam
in partial fulfillment of the requirements for the degree
Doctor of Philosophy in Cognitive Science.

January, 2012

Supervisor: Prof. Dr. Shravan Vasishth

To my parents

Acknowledgements

I am deeply thankful to many people for their contribution in various ways during the course of this thesis. First, I would like to thank my supervisor, Shravan Vasishth, for his continuing support ever since I landed in Berlin. Shravan introduced me to the different aspects of pyscholinguistics, and of research in general. Due to his constant motivation and guidance, my early years as a researcher felt less uncertain. Despite his numerous commitments, he was always accessible. I greatly appreciate his persistent efforts in pushing me to better myself.

I would also like to thank Richard Lewis for all the educative discussions over the years. If Rick's suggestions were illuminating, his criticisms were thought provoking, and left me ruminating for the next couple of weeks.

The various research meetings that I attended, especially AMLaP (2007), CUNY (2008, 2009, 2011), the workshop on processing head-final languages (Leipzig, 2007), and the Ph.D. Colloquiums in the Department of Linguistics, helped me a great deal in maturing as a researcher. I'm grateful to the DFG, German Research Foundation, for providing financial support during the course of my dissertation. I would also like to thank my dissertation committee for generously agreeing to participate.

Next, I would like to express my gratitude to my academic colleagues, and friends: Enkela Begu, Juliane Böhme, Marisa Ferrara Boston, Frank Burchert, Heiner Drenhaus, Felix Engelmann, Susanne Genzel, Sabrina Gerth, Anja Gollrad, Sandra Hanne, Robin Hörnig, Esther Sommerfeld, and Titus von der Malsburg. While academic discussions with them helped to sharpen my understanding of the field, the casual conversations helped to refresh me.

Special thanks go to Mihaela Arambasa, Gerrit Kentner, and Pavel Logačev. The time I spent with them always cheered me up and boosted my morale. Conversations with Mihaela were very encouraging and doubly entertaining. Gerrit's eternally positive attitude and promptness to help without needing to be asked, has always amazed me. Pavel's involvement in every aspect of my life over the last few years has been significant. Professionally, he has been very inspiring, and personally, highly supportive.

The contribution of my dear friend and companion, Rukshin Shaher, has been enormous. Her company made it easier to get through difficult times. She painstakingly proofread every draft of the thesis which has helped immensely in improving the quality of the text.

Finally, I owe a debt of gratitude to my brother, sister-in-law and parents. Without their sustained moral support, love and encouragement, this thesis would not have been possible. I dedicate the thesis to my parents.

Abstract

This thesis evaluates a cognitive architecture of sentence processing, the cue-based retrieval theory, proposed in Lewis and Vasishth (2005) and Lewis, Vasishth, and Van Dyke (2006). First, we show that the architecture can be employed to investigate research questions in psycholinguistics such as processing reflexives and impaired sentence processing in aphasia. We formalize different hypotheses for these processes in the cue-based retrieval architecture and provide a computational evaluation of the hypotheses. Next, we demonstrate the breadth of the architecture by modeling a diverse set of response measures such as: eye movements in reading, eye movements in the visual world paradigm, accuracy of sentence processing, and response time in the sentence-picture matching task. Finally, we assess the robustness of the predictions of the cue-based retrieval theory using two types of validation methods: a controlled empirical study and a study using a corpus of naturally occurring sentences.

Additionally, through these evaluations, we use the cue-based retrieval architecture to gain insights into two important question in psycholinguistics: (i) What type of information is used to resolve the reflexive-antecedent dependency in English? (ii) What is the source of the sentence processing deficit observed in aphasics—a breakdown in the grammar or processing failures? With a combination of modeling studies and an eye tracking study we show that the process of resolving the reflexive-antecedent dependency uses non-structural information along with structural information. In the case of aphasic sentence processing, a set of modeling studies supports those hypotheses that propose a processing breakdown as the source of the deficit in aphasics.

Contents

Acknowledgements **iv**

Abstract **vi**

1 Introduction & Summary **1**

1.1 Aims of the thesis 2

1.2 Cue-based retrieval architecture 4

1.3 Summary 6

1.3.1 Compound effect of probabilistic disambiguation and memory retrievals on sentence processing: Evidence from an eye-tracking corpus (Ch. 2) 6

1.3.2 Retrieval interference in syntactic processing: The case of reflexive binding in English (Ch. 3) 8

1.3.3 A cue-based retrieval model of offline and online sentence processing in aphasia (Ch. 4) 14

2 Compound effect of probabilistic disambiguation and memory retrievals on sentence processing: Evidence from an eye-tracking corpus **19**

2.1 Abstract 20

2.2 Introduction 20

2.2.1 Surprisal 22

2.2.2 Cue-based theory 23

2.3 Experiment 24
2.3.1 Methods 24
2.3.2 Results & Discussion 27
2.4 Conclusions 30

3 Retrieval interference in syntactic processing: The case of reflexive binding in English 33
3.1 Abstract 34
3.2 Introduction 35
3.3 Evidence for the use of non-structural cues in accessing antecedents of reflexives 37
3.4 The cue-based retrieval model 42
3.4.1 Modeling reflexive binding in the cue-based retrieval framework 45
3.4.2 Parametric variability in the model 54
3.4.3 An alternative explanation for the absence of interference effects (E4 & E5) in Sturt 2003 57
3.4.4 A new design 59
3.5 Eye tracking experiment 62
3.5.1 Participants 62
3.5.2 Design and materials 63
3.5.3 Procedure 63
3.5.4 Data analysis 64
3.5.5 Results 65
3.5.6 Discussion 68
3.6 General discussion 71
3.6.1 Strictly structured access as an alternative 74
3.7 Acknowledgements 80

4 A cue-based retrieval model of offline and online sentence processing in aphasia 81

4.1 Abstract . . . 82
4.2 Introduction . . . 83
4.2.1 Explanations for sentence comprehension deficits . . . 84
4.3 Aims of the study . . . 89
4.4 Existing computational models of aphasic sentence comprehension . . . 90
4.5 Cue-based retrieval theory . . . 94
4.6 Cue-based retrieval model of Hanne et al. 2011 . . . 97
4.6.1 Linking hypothesis: Implementing the sentence-picture matching task in the CBR architecture . . . 99
4.6.2 Details of the model of sentence-picture matching . . . 102
4.6.3 Mapping deficit assumptions to the CBR models . . . 103
4.6.4 Estimation of parameter values . . . 104
4.6.5 Results & Discussion . . . 108
4.7 General Discussion . . . 113
4.8 Conclusions . . . 117
4.9 Acknowledgements . . . 118

5 Conclusions 119

A PSC sentences & incremental parses 121

B Stimuli for the experiment in Chapter 3 157

C Predictions of the models for controls and aphasics 165

References 167

List of Figures

3.1 The effects in Sturt 2003 Experiment 1: (a) proportions of ungrammatical interpretations of reflexives in the follow-up study; (b) first fixation durations in the eye tracking study. The error bars in the plot for first fixation duration show 95% confidence intervals. 52

3.2 The twofold predictions of the cue-based retrieval model for Experiment 1 in Sturt 2003, based on the parameter values listed in Table 3.2. The error bars in the plot for retrieval time show 95% confidence intervals. 52

3.3 The predictions of effects E1–E5 for Experiment 1 from Sturt 2003 across a range of parameter values. Each effect is plotted against the parameter that affects it the most. 56

3.4 The predictions of effects E1–E5 for three experimental designs across a range of parameter values. 60

3.5 The variations in effect E4 (match-interference (RT)) across noise and maximum difference parameters for three experimental designs. 62

3.6 Early effect of *interference*. 69

3.7 Late effect of *match*. 69

3.8 Regression contingent FFD. 71

3.9 The predictions of a strictly structural cue-based model and a model with structural and agreement feature based cues. The predictions are for the new experimental design and are generated across a set of parameter values. 78

4.1 Sample pictures used in the Hanne et al. 2011 sentence-picture matching task. For the canonical sentence *Der Sohn fängt den Vater* ('The son is catching the father'), (a) is the correct picture and (b) is the distractor. 98

4.2 The picture chunks created by the model for the semantic representation of pictures used in the Hanne et al. 2011 sentence-picture matching task. Chunk (a) represents the canonical sentence *Der Sohn fängt den Vater* ('The son is catching the father'), and chunk (b) represents the non-canonical sentence *Den Sohn fängt der Vater* ('The father is catching the son'). . 100

4.3 Data from Hanne et al. 2011; accuracy and response time for the sentence-picture matching task. The error bars in the response time plot show 95% confidence intervals. 105

4.4 Accuracy and response time predictions of the model for controls and the models M1, M2 and M3. The predictions of accuracy are in terms of percentages of retrievals of the correct picture at the end of the sentence. The predictions of response time are equal to the models' processing time starting from the first word of the sentence until the retrieval of a picture at the end of the sentence. The error bars on response time show 95% confidence intervals. 109

4.5 Eye movement data from Hanne et al. 2011 and the predictions of models M1, M2 & M3. The x-axis shows different regions in the sentences. NP1 and NP2 denote the two noun phrases in the sentences. For sentences in canonical word order, they denote subject and object respectively, and for sentences in non-canonical word order they denote object and subject respectively. The y-axis on the left-hand side of each plot shows percentages of fixations observed in the data and the y-axis on the right hand side shows differences in the activations of the two pictures in the models. An activation difference of zero is aligned with the 50% mark in the fixation percentages; the horizontal line represents this mark. 111

List of Tables

2.1 Linear model coefficients, standard errors and t-values for surprisal, attachment time and interaction of attachment time and surprisal. An absolute t-value of 2 or greater indicates statistical signicance at $\alpha = 0.05$. 28
2.2 Linear model coefficients, standard errors and t-values for baseline predictors for TRT. 28
2.3 Linear model coefficients, standard error, z-scores and p-values with FPSKIP as the dependent measure. 29
2.4 Deviance Information Criterion values for simpler model (baseline predictors + surprisal) vs. more complex model (simpler model + attachment time). 29

3.1 The match of retrieval cues with the accessible and inaccessible antecedents for the four conditions in Sturt's experiments (cat=category). 49
3.2 The list of parameter values used in the previous studies with CBR models and the values used in the current model. 51
3.3 The range of parameter values used for testing the parametric variability of the cue-based retrieval models. 55
3.4 Mean reading times at the reflexive with standard errors, percentages of first pass regressions from the reflexive, and comprehension question response accuracies across four conditions. 66

3.5 Linear mixed-effects model estimates, standard errors and t-values across reading time measures; 'Interference:Match' denotes the interaction term for the two fixed effects. 67
3.6 Linear model estimates, standard errors and p-values for FPRP; 'Interference : Match' denotes the interaction term for the two fixed effects. 67

4.1 The sequence of steps followed by the model while processing the sentences in (1). 102
4.2 The list of parameter values used in the previous studies with CBR models and the values used in the current models for controls and aphasics. 106
4.3 The range of values of the parameters *default action time* and *utility noise* tested in the model. 106
4.4 Utility values estimated for the productions that process nouns. 108

C.1 The predictions of sentence-picture matching accuracies in percentages across four models. 165
C.2 The predictions of mean response times in milliseconds across four models. 165
C.3 The predictions of eye movements in terms of the differences in the activation of correct pictures and incorrect pictures across four models. 166

Chapter 1

Introduction & Summary

> *Each new machine that is built is an experiment. Actually constructing the machine poses a question to nature; and we listen for the answer by observing the machine in operation and analyzing it by all analytical and measurement means available. Each new program that is built is an experiment. It poses a question to nature, and its behavior offers clues to an answer. Neither machines nor programs are black boxes; they are artifacts that have been designed, both hardware and software, and we can open them up and look inside. We can relate their structure to their behavior and draw many lessons from a single experiment.*

— Allen Newell and Herbert Simon, ACM Turing Award Lecture (1975)

In the spirit of the insight from Allen Newell and Herbert Simon, over the last few decades, the field of psycholinguistics has investigated various computational theories of sentence processing. Here we explore one such theory proposed by Richard Lewis, Julie Van Dyke and Shravan Vasishth (Lewis & Vasishth, 2005; Lewis et al., 2006). We regard this theory as major advancement in developing a computational theory of language processing, mainly because the theory is rooted in a cognitive architecture, ACT-R, which is a unified theory of cognition in the sense of Newell (1990). This thesis attempts to evaluate the theory from various different aspects and use it to advance our understanding of some of the issues in psycholinguistics. We call the theory a "cue-based retrieval architecture"[1] mainly because its predictions about sentence processing are derived from retrievals at each input word using cues specified by the grammar.

1.1 Aims of the thesis

The aim of the thesis is to evaluate the cue-based retrieval (CBR) architecture by validating its predictions, and consequently demonstrating its breadth as a framework for modeling language processing. In particular, we use the following three measures for evaluating the architecture:

1. **Empirical coverage**
   The experiments reported here model various processes and effects in sentence processing like reflexive binding, processing non-canonical structures and impaired sentence comprehension.

   (i) The process of reflexive binding involves retrieving the correct antecedent from a set of elements present in memory. We evaluate the possible set of factors that influence the antecedent retrieval process for reflexives in English. In contrast to earlier claims (e.g., (Phillips, Wagers, & Lau, 2011)), we show that non-syntactic factors may also be employed in the retrieval process.

   (ii) Non-canonical structures are shown to be are harder to process than canonical structures (Ferreira, 2003). The processing difficulty is reflected in slow response time and increased error processing. We suggest a method of modeling the difficulty in processing non-canonical structures in German using the CBR architecture.

[1]Note, the thesis uses "cue-based retrieval architecture" and "cue-base retrieval theory" interchangeably to refer to the same notion.

(iii) Individuals with aphasia are shown to have difficulty in various aspects of sentence processing, especially when processing reversible non-canonical sentences (Grodzinsky, 2000). Various representational and processing accounts have been proposed to explain the processing deficit. We evaluate two such accounts of impaired sentence processing by formalizing them in the CBR architecture.

2. **Response type**

   We model response measures like eye movements in reading, eye movements in the visual world paradigm, accuracy of processing, and response time in a sentence-picture matching task.

   (i) Eye movement patterns in reading reflect the effects of word level as well as sentence level processes (Clifton, Staub, & Rayner, 2007). Leading models of eye movements typically include factors such as word length, frequency and predictability as predictors of eye movements. We show that the CBR theory is also a significant predictor of various "early" and "late" eye movement measures.

   (ii) Eye movements in the visual world paradigm have been shown to reflect the effects of speech perception, memory, and language processing (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). We propose a linking hypothesis for modeling eye movements in the visual world paradigm in the CBR architecture. Using the hypothesis, we model incremental eye movements in a sentence-picture matching task.

   (iii) We model two types of accuracies in sentence processing—the accuracy of a sentence-picture matching task and the proportion of errors in the grammatical interpretation of reflexive pronouns.

   (iv) We model response time in a sentence-picture matching task.

3. **Validation procedure**
   We validate the predictions of the CBR models in two ways—using a controlled eye tracking study and using an eye tracking corpus.

   (i) We formulate a well-studied problem in psycholinguistics, namely reflexive binding, in the CBR architecture, and generate predictions for a new experimental design and validate the predictions using a controlled eye tracking study.

   (ii) Predictions of psycholinguistic theories are normally validated by comparing processing cost at only selected regions in sentences that typically have complex structures. We evaluate predictions of the CBR theory at each individual word in relatively simple sentences from an eye tracking corpus. This demonstrates the capacity of the CBR theory to accurately capture even minute differences in processing.

In sum, on the one hand, the thesis demonstrates the capabilities of the CBR architecture, and on the other hand, it advances our understanding regarding important questions in sentence processing such as the role of non-syntactic information in reflexive binding and possible explanations for impaired sentence comprehension. The rest of the chapter briefly describes the CBR architecture and then summarizes the following chapters in the thesis.

1.2 Cue-based retrieval architecture

The cue-based retrieval architecture derives its architectural assumptions from ACT-R (Adaptive Control of Thought-Rational), a cognitive architecture developed by John Anderson and colleagues (Anderson, Byrne, Douglass, Lebiere, & Qin, 2004; Anderson, 2007). ACT-R is based on three core principles. First, the memory system is divided into a declarative memory and procedural memory system. Declarative memory holds the contents of long-term memory (semantic and episodic memory), as well as elements created at run-time during processing. The procedural memory system describes procedural knowledge in terms of *production rules*, which are condition–action pairs. Second, the design of the ACT-R architecture is modular—each module is specialized for carrying out a different task. And third, the cognitive processes in ACT-R are assumed to be "rational", such that at each choice point ACT-R chooses the option that has the highest expected probability of succeeding and the lowest cost.

In addition to the symbolic system, the behavior of an ACT-R model is modulated by a set of subsymbolic computations.[2] These computations impose constraints on the retrieval of elements from declarative memory and the selection of production rules at each stage. The constraints on retrieval are specified in terms of activation of the memory elements. The activation value of an element determines the probability of its retrieval and the latency of the retrieval. The frequency, recency and prior pattern of retrievals determine the activation value.

Apart from ACT-R's core assumptions, the CBR theory makes two assumptions for linguistic processing. First, based on psycholinguistic evidence, the parser follows the left–corner parsing algorithm (Aho & Ullman, 1972). The grammatical knowledge and parsing rules (the control structure) are maintained in the procedural memory. And second, the incremental syntactic structures created during parsing are assumed to be X–bar structures (Chomsky, 1986) representing maximal projections with features corresponding to X–bar positions (specifier, complement, head) and other grammatical features such as case marking, and person, number and gender agreement. The lexical knowledge and the incremental syntactic structure created during parsing is stored in declarative memory.

[2] Details of subsymbolic computations are covered in later chapters.

1.3 Summary

1.3.1 Compound effect of probabilistic disambiguation and memory retrievals on sentence processing: Evidence from an eye-tracking corpus (Ch. 2)

In Chapter 2, we evaluate the predictions of the cue-based theory using an eye tracking corpus. The theory has been shown to account for processing difficulty at various points in psycholinguistically challenging structures like double center-embeddings, long distance dependencies and negative/positive polarity items, among others (see Lewis & Vasishth, 2005; Vasishth, Brüssow, Lewis, & Drenhaus, 2008). Here we take up the question: Can the theory account for the processing cost at each individual word in sentences involving undemanding structures? This approach should inform us about how sensitive the predictions of the cue-based theory are.

By employing an eye tracking corpus based approach, earlier studies such as Boston, Hale, Kliegl, Patil, and Vasishth (2008), Demberg and Keller (2008), and Frank (2009) successfully demonstrated that theories like surprisal (Hale, 2001; Levy, 2008), Dependency Locality Theory (Gibson, 2000) (DLT) and entropy reduction hypothesis (Hale, 2003, 2006) are significant predictors of different eye movement measures. Through these studies, information theoretic models of sentence processing, namely surprisal and entropy reduction hypothesis, unequivocally emerged as significant predictors of word-by-word reading times. However, the effects of memory-based models of sentence processing are not uniform. Demberg and Keller (2008) showed that DLT, a processing difficulty metric based on memory storage cost, is a significant predictor of reading times only for verbs and nouns, and, Boston, Hale, Vasishth, and Kliegl (2011) showed that a serial dependency parser implementing cue-based retrievals does not model any eye movement measures. They further showed that for retrieval predictions to account for reading time variance, a parallel parser considering more than one candidate analyses has to be used.

The inability of DLT to be a generic model of sentence processing is justifiable given that it assigns an integration cost only to verbs and nouns. However, it is surprising that retrieval predictions from a serial parser cannot account for the processing cost, considering the fact that the theory assigns a retrieval cost to each word in the sentence, and the parsing mechanism in the original proposal is in fact serial, although Lewis & Vasishth, 2005 did not assume a dependency grammar but a phrase structure grammar. While Boston et al. (2011) provided evidence for the CBR theory, they did not retain the original architectural assumptions of the theory by using a parallel parser and a dependency grammar. Clearly, stronger support for the cue-based theory can be provided if we show that retrieval cost can model eye movements even when the original architectural assumptions, viz. serial parsing with phrase structure grammar, are retained.

We implement cue-based retrieval models of a set of sentences from an eye tracking corpus, the Potsdam Sentence Corpus (PSC) (Kliegl, Nuthmann, & Engbert, 2006). Our assumptions are consistent with the original architectural assumptions in the proposal in Lewis & Vasishth, 2005. For testing the predictions, we select 32 sentences from the PSC covering a wide range of syntactic structures.[3] We select "early" and "late" eye movement measures like single fixation duration (SFD), first fixation duration (FFD), first-pass reading time (FPRT), total reading time (TRT) and first-pass skipping probability as dependent measures at each word. We consider unigram and bigram frequency, world length, and Cloze predictability as baseline predictors along with surprisal as a predictor of syntactic processing cost. We use a model comparison approach to test if the processing cost predicted by the CBR theory, which we call *attachment time*, is significant even when other predictors are considered. We fit two linear mixed-effects models: (i) a simpler model with baseline predictors and the syntactic predictor, and (ii)

[3]The set of sentences and the incremental parse created by the model at each word is illustrated in Appendix A.

a complex model with predictors from the simpler model plus the cue-based retrieval cost. A comparison between these two models revealed that the complex model, despite one extra degree of freedom, is a better model of eye movement measures.

We conclude that the processing cost predicted by the cue-based retrieval theory accounts for the observed word-by-word processing difficulty that was not captured by baseline and syntactic predictors. Moreover, retrieval cost influences "early" as well as "late" processing across all word classes. Importantly, assuming a serial parsing mechanism for retrieval predictions was sufficient to account for the processing cost.

1.3.2 Retrieval interference in syntactic processing: The case of reflexive binding in English (Ch. 3)

In Chapter 3, we apply the CBR theory to a specific problem in psycholinguistics—the process of reflexive binding in English. Determining how the antecedents of reflexive pronouns such as *himself/herself* are resolved in online comprehension is important to understanding how short-term memory supports sentence processing.

It has been proposed that the dependency between a reflexive pronoun and its antecedent is resolved using exclusively syntactic constraints (Phillips et al., 2011; Dillon, 2011). Under this *strictly syntactic search* account, Principle A of the binding theory—which requires that the antecedent c-command the reflexive within the same clause that the reflexive occurs in—constrains the parser's search for an antecedent. The parser is thus claimed to be immune to interference from candidate antecedents that might match agreement features of the reflexive (e.g., gender) but are ineligible as potential antecedents because they are in structurally inaccessible positions.

Support for strictly syntactic search is derived from studies reported in Nicol and Swinney (1989), Sturt (2003) and Xiang, Dillon, and Phillips (2009). Sturt (2003) carried out an eye tracking study using material such as (1). Here, the accessible antecedent for the reflexive (*himself* or *herself*) is *surgeon* and the inaccessible antecedent is *Jonathan* or *Jennifer* from Sentence 1. Sturt found that first fixation duration and first pass reading time on the region containing the reflexive were longer when there was a gender mismatch between the anaphor and the gender stereotype associated with the grammatically accessible antecedent (*herself* and *surgeon*) than when the stereotypical gender of the two matched (*himself* and *surgeon*). However, these early reading times were *not* affected by a gender match between the reflexive and the grammatically inaccessible antecedent.

(1)

Sentence 1: {Jonathan/Jennifer} was pretty worried at the City Hospital.
Sentence 2:

a. *Accessible-match/inaccessible-match* (*Match-interference* condition)
He remembered that the surgeon had pricked **himself** with a used syringe needle.

b. *Accessible-match/inaccessible-mismatch* (*Match* condition)
She remembered that the surgeon had pricked **himself** with a used syringe needle.

c. *Accessible-mismatch/inaccessible-match* (*Mismatch-interference* condition)
She remembered that the surgeon had pricked **herself** with a used syringe needle.

d. *Accessible-mismatch/inaccessible-mismatch* (*Mismatch* condition)
He remembered that the surgeon had pricked **herself** with a used syringe needle.

An alternative approach to strictly syntactic search accords no special status to structural constraints: in addition to using Principle A, the parser also uses non-structural cues such as gender to access the antecedent. A large body of work in the domain of dependency resolution in sentence processing has shown that the memory retrieval process utilizes non-syntactic information in addition to syntactic information. Van Dyke and colleagues (Van Dyke & Lewis, 2003; Van Dyke & McElree, 2006) have shown that semantic properties of nouns (e.g., animacy feature) and selectional requirements of verbs are utilized in retrievals. Moreover, the process of binding English reflexives inside picture noun phrases (Runner, Sussman, & Tanenhaus, 2006) and Chinese reflexives (Chen, Jäger, & Vasishth, in press) is shown to be influenced by the agreement features of the grammatically inaccessible antecedent. In light of these results, the strictly syntactic search account seems to be an exception, which calls for a specialized retrieval mechanism to explain only a limited set of results.

We formulate the question, 'what type of retrieval cues are used in the reflexive binding process?', in terms the CBR theory. We implement a CBR model of Sturt's design in (1). The model utilizes both structural and non-structural cues. The effects predicted by the model are in terms of: (i) retrieval error (RE), the percentage of retrievals of *inaccessible* antecedent, and (ii) retrieval time (RT), the amount of time taken to retrieve an antecedent. Five distinct effects are predicted:

E1. Mismatch effect (RE): the retrieval errors for mismatch conditions (c and d in (1)) are higher than for match conditions (a and b in (1)).

E2. Interference effect (RE): the retrieval errors for interference conditions are higher than the other two conditions (a, c vs. b, d).

E3. Mismatch effect (RT): the retrieval times for mismatch conditions are longer than for match conditions.

E4. Match-interference effect (RT): the retrieval times for the match-interference condition are longer than the match condition.

E5. Mismatch-interference effect (RT): the retrieval times for the mismatch-interference condition are shorter than the mismatch condition.

The effect E3 was observed in Sturt's eye tracking experiment. The effects E1 and E2 were observed in a follow-up study with a similar design that explicitly probed for the antecedent of the reflexives, presumably reflecting effects in late processing. The interference effects E4 and E5 were missing in the early eye tracking measures like first fixation duration and first pass reading time.

We suggest alternative explanations for the lack of interference effects in Sturt (2003) and other studies. In terms of the CBR theory, the probability of retrieving an incorrect element is higher if it has a higher activation value at the time of retrieval. The activation value of a memory element is directly dependent on its creation time, retrieval history, and its match with the retrieval cues—the more recently an element is created or retrieved, and the higher the feature overlap it has with the retrieval cues, the better are the chances of it being retrieved. The inaccessible antecedent in (1) above was introduced earlier in the string than the accessible antecedent, and was therefore relatively distant from the reflexive; hence the interferer has less chance of getting retrieved compared to the accessible antecedent. Although, other studies reporting lack of interference effects (e.g., Experiment 2 in Sturt, 2003 and Experiment 1 in Dillon, 2011) did use a local inaccessible antecedent, they used an interferer with an object role (e.g., *manager(s)* in (2) from Dillon, 2011,[4] is the object of the relative clause, which either matches or does not match the number marking on the reflexive, *himself* or *themselves*). Following Principle A of the binding theory, we assume that reflexive binding uses grammatical role as one of the retrieval cues (role = *subject*). With this assumption, an interferer with *object* role has a lesser chance of getting retrieved, because its overlap with the retrieval cues is lower in comparison to the overlap of the accessible antecedent with the retrieval cues. In other words, in these experiments, the inaccessible antecedents may not be strong enough interferers to cause a detectable effect on the retrieval

[4]The conditions are rearranged to keep them consistent with the design discussed earlier.

process. If this reasoning is correct the lack of an interference effect, might be a false negative (a type II error).

(2) a. *Grammatical, interference*
The new executive who oversaw the middle manager apparently doubted himself on most major decisions.

b. *Grammatical, no interference*
The new executive who oversaw the middle managers apparently doubted himself on most major decisions.

c. *Ungrammatical, interference*
The new executive who oversaw the middle managers apparently doubted themselves on most major decisions.

d. *Ungrammatical, no interference*
The new executive who oversaw the middle manager apparently doubted themselves on most major decisions.

Concluding that an absence of an interference effect is evidence that no interference occurs has important consequences for the theory of retrieval processes in sentence comprehension. The lack of interference in processing argument reflexives implies that the retrieval mechanism for reflexive binding is different from other retrieval mechanisms in sentence processing, e.g., subject-verb agreement and agreement attraction, where consistent interference effects have been reported (e.g., Wagers, Lau, & Phillips, 2009; Van Dyke & McElree, 2011). On the other hand, finding an interference effect simplifies the theory of retrieval processes considerably, since no exemption is granted to antecedent-reflexive resolution processes.

In order to increase the strength of the interference effect, we devise a new experimental design (as in (3)) that uses an object relative clause where the inaccessible antecedent is the subject of the clause; this matches one of the retrieval cues (role = *subject*) at the reflexive. It is also closer to the reflexive in terms of linear distance. Under the CBR account with syntactic and non-syntactic retrieval cues, the inaccessible antecedent would be more likely to interfere in the retrieval process than in the two experimental designs mentioned earlier, but under the strictly syntactic search approach, this manipulation should not matter to the reflexive binding process. As expected, when the model's predictions for the new design are compared with the predictions for the previous two designs, all effects (E1-E5) are typically predicted to be stronger for the new design.

(3) a. *Accessible-match/inaccessible-match (match-interference)*
The tough soldier that Fred treated in the military hospital introduced himself to all the nurses

b. *Accessible-match/inaccessible-mismatch (match)*
The tough soldier that Katie treated in the military hospital introduced himself to all the nurses

c. *Accessible-mismatch/inaccessible-match (mismatch-interference)*
The tough soldier that Katie treated in the military hospital introduced herself to all the nurses

d. *Accessible-mismatch/inaccessible-mismatch (mismatch)*
The tough soldier that Fred treated in the military hospital introduced herself to all the nurses

An eye tracking study conducted with the new design supports the predictions of the cue-based retrieval model of reflexive binding, that uses a gender cue. The interference effects E2 and E4 are observed in first pass regression probability. A clear trend for the interference effects E4 and E5 is observed in regression contingent first fixation duration. Importantly, the interference effects are observed in early eye tracking measures, implying that the initial stages of the reflexive binding process are influenced by the inaccessible antecedent. The study also replicates the mismatch effects, E1 and E3, found in Sturt's experiments, in terms of re-reading time, total reading time and regression contingent first fixation duration. In the context of our modeling and experimental results, we further discuss the strictly structured search account which does not predict any differences in reading times across four conditions.

In sum, in Chapter 3 we present a theory, in terms of a CBR model, of the access of antecedents for reflexive pronouns in English. We use this theory to gain insight into empirical studies that have yielded mixed results concerning the putative role of non-structural cues. We used this analysis and the results of further modeling to motivate a new empirical design that formed the basis of an eye tracking study. The results of the eye tracking study are consistent with the model's assumptions concerning the early use of non-structural cues. These results present a challenge for theories advocating the infallibility of the human parser in the case of reflexive resolution, and provide support for the inclusion of agreement features such as gender in the set of retrieval cues.

1.3.3 A cue-based retrieval model of offline and online sentence processing in aphasia (Ch. 4)

In Chapter 4 we extend the cue-based retrieval theory to model aphasic sentence processing. We propose a processing deficit based model of aphasics' offline and online processing in a sentence-picture matching task.

It has been consistently observed in the literature that individuals with aphasia perform at chance level when processing reversible non-canonical sentences (for a review see Grodzinsky, 2000), but their comprehension of canonical sentences is normally unaffected. Two types of theories have been proposed to explain this pattern of behavior: (1) Representational accounts, which attribute this deficit to an impairment in syntactic representations (e.g., Grodzinsky, 1995, 2000, 2006), and (2) Processing accounts, which propose that sentence comprehension is affected by processing deficits such as slow lexical activation, reduction in memory resources, slowed processing and intermittent deficiency among others (e.g., Avrutin, 2006; Caplan, Waters, Dede, Michaud, & Reddy, 2007; Dickey, Choy, & Thompson, 2007; Friederici & Kilborn, 1989; Haarmann, Just, & Carpenter, 1997).

Using the CBR architecture, we test two hypotheses from the processing accounts: *slowed processing* and *intermittent deficiency* (Caplan et al., 2007; Dickey et al., 2007; Friederici & Kilborn, 1989; Hanne, Sekerina, Vasishth, Burchert, & De Bleser, 2011). Hanne et al. (2011) suggested that the chance level performance in aphasics is a result of syntactic processing that is slowed down and affected by intermittent deficiency. They reported a study with a sentence-picture matching task in a visual world paradigm, with agrammatic Broca's aphasics and age-matched controls. During the experiment, participants listened to German reversible canonical and non-canonical sentences as in (4) while they were shown two pictures on the screen. Each trial consisted of listening to either a canonical or a non-canonical sentence and looking at two pictures on the screen—a target picture that matched the sentence and a distractor picture. At the end of the trial, participants were asked to select the correct picture. Participants' response and response time were recorded. Participants' eye movements were also recorded during the whole trial. As a result, the data from Hanne et al. (2011) consists of one online measure (eye movements) and two offline measures (response accuracy and response time) of sentence processing for both, healthy and aphasic individuals.

The accuracy data for aphasics exhibited chance level performance for non-canonical sentences, and the response times for aphasics were slower than for controls. Importantly, the online data (eye movements) showed divergent patterns for aphasics depending on whether the offline response for non-canonical sentences was correct or not: for correctly interpreted trials, their eye movements were qualitatively similar to those of controls, whereas, for incorrect trials, the eye movements were totally different from controls. With these results, Hanne et al. (2011) concluded that "in some instances the aphasic parser processes sentences normally (although often more slowly) and hence arrives at the correct sentence interpretation, whereas in other instances it fails and a comprehension problem arises".

(4) a. **Canonical:**
Der Sohn fängt den Vater
the$_{\text{NOM}}$ son is_catching the$_{\text{ACC}}$ father 'The son is catching the father'

b. **Non-canonical:**
Den Sohn fängt der Vater
the$_{\text{ACC}}$ son is_catching the$_{\text{NOM}}$ father
'The father is catching the son'

We frame the hypotheses of *slowed processing* and *intermittent deficiency* within the CBR architecture. We operationalize slowed processing as slowed procedural memory, so that each processing action is performed slower than normal, and intermittent deficiency as extra noise in the procedural memory, so that the parsing process is more noisy than normal. The models do not assume any representational impairment. Since it is possible that only one of the two assumptions from Hanne et al. (2011) is enough to capture the data, we evaluate these assumptions using three separate models: model M1 assumes only slowed processing, model M2 assumes only intermittent deficiency, and model M3 assumes both deficits.

For modeling the sentence-picture matching task in the CBR architecture, we make a set of assumptions in terms of a linking hypothesis. First, we assume that, while processing a sentence, the model creates two separate semantic representations of the two pictures on the screen. These semantic representations are stored in declarative memory. As in earlier CBR models, at each input word, the parser incrementally updates the partial representation of the sentence that has been built up to that point. Next, we assume that, as the model processes new sentential input, it selects the picture that matches the partial representation of the sentence up to that point. The picture selection is performed by means of a retrieval request for a matching picture representation in memory. As a consequence of a retrieval request, the two picture chunks receive varying amounts of activation boost. The amount of activation boost received by each chunk depends on its match with the partial sentence representation. We further assume that the difference in the activation of the correct and incorrect picture chunk at the time of retrieval determines the likelihood of fixating on the correct picture. This set of assumptions for fixation probabilities in terms of activations is based on the existing evidence in the visual world literature and the architectural constraints on the model. These assumptions are consistent with the linking hypothesis proposed by Altmann and Kamide (2007).

Finally, we assume that the picture that is retrieved at the end of the sentence is the picture that is selected as the response to the sentence-picture matching task, and that the duration between the processing of the first word of the sentence and the retrieval of the picture at the end of the sentence is the response time for the picture matching task.

The modeling results reveal that Models M2 and M3, but not M1, capture the overall processing difficulty for aphasics observed in terms of low accuracy in the sentence-picture matching task. They also capture the classical effect of chance level performance in aphasics for non-canonical structures. On the other hand, models M1 and M3, but not M2, capture the processing difficulty observed in terms of slow response times. Hence, for offline data, the model assuming both deficits (M3) emerges as the best among the three models examined. For eye movement patterns, all three models capture the divergent eye movement patterns observed in the data—normal-like eye movements in correct trials and aberrant eye movements in incorrect trials.

Consequently, the modeling results suggest that assuming *both* deficits—slowed processing and intermittent deficiency—is necessary to account for the sentence processing impairment in aphasia. The results are consistent with the hypothesis proposed by processing or resource reduction accounts, which propose that aphasic sentence processing deficits should be ascribed to processing limitations instead of a breakdown in the underlying grammatical knowledge. Other computational models of aphasia also support processing or resource reduction based explanations, though they differ in terms of the precise processing deficit assumed (e.g., Crescentini & Stocco, 2005; Haarmann & Kolk, 1991a; Haarmann et al., 1997).

Our model goes one step further than previous models by delivering predictions not only for the accuracy data but also for response times and incremental eye movements. It is important to note here that the assumptions about model parameters are made only for the offline data, and predictions of eye movements emerge without fitting any extra parameters for online processing in aphasics. Fitting extra parameters for modeling aphasics' online processing would have meant that the assumptions of slowed processing and intermittent deficiency are not sufficient to capture all aspects of the impairments in aphasics.

Furthermore, the linking hypothesis proposed here for modeling sentence-picture matching data opens new possibilities for modeling data from the visual world paradigm in the CBR architecture. Here we modeled the sentence-picture matching task, but the hypothesis can be extended for modeling other tasks in the visual world paradigm.

Chapter 2

Compound effect of probabilistic disambiguation and memory retrievals on sentence processing: Evidence from an eye-tracking corpus

---

This chapter was published as:

Patil, U., Vasishth, S., & Kliegl, R. (2009). Compound effect of probabilistic disambiguation and memory retrievals on sentence processing: Evidence from an eye-tracking corpus. In A. Howes, D. Peebles, & R. Cooper (Eds.), *9th International Conference on Cognitive Modeling—ICCM2009*. Manchester, UK.

2.1 Abstract

We evaluate the predictions of surprisal and cue-based theory of sentence processing using an eye-tracking corpus, the Potsdam Sentence Corpus. Surprisal is a measure of processing complexity based on a probabilistic grammar and is computed in terms of the total probability of structural options that have been disconfirmed at each input word. The cue-based theory characterizes processing difficulty in terms of working memory costs that derive from decay and interference arising during content-based retrieval requests of previously processed material (e.g., to incrementally build the sentence structure). We show that both surprisal and cue-based parsing independently explain difficulty in sentences processing and interestingly, they have an over-additive effect on processing when combined together.

Keywords: Sentence processing; eye-tracking; cue-based theory; surprisal; memory retrievals

2.2 Introduction

Research in psycholinguistics provides much evidence for probabilistic disambiguation in human language processing at various levels including lexical, syntactic and semantic processing (Jurafsky, 1996, 2003). More frequent words and structures are easier to comprehend than less frequent ones. Surprisal (Hale, 2001) is a proposal which characterizes processing difficulty in terms of the amount of work done in probabilistically disconfirming sentence continuations as a consequence of the information supplied by the current word. Consider, for example, the famous garden path sentence in (1). It has been observed that English speakers hearing this sentence have great difficulty at "fell". Hale (2001) demonstrates using probabilistic context-free grammar that the difficulty occurs because at "fell" the parser has to disconfirm alternatives that together comprise a great amount of the probability mass.

(1) The horse raced past the barn *fell*.

Recent research in computational models of sentence comprehension has shown that surprisal is a significant predictor of eye movements while reading individual sentences and text (Boston et al., 2008; Demberg & Keller, 2008). However, surprisal is likely to furnish only part of the explanation (Levy, 2008). As Lewis (1996) and Gibson (2000) argue, sometimes people take longer to process words that they need to connect to other words processed earlier. Resolving these linguistic relations seems to impose more processing effort even when the constructions are frequent or unsurprising. Grodner and Gibson (2005) provide evidence using self-paced reading study which involved reading sentences like (2) below. They observed monotonically increasing reading time at the verb "supervised" as a function of its distance from the subject "nurse".

(2) a. The nurse *supervised* the ...
b. The nurse from the clinic *supervised* the ...
c. The nurse who was from the clinic *supervised* the ...

This difference between surprisal and integration cost was addressed by Demberg and Keller (2008), who compared the predictions of surprisal with Gibson's (2000) Dependency Locality Theory (DLT), a theory of integration difficulty. They found that DLT's predictions played a limited role in explaining processing difficulty. DLT was a significant predictor only for reading times at nouns and verbs. Here we show that surprisal and retrieval costs unequivocally play a role in determining processing difficulty. More interestingly, we observed a significant interaction of surprisal and memory retrievals, suggesting that a simple additive model of surprisal and retrieval processes will not suffice.

We compared surprisal's predictions to the cue-based retrieval model of (Lewis & Vasishth, 2005) (LV05 henceforth) using the Potsdam Sentence

Corpus (PSC) of German (Kliegl, Nuthmann, & Engbert, 2006). The cue-based retrieval theory characterizes processing difficulty in terms of working memory costs that derive from decay and interference arising during content-based retrieval requests of previously processed material, e.g., to complete dependencies, or to incrementally build structure.

We implemented cue-based retrieval models for sentences from the PSC, closely following the approach taken by LV05 and generated predictions for retrieval cost at each word. We also computed surprisal's predictions using a probabilistic phrase-structure parser. The main findings are that (1) retrieval cost furnishes better models of eye-fixation measures than models based on baseline predictors such as unigram and bigram frequency, word length, Cloze predictability plus surprisal, and (2) surprisal and retrieval cost show a significant interaction in predicting reading times.

2.2.1 Surprisal

Surprisal offers a theoretical reason why a particular word in a sentence should be easier or more difficult to comprehend on the basis of underlying probabilistic grammatical knowledge of the language. The idea of surprisal is to model processing difficulty as a logarithmic function of the probability mass eliminated by the most recently added word. This number is a measure of the information value of the word just seen, as rated by the grammar's probability model; it is nonnegative and unbounded. More formally, the surprisal of the n^{th} word (w_n) in a sentence is defined as the log-ratio of the prefix probability before seeing the word, compared to the prefix probability after seeing it. The prefix probability at word w_n is defined as the total probability of all grammatical analyses that derive the prefix string $w = w_1 \cdots w_n$ which is initial part of the bigger string wv. For grammar G and a set of derivations D the prefix probability α_n at word w_n can be expressed as:

$$prefix\_probability(w, G) = \sum_{d \in D(G,wv)} probability(d) \ = \ \alpha_n$$

Then, the surprisal at w_n is:

$$surprisal(w_n) = \log_2(\frac{\alpha_{n-1}}{\alpha_n})$$

Intuitively, surprisal and hence the difficulty of processing increases when a parser is required to build some low-probability structure.

2.2.2 Cue-based theory

The cue-based theory of sentence processing is derived from the application of independently motivated principles of memory and cognitive skills to the specialized task of sentence parsing. As a result, sentence processing emerges as a series of skilled associative memory retrievals modulated by similarity-based interference and fluctuating activation. The corresponding parsing model is implemented in the cognitive architecture ACT-R (Anderson et al., 2004) which formalizes the cognitive principles mentioned above in terms of the following set of equations:

1. The base activation (B_i) of chunk i, where t_j is the time since the j^{th} retrieval of the item, d is the decay parameter, and the summation is over all n retrievals, is
$$B_i = ln(\sum_{j=1}^{n} t_j^{-d})$$

2. Total activation (A_i) of a chunk i is defined as the summation of its base activation and strength of association. W_j is the amount of activation from the elements j in the goal buffer and S_{ji}s are the strengths of association from elements j to chunk i
$$A_i = B_i + \sum_j W_j S_{ji}$$

3. S_{ji} is defined in terms of fan_j which is the number of items associated with j

$$S_{ji} = S - ln(fan_j)$$

4. Retrieval latency of chunk i is defined in terms of A_i and F, a scaling constant

$$T_i = Fe^{-A_i}$$

The cue-based retrieval theory quantifies the processing difficulty at each word in terms of its *attachment time*, which is the sum of (i) the time required to retrieve the currently-built syntactic structure in order to attach the word into that structure, and (ii) a baseline cost of 100 milliseconds, which is the time required for the execution of the retrieval request and the subsequent attachment of the current word into the existing structure. See LV05 for details about data structures and the parsing algorithm used.

To summarize, the delay in retrieval of a prior syntactic element due to similarity based interference and fluctuating activation is assumed to induce difficulty in processing.

2.3 Experiment

The experiment involved a quantitative evaluation of the predictions of surprisal and cue-based theory using a corpus of eye movements during reading single sentences.

2.3.1 Methods

Data

For the analyses in this paper, we selected 32 sentences from the Potsdam Sentence Corpus (PSC), which is an eye-tracking corpus consisting of fixation durations recorded from 222 persons, each reading 144 German sentences (Kliegl, Nuthmann, & Engbert, 2006). These 32 sentences were selected in a way that enabled us to cover a wide range of syntactic structures.

For generating surprisal values for each word in these selected sentences we used a probabilistic context-free phrase-structure parser from Levy (2008), which is an implementation of Stolcke's Earley parser (Stolcke, 1995). We unlexicalized the parser to avoid overlap of surprisal's predictions with the word frequency effect.

We hand-crafted an ACT-R model for each selected sentence, closely following the approach taken by LV05. The model of each sentence was run for 30 simulations and a prediction of attachment time for every word was generated by averaging across all simulations. All ACT-R parameter values were kept the same as those used by LV05 except for activation noise. In LV05, five out of six simulations were carried out without switching on the activation noise. They also noted from preliminary experiments that adding activation noise did not change their results significantly. Since, one of ACT-R's standard assumptions is that there is always some noise added to the activation value of a chunk at each retrieval which permits modeling various kinds of memory errors, we set its value to 0.45 (this was one of the values used in Vasishth et al., 2008).

Statistical Analyses

The statistical analyses were carried out using linear mixed-effects models (Bates & Sarkar, 2007; Gelman & Hill, 2007) and the *Deviance Information Criterion* or DIC (Gelman & Hill, 2007, 524–527) was used to compare the relative goodness of fit between simpler and complex models. Linear models were fit for the following "early" and "late" eye movements measures:

SFD - fixation duration on a word during first pass if it is fixated only once

FFD - time spent on a word, provided that the word is fixated during the first pass

FPRT - the sum of all fixations on a word during the first pass

TRT - the sum of all fixations

FPSKIP - the probability of skipping the word during the first pass

We considered following baseline predictors in addition to surprisal and attachment time:

unigram - logarithm of token frequency of a word in Das Digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts (DWDS) (Geyken, 2007; Kliegl, Geyken, Hanneforth, & Würzner, 2006)

bigram - logarithm of the conditional likelihood of a word given its left neighbor in DWDS (also called transitional probability)

word length - number of characters in conventional spelling

predictability - empirical predictability as measured in a Cloze task with human subjects (Taylor, 1953; Ehrlich & Rayner, 1981; Kliegl, Grabner, Rolfs, & Engbert, 2004)

Sentences and participants were treated as partially crossed random factors; that is, we estimated the variances associated with differences between participants and differences between sentences, in addition to residual variance of the dependent measures. For the analysis of FPSKIP (coded as a binary response for each word: 1 signified that a skipping occurred at a word, and 0 that it did not), we used a generalized linear mixed-effects model with a binomial link function (Bates & Sarkar, 2007; Gelman & Hill, 2007).

For each reading time analysis reported below, reading times more than three standard deviations away from the mean were removed before the analyses, excluding at most 1.7% of the data. Attachment time and all dependent measures except FPSKIP were log transformed. Word length, surprisal and attachment time were centered in order to render the intercept of the statistical models easier to interpret.

In the initial analyses, as expected, we found collinearity among the baseline predictors. Since collinearity can inflate the estimates of coefficients' standard errors leading to unreliable results, and can also lead to uninterpretable coefficient values, removal of collinearity between predictors was crucial before fitting the linear models for different fixation measures. For removing collinearity, we incrementally regressed each of these predictors against one or more baseline predictors and used residuals of the regressions as the predictors in the subsequent linear models. This was done in the following three steps:

1. Regression of unigram frequency against word length-
   uni.res = residuals (unigram ~ length)

2. Regression of bigram frequency against word length and residual unigram values obtained from step 1-
   bi.res = residuals (bigram ~ length + uni.res)

3. Regression of predictability against word length, residual unigram and bigram obtained from step 1 & 2-
   pred.res = residuals (predictability ~ length + uni.res + bi.res)

As a result, we had four baseline predictors — length, uni.res, bi.res, pred.res — which were completely non-collinear.

2.3.2 Results & Discussion

The results of the mixed-effects models are summarized in tables 2.1 to 2.3. We observed significant main effects of both surprisal and attachment cost across "early" as well as "late" measures and also on FPSKIP. The coefficient for FPSKIP is negative reflecting the fact that the probability of fixating a word increases with increase in surprisal and retrieval cost. These results illustrate that surprisal as well as retrieval cost can account for variance in eye-tracking measures independent of baseline predictors (such as unigram

Table 2.1: Linear model coefficients, standard errors and t-values for surprisal, attachment time and interaction of attachment time and surprisal. An absolute t-value of 2 or greater indicates statistical signicance at $\alpha = 0.05$.

	Coef	SE	t-value
SFD			
surprisal	0.021722	0.001195	18
att. time	0.084338	0.013722	6
att. time:surprisal	0.048706	0.009518	5
FFD			
surprisal	0.018304	0.001032	18
att. time	0.062361	0.012361	5
att. time:surprisal	0.039307	0.008327	5
FPRT			
surprisal	0.021520	0.001217	18
att. time	0.056154	0.014221	4
att. time:surprisal	0.050750	0.009743	5
TRT			
surprisal	0.028558	0.001389	21
att. time	0.058249	0.016197	4
att. time:surprisal	0.055988	0.011128	5

Table 2.2: Linear model coefficients, standard errors and t-values for baseline predictors for TRT.

	Coef	SE	t-value
TRT			
length	0.031052	0.000949	33
uni.res	-0.023228	0.002322	-10
bi.res	-0.011984	0.000879	-14
pred.res	-0.006162	0.002752	-2

Table 2.3: Linear model coefficients, standard error, z-scores and p-values with FPSKIP as the dependent measure.

	Coef	SE	z-score	p-value
att. time	-0.51588	0.09401	-5.5	<0.001
surprisal	-0.18235	0.01000	-18.2	<0.001
att. time:surp	-0.12521	0.08067	-1.6	0.121

Table 2.4: Deviance Information Criterion values for simpler model (baseline predictors + surprisal) vs. more complex model (simpler model + attachment time).

	Simpler model	Complex model
SFD	8624.7	8576.5
FFD	9908.0	9873.1
FPRT	22606.0	22581.9
TRT	30695.5	30674.6
FPSKIP	36140.8	36111.5

and bigram frequency, word length, Cloze predictability, etc.). For comparison, coefficients of baseline predictors for TRT are listed in table 2.2; similar coefficient values were obtained for other reading time measures.

The interaction of attachment time and surprisal is significant for all measures except for FPSKIP (though even in this case the coefficient has the expected sign), which indicates that there is a disproportionate increase in reading difficulty when both surprisal and retrieval cost are high.

Table 2.4 compares the DIC values for simpler models (baseline predictors + surprisal) and complex models (baseline predictors + surprisal + attachment time). For all dependent measures the predictive error (DIC value) was lower in the more complex model that included attachment time, which means that the complex models should be preferred to the simpler ones.

Retrieval cost, surprisal and their interaction show effects on "early" as well as "late" measures. This suggests that structure-building and retrieval processes start very soon after lexical access begins.

Implications for eye movement models

Besides the contribution to psycholinguistic theories, this work can contribute towards extending models of eye movement control such as E-Z Reader (Pollatsek, Reichle, & Rayner, 2006) and SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005) which despite being the two most fully developed models of eye movements, do not incorporate any theory of language processing. The latest version of E-Z Reader (Reichle, Warren, & McConnell, 2009) makes an attempt in this direction by augmenting the model with a post-lexical integration stage, named I. This stage is assumed to reflect all of the post-lexical processing like linking the word into a syntactic structure, generating a context-appropriate semantic representation, and incorporating its meaning into a discourse model. However, the amount of time to complete I, t(I), is independent of the language processing demands at that word; instead t(I) is sampled from a gamma distribution having a mean of 25 msec and standard deviation of 0.22. Models of sentence processing like the two evaluated here or, preferably, a systematic combination of them would offer a more realistic way of computing t(I). A similar approach of incorporating post-lexical processes can be taken in other eye movement models depending on the particular architecture of each model.

2.4 Conclusions

This work evaluated the combined contribution of two theories of sentence processing, viz., surprisal and cue-based retrieval theory. The two approaches capture different aspects of sentence processing, namely instantaneous probabilistic disambiguation and processing constraints due to memory retrievals. It was shown that when effects of these theories were combined together to predict eye movements measures, they emerged as significant predictors even when word length, n-gram frequency and Cloze predictability were taken into account. Moreover, they showed an over-additive effect on several eye movements measures. This needs to be taken into account in future models of sentence processing that integrate surprisal and retrieval costs. Also, models of eye movement could benefit from this work. Although the size of the evaluation corpus is small (total 32 sentences and 222 participants) and models of cue-base parsing were hand-crafted, this work serves as a first step towards developing a broad coverage model of sentence processing that combines the two processes – probabilistic disambiguation and memory retrieval – in a principled way.

Chapter 3

Retrieval interference in syntactic processing: The case of reflexive binding in English

This chapter has been submitted as a paper to the *Journal of Memory and Language* on January 21, 2012 as:

Patil, U., Lewis, R. L., & Vasishth, S. (submitted). Retrieval interference in syntactic processing: The case of reflexive binding in English.

3.1 Abstract

It has been proposed that in online sentence comprehension the dependency between a reflexive pronoun such as *himself/herself* and its antecedent is resolved using exclusively syntactic constraints. Under this *strictly syntactic search* account, Principle A of the binding theory—which requires that the antecedent c-command the reflexive within the same clause that the reflexive occurs in—constrains the parser's search for an antecedent. The parser thus ignores candidate antecedents that might match agreement features of the reflexive (e.g., gender) but are ineligible as potential antecedents because they are in structurally illicit positions. An alternative possibility accords no special status to structural constraints: in addition to using Principle A, the parser also uses non-structural cues such as gender to access the antecedent. According to cue-based retrieval theories of memory (e.g., Lewis and Vasishth, 2005), the use of non-structural cues should result in increased retrieval times and occasional errors when candidates partially match the cues, even if the candidates are in structurally illicit positions. In this paper, we first show how the retrieval processes that underlie the reflexive binding are naturally realized in the Lewis and Vasishth (2005) model. We present the predictions of the model under the assumption that both structural and non-structural cues are used during retrieval, and provide a critical analysis of previous empirical studies that failed to find evidence for the use of non-structural cues, suggesting that these failures may be Type II errors. We use this analysis and the results of further modeling to motivate a new empirical design that we use in an eye tracking study. The results of this study confirm the key predictions of the model concerning the use of non-structural cues, and are inconsistent with the strictly syntactic search account. These results present a challenge for theories advocating the infallibility of the human parser in the case of reflexive resolution, and provide support for the inclusion of agreement features such as gender in the set of retrieval cues.

Keywords: sentence processing; anaphor resolution; memory retrieval; interference; computational modeling; eye tracking

3.2 Introduction

The task of sentence comprehension involves, among other things, recovering hierarchical structure from an input string of words (e.g., Frazier, 1979). Such recovery requires the online application of grammatical constraints that delimit the possible relationships between various elements of the sentence. For example, to understand a sentence like (1), the pronoun *himself* has to be resolved to a referent of an earlier noun *surgeon*; the reflexive cannot be associated with *Jonathan* due to Principle A of the binding theory (Chomsky, 1981).[1]

(1) The surgeon who treated Jonathan had pricked himself.

Establishing a relationship between two non-adjacent elements in a sentence requires maintaining some memory of the immediate past. The question we are concerned with here is: how is access to the past constrained by structural grammatical principles? The binding of reflexive pronouns is a particularly informative case, because the configurational and agreement constraints are relatively clear, and the structure admits of manipulations of distance and distracting candidate antecedents (Sturt, 2003).

One proposal for how structural constraints are implicated in dependency resolution is due to Phillips et al. (2011): in the case of reflexives, antecedent relations are computed exclusively using structural constraints—Principle A of the binding theory constrains the parser's antecedent resolution process

[1]Principle A specifies a structural constraint on the interpretation of reflexives in English: a reflexive must be bound by an antecedent in the local domain (the current clause). An antecedent X can bind a reflexive Y, if X and Y are coindexed, and X c-commands Y. The term c-command defines a hierarchical relationship between two constituents in a syntax tree. A constituent c-commands its sister constituent and every constituent *below* the sister constituent in the syntax tree. In (1), the reflexive *himself* is bound by *surgeon*; the noun *Jonathan* cannot bind the reflexive because it does not c-command it.

so that the correct antecedent is retrieved. Under this account, the parser ignores candidate antecedents that might match agreement features of the reflexive (e.g., gender) but which are structurally disallowed due to Principle A. An alternative possibility accords no special status to the structural constraints: in addition to using Principle A, the parser also uses non-structural cues such as gender to access the antecedent. For example, in (1), it is possible that the parser considers a relation between *Jonathan* and *himself*, due to a gender-feature match, and perhaps also due to the relative proximity of *Jonathan* compared to *surgeon*. Under at least one cue-based retrieval theory (Lewis & Vasishth, 2005; Lewis et al., 2006), this should result in overall increased retrieval times and occasional errors when candidates partially match the cues, even if they are in structurally illicit positions. (We describe below precisely how such effects might arise.)

Recently, Phillips et al. (2011), Dillon, Mishler, Sloggett, and Phillips (2011) and Dillon (2011) have formulated a memory retrieval account of reflexive binding, but have argued that only structural cues are part of the retrieval cue set, and not agreement features like person, number and gender. This constrained strategy is thought to facilitate *selective retrieval* of an antecedent in a c-commanding position relative to the reflexive (e.g., *soldier* in (2)) without interference from other NPs that are in structural positions that do not satisfy Principle A (e.g., *Fred* in (2)). The relative adaptive or functional value of using structural cues alone vs. structural-plus-agreement cues is an important open theoretical and empirical question that will require further work to resolve; we set this question aside for now and focus on clarifying the answer to the question of what cues are in fact used. We will follow the literature in referring to the correct antecedent as stipulated by Principle A as the *grammatically accessible* antecedent and the antecedent that is incorrect following Principle A as the *grammatically inaccessible* antecedent. Occasionally, we abbreviate these terms to *accessible* and *inaccessible* antecedents. It is important to keep in mind that under the model we advocate in this paper, the grammatically inaccessible antecedent is in fact "accessible" for memory retrieval; a more appropriate term would have been "incorrect antecedent", since this does not presuppose that the non-c-commanding antecedent is inaccessible.

(2) The tough soldier that Fred treated in the military hospital introduced himself to all the nurses.

The remainder of this paper has the following structure and underlying logic. We first summarize existing empirical studies concerning the use of non-structural cues in binding reflexives. We then briefly describe an existing computational model of cue-based parsing (Lewis & Vasishth, 2005) and apply it to one of the key extant empirical paradigms, drawing out the principal qualitative predictions, and demonstrating that these predictions are robust against substantial variation in the quantitative parameters. We then use the theoretical perspective provided by the model to formulate conjectures for why some of the existing empirical work may have failed to detect evidence for the use of non-structural cues. Based on this analysis we advance a new experimental design which is intended to be more sensitive, and demonstrate that for many of the predictions the new design yields larger effects in modeling simulations. We next present an eye tracking study based on the new design, yielding several results that confirm the early use of non-structural cues in a manner consistent with the model. The paper concludes with discussion of the implication of these new results for some current theoretical approaches to dependency resolution.

3.3 Evidence for the use of non-structural cues in accessing antecedents of reflexives

The support for selective (structurally-constrained) retrieval is derived mainly from studies reported in Nicol and Swinney (1989), Sturt (2003) and Xiang et al. (2009). These studies found either no effect of interference, or a late effect of it. Nicol and Swinney (1989) proposed that when comprehenders process a pronoun or anaphor, the initial set of candidate antecedents contains only those nouns that are in structurally licit positions as stipulated by Binding Theory (Chomsky, 1981). They presented evidence from a series of experiments that employed the cross modal priming paradigm. Participants listened to sentences similar to those shown in (3) and responded to visually presented probe words. A probe word was presented immediately following the reflexive *himself*. The probe word was either semantically related or unrelated to one of the three previously occurring nouns: *boxer*, *skier*, or *doctor*. Participants judged whether the probe word was an English word or non-word. A significant priming effect was observed when probe words were related to nouns that were grammatically accessible as antecedents (e.g., *doctor* in (3)). When probe words were related to nouns that were grammatically inaccessible antecedents (e.g., *boxer* and *skier* in (3)), no priming was found. Nicol and Swinney (1989) concluded that no priming was observed for words related to grammatically inaccessible antecedents because they had not been considered during co-reference resolution.

(3) The boxer told the skier that the doctor for the team would blame himself for the recent injury.

Further evidence for an absence of interference from the grammatically inaccessible antecedent was provided in (Sturt, 2003). Sturt (2003) reported eye tracking studies using materials such as (4) and (5). He found that first fixation duration and first pass reading time on the region containing the reflexive *himself* or *herself* were longer when there was a gender mismatch between the anaphor and the gender stereotype associated with the grammatically accessible antecedent than when the stereotypical gender of the two matched (e.g., *herself* and *surgeon* (mismatch) vs. *himself* and *surgeon* (match)). Early reading times were *not* affected by gender match between the reflexive and the grammatically inaccessible antecedent (*Jonathan* or *Jennifer*). Second pass reading time at the reflexives showed an interaction for gender match between the reflexive and the two antecedents: when the accessible antecedent matched in gender with the reflexive, second pass reading time two regions downstream from the reflexive was longer when the *in*accessible antecedent mismatched in gender, suggesting that in later interpretation stages (but, crucially, not in earlier processing stages)[2] the inaccessible antecedent is part of the candidates being considered as antecedents (when the accessible antecedent mismatched the reflexive in gender, no effect of the inaccessible antecedent's gender was found).

To gain further insight into this late-stage interpretation of the sentences, Sturt (2003) also ran a follow-up study, where a sentence-by-sentence self-paced reading task was followed by a question that directly probed for the antecedent of the reflexive. This study showed a significant interference effect, with more ungrammatical interpretations when the grammatically inaccessible antecedent matched for gender with the reflexive; the effect was bigger when the accessible antecedent mismatched with the reflexive. On the basis of these results, Sturt (2003) concluded that grammatical constraints are applied very early in processing, but interference from the grammatically inaccessible antecedent occurs during later processes that are related to recovery strategies, rather than during processes related to the initial interpretation of the reflexive.

(4) {Jonathan/Jennifer} was pretty worried at the City Hospital. {He/She} remembered that the surgeon had pricked {himself/herself} with a used syringe needle. There should be an investigation soon.

(5) {Jonathan/Jennifer} was pretty worried at the City Hospital. The surgeon who treated {Jonathan/Jennifer} had pricked {himself/herself} with a used syringe needle. There should be an investigation soon.

[2] In this paper, we follow the literature (see e.g., Sturt, 2003) in assuming that so-called early and late measures in eye tracking data map onto processes that occur (respectively) in early and late stages of parsing.

Xiang et al. (2009) reported similar results in an ERP study using materials like (6), where they found that a P600 is elicited by a reflexive that mismatches the stereotypical gender of the local subject, and is not attenuated by the presence of a matching subject NP in a grammatically inaccessible position.

(6) The tough soldier that {Fred/Katie} treated in the military hospital introduced {himself/herself} to all the nurses.

Based on results from these studies, Phillips et al. (2011) suggest that:

"... *argument reflexives are immune to interference from structurally inaccessible antecedents because antecedents are retrieved using only structural cues. In effect, we are suggesting that the person, gender, and number features of reflexives like himself, herself, and themselves play no role in the search for antecedents* ..."

Evidence for earlier interference from gender-matched distractors was obtained by Badecker and Straub (2002), who reported such an effect in a word-by-word moving-window self-paced reading experiment using sentences as in (7). They found that reading times two words beyond a reflexive were slowed by the presence of a gender matching NP in a grammatically inaccessible subject position.

(7) {Jane/John} thought that Bill owed himself another opportunity to solve the problem.

Recently, Cunnings and Felser (in press) added stronger evidence for early retrieval interference in reflexive binding. They reported two eye tracking studies that evaluated how application of Principle A varies across readers with different working memory capacities. They tested low and high working memory span readers with experimental manipulations similar to those used in Sturt (2003). In the first study they found a late effect of inaccessible antecedent, emerging only at regions following the reflexive region. However, in the second study where the inaccessible antecedent was closer in the surface string to the reflexive, the effect of inaccessible antecedent was observed in an early eye movement measure, namely first fixation duration, at the reflexive itself, although this effect was limited to low span readers. Consequently, Cunnings and Felser (in press) conclude that "lower span participants were more likely to consider both potential antecedents of the reflexive early on during processing, before converging on the structurally accessible antecedent later on."

Further evidence for the effect of interference from grammatically inaccessible antecedent comes from an eye tracking study in a visual world paradigm reported by Choy and Thompson (2010) and Thompson and Choy (2009). Although this study was targeted at aphasics' processing deficits with binding constructions, for present purposes we focus on the data from unimpaired participants. Choy and Thompson (2010) recorded eye movement patterns while the participants listened to a story as in (8), with the critical sentence containing a pronoun or a reflexive (e.g., *him* or *himself*). The visual stimuli consisted of pictures of two persons, one of which was grammatically accessible and the other inaccessible (e.g., *soldier* and *farmer*); a human-referring distractor; and an inanimate-referring noun mentioned in the story (e.g., *glasses*). The data for the reflexive condition from unimpaired participants showed an increase in the proportion of fixations to the inaccessible antecedent in the reflexive and post-reflexive regions compared to the pre-reflexive region (the verb *shave* in (8)). Although the proportion of fixations to the accessible antecedent was higher than the fixations to the inaccessible antecedent in most of the regions, the increase in fixations to the inaccessible antecedent from the onset of the reflexive indicates that participants considered the inaccessible antecedent as the potential antecedent of the reflexive, albeit less often than the accessible antecedent.

(8) Some soldiers and farmers were in a house.
The soldier told the farmer with glasses to shave {himself/him} in the bathroom.
And he did.

In summary, the effect of interference from grammatically inaccessible antecedents is sometimes observed in early processing and sometimes in late processing, and in some studies the effect is completely absent. (Though we do not review the details here, evidence of interference from inaccessible antecedents is also found in English reflexives inside picture noun phrases (e.g., *Harry's picture of himself*; Runner et al., 2006) and in Chinese reflexives; (e.g., Chen et al., in press)). The goal of this paper is to provide new and arguably more sensitive empirical tests of the competing claims about the reflexive binding process, and in doing so provide further insight into the nature of the short-term memory processes that underlie incremental comprehension. We motivate the specific design of these new empirical tests by framing the question within a detailed computational model of sentence processing that crucially relies on cue-based retrieval to resolve dependencies. This is the cue-based retrieval (hereafter, CBR) architecture of Lewis and Vasishth (2005). This architecture provides one characterization of how retrieval cues are used to complete linguistic dependencies, making it a potentially useful tool for understanding the consequences of the two different theoretical positions under discussion. We turn next to a brief overview of the details of the CBR architecture.

3.4 The cue-based retrieval model

A more detailed presentation of the model is in Lewis and Vasishth (2005), Lewis et al. (2006) and Vasishth and Lewis (2006). Here we briefly describe the main features of the architecture relevant for the present discussion. The cue-based parsing theory is derived from (i) independent theory concerning the general principles of memory (McElree, 1993, 1998; McElree & Dosher, 1989), as realized in the ACT-R architecture (Anderson et al., 2004), (ii) parsing assumptions in psycholinguistics, and (iii) representational assumptions in theoretical syntax.

In ACT-R, cognitive processes can be defined for specific tasks by means of interactions between a declarative memory system and a procedural memory system. The declarative memory system serves as a long-term memory (semantic and episodic memory) but also serves to store the transient products of processing (in the context of parsing, this could be, for example, representations of phrases, and incremental trees). Each item in declarative memory, called a chunk, is a set of feature-value pairs. Procedural memory contains procedural knowledge specified in terms of *production rules*, which are condition–action associations.

In the cue-based retrieval architecture, lexical knowledge is stored in declarative memory, and grammatical knowledge is held in procedural memory as a set of production rules that specify how to apply the grammatical knowledge (the control structure) to incrementally parse sentences. Production rules are specified such that sentence parsing happens according to the left-corner parsing algorithm (Aho & Ullman, 1972). The novel structures constructed incrementally during sentence processing are stored in terms of chunks in the declarative memory. Each such chunk is an X-bar structure (Chomsky, 1986) representing a maximal projection with features corresponding to X-bar positions (specifier, complement, head) and other grammatical features such as person, number, gender, case and agreement. Sentence processing unfolds as a sequence of production rule firing, retrieval of memory chunks and update of the current parse tree. The parse tree is updated by creating new chunks and attaching them to the parse tree.

Apart from the symbolic system (i.e., procedural and declarative memory), the model's behavior depends on constraints imposed on the retrieval of chunks from memory. These constraints are defined in ACT-R in terms of a set of sub-symbolic computations that affect the activation of chunks.

The activation value influences the retrieval probabilities and retrieval latencies of chunks. The activation fluctuates as a function of frequency, recency and prior pattern of retrievals of chunks. The total momentary activation of a chunk i is given by Equation 3.1, which is the sum of base level activation (B_i), the spreading activation received through retrieval cues (the first summation component), activation received due to partial match between retrieval cues and corresponding feature values in the chunks (the second summation component), and stochastic noise (ϵ).

$$A_i = B_i + \sum_{j=1}^{m} W_j S_{ji} + \sum_{k=1}^{p} P M_{ki} + \epsilon \tag{3.1}$$

The base-level activation of a chunk is calculated in terms of Equation 3.2. Here, t_j is the time since the j^{th} successful retrieval of chunk i and d is the decay parameter.

$$B_i = \ln\left(\sum_{j=1}^{n} {t_j}^{-d}\right) \tag{3.2}$$

The spreading activation that a chunk i receives (the first summation component in Equation 3.1) is computed using W_j and S_{ji} values. W_j is typically equal to $1/m$, where m is normally the number of retrieval cues. S_{ji} is the strength of association from an element (typically a retrieval cue) j to chunk i and it is computed using Equation 3.3. Here, S is the maximum associative strength parameter and $ch3fan_j$ is the number of items associated with cue j. Associative retrieval interference arises because the strength of association from a cue is reduced as a function of the "fan" of the retrieval cue.

$$S_{ji} = S - \ln(fan_j) \tag{3.3}$$

Activation received by means of a partial match (the second summation component in Equation 3.1) is computed using P and M_{ki} over p retrieval cues. P is the match scaling parameter, and M_{ki} refers to the similarity between the retrieval cue k and the corresponding value in chunk i. The range for similarity values is specified in terms of maximum similarity and maximum difference parameters. By default, similarity between a cue and a chunk is equal to the maximum similarity value if the two are the same, and is equal to maximum difference otherwise.

Finally, the mapping from activation A_i to retrieval latency T_i for a chunk i is obtained in terms of Equation 3.4. Here F is the scaling parameter, called the latency factor. A chunk can be retrieved only if its activation is above a certain lower limit, defined in terms of the retrieval threshold parameter (τ).

$$T_i = Fe^{-A_i} \tag{3.4}$$

For present purposes, what is useful about ACT-R and the CBR theory of Lewis and Vasishth (2005) is that it provides a well-specified computational realization of the idea of memory retrieval as a noisy process of discriminating targets against a background of potentially similar distractors in short-term memory—a computational realization that has been used to build a functional (if limited) parser and applied to other sentence processing phenomena (Vasishth & Lewis, 2006; Vasishth et al., 2008).

3.4.1 Modeling reflexive binding in the cue-based retrieval framework

The cue-based retrieval architecture provides a natural characterization of the retrieval steps triggered in the process of reflexive resolution. We begin by presenting a model of Experiment 1 and its follow-up in Sturt (2003), which will provide insight into the predicted effects and their robustness against parametric variation, and provide motivation for the new paradigm used in the eye tracking study reported here. The emphasis of the model described here is not on parsing the entire sentence, but on detailed modeling of the retrieval process carried out at the reflexive.

Experiment 1 in Sturt (2003) included an eye tracking experiment in which participants were required to read short texts consisting of three sentences. An example is given in (9), showing the four experimental conditions. A named referent (*Jonathan* or *Jennifer*) is introduced in the first sentence, and this referent is subsequently referred to using a pronoun (*he* or *she*) in the second sentence. The second sentence also introduces a second referent *the surgeon*, and includes a reflexive anaphor (*himself* or *herself*). The first named referent is not a grammatically accessible antecedent for the reflexive in terms of binding theory, while the second referent (*the surgeon*) is a grammatically accessible antecedent. Accessible and inaccessible antecedents either matched or did not match the gender of the reflexive.

In conditions (a) and (b) in (9) the accessible antecedent matches the gender requirement of the reflexive and in conditions (c) and (d) in (9) it does not. Furthermore, in conditions (a) and (c) the inaccessible antecedent matches the gender of the reflexive. Henceforth, we will refer to conditions (a) and (b) as *match* conditions, and conditions (c) and (d) as *mismatch* conditions, reflecting the fact that the accessible antecedent matches the gender of the reflexive for one pair and does not for the other. We will refer to conditions (a) and (c) as the *interference* conditions because the gender of the inaccessible antecedent matches that of the reflexive—potentially causing interference.

(9)

Sentence 1: {Jonathan/Jennifer} was pretty worried at the City Hospital.
Sentence 2:

a. *Accessible-match/inaccessible-match* (*Match-interference*)
He remembered that the surgeon had pricked **himself** with a used syringe needle.

b. *Accessible-match/inaccessible-mismatch* (*Match*)
She remembered that the surgeon had pricked **himself** with a used syringe needle.

c. *Accessible-mismatch/inaccessible-match* (*Mismatch-interference*)
She remembered that the surgeon had pricked **herself** with a used syringe needle.

d. *Accessible-mismatch/inaccessible-mismatch* (*Mismatch*)
He remembered that the surgeon had pricked **herself** with a used syringe needle.

Sentence 3: There should be an investigation soon.

This eye tracking study showed an early effect of the accessible antecedent (Figure 3.1). First fixation duration and first pass reading time were faster when the gender of the anaphor matched the gender of the accessible antecedent than when they did not. But no effect of the inaccessible antecedent was found in these early measures. In one of the late measures, second pass reading time, longer reading times were found when the inaccessible antecedent matched the reflexive in gender (this was the case only when the accessible antecedent also matched the reflexive in gender).

As mentioned above, Sturt (2003) also conducted a follow-up study to find out the participants' final interpretation of the reflexive. This was a sentence-by-sentence self-paced reading with the same sentences as in (9) but, instead of sentence 3, there was a question that explicitly probed for the antecedent of the reflexive (e.g., a question like *Who had been pricked with a used needle?* with possible answers, for example, for condition (a) as *Jonathan* or *surgeon*). The follow-up study showed a main effect of accessible antecedent, inaccessible antecedent and also an interaction between these two factors (see Figure 3.1). When the accessible antecedent did not match the gender of the reflexive, participants made a higher proportion of errors in selecting the correct antecedent. In addition, when the inaccessible antecedent matched the gender of the reflexive participants made more errors than when it did not. Moreover, the increase in error due to gender match with the inaccessible antecedent was larger when the accessible antecedent did not match the gender, resulting in the interaction between the two factors.

Thus, there are four major findings in Sturt's Experiment 1 and his follow-up study. First, gender mismatch with the default gender of the accessible antecedent resulted in lower question-response accuracies. Second, gender match with the inaccessible antecedent resulted in lower question-response accuracies. Third, early reading time measures (first fixation duration and first pass reading time) increased when the gender specification of the accessible antecedent mismatched that of the reflexive. Fourth, second pass reading time (re-reading time) was *longer* when the gender of the inaccessible antecedent matched the gender of the reflexive (this occurred in the case where the accessible antecedent matched the reflexive in gender).

Interestingly, the majority (the first three of four) effects can be explained by simply assuming that the search for an antecedent includes a gender feature. Therefore, we begin our modeling by assuming that both grammatical knowledge about antecedent resolution and gender matching is used when resolving antecedents in English. For simplicity, we model the grammatical constraint by assuming that the antecedent should be a noun and should be the subject of the clause containing the reflexive.[3] This choice of re-

[3]There are more sophisticated ways to encode the c-command constraint but these implementation details are orthogonal to the present discussion. For example, as syntactic

Table 3.1: The match of retrieval cues with the accessible and inaccessible antecedents for the four conditions in Sturt's experiments (cat=category).

Conditions	Accessible	Inaccessible
a (match-interference)	gender, cat, role, clause	gender, cat, role
b (match)	gender, cat, role, clause	cat, role
c (mismatch-interference)	cat, role, clause	gender, cat, role
d (mismatch)	cat, role, clause	cat, role

trieval cues is motivated by the conjecture that including agreement features in general may be an adaptive feature of the parser, though attempting to establish this is not the purpose of the work reported here. As a result the set of retrieval cues for the reflexives *himself* and *herself* are {gender = masculine/feminine, category = noun, role = subject, clause = current-clause}, differing only in the value of the *gender* feature. See Table 3.1 for the list of cues matched by the two antecedents across the four conditions in Sturt's experiments.

The accessible antecedent matches all four cues in conditions (a) and (b) (the 'match' conditions), but only three cues in conditions (c) and (d) (the 'mismatch' conditions), since the stereotypical gender of *surgeon* is masculine which does not match the gender retrieval cue at *herself* (gender = feminine). The inaccessible antecedent matches three cues (gender, category, role) out of a total of four cues in conditions (a) and (c), and in conditions (b) and (d) it matches two cues (category, role). As a result, interference for retrieving the antecedent will be higher in conditions (a) and (c) (the 'interference' conditions) as compared to conditions (b) and (d). Note that the alternative possibility, as suggested by Phillips, Dillon, and colleagues, is that gender

trees are incrementally built, each node built could be associated with an incrementally expanding table that tracks its c-commanders. This would allow the parser to dynamically mark, by table lookup, all c-commanding nodes for any given node; the c-commanders would then be marked as such, and this dynamically specified feature could be used when setting retrieval cues at a particular point in the sentence. Implementing this idea is technically interesting, but the end-product of such a dynamic c-command marking would be identical to the assumptions we make here about features used in retrieval.

plays no role in retrieval; in that case, the cues for the match (a vs. b) and mismatch conditions (c vs. d) would be identical, leading to no interference.

The cue-based retrieval model predicts that similarity-based interference (SBI) arises at the moment of retrieval. SBI in reflexive binding is manifested in terms of delay in retrieval of the correct antecedent or an error in retrieving the correct antecedent. The delay in retrieval of the correct antecedent is a result of the fan assumption (see Equation 3.3) that reduces the strength of association between a cue and a target as a function of the number of items associated with that cue. Reduced strength of association means reduced activation boost, which produces higher latencies. On the other hand, the error in retrieval of the correct antecedent is a combined result of activation fan and partial match. Reduction in activation boost of the accessible antecedent due to activation spreading, and partial matching between retrieval cues (the second summation component in Equation 3.1) and any inaccessible antecedents can lead (probabilistically as a function of activation noise) to higher activation of the inaccessible antecedents. As a result, the probability of retrieving the inaccessible antecedent increases. The greater the partial match with inaccessible antecedents, the higher the percentage of errors in retrieving the accessible antecedent.

We model retrieval in sentence 2 from (9); this is the crucial sentence for generating predictions about the reflexive binding process. The predictions of the model are generated by running 1000 simulations for each condition. All model parameters are set to the values that have been used in the previous models from Lewis and Vasishth (2005); Vasishth and Lewis (2006); Vasishth et al. (2008). A list of all the parameter values that we use is given in Table 3.2.

The predicted retrieval error percentages accurately capture the pattern found in the Sturt (2003) follow-up study: There is a main effect of accessible antecedent, inaccessible antecedent, and an interaction between these two factors, exactly as in Sturt's follow-up study's response accuracies. First, when the accessible antecedent does not match the gender of the reflexive the

Table 3.2: The list of parameter values used in the previous studies with CBR models and the values used in the current model.

Parameter	Previous models	Current model
Noise (ϵ)	0, 0.15, 0.30, 0.45	0.30
Decay (d)	0.50	0.50
Maximum associative strength (S)	1.50	1.50
Match scale (P)	1	1
Maximum similarity	0	0
Maximum difference	−0.60	−0.60
Latency factor (F)	0.14, 0.46	0.14
Retrieval threshold (τ)	−1.50	−1.50

model makes a higher number of errors in retrieving the correct antecedent (the mismatch effect in response accuracy). Second, when the inaccessible antecedent matches the gender of the reflexive the model makes more errors than when it does not (the interference effect in response accuracy). Third, the increase in error due to gender match with the inaccessible antecedent is greater in the mismatch conditions.

The retrieval times predicted by the model (shown in Figure 3.2) show a main effect of matching in the accessible antecedent: When the accessible antecedent does not match the gender of the reflexive, the retrieval times are higher than when it does. The model also predicts a match $\times$ interference interaction—retrieval times are predicted to be higher in the *match-interference* condition (198 ms) than in the *match* condition (194 ms); however, retrieval times are predicted to be lower in the *mismatch-interference* condition (274 ms) compared to the *mismatch* condition (295 ms).

In order to compare the predictions to the data, we use the following terminology: the *mismatch effect* is the difference between the *match* conditions and the *mismatch* conditions; the *interference effect* is the effect between the two interference conditions and the other two conditions; the *match-interference effect* is the effect of interference in the two *match* conditions; and the *mismatch-interference effect* is the effect of interference in the

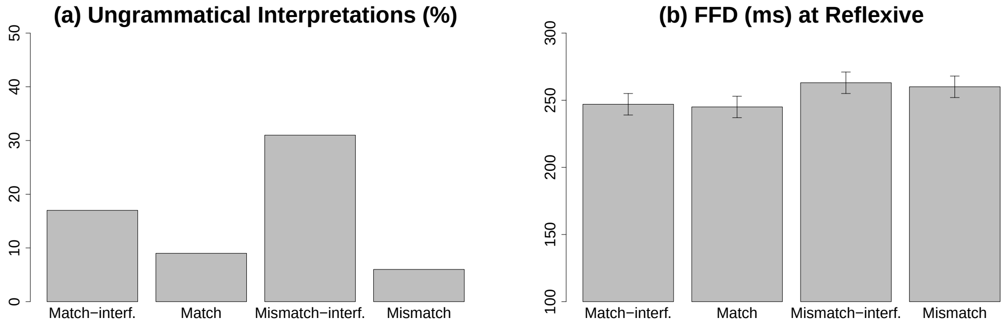


Figure 3.1: The effects in Sturt 2003 Experiment 1: (a) proportions of ungrammatical interpretations of reflexives in the follow-up study; (b) first fixation durations in the eye tracking study. The error bars in the plot for first fixation duration show 95% confidence intervals.

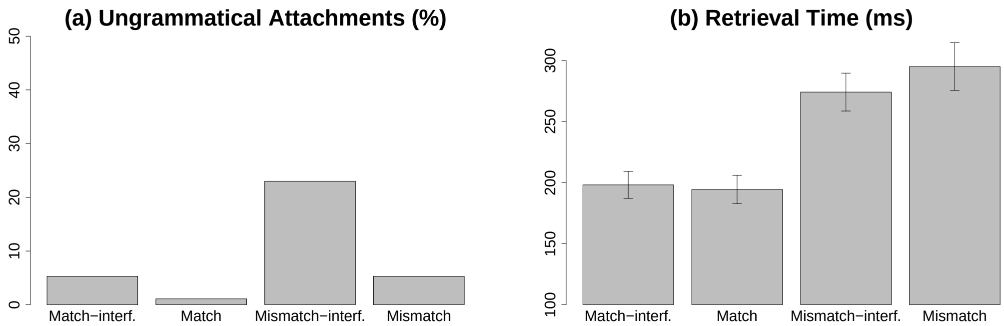


Figure 3.2: The twofold predictions of the cue-based retrieval model for Experiment 1 in Sturt 2003, based on the parameter values listed in Table 3.2. The error bars in the plot for retrieval time show 95% confidence intervals.

two *mismatch* conditions.

The predicted ungrammatical attachments accurately model the ungrammatical interpretations observed in the Sturt (2003) follow-up study. However, the predicted retrieval times accurately capture only the mismatch effect observed in the first fixation duration (FFD) in the eye tracking study. The interaction predicted between the *mismatch* effect and *interference* effect is not observed in the data. Thus, the model accurately captures the question-response accuracy data, but only partly characterizes the first fixation duration data.

The divergent patterns between the model's retrieval times and the first fixation durations in Sturt's study come from the differences in the patterns seen in the predicted match-interference effect and the mismatch-interference effect. The predicted match-interference effect is a consequence of spreading of activation of the gender cue which is matched by both the accessible and inaccessible antecedent. As described earlier, activation spreading reduces the strength of association between the cue and the target, causing longer retrieval latencies in the match-interference condition than the match condition. On the other hand, the mismatch-interference effect is a consequence of partial match between the cues and inaccessible antecedent: the inaccessible antecedent matches the gender cue which is not matched by the accessible antecedent (see Table 3.1), leading to higher probability of retrieving the inaccessible antecedent. This can be seen in the predicted retrieval error percentages in Figure 3.2. Moreover, in the *mismatch-interference* condition the inaccessible antecedent receives more activation from retrieval cues than in the *mismatch* condition as it matches more retrieval cues in the *mismatch-interference* condition. A substantially higher number of incorrect retrievals occur due to higher activation from the retrieval cues, and the retrieval times in the *mismatch-interference* condition are faster than the retrieval times in the *mismatch* condition, contrary to the findings in Sturt's study (see Figs. 1 vs. 2). We return to this issue in the discussion section.

To summarize, the model predicts the following effects for retrieval errors (RE) and retrieval times (RT) at the reflexive:

E1. Mismatch effect (RE): the retrieval errors for mismatch conditions are higher than for match conditions.

E2. Interference effect (RE): the retrieval errors for match-interference and mismatch-interference conditions are higher than the other two conditions.

E3. Mismatch effect (RT): the retrieval times for mismatch conditions are longer than for match conditions.

E4. Match-interference effect (RT): the retrieval times for the match-interference conditions are longer than the match condition.

E5. Mismatch-interference effect (RT): the retrieval times for the mismatch-interference conditions are shorter than the mismatch condition.

Only the effects E1, E2 and E3 were observed in Sturt's experiment; the interference effects E4 and E5 were missing in the early measures (first fixation duration and first pass reading time) of the Sturt (2003) eye tracking study.

3.4.2 Parametric variability in the model

We did not estimate any parameter values for the current model. All existing parameters were set to the values that have been used in previous published versions of the cue-based retrieval model. It is possible, however, that the predictions of the model are valid only for the specific parameter values that we used here; this could be the reason behind the lack of effects E4 and E5 in the data—these effects might emerge only for a particular combination of parameter values. Conversely, the correct predictions of effects E1, E2 & E3 might depend on the specific values used by the model. To gain a better understanding of the range of possible predictions of the model, we ran the

Table 3.3: The range of parameter values used for testing the parametric variability of the cue-based retrieval models.

Parameter	Range of values
Noise	0.05 to 0.4, in steps of 0.05
Maximum associative strength	1 to 4, in steps of 0.25
Maximum difference	-1 to 0, in steps of 0.1

model for a range of values of three crucial ACT-R parameters: *noise*, *maximum associative strength* and *maximum difference*. The noise parameter controls the amount of instantaneous activation noise added to each chunk at retrieval; maximum associative strength is the constant 'S' in Equation 3.3; and the maximum difference parameter controls the penalty due to a mismatch between a retrieval cue and a feature value of a chunk. For each of these parameters, the range of values over which the predictions are generated is given in Table 3.3. The predictions are generated by running 1000 simulations for each combination of values of the three parameters. The total number of combinations of the three parameter values are 1287 (see Table 3.3). The predictions of effects E1–E5 across these sets of parameter values are plotted in Figure 3.3. Each effect is plotted against the parameter along which it varies the most. Effects E1 and E2 are influenced the most by noise, E3 and E5 are influenced the most by the maximum difference parameter, and E4 is influenced the most by the maximum associative strength parameter. Each point in the plots represents a mean over all values of the other two parameters. In short, Figure 3.3 illustrates how the size of each effect varies across different parameter values.

The effect E1 varies from 0% to 23.9%, the effect E2 varies from 0% to 17.15%, the effect E3 varies from -1.55 ms to 228.4 ms, the effect E4 varies from -6.53 ms to 17.34 ms and the effect E5 varies from from -36.01 ms to 3.21 ms. The effects E1 and E2 are zero when the instantaneous activation noise is zero, which essentially means that the model doesn't make any mistake in retrieving the accessible antecedent when there is no noise added to

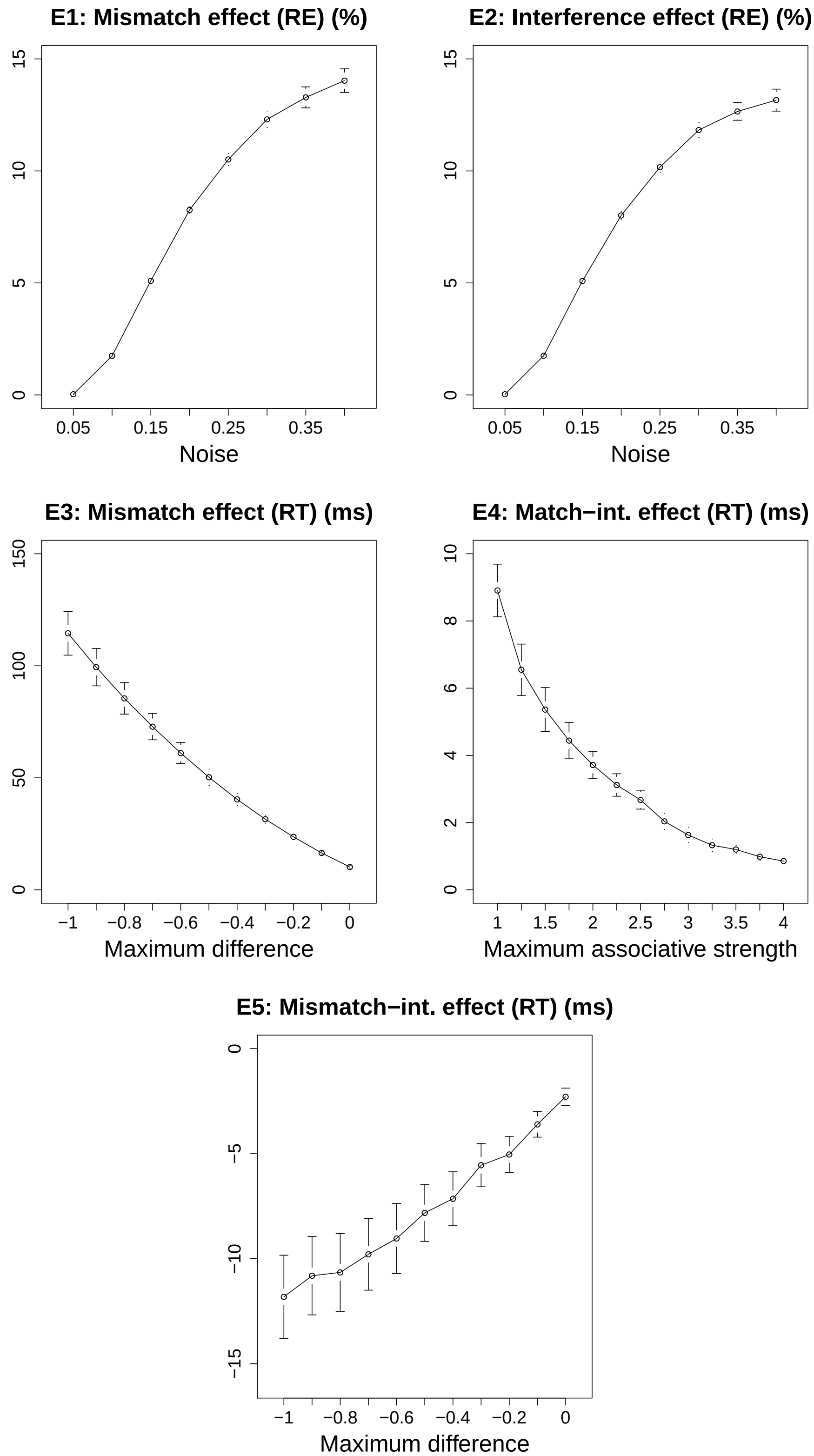


Figure 3.3: The predictions of effects E1–E5 for Experiment 1 from Sturt 2003 across a range of parameter values. Each effect is plotted against the parameter that affects it the most.

the activations of chunks. But, in general, a non-zero value for noise parameter is necessary for modeling memory errors and response time distribution. Overall, although all effects show variation across different parameter combinations, they all remain mainly non-zero and have the same numerical sign as the predicted effects with the predefined parameter values. These results show that the model's predictions for E1-E3 are not crucially dependent on the fixed parameter values we used.

3.4.3 An alternative explanation for the absence of interference effects (E4 & E5) in Sturt 2003

Although the lack of an interference effect in Sturt (2003) Experiment 1 could imply that non-structural cues like gender are not used in retrieval, Sturt noticed that the absence of an effect could be due to the non-local linear position of the interferer (inaccessible antecedent) with respect to the reflexive (see (4) above). The accessible antecedent was introduced later in the string than the inaccessible antecedent, and was therefore closer to the reflexive. In his Experiment 2, Sturt (2003) modified this design by using stimuli as in (5), where the linear positions of the binding accessible and inaccessible antecedents are reversed with relation to Experiment 1, while their accessibility with respect to the binding theory is kept constant. However, this experiment also did not show any interference effect.

In addition to surface-string locality, we consider now another possibility for the apparent lack of interference: the degree of overlap between potential distractors and retrieval cues. We hypothesized above that reflexive binding uses grammatical category (noun, verb etc.), grammatical role (subject, object etc.) and gender as the retrieval cues to retrieve the correct antecedent. In the CBR model, the overlap of these cues with grammatically inaccessible antecedents leads to an interference effect in both retrieval latency and retrieval errors. This formulation in the model leads to the following alternative explanation for the lack of interference effect in Sturt's Experiment 2: the interfering antecedent was the object of the relative clause (see (4) above), and hence did not match the grammatical role cue for retrieval at the reflexive. In fact, Van Dyke and McElree (2011) have recently proposed that although distractors with matching semantic cues exert interference, cues like grammatical role are weighted heavily in the retrieval process. They found that the interference effect due to the semantic match was present only when the distractors matched the grammatical cues as well. These results can also explain the lack of interference effect in Sturt's Experiment 2.

In terms of activations of memory elements, the probability of retrieving an incorrect element is higher if it has a higher activation value at the time of retrieval. The activation value of a memory element is directly dependent on its creation time, retrieval history, and its match with the retrieval cues—the more recently an element is created or retrieved, and the higher feature overlap it has with the retrieval cues, the better chances it has of being retrieved. Consequently, in Sturt's Experiment 1 the interferer has less chance of getting retrieved due to its less recent creation time with respect to the accessible antecedent, and in Experiment 2 the interferer has less chance of getting retrieved because its overlap with the retrieval cues is lower in comparison to the overlap of the accessible antecedent with the retrieval cues. In other words, in Sturt's experiments the inaccessible antecedents may not be strong enough interferers to detect their effect on the retrieval process. If this reasoning is correct, then the effect, or rather the lack of an interference effect, might be a false negative (a type II error). Concluding that an absence of an interference effect is evidence that no interference occurs has important consequences to the theory of retrieval processes in sentence comprehension. No interference in processing argument reflexives implies that the retrieval mechanism for reflexive binding is different from other retrieval mechanisms in sentence processing, e.g., subject-verb agreement, and agreement attraction. On the other hand, finding an interference effect simplifies the theory of retrieval processes considerably, since no exemption is granted to antecedent-reflexive resolution processes.

3.4.4 A new design

In order to increase the strength of the interference effect, we can use an object relative clause (see (10)) where the inaccessible antecedent has the subject role in the clause. It is also closer to the reflexive in terms of linear distance. Under the CBR account, the inaccessible antecedent would be more likely to interfere in the retrieval process than in the two experimental designs in Sturt (2003) — but under the structurally-constrained approach, this manipulation should not matter to the reflexive binding process. In fact, Xiang et al. (2009) used this design in their ERP study, but they left out the crucial *match* condition.

(10) a. *Accessible-match/inaccessible-match (match-interference)*
The tough soldier that Fred treated in the military hospital introduced himself to all the nurses

b. *Accessible-match/inaccessible-mismatch (match)*
The tough soldier that Katie treated in the military hospital introduced himself to all the nurses

c. *Accessible-mismatch/inaccessible-match (mismatch-interference)*
The tough soldier that Katie treated in the military hospital introduced herself to all the nurses

d. *Accessible-mismatch/inaccessible-mismatch (mismatch)*
The tough soldier that Fred treated in the military hospital introduced herself to all the nurses

We implemented a CBR model for the new design described in (10) as well as for Experiment 2 in Sturt (2003). The goal of this modeling is to compare the predictions of the CBR theory for the five effects (E1–E5) across three designs—Experiment 1 (including the follow-up study; we count the eye tracking study and the follow-up study as one experiment, following Sturt), Experiment 2 from Sturt (2003), and the new design. The modeling assumptions are the same as in the model described above.

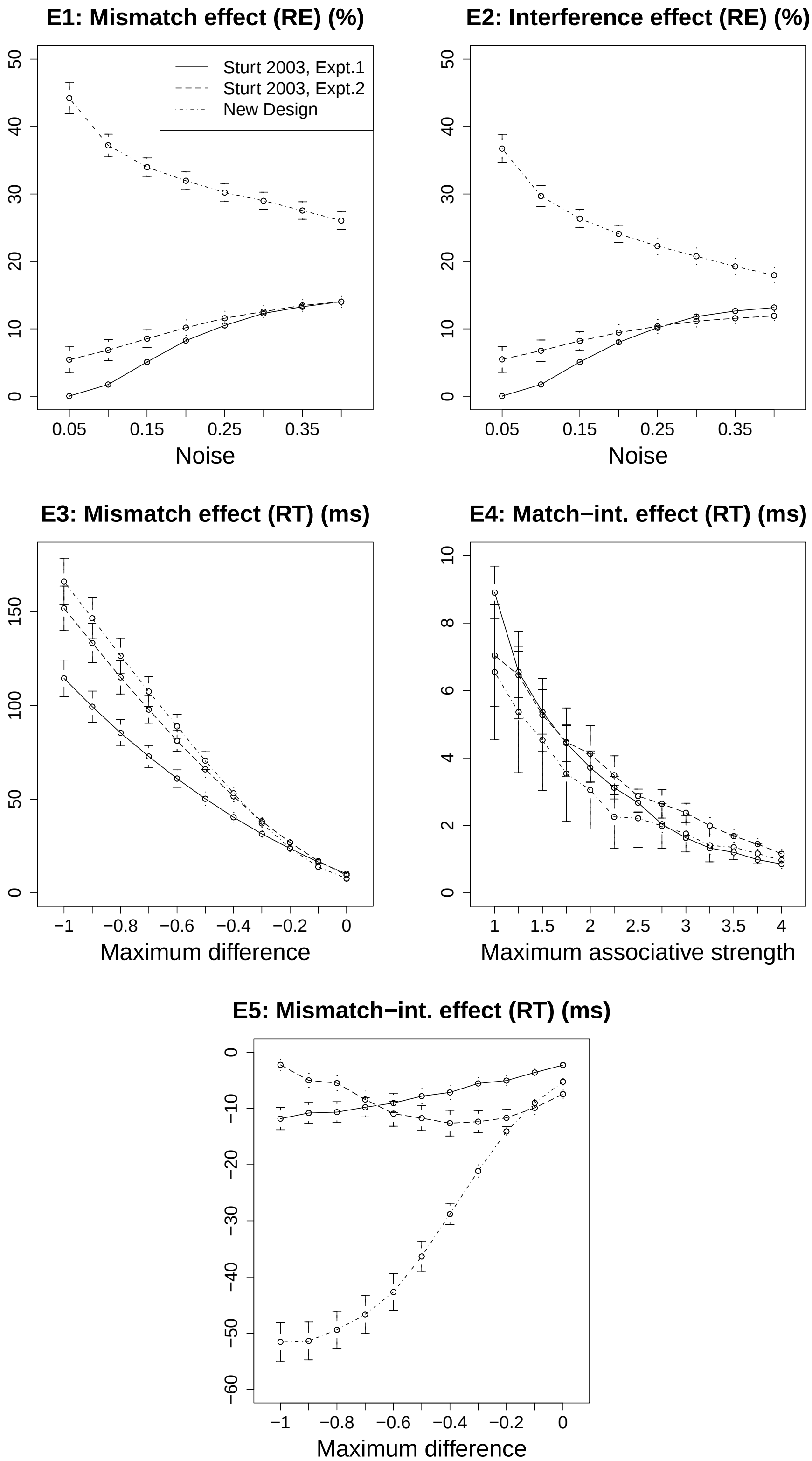


Figure 3.4: The predictions of effects E1–E5 for three experimental designs across a range of parameter values.

Figure 3.4 compares the predictions for effects E1–E5 across the three experimental designs. The predictions are generated for a range of parameter values listed in Table 3.3. The pattern of effects E1–E5 for Experiment 2 and the new design are similar to that for Experiment 1. Across a range of parameter values, the predictions for effects E1, E2 and E5 are clearly stronger (higher numerical value) for the new design than for Experiment 1 and 2 in Sturt (2003). Although the predictions for effect E3 are almost identical for the new design and Experiment 2, they are nevertheless stronger than for Experiment 1. In contrast, the predictions for effect E4 are not distinguishable across the three designs. To gain better insight into the predictions for E4, we compared effect E4 across variations of the other two parameters (noise and maximum difference); see Figure 3.5. For the maximum difference parameter, effect E4 is stronger in the new design and Experiment 2 than in Experiment 1 when the difference penalty is high (more negative), and it is weaker when the maximum difference penalty is low. For the noise parameter, effect E4 is stronger in the new design and Experiment 2 than in Experiment 1 when the noise is low, and it is weaker when the noise is high. These patterns show that the predicted strength of effect E4 is dependent on the specific value or a range of values that are selected for these parameters. The noise parameter is a frequently modified parameter across various models (Wong, Cokel, & Schooler, 2010), which is suggestive of uncertainty regarding its value across diverse cognitive tasks (cf. the decay parameter, which is usually kept fixed). The best way to estimate or, at least, restrict the noise parameter's value would be to empirically validate predictions of various models. In contrast to noise, the maximum difference penalty parameter is seldom modified, and is set to its default value of -1. For the default value of this parameter, the model clearly predicts a stronger E4 effect for the new design and Experiment 2.

In sum, the predictions for the new design and Experiment 2 show stronger effects than Experiment 1 across a range of parameter values. For the most part—and as expected—the effects for the new design are much stronger

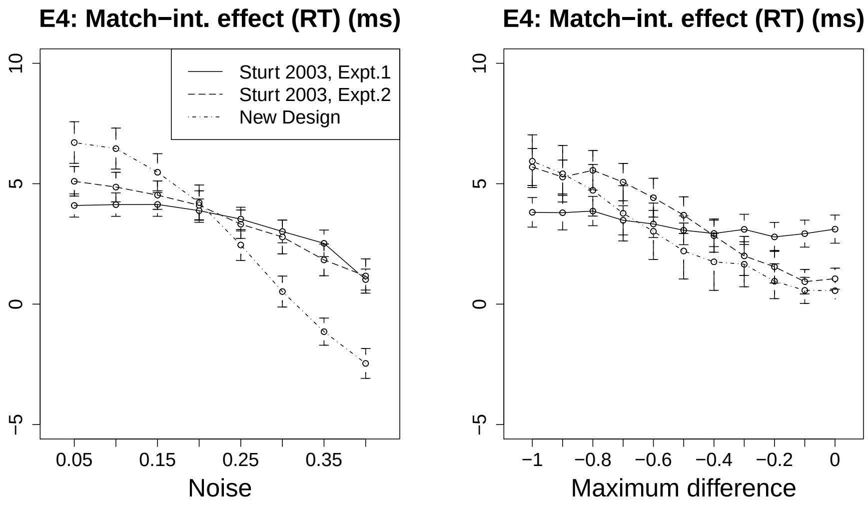


Figure 3.5: The variations in effect E4 (match-interference (RT)) across noise and maximum difference parameters for three experimental designs.

than the other two designs. Next, we report an eye tracking study that we ran with the new design (10). The goal of this study is to evaluate the predictions of our model that diverge from Sturt's findings (specifically, effects E4 and E5), as well as to replicate the effects E1-E3 that Sturt (2003) found.

3.5 Eye tracking experiment

3.5.1 Participants

Forty English native speakers residing in Berlin, Germany participated in the eye tracking study. Data from one participant was excluded due to less than 40% accuracy on the sentence comprehension questions on all trials including experimental and filler trials. The remaining 39 participants included 20 female participants and had a mean age of 29.5 years. The 39 native English speakers consisted of 14 British, 13 American, 8 Australian, 3 Canadian and 1 New Zealander. All participants had normal or corrected-to-normal vision and were paid 10 Euros for their participation. The experiment had a duration of approximately 45 minutes, including set-up time.

3.5.2 Design and materials

Twenty-four stimuli were selected from the Xiang et al. (2009) study and constructed as per (9) by adding an extra *match* condition (see Appendix B for the list of stimuli). Of these 24 stimuli, 12 used stereotypically male nouns and 12 used stereotypical female nouns for the binding-accessible antecedent. There were 4 lists that comprised different item-condition combinations according to a Latin Square. Each list contained 54 filler sentences. Two-third of the target items and all fillers contained a comprehension question, and these were equally distributed across yes and no answers. In all, each participant answered 70 comprehension questions.

3.5.3 Procedure

Participants were seated 60 cm from an NEC Multisync 2080UX screen color monitor with 1600 $\times$ 1200 pixel resolution. They were asked to sit comfortably in front of an EyeLink 1000 eye tracker (SR Research) running at 500 Hz sampling rate (0.01° tracking resolution, and $<$ 0.5° gaze position accuracy). Though the viewing was binocular, only the participant's right eye was tracked. The distance between the camera and the eye was 50 cm.

Participants were asked to position their head in a frame that stabilized their forehead and chin. They were asked to avoid large head movements throughout the experiment and to avoid blinking while reading the sentences. A 7-button Microsoft Sidewinder game pad was used to record button responses. The presentation of the materials and the recording of responses was controlled by two separate PCs, one running internally developed software (this is called EyeScript, and was originally developed in Richard Lewis' lab by Mason Smith, and later in Shravan Vasishth's lab by Felix Engelmann, Titus von der Malsburg, and Tobias Günther; the software is open source and available from the authors) and the other running SR

Research's proprietary software.

Each participant was randomly assigned one of four different stimulus lists. The list was randomized for every subject. At the start of the experiment, a standard calibration procedure was performed which involved participants looking at a grid of 13 fixation targets in random succession, in order to validate their gazes. Calibration and validation were repeated if the experimenter noticed that measurement accuracy was poor, and if participants took a break during the experiment.

Each trial consisted of the following steps: First, a fixation target in the same position as the first character of the text display was presented; two 200 ms fixations followed by one 400 ms fixation on this target triggered the presentation of the sentences (this procedure ensured that the participants always started reading in the left-most character position and helped the experimenter ensure the accuracy of calibration). Participants were instructed to read the sentence at a normal pace and to move their gaze to a dot at the bottom right of the screen after finishing the sentence. This triggered the presentation of a comprehension question in two-thirds of the trials, and in the rest it triggered the presentation of the next trial. The comprehension questions were included in order to ensure that the participants attended to the content of the sentences.

3.5.4 Data analysis

All data processing and analyses were carried out in GNU-R (R Development Core Team, 2009). Fixations were detected using the algorithm described by Engbert and Kliegl (2003); an open source R package, *saccades*, developed by Titus von der Malsburg was used to carry this step out. Fixation and regression-based measures were extracted using another open source R package, *em*, written by Logačev and Vasishth (2006). All fixations 30 pixels above and below the sentence were included in the sentence. Fixations in the blank spaces between words were also counted; fixations in the first half of the space were included in the fixations on the preceding word and fixations in the second half were included in the fixations on the following word. All other fixations outside these regions were excluded.

Effects of accessible antecedent gender match (henceforth, *match*) and inaccessible antecedent gender match (*interference*) with the gender of the reflexive were evaluated across various eye movement measures. Data analysis was carried out using linear mixed models (Bates & Sarkar, 2007; Gelman & Hill, 2007). Linear mixed models were fit for the following eye movement measures at the reflexive. *Single Fixation Duration* (SFD), the time spent in a region during first pass if it is fixated only once; *First Fixation Duration* (FFD), the time spent during the first fixation during the first pass; *First Pass Reading time* (FPRT), the sum of all fixations during the first pass; *Re-reading Time* (RRT), the sum of all fixations in a region that occurred after first pass (including zero RRTs); *Total Reading Time* (TRT), the sum of all fixations in a region; and *First Pass Regression Probability* (FPRP), the probability of regressing from a region after fixating in that region during first pass. Apart from *match* and *interference*, trial number was used as a (centered) predictor in the linear models. An interaction term for *match* and *interference* was also included in the models. Participants and items were crossed random intercepts. All reading times were log transformed before fitting the linear models. For FPRP, a generalized linear mixed models was fit with a binomial link function.

3.5.5 Results

All mean reading times at the reflexive along with standard errors, FPRP from the reflexive and comprehension accuracy percentages are summarize in Table 3.4. The results of the statistical analysis are summarized in Table 3.5 and Table 3.6.

Table 3.4: Mean reading times at the reflexive with standard errors, percentages of first pass regressions from the reflexive, and comprehension question response accuracies across four conditions.

Condition	SFD	FFD	FPRT	RRT	TRT	FPRP	Accuracy
a. match-interf.	252 (8)	258 (6)	280 (8)	108 (14)	410 (18)	13.10%	83.97%
b. match	260 (8)	263 (7)	292 (10)	79 (11)	389 (17)	6.63%	89.74%
c. mismatch-interf.	279 (12)	272 (8)	295 (10)	149 (20)	473 (24)	11.17%	82.69%
d. mismatch	260 (10)	266 (9)	284 (9)	145 (15)	468 (20)	10.47%	79.49%

Table 3.5: Linear mixed-effects model estimates, standard errors and t-values across reading time measures; 'Interference:Match' denotes the interaction term for the two fixed effects.

	Effect	Estimate	Std. Error	t-value
SFD	Intercept	5.518	0.037	148.566
	Trial	0.000	0.001	0.336
	Interference	−0.037	0.039	−0.945
	Match	−0.013	0.043	−0.304
	Interference : Match	0.079	0.060	1.325
FFD	Intercept	5.511	0.034	163.546
	Trial	0.001	0.001	1.700
	Interference	−0.022	0.032	−0.698
	Match	0.002	0.032	0.061
	Interference : Match	0.044	0.045	0.973
FPRT	Intercept	5.585	0.036	155.614
	Trial	0.001	0.001	2.163 *
	Interference	−0.033	0.036	−0.935
	Match	−0.020	0.036	−0.545
	Interference : Match	0.063	0.051	1.247
RRT	Intercept	−5.656	0.637	−8.884
	Trial	−0.019	0.009	−1.989
	Interference	0.904	0.591	1.530
	Match	2.759	0.591	4.671 *
	Interference : Match	−2.027	0.835	−2.427 *
TRT	Intercept	5.809	0.057	102.112
	Trial	0.000	0.001	0.475
	Interference	0.036	0.048	0.751
	Match	0.171	0.049	3.514 *
	Interference : Match	−0.052	0.068	−0.766

Table 3.6: Linear model estimates, standard errors and p-values for FPRP; 'Interference : Match' denotes the interaction term for the two fixed effects.

	Effect	Estimate	Std. Error	p-value
FPRP	Intercept	−3.067	0.360	0.000
	Trial	0.004	0.005	0.430
	Interference	0.782	0.377	0.038 *
	Match	0.483	0.396	0.223
	Interference : Match	−0.671	0.515	0.193

Question-response accuracy

Overall average accuracy for trials that included a comprehension question was 88.32% and average accuracy for target items was 83.97%. Accuracy values for comprehension questions across four conditions are listed in Table 3.4, but are not theoretically interpretable because the questions targeted different parts of the critical sentence, not just the antecedent-reflexive relation as in the Sturt follow-up study. We present these mean accuracies only for completeness.

Eye tracking dependent measures

A statistically significant main effect of *interference* was observed in FPRP, with the high interference conditions showing more regressions than the low interference conditions. A main effect of *match* was significant for RRT and TRT; i.e., the conditions in which the stereotypical gender of the accessible antecedent did not match the gender of the reflexive were read more slowly than the conditions where it matched. For the other eye movement measures, the main effect of *interference* and *match* and their interaction did not reach significance.

3.5.6 Discussion

Early effects

The results outlined above show an early effect of interference (effects E2 and E4) from the inaccessible antecedent in first pass regression probability, such that a gender match between the reflexive and the inaccessible antecedent leads to a higher number of first pass regressions from the reflexive (see Figure 3.6). The occurrence of a regression from a word reflects some difficulty in integrating the word when it is fixated and hence it is plausibly an early effect (Clifton et al., 2007). First pass regressions cannot reflect the late processes triggered at the end of a sentence or the processes reflected by late measures

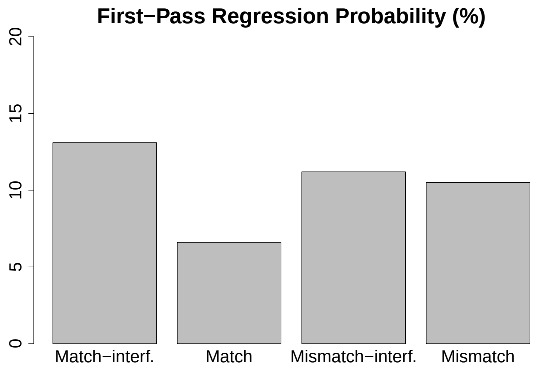


Figure 3.6: Early effect of *interference*.

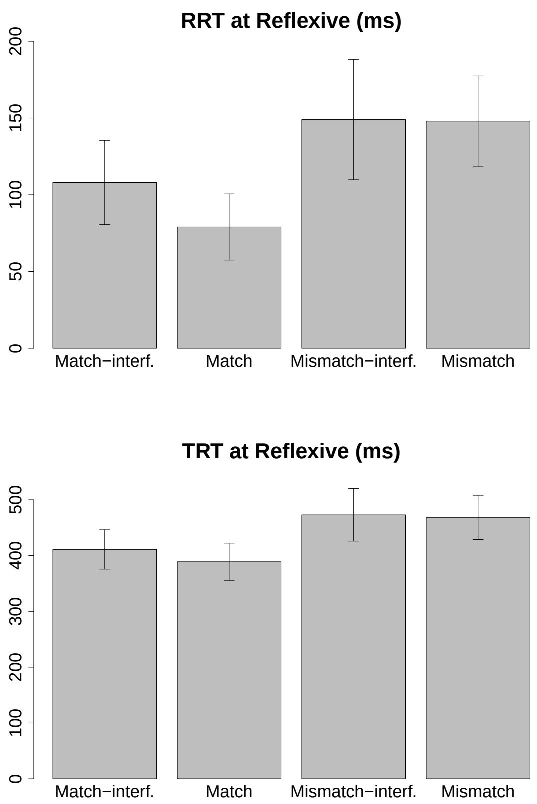


Figure 3.7: Late effect of *match*.

such as second pass reading time. Assuming that first pass regressions reflect processing difficulty triggered relatively early during the first contact with the critical word, the interference effect is inconsistent with the conclusion of Sturt (2003), that the online application of Principle A is not affected by interference from the inaccessible antecedent at early stages of processing. Conclusions derived in Nicol and Swinney (1989) and Xiang et al. (2009) are also not compatible with these results. As a result, this study challenges the claim from Phillips et al. (2011) and Dillon (2011) that an antecedent for a reflexive is retrieved using only structural cues without considering the gender feature. Our findings are of course consistent with those of Badecker and Straub (2002), Choy and Thompson (2010), Cunnings and Felser (in press), Thompson and Choy (2009).

Regression contingent effects in FFD

As an exploratory data analysis, we analyzed FFD contingent on the first pass regressions—separate analysis for FFD followed by regressions and FFD not followed by regressions. The two patterns are plotted in Figure 3.8. FFD followed by regressions show a pattern consistent with the retrieval times predicted by the model (see Figure 3.2). Although the match-interference effect (E4) ($t=-1.77$) and mismatch-interference effect (E5) ($t=1.70$) do not quite reach conventional significance levels, they crucially show the trend of interference effect in the opposite direction between match and mismatch conditions, as predicted by the model. These FFDs also show the main effect of mismatch (E1 and E3) ($t=2.78$) which is consistent with the early mismatch effect in Sturt (2003). FFD not followed by regressions did not show this effect.

Late effects

The effect of accessible antecedent gender match (E1 and E3) was also observed in the RRT and TRT (see Figure 3.7) such that these reading times

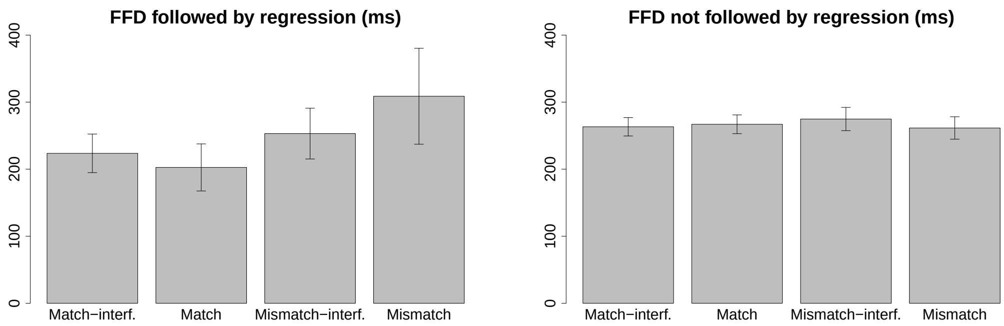


Figure 3.8: Regression contingent FFD.

were elevated when the accessible antecedent did not match the gender of the reflexive. The absence of an early effect of accessible antecedent is different from the finding of Sturt (2003), where the effect appeared at FFD.

In sum, the eye tracking study supported the predictions of the cue-based retrieval model of reflexive binding. The interference effects E2 and E4 were observed in first pass regression probability. A clear trend for the interference effects E4 and E5 was observed in regression contingent first fixation duration. The presence of effect E4, which was absent in Sturt's experiments, is consistent with the model's predictions with a higher maximum difference penalty (more negative values) and lower noise (see Figure 3.5). The eye tracking study also replicated the mismatch effects, E1 and E3, found in Sturt's experiments, in terms of re-reading time, total reading time and regression contingent first fixation duration.

3.6 General discussion

In this paper, we investigated the question: what kinds of cues are used initially by the parser when resolving antecedent-reflexive relations? The two positions on this question are: structural cues only (Nicol & Swinney, 1989; Sturt, 2003; Phillips et al., 2011; Xiang et al., 2009; Dillon, 2011), or structural as well as other cues such as gender marking (Badecker & Straub, 2002; Cunnings & Felser, in press). We framed the theoretical question within a computational model of sentence processing, the cue-based retrieval model of Lewis and Vasishth (2005); Lewis et al. (2006), and showed that if we assume that cue-based retrieval involves structural as well as non-structural cues, the model makes five predictions, repeated below:

E1. Mismatch effect (RE): the retrieval errors for mismatch conditions are higher than for match conditions.

E2. Interference effect (RE): the retrieval errors for match-interference and mismatch-interference conditions are higher than the other two conditions.

E3. Mismatch effect (RT): the retrieval times for mismatch conditions are longer than for match conditions.

E4. Match-interference effect (RT): the retrieval times for the match-interference conditions are longer than the match condition.

E5. Mismatch-interference effect (RT): the retrieval times for the mismatch-interference conditions are shorter than the mismatch condition.

Effects E1-E3 are attested in Sturt's studies; but effects E4 and E5 are not. We hypothesized that Sturt failed to find effect E4 because the inaccessible antecedent had a different grammatical role (object) than the accessible antecedent (subject); i.e., it was distinct enough from the accessible antecedent to be rejected successfully during search. We predicted that if both the accessible and inaccessible antecedents had the subject role, then a match-interference (E4) effect would occur; moreover if grammatical cues are weighted heavily in the retrieval process (Van Dyke & McElree, 2011), a subject distractor will induce a higher interference effect. We then conducted an eye tracking study in which both the accessible and inaccessible antecedents had the subject role, thereby increasing their similarity; we showed that in first pass regression probability a match-interference effect is indeed seen, as predicted by the model. In addition, when we separately analyzed the first fixation duration contingent on regressions, the first fixation durations that were followed by regressions showed a clear trend consistent with the two interference effects E4 and E5. These first fixation durations also confirmed the effect E3—when the stereotypical gender of the accessible antecedent mismatched the gender of the reflexives, reading times were higher. The effect E3 was observed in re-reading time and total reading time as well. This result is consistent with the model's predicted mismatch effect in retrieval times, and the predicted mismatch effect in retrieval errors. In sum, the eye tracking study provided empirical evidence for all the effects predicted by the model, including the interference effects that were not observed in the earlier studies like Sturt (2003) and Xiang et al. (2009).

Why wasn't the effect E5 (mismatch-interference) found in earlier studies? The absence of the effect could just be a failure to find an effect that in fact exists. For example, the effect could be masked by other confounding variables. Indeed, Cunnings and Felser (in press, p. 23) found that participants with high working memory spans show (in first fixation duration) exactly the direction predicted by the cue-based retrieval model. It is low-spans that show longer first fixation duration in the interference conditions in the mismatch cases. If one were to ignore the working memory span in the Cunnings and Felser data, the two differently-signed effects by span would cancel out, showing no difference between the interference and no-interference condition in the mismatch cases, exactly as found in the literature. Thus, since our data and all previous experiments (except, of course, Cunnings and Felser's) do not take working memory capacity into account as a variable, it is quite possible that we are missing an effect that is correctly predicted by the model. Of course, this raises the question that the ACT-R model as currently implemented does not explicitly model high working memory capacity participants. Nevertheless, in future work we intend to explore the role of working memory capacity in triggering the mismatch interference effect.

3.6.1 Strictly structured access as an alternative

Here, we discuss the strictly structured retrieval approach proposed in Dillon (2011) for resolving reflexive-antecedent dependency and examine its claims in the light of existing experimental and modeling findings in this domain.[4] Dillon (2011) proposed a theory of structured access with a set of elaborate computational and experimental studies involving English and Chinese reflexives and English subject-verb agreement. Experiment 1 was an eye tracking study where he performed a direct comparison between processing English reflexives and English subject-verb agreement with sentences as in (11). The manipulations were sentence type (reflexives vs. subject-verb agreement), grammaticality (correct vs. incorrect number marking on the subject) and presence of interference (interference vs. no interference) from a grammatically illicit noun. The conditions were similar to the experiment reported above, with the difference that four additional subject-verb agreement conditions and ungrammatical sentences were used instead of the gender mismatch. In the reflexive conditions, the object noun in the relative clause *the middle manager(s)* either matched the number of the reflexive or did not, and in subject-verb agreement conditions, it either matched the number marking on the matrix verb or did not, resulting in the interference and no interference conditions. The number marking on the matrix subject resulted in the grammatical and ungrammatical conditions.

(11) a. *Reflexive, grammatical, no interference*
The new executive who oversaw the middle managers apparently doubted himself on most major decisions.

b. *Reflexive, grammatical, interference*
The new executive who oversaw the middle manager apparently doubted himself on most major decisions.

[4] Although Dillon (2011) refers to the mechanism as structured access, we refer to it as a *strictly* structured access to emphasize the point that the approach suggested in this paper doesn't ignore the structural constraints, but it includes other constraints as well.

c. *Reflexive, ungrammatical, no interference*
The new executive who oversaw the middle manager apparently doubted themselves on most major decisions.

d. *Reflexive, ungrammatical, interference*
The new executive who oversaw the middle managers apparently doubted themselves on most major decisions.

e. *Subject-verb, grammatical, no interference*
The new executive who oversaw the middle managers apparently was dishonest about the company's profits.

f. *Subject-verb, grammatical, interference*
The new executive who oversaw the middle manager apparently was dishonest about the company's profits.

g. *Subject-verb, ungrammatical, no interference*
The new executive who oversaw the middle manager apparently were dishonest about the company's profits.

h. *Subject-verb, ungrammatical, interference*
The new executive who oversaw the middle managers apparently were dishonest about the company's profits.

In the eye movement measures, in the critical region (*himself/themselves on* for reflexive sentences and *was/were dishonest* for subject-verb agreement sentences), Dillon (2011) found a main effect of grammaticality; ungrammatical sentences were read more slowly than grammatical sentences. Additionally, for subject-verb agreement sentences there was a significant interaction between grammaticality and interference, and this interaction was driven by the interference effect in ungrammatical conditions with the interference condition being read *faster* than the no interference condition. In contrast, no reliable interference effect was observed for any measures at any region in reflexive sentences. This experiment, essentially, replicated the *interference asymmetry* effect from Wagers et al. (2009) and the absence of interference effect in processing English reflexives from Sturt (2003). Dillon (2011) replicated these results in Experiment 2 and 3. Based on these results Dillon concluded that agreement dependency and reflexive dependency employ different retrieval mechanisms for resolving the dependencies—agreement dependencies are resolved using morphological features of the target noun phrase whereas the antecedent for a reflexive is retrieved using *only* structural constraints. Dillon supports his hypothesis of a strictly structured access mechanism for reflexives with a set of computational modeling experiments (Experiment 4 and 5 in Dillon (2011)). The computational models simulated the step-by-step retrieval and constituent creation processes of a CBR model of the sentences in Experiment 1–3 described above.

From the perspective of the current discussion, the more relevant study was Experiment 5 which modeled data from Experiment 1 and 3. It directly compared the predictions of a strictly structural cue based retrieval model of reflexives to a model utilizing mixed cues—structural as well as agreement. The mixed cue model retrieved the antecedent using the cues *local subject*, *number* and *gender*, whereas the strictly structural cue based model used only *local subject* as the retrieval cue. The predictions were generated across 324 combinations of various parameter values. The predictions of the model were in terms of rates of incorrect retrievals and the retrieval latencies. The mixed cue model predicted an interference effect in retrieval errors (similar to E2 above) and a mismatch-interference effect in retrieval times (similar to E5 above); the prediction of the match-interference effect (E4) was not reliably non-zero. The structural cue based model predicted no interference effect in either retrieval errors or retrieval times. The mismatch effects in retrieval errors and retrieval times (E1 and E3) were not discussed in Dillon (2011). On the basis of these predictions Dillon (2011) concluded that the strictly structured access model captures the reflexive binding data from Experiment 1 and 3 better than the mixed cue model.

Although Dillon (2011) replicated the findings in Sturt (2003), the lack of interference effect is subject to the same alternative explanation that we suggested for Experiment 2 in Sturt (2003): We hypothesized that reflexive binding uses the grammatical role *subject* as one of the retrieval cues for retrieving the correct antecedent. The absence of the interference effect could be due to the fact that the interfering antecedent had an *object* role in the experiments above which does not match one of the retrieval cues reducing the strength of interference. Badecker and Straub (2002) also reported that the interference effect is found when the interferer is in the subject position. Moreover, Van Dyke and McElree (2011) found that, in subject-verb agreement, the interference effect due to the semantic match was present only when the distractors matched the grammatical cues along with semantic cues. If the retrieval process gives higher weight to grammatical cues than semantic cues, the absence of interference effect could simply be due to the absence of matching a grammatical role in the inaccessible antecedent.

The predictions of the structured access model hold only for a limited set of experiments and a limited set of effects in those experiments. The mismatch-effects (E1 and E3) that have been replicated in various studies like Sturt (2003), Cunnings and Felser (in press) and also the one reported in this paper cannot be explained by this model. The structured access model predicts no difference between match and mismatch conditions (see Figure 3.9). Furthermore, the interference-effects (E2, E4 and E5) observed in various reflexives studies like Badecker and Straub (2002) Experiment 3, Sturt (2003) follow-up study, Cunnings and Felser (in press) Experiment 2 and the one reported here cannot be explained by this model. Consequently, a model assuming structural as well as agreement features as retrieval cues predicts a broader set of data than a strictly structured access model.

Dillon further provided theoretical arguments for what effect can qualify to be an interference effect. According to Dillon (2011), the match-interference effect (E4)—higher reading times when the inaccessible noun matches gender or number of the reflexive—is not reliable evidence for interference from the grammatically inaccessible antecedent, for mainly two reasons: (1) The CBR model cannot reliably predict a non-zero match-interference effect in retrieval times, since on the one hand the cue-overlap

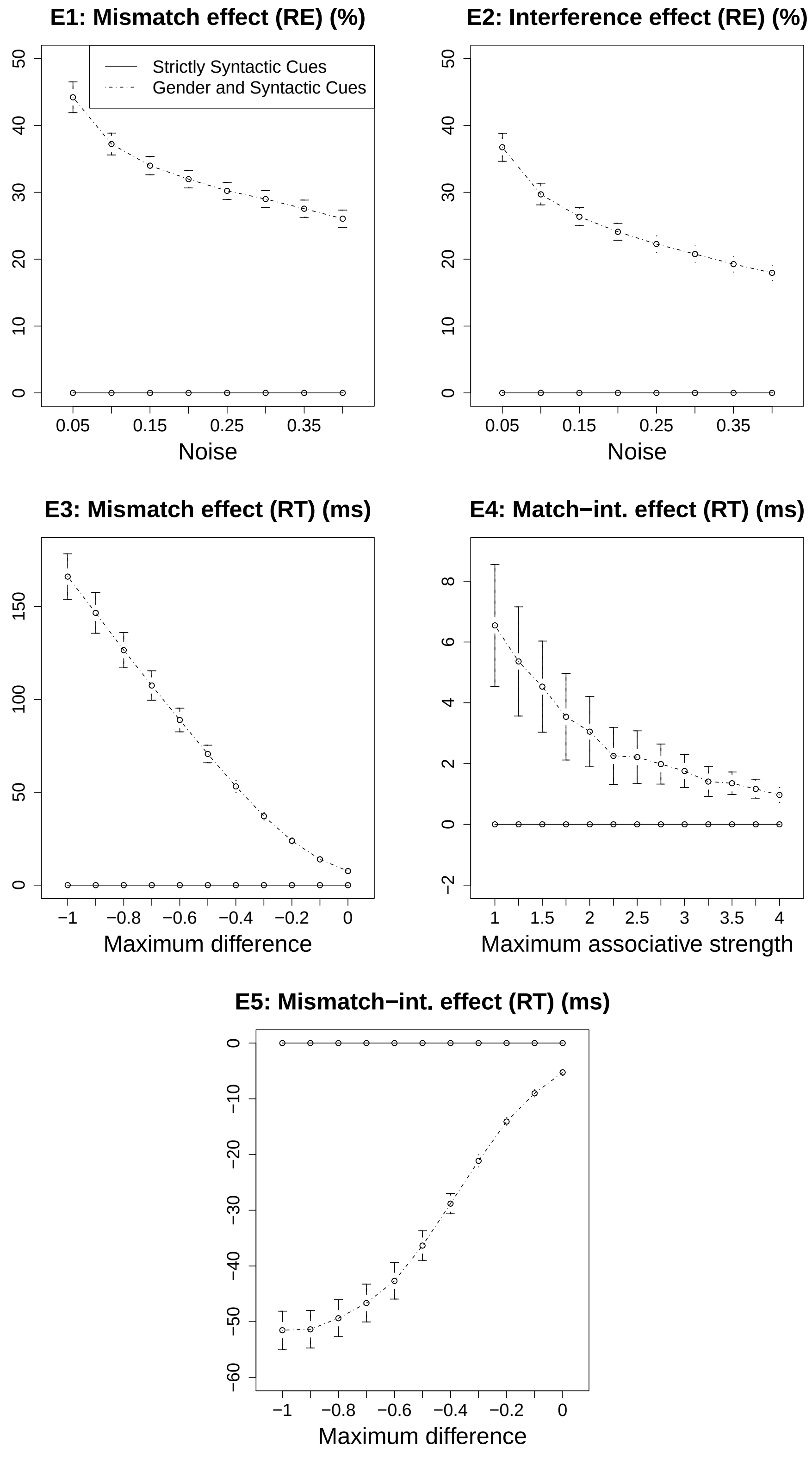


Figure 3.9: The predictions of a strictly structural cue-based model and a model with structural and agreement feature based cues. The predictions are for the new experimental design and are generated across a set of parameter values.

(gender and number) between accessible and inaccessible antecedents leads to an inhibitory effect, and on the other hand the retrieval of the inaccessible antecedent leads to a facilitatory effect; (2) The match-interference effects can be explained in terms of feature-overwriting (Nairne, 1990; Oberauer & Kliegl, 2006; Gordon, Hendrick, & Johnson, 2001, 2004; Gordon, Hendrick, Johnson, & Lee, 2006) instead of interference at the time of retrieval. Consequently, Dillon proposed that only a facilitatory effect in mismatch-interference can be considered as evidence for retrieval interference.

As far as the first argument is concerned, we, in fact, show that the CBR model with mixed cues consistently predicts a positive match-interference effect across a set of parameter values for Sturt's two experiments (see Figure 3.4 and 3.5). Although the effect for the new design is not predicted to be positive for all combinations of parameter values, for a certain set of combination of values the effect is *reliably* non-zero and positive, and only for a very small set of parameters values is the effect predicted to be zero.

The second argument presents a valid alternative explanation for the match-interference effect. If we assume a memory architecture as in ACT-R, then the inhibition effect is a result of cue overlap between accessible and inaccessible antecedents. But under other models of memory (Nairne, 1990; Oberauer & Kliegl, 2006) which cast inhibition in terms of feature-overwriting, the interference effect cannot be claimed to be due to specific retrieval cues. Further planned studies have to be carried out to tease apart these two explanations.[5] The facilitatory effect predicted in the mismatch-interference condition is observed in Cunnings and Felser (in press) Experiment 2. They found that participants with a high working memory spans showed a facilitation due to interference in the mismatch conditions. Possibly the effect in low working memory span participants was masked by other processes. The eye tracking study reported here also showed a strong tendency towards a facilitatory mismatch-interference effect in the regression

[5] We are collaborating with Dillon to carry out just such a study.

contingent analysis of first fixation duration.

In summary, we have presented a theory and computational model of the access of antecedents for reflexive pronouns in English, and used this theory to gain insight into empirical studies that have yielded mixed results concerning the putative role of non-structural cues. We used this analysis and the results of further modeling to motivate a new empirical design that formed the basis of an eye tracking study. The results of the eye tracking study are consistent with the model's assumptions concerning the early use of non-structural cues. These results present a challenge for theories advocating the infallibility of the human parser in the case of reflexive resolution, and provide support for the inclusion of agreement features such as gender in the set of retrieval cues. In general, the results provide further support for the deployment of a rapid, parallel cue-based access mechanism in service of sentence parsing (McElree, 2000; McElree, Foraker, & Dyer, 2003; Lewis & Vasishth, 2005; Lewis et al., 2006), and help to sharpen deeper explanatory questions concerning the utility and selection of cues.

3.7 Acknowledgements

We would like to thank Brian Dillon, Colin Phillips, and Ming Xiang for helpful comments. The eye tracking experiment reported here was run at the ZAS in Berlin; many thanks to Manfred Krifka for providing space to run this study. Umesh Patil is grateful to Felix Engelmann, Pavel Logačev, Titus von der Malsburg and Rukshin Shaher for their help at various stages of the eye tracking study and to William Badecker for comments about the experiment design.

Chapter 4

A cue-based retrieval model of offline and online sentence processing in aphasia

This chapter has been submitted as a paper to *Cognitive Science* on January 20, 2012 as:

Patil, U., Hanne, S., Vasishth, S., Burchert, F., & De Bleser, R. (submitted). A cue-based retrieval model of offline and online sentence processing in aphasia.

4.1 Abstract

Individuals with aphasia perform at chance level when processing reversible non-canonical sentences. Representational accounts (e.g., Grodzinsky, 1995, 2000, 2006) attribute this deficit to an impairment in syntactic representations. Processing accounts (e.g., Haarmann et al., 1997; Caplan et al., 2007), on the other hand, propose that the underlying structural representations are unimpaired, but sentence comprehension is affected by processing deficits such as slow lexical activation, reduction in memory resources, slowed processing and intermittent deficiency among others. Hanne et al. (2011) using a sentence-picture matching study suggested that the chance level performance in aphasics is a result of syntactic processing that is slowed down and affected by intermittent deficiency. They reported that the online processing (eye movements) in aphasics showed divergent patterns depending on whether the non-canonical sentences were processed correctly or not. In this paper, we frame the claims of Hanne et al. (2011) in a computational architecture for sentence processing, namely the cue-based retrieval theory, proposed by Lewis and Vasishth (2005). We operationalize the assumption of slowed processing as slow procedural memory, so that each processing action is performed slower than normal, and we operationalize intermittent deficiency as extra noise in the procedural memory, so that the parsing process is more noisy than normal. No representational impairments are assumed. We test the predictions of three models, the first of which assumes only slowed processing, the second only intermittent deficiency and the third assumes both deficits. The model assuming both deficits emerges as the best model of offline processing—sentence-picture matching accuracy and response time. For modeling eye movement data, we propose a linking hypothesis that maps activation levels of picture stimuli to eye fixation likelihood. All three models capture the divergent eye movement patterns observed in Hanne et al. (2011).

Keywords: aphasia; non-canonical sentences; sentence-picture matching; eye movements; computational modeling; cognitive architecture

4.2 Introduction

Impaired comprehension of non-canonical sentences is a symptom for some patients with aphasia (for reviews see Grodzinsky, 2000; Mitchum & Berndt, 2008). Although the deficit is often claimed to be associated with Broca's aphasia (e.g., Grodzinsky & Santi, 2008), it has been documented across various aphasic syndrome classifications, ranging from Broca's to Wernicke's and global aphasia (see Caplan, Waters, & Hildebrandt, 1997; Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004). For these patients, comprehension of single words and simple canonical sentences is usually preserved. Furthermore, they can understand irreversible non-canonical sentences properly, because, for these, world knowledge and semantically-based comprehension strategies are sufficient to correctly understand their meaning. However, patients with sentence processing deficits tend to experience comprehension problems whenever reliance on syntactic structure is necessary in order to derive the correct sentence interpretation (Caramazza & Zurif, 1976). This leads to difficulties in comprehending reversible non-canonical sentence structures like passives, object relatives, object clefts, and object-topicalized sentences (Mitchum & Berndt, 2008). This particular sentence comprehension deficit has been observed cross-linguistically in various studies involving patients with aphasia (Burchert, De Bleser, & Sonntag, 2003; Caplan et al., 2007, 1997; Caplan & Hildebrandt, 1988; Gavarro & Romeu, 2010; Grodzinsky, 2000; Grodzinsky, Piñango, Zurif, & Drai, 1999; Schwartz, Linebarger, Saffran, & Pate, 1987).

Grodzinsky (2000) and Grodzinsky et al. (1999) provide overviews of studies that evaluated aphasic patients' performance on canonical and various non-canonical sentences relative to chance level performance. Although the theoretical and clinical interpretation of chance performance is controversial (see, for example, Caplan et al., 2007; Caplan, 2001; De Bleser, Schwarz, & Burchert, 2006; Caramazza, Capasso, Capitani, & Miceli, 2005), Grodzinsky and colleagues' studies nevertheless provide evidence that patients experience much more difficulties with non-canonical compared to canonical sentence structures.

4.2.1 Explanations for sentence comprehension deficits

There exist several theories to explain sentence comprehension disorders in aphasia (for an overview, see Caplan, 2009). According to the representational or specific deficit accounts, the impairment is due to disturbances in underlying syntactic representations, i.e., it is assumed that patients suffer from a *breakdown in declarative knowledge of grammar*. For example, the Trace Deletion Hypothesis (TDH, Grodzinsky, 1995, 2000, 2006) proposes that aphasic patients have lost the ability to represent traces of syntactic movement in their parsing structures. Because of this trace deletion, they have no information about the theta-role of a moved NP in a non-canonical sentence and they need to rely on a default cognitive strategy assigning the theta-role of agent to the first NP in a sentence. However, patients' syntactic representation also contains another agent, the one assigned to the unmoved subject NP in base position, and hence they have to guess which of the two possible interpretations of a reversible sentence is the correct one. Besides the TDH, other accounts exist that relate syntactic comprehension disorders to disruptions in constructing fully intact syntactic representations (for example, Beretta & Munn, 1998; Hickok & Avrutin, 1995; Mauner, Fromkin, & Cornell, 1993). These accounts ascribe patients' sentence comprehension deficits to a breakdown in syntactic chain formation.

In contrast to representational explanations, processing or resource reduction accounts assume that underlying grammatical knowledge, including syntactic structure, is preserved, but the syntactic processing system is affected by processing (or capacity) limitations. Thus, these theories ascribe patients' difficulties with non-canonical sentences to a *procedural breakdown in parsing*. The different processing accounts vary in how they exactly conceptualize processing limitations in syntactic parsing. While some authors assume a processing weakness (Avrutin, 2006; Saffran, Schwartz, & Linebarger, 1998), which could result in slowed processing or intermittent breakdowns of the parser (e.g., Dickey et al., 2007; Friederici & Kilborn, 1989), others propose a slow activation or a too fast decay of structural syntactic information (e.g., Haarmann et al., 1997). The last mentioned idea of slow syntax has also been elaborated by Burkhardt, Mercedes Piñango, and Wong (2003), who claimed that patients suffer from a slowdown in the online assembly of phrase structure. Haarmann and Kolk (1991b) developed a processing account in which a timing disorder disrupts phrase structure building, resulting in a processing failure. Caplan and colleagues (Caplan et al., 2007; Caplan, 2006; Caplan & Waters, 2003; Caplan et al., 1997; Caplan & Hildebrandt, 1988) propose that patients might be unable to deal with several syntactic operations simultaneously and that this inability could be due to a pathological variability and intermittent reductions in available processing resources. Evidence in favor of such a resource variability comes from factor analyses, principal components analyses, and Rasch models of aphasic offline performance on different sentence types (Gutman, DeDe, Michaud, Liu, & Caplan, 2010; Caplan et al., 2007; Caplan, Baker, & Dehaut, 1985), and from studies in which normal control participants had to deal with stressful conditions that are assumed to evoke reduction in processing resources normally available by, for example, increasing concurrent load or stimulus presentation speed during language processing (e.g., Blackwell & Bates, 1995; Dick et al., 2001; Kilborn, 1991; Miyake, Carpenter, & Just, 1995). Moreover, in a self-paced listening study, which was combined with a sentence-picture matching and grammaticality judgement task, Caplan et al. (2007) found normal *online* performance for aphasic patients when they provided correct offline responses. In contrast, incorrect offline responses were associated with abnormal online performance. This result is unexpected under the *guessing* *view* of chance performance (which would be predicted by the TDH). Caplan and colleagues concluded that patients cannot be suffering from constant impairments in an underlying grammatical structure (e.g., deleted traces), or from a total breakdown in specific parsing operations (e.g., associating a trace with a filler). Instead, they argue, sentence comprehension deficits should better be conceptualized as reflecting intermittent deficiencies in resources necessary for syntactic parsing. These intermittent reductions are then seen in divergent self-paced listening data and lead the patient to end up with an erroneous sentence interpretation. The insights from this investigation also supported results of former sentence processing studies by Caplan and colleagues (Caplan & Waters, 2003, 1995) and shed a new light on the underlying cause of sentence comprehension disorders: Aphasic patients still have their grammatical knowledge and syntactic processing strategies available. Therefore, patients' failures in sentence comprehension must be associated with disruptions in otherwise available sentence processing resources.

Further evidence consistent with this view has emerged from studies looking at online sentence processing in aphasia using eye tracking (e.g., Dickey et al., 2007; Hanne et al., 2011; Thompson & Choy, 2009). These studies employed the visual world paradigm (Cooper, 1974; Tanenhaus et al., 1995) in order to look at patients' online parsing abilities. Dickey et al. (2007) let participants listen to short stories while looking at four pictures, of which three were mentioned in the story. Comprehension probes (either a yes-no or a wh-question) were obtained at the end of each trial. For wh-questions, controls showed anticipatory eye movements to a potential (visually presented) filler for the gap, which was signaled by the moved constituent (the object wh-question word) at sentence beginning, when they heard the verb (see also Sussman & Sedivy, 2003). According to the authors, these anticipatory eye movements reflect participants' incremental, automatic gap-filling during sentence comprehension. Interestingly, the same eye movement pattern was found for aphasic participants when they gave a correct response (when the offline response was incorrect, they showed increased looks to a subject competitor during late regions of the sentence). This suggests that resolving wh-dependencies was relatively unimpaired in the patients with aphasia (Dickey et al., 2007, p. 14). Moreover, because patients' eye movements in their correct responses were so similar to controls' in speed and their overall pattern, the results are inconsistent with a slow-down in aphasics' online processing. Therefore, Dickey and colleagues, referring to Avrutin (2006), argue for a weakened-syntax view of sentence comprehension disorders: Syntactic representations and processing operations in aphasia are undamaged and operate with the same speed as in listeners without language impairment, but their output is weakened and hence vulnerable to competing heuristics.

In a study combining a classical sentence-picture matching task with the visual world paradigm, Hanne et al. (2011) came to a similar conclusion about the underlying deficit in sentence comprehension. This study investigated processing of German reversible canonical, subject-verb-object (SVO), sentences and their non-canonical counterpart, object-verb-subject (OVS), sentences. Participants listened to a spoken sentence and saw two pictures on a screen; one target picture depicting the action mentioned in the spoken sentence correctly and a foil picture showing the same action but with reversed thematic roles of the actors. The task was to identify the picture that matches the spoken sentence. Data on online processing (eye movements) and offline performance (accuracy and response time) were collected simultaneously during the whole task.

Patients' offline accuracy reflected the expected pattern: On average, they performed worse than controls, and comprehension of non-canonical (i.e. OVS) sentences was significantly lower than on canonical (i.e. SVO) sentences. Reaction times were significantly higher in patients than in controls, and non-canonical sentences elicited higher latencies than canonical ones. Furthermore, there was a strong effect of response accuracy on patients' response times: incorrect responses elicited significantly higher response times than correct ones. Moreover, systematic differences in fixation patterns that led to correct vs. incorrect offline comprehension of sentences were found. For correctly answered trials, patients' eye movement patterns were very similar (in terms of relative fixation probability on the two pictures presented) compared to controls. In addition, although there was no qualitative difference observable between patients' and controls' online processing, patients' pattern was delayed, pointing to a slowdown in online sentence processing. In contrast, for incorrect offline responses, patients' eye movements were clearly deviant from controls'.

To sum up, Hanne et al. replicated the findings both by Caplan and colleagues and Dickey and colleagues: For correct sentence interpretations, aphasics' online sentence processing is unimpaired and conveys no evidence for structural impairments in declarative knowledge of grammar. Following Caplan and colleagues, they came to the conclusion that these preserved processing routines are not always available because of intermittent deficiencies of parsing operations. In contrast to Dickey et al., the data of Hanne and colleagues also point to processing limitations in terms of a slow-down in syntactic processing in aphasia. However, as the methodology used was different, the results of Dickey et al. and Hanne et al. are not straightforwardly comparable.

The accounts on sentence comprehension disorders mentioned above clearly emphasize a syntactical view on the deficit, i.e., they seek to locate the origin of the sentence comprehension problems explicitly at the sentence level of language processing. However, some authors take a different view; they trace back patients' syntactic difficulties to a more fundamental stage of language processing: the level of lexical access (Love, Swinney, Walenski, & Zurif, 2008; Prather, Zurif, Love, & Brownell, 1997). Using a cross-modal lexical priming paradigm with Broca's aphasics, Love et al. (2008) found delayed lexical activation for moved NPs in a sentence and, moreover, delayed lexical reactivation of the respective NPs at their gap site. They interpreted their findings as evidence that "the formation of a syntactic dependency involving a moved constituent is selectively vulnerable, not because it's a syntactic operation, but because if lexical reactivation is not accomplished within a normal time frame, a non-grammatical heuristic kicks in to provide a conflicting interpretation" (Love et al., 2008, p. 216). The evidence in favor of the lexical access hypothesis shows that although most accounts focus on the syntactic level of language comprehension, it is still a matter of debate at which language processing stage syntactic comprehension disorders are localized.

4.3 Aims of the study

Overall, the results of the aphasic sentence processing studies discussed above point to a processing deficit explanation of sentence comprehension disorders. However, as the above discussion shows, there are still many open questions. We think that an important tool for understanding aphasic sentence processing is *computational modeling*. Recently, the idea of implementing computational models to inform our understanding of neuropsychological impairments has gained more and more attention (for an overview see Dell & Caramazza, 2008). As Martin (2006, p. 91) points out, "attempts to model patient performance would lead to important predictions and new empirical tests regarding the nature of the patients' deficits."[1]

In this paper, we build on the computational modeling approach in order to test some of the assumptions on sentence comprehension disorders put forward by processing accounts. More specifically, we computationally explore whether a potential cause for breakdown at the sentence level is (i) slowed syntactic processing, (ii) intermittent deficiencies of the parser, or (iii) a combination of both these impairments. Before describing our own simulations,

[1] Some models of patient performance do already exist for single-word processing and the respective impairments. Among these are modeling accounts which concentrate on impairments of auditory single-word comprehension and production (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Mirman, Yee, Blumstein, & Magnuson, 2011; Rogers et al., 2004), on acquired dyslexias (for example, Coltheart, 2006; Plaut, McClelland, Seidenberg, & Patterson, 1996; Woollams, Ralph, Plaut, & Patterson, 2007), on optic aphasia (Plaut, 2002) and on single-word processing in bilingual aphasia (Grasemann, Kiran, Sandberg, & Miikkulainen, 2011). Penke and Westermann (2006) presented an implementation of a model designed to account for impairments in morphological processing.

it may be useful to briefly review existing computational models of aphasic sentence comprehension disorders (Crescentini & Stocco, 2005; Gigley, 1983; Clifton, Staub, & Rayner, 1988; Haarmann & Kolk, 1991a; Just & Carpenter, 1992; Vosse & Kempen, 2000).

4.4 Existing computational models of aphasic sentence comprehension

Crescentini and Stocco (2005) presented an ACT-R (Anderson et al., 2004) based model of aphasic sentence processing. In line with the Slow Syntax Hypothesis (Piñango, 2000), they assumed that lexical activation in agrammatic aphasics is slower than normal.

They implemented this hypothesis in terms of reduced attentional resources in the ACT-R model and simulated aphasic sentence processing using non-canonical sentences such as passive sentences. In the next section, we examine the computational models of aphasic sentence comprehension disorders proposed by Haarmann and Kolk (1991a) and Haarmann et al. (1997); these models are the closest in terms of their assumptions to the model that we present in this paper.

Haarmann and Kolk (1991a) proposed a computational model of aphasic language breakdown, called SYNCHRON. The model is based on the Kolk and Van Grunsven (1985) hypothesis of sentence comprehension deficits in aphasia. According to this hypothesis, parsing fails in agrammatic aphasics since syntactic representational elements that need to be simultaneously active in working memory are often not coactive because of disturbances in timing due to brain damage.

SYNCHRON is a model of the time course of parsing rather than the actual parsing process of constructing a parse tree from an input string of words. The model is provided with a predefined phrase-structure representation of a sentence. It determines whether the complete construction of this phrase-structure representation is possible given a set of temporal constraints in the model. A failure to compute the specified phrase-structure representation is assumed to be a failure in processing the sentence. The phrase-structure representation is constructed bottom-up as a chain of retrievals starting at input words and ending in phrasal categories—input words cause the retrieval of word forms, word forms prompt the retrieval of associated lexical categories and lexical categories lead to the retrieval of phrasal categories. The retrieval of a phrasal category is possible only if all its constituent categories are in their *memory time* phase, which is the *computational simultaneity* constraint in the model. *Memory time* along with *retrieval time* provides a temporal constraint on the construction of the phrase-structure representation. Retrieval time specifies the time required to retrieve an element, and memory time specifies the amount of time a retrieved element can remain available for further processing. In SYNCHRON, aphasics are assumed to have a temporal disorder—either the retrieval time is longer than normal or the memory time is shorter than normal. This temporal disorder disrupts computational simultaneity among elements of the phrasal category causing parsing failures. Haarmann and Kolk (1991a) showed that assuming a temporal disorder is sufficient to model the combined effects of the degree of severity and sentence complexity in agrammatic aphasics described in Schwartz, Saffran, and Marin (1980) and in a replication study by Kolk and Van Grunsven (1985).

Although SYNCHRON was successful in modeling aphasic behavior on simpler sentence types, its capabilities were limited due to the absence of an elaborate parsing process. Moreover, it lacked a mechanism of thematic role assignment. Since correct thematic role assignment is a crucial test of aphasic sentence processing, SYNCHRON was left incomplete without it. Haarmann et al. (1997) proposed an improved model of aphasic sentence processing in terms of a Capacity Constrained Resource Deficit model (for convenience CCRD henceforth). CCRD is derived from the Resource Reduction Hypothesis illustrated in Miyake, Carpenter, and Just (1994). The hypothesis proposes that the sentence processing impairment in aphasia is an extension of the sentence processing limitations in low working memory capacity of unimpaired individuals. The CCRD model is implemented in the 3CAPS (Just & Carpenter, 1992) architecture, which is a production system involving activation propagation. 3CAPS imposes processing constraints in terms of the working memory capacity. Both storage and computation of information put demands on the available activation. The maximum possible activation in the system is equal to the working memory capacity, a finite resource available for information processing. If the storage and computational demands in processing a sentence are not fulfilled, it leads to a breakdown in comprehending that sentence. The activation resource consumed in sentence processing essentially involves lexical, syntactic, semantic, and inferential processes. The hypothesis for aphasic patients is that they share a common deficit—pathologically reduced working memory capacity. A more complex sentence puts higher storage and computational demands, and the reduction in the available activation resource in aphasics induces a breakdown in processing the sentence.

CCRD focuses on deriving thematic roles assigned by the verb in the sentence. CCRD is composed of three main subsystems that accomplish thematic role assignment by carrying out three different sub-tasks in sentence processing—performing lexical access, constructing the parse tree and mapping thematic roles. The functionality of each component subsystem is achieved through a set of production rules. Production rules temporarily activate the working memory elements capturing various sentential representations. The rules in the thematic role component utilize the parse tree representation of a sentence to generate thematic roles between words in the sentence. Once the processing of a sentence is completed the level of activation of the working memory elements representing the thematic role bindings are recorded. The sentence comprehension accuracy is indicated by the average activation of these memory elements. The assumed resource reduction in aphasia is induced in the model by decreasing working memory capacity. Several other parameters, along with the memory capacity parameter, are set by hand. The model is shown to reproduce the sentence complexity effect obtained by Caplan et al. (1985) across nine sentence types, as well as the interaction between sentence complexity effect and the degree of severity of aphasia in the data from Kolk and Van Grunsven (1985). All simulations involved modeling the offline measure of sentence comprehension accuracy.

Apart from CCRD and SYNCHRON, two other attempts of modeling aphasic sentence comprehension have been proposed in the literature: the HOPE model (Gigley, 1983; Clifton et al., 1988) and the UNIFICATION SPACE model (Kempen & Vosse, 1989; Vosse & Kempen, 2000). While these models differed considerably in their details they are consistent with the assumption that aphasic sentence comprehension is not a result of any breakdown in the declarative knowledge of the grammar, but rather a deficit in the processing of this knowledge.

As observed by Haarmann et al. (1997, p. 82), all these previous models share the common assumptions that "(i) knowledge representation and processing are activation driven, (ii) successful sentence comprehension requires the co-activation of certain critical representational elements, and (iii) in aphasia, co-activation is disturbed by an immediate or emergent timing deficit."

We consider these assumptions to be strengths of these models and retain them in our model. Though effective, all the existing models are limited to the modeling of only offline measures like sentence comprehension accuracy. We present a computational model (Lewis & Vasishth, 2005; Lewis et al., 2006) that can model online incremental sentence processing in controls versus aphasics. We employed the cue-based retrieval (henceforth, CBR) architecture (Lewis & Vasishth, 2005; Lewis et al., 2006) to model sentence processing data from aphasics and unimpaired individuals. CBR theory has already been shown to account for several key sentence comprehension phenomena in healthy individuals (Vasishth & Lewis, 2006; Vasishth et al., 2008; Boston et al., 2011; Patil, Vasishth, & Lewis, 2011).

We modeled data from a visual world paradigm study in Hanne et al.

(2011). The data consist of the eye movement patterns of controls and aphasics during a sentence-picture matching task for German reversible canonical and non-canonical sentences. We also modeled offline measures (accuracy and response time) for the sentence-picture matching task. As far as we know, this is the first attempt to model eye movement data from aphasics. In the next section, we describe the CBR architecture and then we illustrate an elaborate model of the data from Hanne et al. (2011).

4.5 Cue-based retrieval theory

The complete details of the Cue-based retrieval (CBR) theory are described in Lewis and Vasishth (2005), Lewis et al. (2006) and Vasishth and Lewis (2006). Here we briefly describe the important features of the theory. The CBR theory is derived from (i) the architectural assumptions of ACT-R (Adaptive Control of Thought-Rational, Anderson et al., 2004), (ii) assumptions about the parsing process based on psycholinguistic evidence, and (iii) representational assumptions from generative grammar.

ACT-R is a generic cognitive architecture developed for modeling various cognitive processes. The two main components of the architecture are the declarative memory system and the procedural memory system. Declarative memory holds the contents of long-term memory (semantic and episodic memory), as well as phrases created at run-time during processing. The elements in declarative memory are called *chunks*. Each chunk is a set of feature–value pairs. The procedural memory system describes procedural knowledge in terms of *production rules*, which are condition–action pairs. These two memory systems also form the core of the CBR architecture. In the CBR architecture, declarative memory maintains the lexical knowledge. The grammatical knowledge and parsing rules (the control structure) are maintained in the procedural memory in terms of production rules. Production rules are specified such that sentence parsing occurs as per the left–corner parsing algorithm (Aho & Ullman, 1972). The incremental syntactic structure created during parsing is stored in declarative memory as chunks. These chunks are X–bar structures (Chomsky, 1986) representing maximal projections with features corresponding to X–bar positions (specifier, complement, head) and other grammatical features such as case marking, and person, number, gender agreement. Traces are implemented as co-indexation inside maximal projections. The parse tree is updated at each input word by creating new chunks and attaching them to the existing parse tree representation. Sentence parsing takes place as an iterative sequence of production rule firing, retrieval of memory chunks and update of the current parse tree.

In addition to the symbolic system (i.e., procedural and declarative memory), the behavior of a model is modulated by a set of subsymbolic computations. These computations impose constraints on the retrieval of chunks from memory and the selection of production rules at each stage. The constraints on retrieval are specified in terms of activation of chunks. The activation value of a chunk determines the probability of its retrieval and the latency of the retrieval. The frequency, recency and prior pattern of retrievals determine the activation of a chunk. Equations (4.1-4.3) give the details of computing the activation value for each chunk at every retrieval.

$$A_i = B_i + \sum_{j=1}^{m} W_j S_{ji} + \sum_{k=1}^{p} PM_{ki} + \epsilon \tag{4.1}$$

$$B_i = \ln\left(\sum_{k=1}^{n} {t_k}^{-d}\right) \tag{4.2}$$

$$S_{ij} = S - ln(\text{fan}_j) \tag{4.3}$$

Equation 4.1 specifies the total activation of a chunk i (A_i), which is the sum of the base level activation (B_i), the spreading activation received through retrieval cues (first summation component), the activation received due to partial match between retrieval cues and corresponding feature values in the chunks (second summation component), and stochastic noise (ϵ). The base–level activation of a chunk is calculated using Equation 4.2. Here, t_k is the time since the k^{th} successful retrieval of chunk i and d is the decay parameter. The spreading activation that a chunk i receives (first summation component in Equation 4.1) is computed using W_j and S_{ji} values. W_j is normally equal to $1/n$, where n is the number of retrieval cues. S_{ji} is the strength of association from an element (typically a retrieval cue) j to chunk i and it is computed using Equation 4.3. S is a parameter defining the maximum associative strength and fan_j is the number of items associated with cue j. The strength of association from a cue is reduced as a function of the "fan" of the retrieval cue, resulting in an associative retrieval interference. The second summation component in Equation 4.1 specifies the activation received through a partial match. It is computed over p retrieval cues using P and M_{ki} values. P is the match scaling parameter and M_{ki} refers to the similarity between the retrieval cue k and the corresponding value in chunk i.

The mapping from activation A_i to retrieval latency T_i for a chunk i is obtained using Equation 4.4. F is the scaling parameter. The higher the activation of the chunk the faster the retrieval.

$$T_i = Fe^{-A_i} \tag{4.4}$$

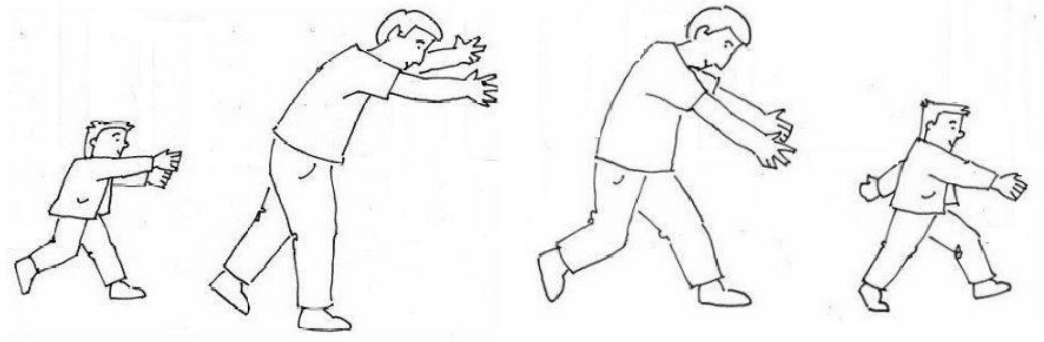
Another subsymbolic computation that affects the behavior of a model is the calculation of the utility value for each production. If there are multiple productions that can fire at a certain stage then the production with the highest utility value is the one that gets selected. The utility value of each production is set depending on the significance of that production in the given context. Just like chunk activations, utilities have stochastic noise added to them; the distribution of this noise is controlled by the utility noise parameter. If there are a number of productions competing with expected utility values U_j the probability of choosing production i is described by the Equation 4.5; s is the noise parameter that controls the variance of the noise distribution. Once a production is selected it takes a constant amount of time for it to "fire" and accomplish the actions assigned to it. The value of this constant time can be modified using a parameter called *default action time.*

$$Probability(i) = \frac{e^{U_i/\sqrt{2}s}}{\sum_j e^{U_j/\sqrt{2}s}} \tag{4.5}$$

From the perspective of modeling aphasic sentence processing, the CBR architecture provides a framework with a set of tight constraints on the memory processes involved in parsing sentences, while at the same time offering the flexibility, in terms of possible memory representations and the control structure, for extending the architecture to model other tasks. The set of modifiable parameters makes it possible to extend the framework to impaired sentence processing. The next section describes how we harness this flexibility to extend the CBR architecture of unimpaired sentence processing to the modeling of a sentence-picture matching task with aphasics.

4.6 Cue-based retrieval model of Hanne et al. 2011

Here we present the details of the CBR model of the data from Hanne et al. (2011). The study reported in Hanne et al. (2011) was a sentence-picture matching task in a visual world paradigm. Participants listened to German reversible canonical and non-canonical sentences as in (1) while they were presented with two pictures (see Figure 4.1) on the screen. Each trial consisted of listening to a sentence in one of the two word orders; one of pictures on the screen matched the sentence and the other did not. At the end of the trial participants were asked to select the correct picture. Participants' response and response time were recorded. Participants' eye movements were also recorded during the whole trial.



(a) (b)

Figure 4.1: Sample pictures used in the Hanne et al. 2011 sentence-picture matching task. For the canonical sentence *Der Sohn fängt den Vater* ('The son is catching the father'), (a) is the correct picture and (b) is the distractor.

The experiment was carried out with seven individuals with agrammatic Broca's aphasia and eight age-matched controls without any history of neurological impairment. As a result, Hanne et al. (2011) provides one online and two offline measures of sentence processing of both, healthy and aphasic individuals.

(1) a. **Canonical:**
Der Sohn fängt den Vater
the$_{\mathrm{NOM}}$ son is_catching the$_{\mathrm{ACC}}$ father
'The son is catching the father'

b. **Non-canonical:**
Den Sohn fängt der Vater
the$_{\mathrm{ACC}}$ son is_catching the$_{\mathrm{NOM}}$ father
'The father is catching the son'

In the past, the CBR theory has been predominantly used for explaining reading time data (Lewis & Vasishth, 2005; Vasishth & Lewis, 2006; Van

Dyke & Lewis, 2003; Patil, Vasishth, & Kliegl, 2009; Wagers et al., 2009). The predictions of CBR models are in terms of retrieval times, which have been used as the model's estimate of response times. We retain this assumption in the current model for simplicity. The CBR theory also makes predictions in terms of retrieval failures and parsing failures. These predictions can be easily mapped to errors in sentence processing (Vasishth et al., 2008) or comprehension question-response accuracies (Patil et al., 2011). However, the CBR theory has never been used to model data from cross-modal studies and there is no CBR model which targets online (eye movement patterns) and offline (picture matching accuracies and response times) data in a sentence-picture matching task. To model the sentence-picture matching data from Hanne et al. (2011), here we make an extra assumption in terms of a *linking hypothesis* between predictions of the CBR theory and cross-modal tasks.

4.6.1 Linking hypothesis: Implementing the sentence-picture matching task in the CBR architecture

A sentence-picture matching task typically involves listening to a sentence while watching two pictures on the screen—a target and a distractor picture. Eye movements are recorded during sentence presentation, and at the end of each sentence participants select the picture that matches the sentence. Modeling the sentence-picture matching task entails modeling eye movements across the two pictures during sentence presentation, modeling response accuracy and modeling response time. To accomplish this, we make a set of new assumptions in the CBR architecture.

First, we assume that, while processing a sentence, the model creates two separate semantic representations of the two pictures on the screen after the first noun is processed. These semantic representations are stored as chunks in declarative memory (see Figure 4.2 showing chunks with semantic representation of pictures in Figure 4.1). As in earlier CBR models, at each input word, the parser incrementally updates the partial representation of

(a) Chunk: *Sohn_nom-Vater_acc*

AGENT : Sohn
THEME : Vater

(b) Chunk: *Vater_nom-Sohn_acc*

AGENT : Vater
THEME : Sohn

Figure 4.2: The picture chunks created by the model for the semantic representation of pictures used in the Hanne et al. 2011 sentence-picture matching task. Chunk (a) represents the canonical sentence *Der Sohn fängt den Vater* ('The son is catching the father'), and chunk (b) represents the non-canonical sentence *Den Sohn fängt der Vater* ('The father is catching the son').

the sentence. Table 4.1 lists the steps followed by the model to create syntactic and semantic representations of sentences in (1). The assumption that the semantic representations of pictures are stored as chunks is necessary for two reasons: firstly, chunks are the only representational units in declarative memory, hence this is the only way the model can operate on picture objects; and, secondly, the task of matching a sentence to a picture can be accomplished if they have comparable representations.

Next, we assume that, as the model processes new sentential input, it selects the picture that matches the partial representation of the sentence up to that point. The picture selection is performed by means of a retrieval request for a matching picture chunk. As a consequence of a retrieval request the two picture chunks receive varying amounts of activation boost. The amount of activation boost received by each chunk depends on its match with the partial sentence representation. We further assume that the difference in the activation of the correct and incorrect picture chunk at the time of retrieval determines the likelihood of fixating on the correct picture. The difference in activation is calculated by subtracting the activation of the incorrect picture from the activation of the correct picture. A positive sign on the difference denotes a fixation on the correct picture and a negative sign denotes a fixation on the incorrect picture; the higher the value of the difference the more likely it is that the the correct picture is fixated. For simplicity, we assume that the mapping between activation difference and fixation probability is linear. This set of assumptions for fixation probabilities in terms of activations is based on the existing evidence in the visual world literature and the architectural constraints on the model. Our assumptions are consistent with the linking hypothesis proposed by Altmann and Kamide (2007). Their account proposed a linkage between language processing and eye movements in the visual world paradigm. On the basis of studies reported in Dahan and Tanenhaus (2005); Huettig and Altmann (2005); Myung, Blumstein, and Sedivy (2006); Altmann and Kamide (2007), Altmann and Kamide (2007) proposed that a conceptual overlap between the linguistic input and visual objects results in an increase in activation of the memory representations of those objects. The increase in the activation of an object constitutes a shift in the attentional state of the cognitive system, and this shift in attention increases the likelihood of eye movements towards the spatial location of that object. We implement the effect of activation boost from linguistic input in terms of multiple retrievals of the picture chunks. The reasoning behind this is that retrieval is the only process in the CBR architecture that induces an increase in the activation of an existing memory chunk. Memory retrievals cause an increase in activation through activation spreading and a boost in the base-level activation (see the explanation for Equation 4.1 and Equation 4.2 in the earlier section).

Finally, we assume that the picture that is retrieved at the end of the sentence is the picture that is finally selected as the response to the sentence-picture matching task, and that the duration between the processing of the first word of the sentence and the retrieval of the picture at the end of the sentence is the response time for the picture matching task.

Table 4.1: The sequence of steps followed by the model while processing the sentences in (1).

Step no.	Processing steps	
	Canonical	Non-canonical
1	Process nominative determiner	Process accusative determiner
2	Process noun-1 {use case info/ignore case info}	Process noun-1 {use case info/ignore case info}
3	Create 2 picture chunks	Create 2 picture chunks
4	Retrieve picture chunk	Retrieve picture chunk
5	Process verb	Process verb
6	Retrieve picture chunk	Retrieve picture chunk
7	Process accusative determiner	Process nominative determiner
8	Process noun-2 {use case info/ignore case info}	Process noun-2 {use case info/ignore case info}
9	Retrieve picture chunk	Retrieve picture chunk
10	Retrieve picture chunk	Retrieve picture chunk

4.6.2 Details of the model of sentence-picture matching

Modeling the sentence-picture matching task is dependent on accurately creating a representation of the input sentence and retrieving the picture chunk from memory (refer to Table 4.1 for the sequence of steps followed by the model to perform the task in Hanne et al. (2011)). Creating a syntactic and semantic representation of the sentence involves cue-based retrieval of existing syntactic structures like verb phrase (VP) and noun phrase (NP) (see Table 4.1). Retrievals of these structures and the picture chunks from memory is subjected to retrieval constraints—*decay*, *interference* and *partial match*—of the CBR architecture. These constraints are enforced using Equations 4.1–4.4. For example, at the input object noun *Vater* in the OVS sentence, the model sets a retrieval request like "retrieve a syntactic structure with CATEGORY = DP and CASE = accusative" to retrieve an existing determiner phrase (DP) marked with accusative case. The accuracy and latency of retrieving the requested DP are affected by the decay of activation (Equation 4.2) of the DP, spreading of activation due to the fan of the retrieval cues (the first summation component in Equation 4.1 and Equation 4.3), the partial match of the cues (the second summation component in Equation 4.1) and the noise component (ϵ in Equation 4.1). The picture chunks are retrieved with cues specifying the thematic roles—agent and theme—in the sentence; for example, a picture retrieval request is of the form: "retrieve a picture representation with AGENT = Sohn and THEME = Vater". The picture retrieval requests that are carried out before the second noun is processed (e.g., steps 4 and 6 in Table 4.1) have an empty value for one of the features—AGENT or THEME—depending on the word order. As in syntactic structure retrievals, the accuracy and latency of picture retrievals are subject to the same retrieval constraints.

4.6.3 Mapping deficit assumptions to the CBR models

As discussed earlier, sentence processing disorders in aphasia have been mainly explained either in terms of *representational accounts* or *processing accounts*. Our goal in the current study is to determine whether a processing deficit in the model for unimpaired individuals can be used to model aphasic sentence processing; for this reason, we test some of the proposals put forward by processing accounts using the cue-based retrieval architecture discussed above.

Processing accounts assume that the underlying grammatical knowledge is unimpaired in aphasics, but the overall processing system is impaired in various ways. Using CBR models, we tested the hypotheses of *slowed processing* (Haarmann & Kolk, 1991a; Friederici & Kilborn, 1989) and *intermittent deficiency* (Caplan et al., 2007). Slowed processing is proposed to be a pathological slowdown in the processing system, and intermittent deficiency is suggested to be a reduction in the strength of the resources available for processing.

In the long run, it will be necessary to develop computational models of representational accounts as well, but this lies beyond the scope of the present work.

Consistent with the processing accounts, we assume that there is no impairment in the grammatical knowledge of the aphasics. This means that the set of production rules (the procedural memory) is the same across models for controls and aphasics. Since it is possible that only one of the two assumptions, slowed processing and intermittent deficiency, is enough to explain the data from Hanne et al. (2011), we evaluated these assumptions using three separate models:

(M1) Only slowed processing

Slowed processing in the CBR architecture is implemented in terms of a higher value for the *default action time* (DAT) parameter. DAT controls the amount of time required for one production rule to fire and hence the amount of time required to processing. The default value of the parameter is 50 milliseconds.

(M2) Only intermittent deficiency

Intermittent deficiency in the CBR architecture is implemented in terms of a higher value for the *utility noise* parameter. The utility noise value controls the variance in the utility values of productions. The utility value of productions describes the probability of selecting a particular production when multiple productions are competing to get fired.

(M3) Both

This model is affected by both—slowed processing and intermittent deficiency.

4.6.4 Estimation of parameter values

We estimated ACT-R parameter values only for the offline data from aphasics (see Figure 4.3); this estimation was done by visual inspection, no optimiza-

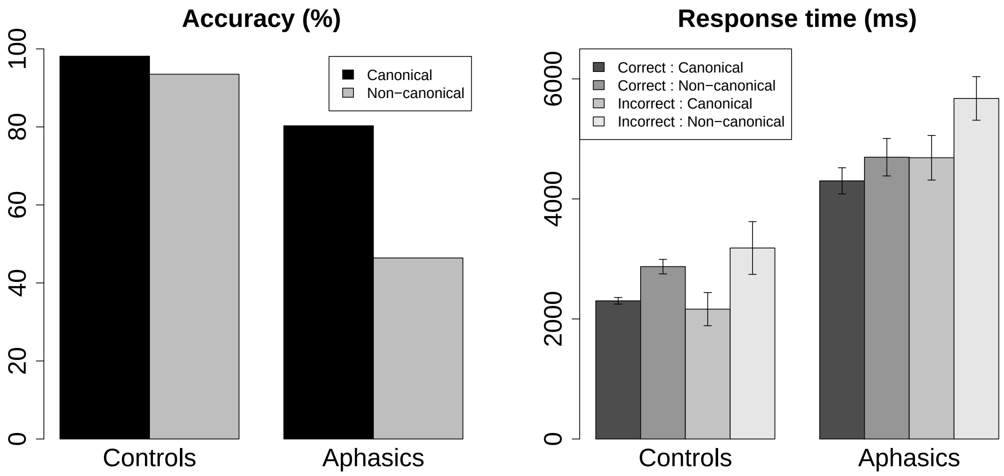


Figure 4.3: Data from Hanne et al. 2011; accuracy and response time for the sentence-picture matching task. The error bars in the response time plot show 95% confidence intervals.

tion procedures were used for parameter estimation. For data from controls, all existing parameters were set to the values that have been used in the previous models from Lewis and Vasishth (2005); Vasishth and Lewis (2006); Vasishth et al. (2008); Patil et al. (2011). Once parameters were estimated for aphasic offline data, predictions for eye movement data were generated without any changes to the estimated parameter values. A list of all the parameter values that we used for controls' data is given in Table 4.2.

For model M1 we estimated the value for the DAT parameter by generating predictions of the model for a range of values of the parameter. The predictions were generated for the picture-matching accuracy and the response time. Similarly, for model M2 we estimated the value of the *utility noise* parameter by generating predictions on a range of values of the parameter. The range of values of DAT and *utility noise* tested in the model are listed in Table 4.3. For model M3 we used the new estimated values of both the parameters simultaneously (see assumptions for M1, M2 & M3 in the preceding section). The final estimated values for DAT and *utility noise*

Table 4.2: The list of parameter values used in the previous studies with CBR models and the values used in the current models for controls and aphasics.

Parameter	Previous Models	Current Model (Controls)	Current Model (Aphasics)
Decay (d)	0.50	0.50	0.50
Maximum associative strength (S)	1.50	1.50	1.50
Retrieval Threshold (T)	−1.50	−1.50	−1.50
Maximum difference	−0.60	−0.60	−0.60
Latency Factor (F)	0.14, 0.46	0.14	0.14
Noise (ϵ)	0, 0.15, 0.30, 0.45	0.15	0.15
Default action time	0.05	0.05	0.07
Utility noise	0	0.01	0.1

Table 4.3: The range of values of the parameters *default action time* and *utility noise* tested in the model.

Parameter	Range	Selected Value
Utility noise	0, 0.02, 0.04, 0.06, 0.08, 0.1, 0.12, 0.14, 0.16	0.1
Default action time	0.05, 0.07, 0.09, 0.11, 0.13	0.07

parameters for the aphasics' model M3 are specified in the last column of Table 4.2 along with other parameter values.

In the current model we also adjusted an ACT-R parameter that was not used in the previously published CBR models, the *utility values* of productions.[2] As described earlier, the utility value of a production determines the probability of selecting that production. It is evident from the Hanne et al. (2011) data (see Figure 4.3) and other psycholinguistic studies (Matzke, Mai, Nager, Rüsseler, & Münte, 2002; Ferreira, 2003) that non-canonical sentences are harder to process even for healthy individuals. The difficulty in processing non-canonical sentences is reflected in slow response time and increased error in the picture matching task. We adjusted the utility values of productions that process the case marking on the subject and object nouns. The assumption that we make here is that the error in processing non-canonical sentences is a consequence of error in utilizing the case marking information. When the information of case marking on the nouns is ignored, the parser fails to correctly mark the *agent* and *theme* roles of the sentence which, in turn, leads to an error in comprehending the sentence. We estimated the utility values of the productions that process nouns in the sentence. There were two alternative productions that could process the nouns—one that utilizes the case information and the other that does not (see Table 4.1). For each simulation, one of the two productions is selected depending on its utility value. The probability of selecting a production i given its utility value U_i is calculated using Equation 4.5.

We set the utility values of these productions such that non-canonical sentences were processed with accuracy observed for controls in the Hanne et al. (2011) data. The utility values that we used for these productions are listed in Table 4.4. The utility values were estimated *only* for the data from controls and then kept constant while modeling data from aphasics. The

[2] The utility value and utility noise value are different parameters. The utility value is associated with each production and can be different for different productions. On the other hand, the utility noise value is globally defined for the procedural memory system and it controls the variance in the utility value of each parameter.

reason behind not estimating utility values separately for aphasics is that we assume here that the grammatical knowledge of aphasics is the same as that of controls; i.e., no impairment in syntactic representations. This means the utility values of these productions were the same for models M1, M2, M3 and for the model for controls. Next we report the predictions of these models. Predictions for both offline and online responses are generated by averaging across 1000 runs of each model for each condition.

Table 4.4: Utility values estimated for the productions that process nouns.

Sentence Type	Production (task performed)	Utility Value
SVO	Process noun-1 use case info	0.6
	Process noun-1 ignore case info	0.4
	Process noun-2 use case info	0.6
	Process noun-2 ignore case info	0.4
OVS	Process noun-1 use case info	0.52
	Process noun-1 ignore case info	0.48
	Process noun-2 use case info	0.52
	Process noun-2 ignore case info	0.48

4.6.5 Results & Discussion

The accuracies and response times of the three models for the sentence-picture matching task are plotted in Figure 4.4. The exact values predicted by the models are listed in Table C.1 and Table C.2 in Appendix A. If we compare the accuracy values and response times in the data (Figure 4.3) with the models' predictions, model M1 with only slowed processing can qualitatively fit the longer response times observed in aphasics. However, the pattern for response accuracies in M1 is very different from that in the data—aphasics' model performs as accurately as controls' data across canonical and non-canonical conditions. On the other hand, model M2 captures the fact that aphasics make more errors in the sentence-picture matching task. Importantly, M2 also captures the classical chance level performance

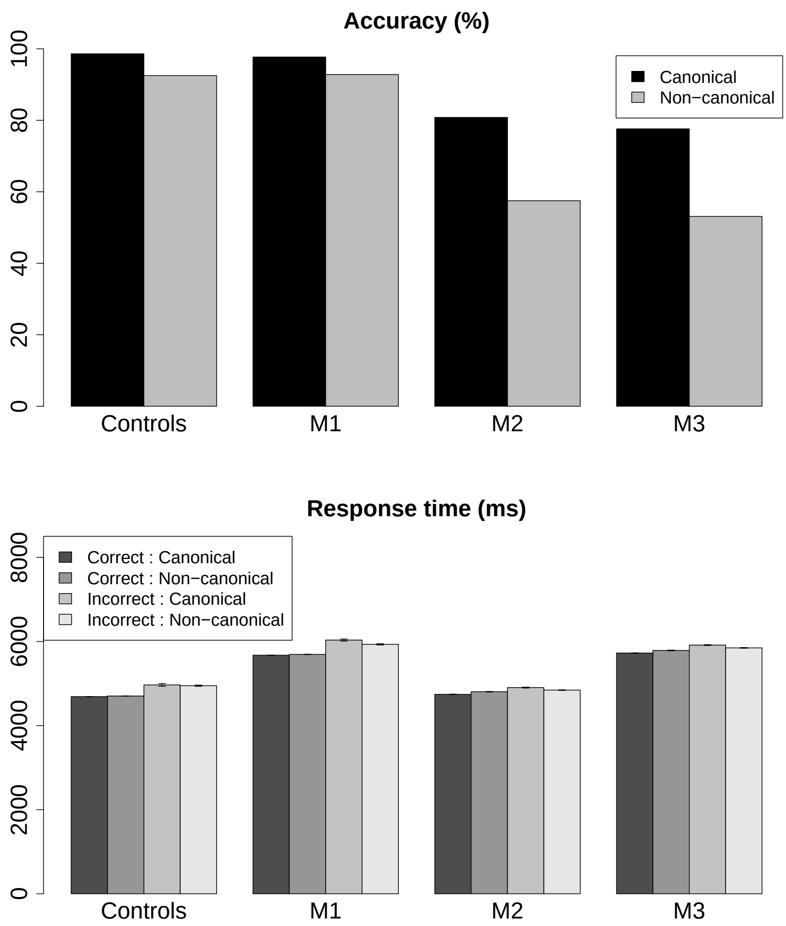


Figure 4.4: Accuracy and response time predictions of the model for controls and the models M1, M2 and M3. The predictions of accuracy are in terms of percentages of retrievals of the correct picture at the end of the sentence. The predictions of response time are equal to the models' processing time starting from the first word of the sentence until the retrieval of a picture at the end of the sentence. The error bars on response time show 95% confidence intervals.

in aphasics for non-canonical sentences—the model's accuracy of sentence-picture matching for non-canonical sentences is close to the 50% mark.

The response times predicted by model M2, however, are as fast as controls' response times, therefore M2 does not capture the delay in the response times of aphasics. In contrast, model M3, which combines slowed processing and intermittent deficiency, performs better in terms of modeling both response accuracies and response times. M3 captures the high percentage of errors in the sentence-picture matching task, the chance level performance for non-canonical structures, as well as the delay in response times found in the aphasics' data. Consequently, model M3 emerges as the best model of offline measures of aphasic sentence processing.

Next, we model the eye movement data using the linking hypothesis proposed above. According to this hypothesis, the models carry out retrievals of picture chunks from memory based on the semantic representation of the sentence at various stages of the input. The models' predictions about the percentages of fixations on the correct picture are derived from the difference in activations of the correct and incorrect picture at the time of these retrievals. The predictions of correct fixations are calculated from a total of 1000 runs of the model. The predictions for eye movement patterns for controls and aphasics are illustrated in Figure 4.5 (see Table C.3 in Appendix A for exact predicted values). The eye movements show percentage of fixations on the correct picture at various points in the sentences—'NP1', 'verb', 'NP2', 'silence'. For canonical sentences 'NP1' is the subject and 'NP2' is the object, and vice versa for the non-canonical sentences. 'Silence' is the time after presentation of the spoken sentence during which participants still look at both pictures until they respond by pressing a button to select the correct picture. Controls' data consisted of correct responses in the sentence-picture matching task, and since they performed close to ceiling, this constitutes most of their data. Aphasics' data was partitioned into correct and incorrect trials.

The predictions of models M1, M2 and M3 for eye movement patterns

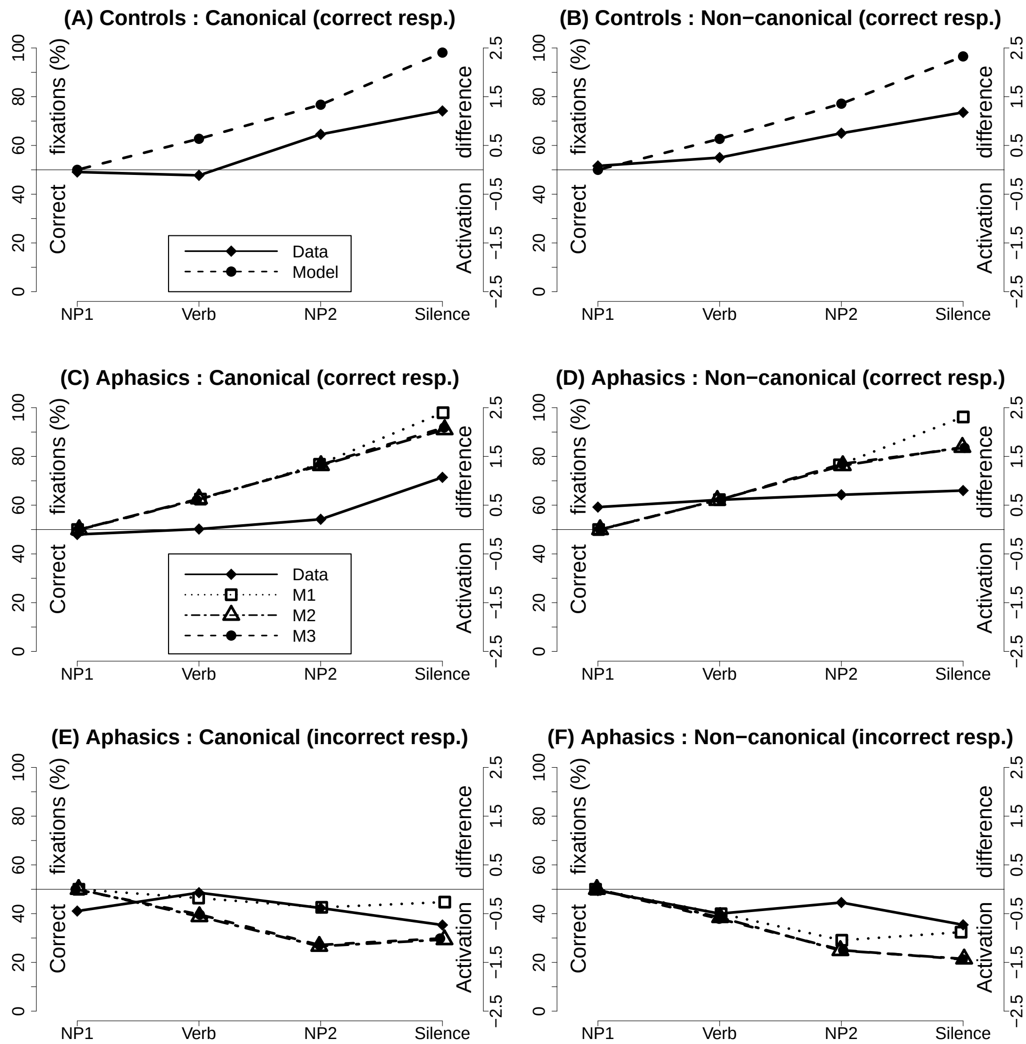


Figure 4.5: Eye movement data from Hanne et al. 2011 and the predictions of models M1, M2 & M3. The x-axis shows different regions in the sentences. NP1 and NP2 denote the two noun phrases in the sentences. For sentences in canonical word order, they denote subject and object respectively, and for sentences in non-canonical word order they denote object and subject respectively. The y-axis on the left-hand side of each plot shows percentages of fixations observed in the data and the y-axis on the right hand side shows differences in the activations of the two pictures in the models. An activation difference of zero is aligned with the 50% mark in the fixation percentages; the horizontal line represents this mark.

and the Hanne et al. (2011) eye movement data are compared in Figure 4.5. The data for controls and aphasics for correct trials shows an increase in fixations to the correct picture (see Figure 4.5 (A)-(D)) as more linguistic input is provided to the participants. We observe similar patterns in the predictions of the model for controls and all the three models for aphasics—the more the linguistic input, the higher the proportions of predicted correct fixations. When the data for correct responses is compared across the four conditions (sentence type (canonical vs. non-canonical) $\times$ participant type (control vs. aphasic)), it shows some variations in eye movements, especially in terms of the rate and point at which the increase in correct fixation takes place. In canonical sentences, controls start fixating on the correct picture at 'NP2' and in non-canonical sentences, they start fixating on the correct picture at 'verb'. The model's predictions for controls across both sentence types show a steeper increase in correct fixations after 'NP1' (at 'verb'); the model utilizes case marking information more accurately than the participants to disambiguate between the two pictures. The aphasics, on the other hand, fixate on the correct picture only in the 'silence' region for canonical sentences, but for non-canonical sentences they start fixating on the correct picture in the 'verb' region. When the predictions of the three models for correct responses are compared with the data (plots (C) & (D) in Figure 4.5), all three models appear to make similar predictions; in fact, models M2 and M3 make identical predictions. All models capture the pattern of increasing correct fixations with more linguistic input.

For incorrect trials in aphasics, the main pattern observed in the data was that the proportion of fixations on the correct picture were mostly below the 50% mark (see plots (E) & (F) in Figure 4.5). The predictions of the three models show similar patterns—the proportions of fixations on the correct picture at 'NP1' are at the 50% mark and for the following regions they go below the 50% mark. Models M2 and M3 again show identical patterns. On the whole, all three models for aphasics predict the critical pattern observed in eye movement data in Hanne et al. (2011)—the correct responses in aphasics are associated with normal online processing as in controls and incorrect responses are associated with incorrect online processing.

In sum, the model that assumes both slowed processing and intermittent deficiency, i.e. model M3, can capture the pattern of sentence-picture matching accuracy and the response time better than the model assuming only slowed processing (M1) or only intermittent deficiency (M2). Model M3 captures the main effect of participant type in the accuracy data: aphasics are predicted to have significantly lower accuracy than controls, and the main effect of word order: non-canonical sentences result in lower accuracy in the picture matching task, and, importantly, it captures the classical effect of chance level performance for non-canonical structures in agrammatic aphasia. Model M3 also captures the main effect of participant type in the response time data: aphasics are predicted to have higher response times than controls. No model captures the main effect of word order and the interaction between word order and response (correct or incorrect). For the eye movement data, all three models perform equally well—they capture the divergent eye movement patterns in correct vs. incorrect responses. Consequently when the predictions for both, online and offline, measures are considered, model M3 emerges as a better model of aphasic sentence processing than the other two.

4.7 General Discussion

The main goal of the modeling work reported in this paper was to employ the CBR architecture for modeling the sentence-picture matching study reported in Hanne et al. (2011). The reported study consists of offline and online data from aphasics and age-matched controls. In this modeling effort, we explored different assumptions put forward by processing accounts about sentence comprehension disorders in aphasia. In order to computationally investigate the claims made in Hanne et al. (2011), we made the same assumptions about the impairments in aphasic sentence processing as

Hanne and colleagues did. Hanne et al. (2011), along with others (Friederici & Kilborn, 1989; Caplan et al., 2007; Dickey et al., 2007), have proposed that the sentence comprehension disorder in aphasia is a result of processing impairments like *slowed processing* or *intermittent deficiency* rather than representational deficits. These assumptions were operationalized in the CBR architecture in terms of slowed procedural memory and extra noise in the utilities values of the parsing rules (productions in ACT-R). For offline measures, the modeling results revealed that assuming only slowed processing (M1) or only intermittent deficiency (M2) was not enough to model the accuracy and response time data. The model assuming both deficits (M3) outperformed models assuming only one of the two deficits.

For modeling online data—incremental eye movement patterns—we proposed a linking hypothesis. The hypothesis states that the model's predictions about the picture fixations are derived from the activation values of the two pictures presented on the screen. After each phrase is processed, based on the current partial representation of the sentence, the model retrieves one of the two existing picture chunks (each representing a picture on the screen) from the declarative memory. The difference in activation of the correct and incorrect picture is assumed to predict the likelihood of eye fixations on the pictures. The eye movement data for aphasics revealed critical differences between correctly answered trials and incorrectly answered trials. Aphasics' eye movements in correct trials were similar to those in controls, whereas, their eye movements in incorrect trials showed deviant online processing right from the first noun phrase in the sentence. These divergent patterns show that correct offline responses are associated with normal-like online processing and incorrect offline responses are associated with aberrant online processing. All three models predicted this divergent pattern in eye movements. It is important to note here that the assumptions about model parameters are made only for the offline data, and predictions of eye movements emerge without fitting any extra parameters for online processing in aphasics. Fitting extra parameters for modeling aphasics' online processing would have meant that the assumptions of slowed processing and intermittent deficiency are not sufficient to capture all aspects of the impairments in aphasics. The results of modeling online data, in fact, shows that any one of the two assumptions is enough to model eye movement data. On the whole, when both offline and online data are considered, the model assuming slowed processing *and* intermittent deficiency (M3) proved to be the best model of sentence processing in aphasia.

Though the aphasics' model captured some of the observed patterns in the data, not all the effects from Hanne et al. (2011) were accurately captured by the model. In Hanne et al. (2011), incorrect responses elicited significantly higher response times than correct responses; this effect was not predicted by any of the aphasics' models. The predictions for these data points could be improved if we adjusted other available parameters in the CBR architecture. However, since this was a first step towards a complete model of aphasic sentence processing, we parsimoniously adjusted values of only two parameters. We leave more precise parameter estimation for future work.

The modeling results are consistent with the hypothesis proposed by processing or resource reduction accounts of sentence comprehension impairments in aphasia. These accounts propose that aphasic sentence processing deficits should be ascribed to processing limitations and not to a breakdown in the underlying grammatical knowledge. In our model for aphasics, the grammatical knowledge, realized in terms of production rules, is kept the same as in the model for controls. Consequently, assuming processing deficits alone was sufficient to account for the accuracy and response time data. As observed in Caplan et al. (2007), Dickey et al. (2007) and Hanne et al. (2011), aphasic participants' correct offline responses are associated with normal online processing and incorrect responses with abnormal performance. Interestingly, without making extra processing or representational assumptions for online processing, the model's predictions for eye movements showed exactly these divergent patterns for correct and incorrect responses.

Other computational models of aphasia also support processing or resource reduction based explanations (Crescentini & Stocco, 2005; Haarmann & Kolk, 1991a; Haarmann et al., 1997), though they differ in terms of the precise processing deficit assumed. Crescentini and Stocco (2005) assume slow lexical activation, Haarmann and Kolk (1991a) assume temporal disruption, and Haarmann et al. (1997) assume reduction in memory resources for aphasics. The model we propose here also differs from these models in terms of the deficits assumed. But despite the differences in assumptions, we successfully modeled the classical effects in accuracy data—chance level performance in non-canonical structures. In fact, our model goes further than previous models because it can deliver predictions for response times as well as incremental eye movements. To our knowledge, this is the first effort in modeling online measures (fixation probabilities in the visual world paradigm) of aphasic sentence processing.

Furthermore, the linking hypothesis proposed here for modeling sentence-picture matching data opens new possibilities for modeling data from the visual world paradigm which is a frequently used methodology in psycholinguistics. Here we modeled mainly the task of sentence-picture matching, but the hypothesis can be extended for modeling other tasks in the visual world paradigm, such as the task described in Altmann and Kamide (1999), where participants listen to sentences while they are shown a semi-realistic scene containing mentioned and distractor objects, or the task described in Choy and Thompson (2010) where participants are shown an array of mentioned and distractor objects instead of a scene, or even the task described in Huettig and McQueen (2007) where participants are shown objects as textual words on the screen. For modeling these tasks, a more generic statement of the linking hypothesis can be developed: The chunk in declarative memory that receives higher activation as a consequence of processing the current sentence fragment is the object from the visual stimuli that is more likely to be fixated by the participants. This claim is consistent with the proposal about the linkage between language processing and visual attention presented in

Altmann and Kamide (2007).

In the current model, we limited our scope to testing the assumptions of slowed processing and intermittent deficiency accounts on aphasic sentence comprehension for non-canonical sentences in German. However, there are distinct claims about the underlying processing deficit in aphasia, such as a timing disorder proposed in Haarmann and Kolk (1991a), delayed lexical reactivation in Love et al. (2008) and lexical integration failure in Thompson and Choy (2009). Additionally, the experimental work in aphasia provides data for a wide range of linguistic structures like passives, clefts, subject/object relative clauses, etc. The goal of this modeling work was to test the claims from Hanne et al. (2011) and, importantly, to provide a computational model of online processing in aphasia that can investigate these claims in a detailed model of parsing. We acknowledge that a complete model of sentence processing in aphasia should compare predictions of various hypotheses across a range of linguistic structures. Considering that the CBR architecture, being grounded in ACT-R, provides a flexible environment for testing various assumptions about deficits in aphasia, this work provides a starting point for future work in the direction of delivering a complete model of sentence comprehension disorders in aphasia.

4.8 Conclusions

In this paper, we evaluated two hypotheses of aphasic sentence processing proposed by Hanne et al. (2011) using a computational modeling approach. The two hypotheses, namely slowed processing and intermittent deficiency, were framed in the cue-base retrieval architecture of sentence processing (Lewis & Vasishth, 2005). Both hypotheses were operationalized as deficits in the processing mechanism (the procedural memory system)—slow rate of production execution and high noise in the parsing process. No other impairments were assumed. Three separate models assuming either one of the two impairments or both were implemented. The predictions of the three models were evaluated using offline (sentence-picture matching accuracy and response time) and online (eye movements) data from the sentence-picture matching study reported in Hanne et al. (2011). For modeling online data, a new hypothesis was proposed. The modeling results showed that assuming only the two processing deficits was sufficient to capture the varied data patterns in aphasic offline and online processing.

4.9 Acknowledgements

We would like to thank David Caplan for elaborate discussions of the work reported here, and also Richard Lewis for helpful comments. Comments received during various meetings like the Academy of Aphasia 2010, 2011; CUNY 2011; and the Science of Aphasia conference 2011 were of great help.

Chapter 5

Conclusions

In this thesis, we evaluated a cognitive architecture of sentence processing, the cue-based retrieval theory, proposed in Lewis and Vasishth (2005) and Lewis et al. (2006). We assessed the architecture along three dimensions: (i) the empirical coverage of the theory, (ii) the response types that can be modeled in the architecture, and (iii) the types of validation tests for which the predictions of the theory hold. Furthermore, through these evaluations, we contributed to the advancement of two important issues in psycholinguistics: (i) the role of non-syntactic information in reflexive binding, and (ii) the possible explanations for impaired sentence comprehension. Next, we provide more details about each of the main contributions of the thesis.

First, we demonstrated the empirical coverage of the CBR theory. The set of modeling studies reported in Chapter 3 showed that the competing hypotheses about the retrieval process involved in reflexive binding can be evaluated in the CBR architecture. The analysis of these models led to a set of detailed predictions in terms of retrieval errors and retrieval latencies, which can be empirically evaluated to gain a better understanding of the reflexive binding process. Then, in Chapter 4, we demonstrated that two hypotheses about impaired sentence processing in aphasia, *slowed processing* and *intermittent deficiency*, can be formalized in a CBR architecture. Chapter 4 also suggested a method for modeling the processing difficulty associated with non-canonical structures.

Second, we demonstrated that the CBR architecture can be used to model a diverse set of response measures from empirical studies. The studies reported in Chapter 2 and 3 modeled eye movement patterns in reading, while the studies in Chapter 4 modeled eye movement patterns in the visual world paradigm as well as the response time measure in a sentence-picture matching task. In Chapter 3 and Chapter 2, we modeled two types of accuracy measures of sentence comprehension.

Third, we demonstrated the robustness of the predictions of the CBR theory with two types of validation methods: (i) a controlled empirical study, and (ii) a corpus study. In Chapter 3, we verified the predictions of a CBR model of reflexive binding using a controlled eye tracking study, and, in Chapter 2 we verified the predictions of the theory using an eye tracking corpus of naturally occurring sentences.

Finally, we used these assessments of the CBR theory to further our understanding of two issues in psycholinguistics. In Chapter 3, with a combination of modeling and empirical studies we provided support for the hypothesis that non-syntactic information is used in resolving the reflexive-antecedent dependency in English. In Chapter 4, we showed that assuming slowed processing and intermittent deficiency we can capture various aspects of aphasic sentence processing, and as a result, we provided support for processing deficit based accounts which propose that aphasic sentence processing deficits should be ascribed to processing limitations and not to a breakdown in the underlying grammatical knowledge.

Appendix A

PSC sentences & incremental parses

This appendix lists the set of 32 sentences from the Potsdam Sentence Corpus,[1] and the corresponding incremental parses, that the cue-based retrieval models generate (see Chapter 2) during processing.

1. **Sentence:** Den Ton gab der Küstler seinem Gehilfen.

   Input 1: Den

   Parse 1: [CP [SpecCP [DP$_j$ [D' [D Den] [NP [N' [N]]]]]] [C' [C V$_k$] [IP [I' [VP [SpecVP [DP]] [V t$_j$ t$_k$]]]]]]

   Input 2: Ton

   Parse 2: [CP [SpecCP [DP$_j$ [D' [D Den] [NP [N' [N Ton]]]]]] [C' [C V$_k$] [IP [I' [VP [SpecVP [DP]] [V t$_j$ t$_k$]]]]]]

   Input 3: gab

   Parse 3: [CP [SpecCP [DP$_j$ [D' [D Den] [NP [N' [N Ton]]]]]] [C' [C gab$_k$] [IP [I' [VP [SpecVP [DP]] [V' [DP] [V t$_j$ t$_k$]]]]]]

   Input 4: der

   Parse 4: [CP [SpecCP [DP$_j$ [D' [D Den] [NP [N' [N Ton]]]]]] [C' [C gab$_k$] [IP [I' [VP [SpecVP [DP [D' [D Der] [NP [N' [N]]]]]] [V' [DP] [V t$_j$ t$_k$]]]]]]

   Input 5: Küstler

   Parse 5: [CP [SpecCP [DP$_j$ [D' [D Den] [NP [N' [N Ton]]]]]] [C' [C gab$_k$] [IP [I' [VP [SpecVP [DP [D' [D Der] [NP [N' [N Küstler]]]]]] [V' [t$_j$] [V t$_k$]]]]]]]

[1] The sentence numbering used here is different from the numbering used in the Potsdam Sentence Corpus.

Input 6: seinem

Parse 6: [CP [SpecCP [DP$_j$ [D' [D Den] [NP [N' [N Ton]]]]]] [C' [C gab$_k$] [IP [I' [VP [SpecVP [DP [D' [D Der] [NP [N' [N Küstler]]]]]] [V' [DP [D' [D seinem] [NP[N' [N]]]]] [V t$_j$ t$_k$]]]]]]]

Input 7: Gehilfen

Parse 7: [CP [SpecCP [DP$_j$ [D' [D Den] [NP [N' [N Ton]]]]]] [C' [C gab$_k$] [IP [I' [VP [SpecVP [DP [D' [D Der] [NP [N' [N Küstler]]]]]] [V' [DP [D' [D seinem] [NP[N' [N Gehilfen]]]]] [V t$_j$ t$_k$]]]]]]]

2. **Sentence:** Der Hof lag weit außerhalb des eigentlichen Dorfes.

Input 1: Der

Parse 1: [CP [SpecCP [DP$_j$ [D' [D Der] [NP [N' [N]]]]]] [C' [C V$_k$] [IP [I' [VP [SpecVP t$_j$] [V t$_k$]]]]]]

Input 2: Hof

Parse 2: [CP [SpecCP [DP$_j$ [D' [D Der] [NP [N' [N Hof]]]]]] [C' [C V$_k$] [IP [I' [I [Aux]] [VP [SpecVP t$_j$] [Vt$_k$]]]]]]

Input 3: lag

Parse 3: [CP [SpecCP [DP$_j$ [D' [D Der] [NP [N' [N Hof]]]]]] [C' [C lag$_k$] [IP [SpecIP t$_j$] [I' [VP [SpecVP t$_j$] [V' PP [V t$_k$]]]]]]]

Input 4: weit

Parse 4: [CP [SpecCP [DP$_j$ [D' [D Der] [NP [N' [N Hof]]]]]] [C' [C lag$_k$] [IP [I' [VP [SpecVP t$_j$] [V' [AdvP [Adv' [AP [A' [A weit]]] [Adv]]] [V t$_k$]]]]]]]

Input 5: außerhalb

Parse 5: [CP [SpecCP [DP$_j$ [D' [D Der] [NP [N' [N Hof]]]]]] [C' [C lag$_k$] [IP [SpecIP t$_j$] [I' [VP [SpecVP t$_j$] [V' [AdvP [Adv' [AP [A' [A weit]]] [Adv außerhalb]]] V t$_k$]]]]]

Input 6: des

Parse 6: [CP [SpecCP [DP$_j$ [D' [D Der] [NP [N' [N Hof]]]]]] [C' [C lag$_k$] [IP [SpecIP t$_j$] [I' [VP [SpecVP t$_j$] [V' [AdvP [Adv' [AP [A' [A weit]]] [Adv außerhalb]] [DP [D' [D des] [NP [N' [N]]]]] [V t$_k$]]]]]]]

Input 7: eigentlichen

Parse 7: [CP [SpecCP [DP$_j$ [D' [D Der] [NP [N' [N Hof]]]]]] [C' [C lag$_k$] [IP [SpecIP t$_j$] [I' [VP [SpecVP t$_j$] [V' [AdvP [Adv' [AP [A' [A weit]]] [Adv außerhalb]] [DP [D' [D des] [NP [AP [A' [A eigentlichen]]][N' [N]]]]] [V t$_k$]]]]]]]

Input 8: Dorfes

Parse 8: [CP [SpecCP [DP$_j$ [D' [D Der] [NP [N' [N Hof]]]]]] [C' [C lag$_k$] [IP [SpecIP t$_j$] [I' [VP [SpecVP t$_j$] [V' [AdvP [Adv' [AP [A' [A weit]]] [Adv außerhalb]] [DP [D' [D des] [NP [AP [A' [A eigentlichen]]][N' [N Dorfes]]]]]] [V t$_k$]]]]]]]

3. **Sentence:** Die Wanderer sahen Rehe auf einer Lichtung im Wald äsen.

Input 1: Die

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N]]]]]] [C' [C t$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V t$_j$]] [I t$_j$]]]]]

Input 2: Wanderer

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Wanderer]]]]]] [C' [C t$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V t$_j$]] [I t$_j$]]]]]

Input 3: sahen

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Wanderer]]]]]] [C' [C sahen$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V' [DP] [V t$_j$]]] [I t$_j$]]]]]

Input 4: Rehe

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Wanderer]]]]]] [C' [C sahen$_j$] [IP [I' [VP [SpecVP t$_i$] [V' [DP [D' [D Ø] [NP [N' [N Rehe]]]]] [V t$_j$]]] [I t$_j$]]]]]

Input 5: auf

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Wanderer]]]]]] [C' [C sahen$_j$] [IP [I' [VP [SpecVP t$_i$] [V' [DP [D' [D Ø] [NP [N' [N Rehe]]]]] [V' [PP [P' [P auf] [DP]]] [V t$_j$]]]] [I t$_j$]]]]]

Input 6: einer

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Wanderer]]]]]] [C' [C sahen$_j$] [IP [I' [VP [SpecVP t$_i$] [V' [DP [D' [D Ø] [NP [N' [N Rehe]]]]] [V' [PP [P' [P auf] [DP [D' [D einer] [NP [N' [N]]]]]] [V t$_j$]]]] [I t$_j$]]]]]

Input 7: Lichtung

Parse 7: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Wanderer]]]]]] [C' [C sahen$_j$] [IP [I' [VP [SpecVP t$_i$] [V' [DP[D' [D Ø] [NP [N' [N Rehe]]]]] [V' [PP [P' [P auf] [DP [D' [D einer] [NP [N' [N Lichtung]]]]]]] [V t$_j$]]]] [I t$_j$]]]]]

Input 8: im

Parse 8: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Wanderer]]]]]] [C' [C sahen$_j$] [IP [I' [VP [SpecVP t$_i$] [V' [DP [D' [D Ø] [NP [N' [N Rehe]]]]] [V' [PP [P' [P auf] [DP [D' [D einer] [NP [N' [N Lichtung] [PP [P' [P' [P in] [X$_k$ + (de)m]] [DP [D' [D t$_k$] [NP [N' [N]]]]]]]]]]]] [V t$_j$]]]] [I t$_j$]]]]]

Input 9: Wald

Parse 9: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Wanderer]]]]]] [C' [C sahen$_j$] [IP [I' [VP [SpecVP t$_i$] [V' [DP [D' [D Ø] [NP [N' [N Rehe]]]]] [V' [PP [P' [P auf] [DP [D' [D einer] [NP [N' [N Lichtung] [PP [P' [P' [P in] [X$_k$ + (de)m]] [DP [D' [D t$_k$] [NP [N' [N Wald]]]]]]]]]]]] [V t$_j$]]]] [I t$_j$]]]]]

Input 10: äsen

Parse 10: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Wanderer]]]]]] [C' [C sahen$_j$] [IP [I' [VP [SpecVP t$_i$] [V' [VP [V' [DP [D' [D Ø] [NP [N' [N Rehe]]]]] [V' [PP [P' [P auf] [DP [D' [D einer] [NP [N' [N Lichtung] [PP [P' [P' [P in] [X$_k$ + (de)m]] [DP [D' [D t$_k$] [NP [N' [N Wald]]]]]]]]]]]] [V äsen]]]] [V t$_j$]]] [I t$_j$]]]]]

4. **Sentence:** Den Kopf hieb man früher nur Mördern und Verrätern ab.

Input 1: Den

Parse 1: [CP [SpecCP [DP$_k$ [D' [D Den] [NP [N' [N]]]]] [C' [C V$_j$] [IP [SpecIP DP$_i$] [I' [VP [SpecVP t$_i$] [V' t$_k$ [V t$_j$]]] [I [V t$_j$]]]]]]

Input 2: Kopf

Parse 2: [CP [SpecCP [DP$_k$ [D' [D Den] [NP [N' [N Kopf]]]]]] [C' [C V$_j$] [IP [SpecIP DP$_i$] [I' [VP [SpecVP t$_i$] [V' t$_k$ [V t$_j$]]] [I [V t$_j$]]]]]]

Input 3: hieb

Parse 3: [CP [SpecCP [DP$_k$ [D' [D Den] [NP [N' [N Kopf]]]]]] [C' [C hieb$_j$] [IP [SpecIP DP$_i$] [I' [VP [SpecVP t$_i$] [V' t$_k$ [V [Part] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 4: man

Parse 4: [CP [SpecCP [DP$_k$ [D' [D Den] [NP [N' [N Kopf]]]]]] [C' [C hieb$_j$] [IP [SpecIP [DP$_i$ [D' [D Ø] [NP [N' [N man]]]]]] [I' [VP [SpecVP t$_i$] [V' t$_k$ [V [Part] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 5: früher

Parse 5: [CP [SpecCP [DP_k [D' [D Den] [NP [N' [N Kopf]]]]]] [C' [C $hieb_j$] [IP [SpecIP [DP_i [D' [D Ø] [NP [N' [N man]]]]]] [I' [VP [AdvP [Adv' [Adv früher]]] [VP [SpecVP t_i] [V' t_k [V [Part] [V t_j]]]]] [I [V t_j]]]]]]

Input 6: nur

Parse 6: [CP [SpecCP [DP_k [D' [D Den] [NP [N' [N Kopf]]]]]] [C' [C $hieb_j$] [IP [SpecIP [DP_i [D' [D Ø] [NP [N' [N man]]]]]] [I' [VP [AdvP [Adv' [Adv früher]]] [VP [SpecVP t_i] [V' [Part nur] [V' t_k [V [Part] [V t_j]]]]]] [I [V t_j]]]]]]

Input 7: Mördern

Parse 7: [CP [SpecCP [DP_k [D' [D Den] [NP [N' [N Kopf]]]]]] [C' [C $hieb_j$] [IP [SpecIP [DP_i [D' [D Ø] [NP [N' [N man]]]]]] [I' [VP [AdvP [Adv' [Adv früher]]] [VP [SpecVP t_i] [V' [Part nur] [V' [DP [D' [D Ø] [NP [N' [N Mördern]]]]] [V' t_k [V [Part] [V t_j]]]]]]] [I [V t_j]]]]]]

Input 8: und

Parse 8: [CP [SpecCP [DP_k [D' [D Den] [NP [N' [N Kopf]]]]]] [C' [C $hieb_j$] [IP [SpecIP [DP_i [D' [D Ø] [NP [N' [N man]]]]]] [I' [VP [AdvP [Adv' [Adv früher]]] [VP [SpecVP t_i] [V' [Part nur] [V' [DP [D' [D Ø] [NP [N' [NP [N' [N Mördern]]] [COORD und] [NP]]]]] [V' t_k [V [Part] [Vt_j]]]]]]] [I [V t_j]]]]]]

Input 9: Verrätern

Parse 9: [CP [SpecCP [DP_k [D' [D Den] [NP [N' [N Kopf]]]]]] [C' [C $hieb_j$] [IP [SpecIP [DP_i [D' [D Ø] [NP [N' [N man]]]]]] [I' [VP [AdvP [Adv' [Adv früher]]] [VP [SpecVP t_i] [V' [Part nur] [V' [DP [D' [D Ø] [NP [N' [NP [N' [N Mördern]]] [COORD und] [NP [N' [N Verrätern]]]]]]] [V' t_k [V [Part] [V t_j]]]]]]] [I [V t_j]]]]]]

Input 10: ab

Parse 10: [CP [SpecCP [DP_k [D' [D Den] [NP [N' [N Kopf]]]]]] [C' [C $hieb_j$] [IP [SpecIP [DP_i [D' [D Ø] [NP [N' [N man]]]]]] [I' [VP [AdvP [Adv' [Adv früher]]] [VP [SpecVP t_i] [V' [Part nur] [V' [DP [D' [D Ø] [NP [N' [NP [N' [N Mördern]]] [COORD und] [NP [N' [N Verrätern]]]]]]] [V' t_k [V [Part ab] [V t_j]]]]]]] [I [V t_j]]]]]]

5. **Sentence:** Vorne am Bug sah man eine prächtige Galionsfigur.

   Input 1: Vorne

   Parse 1: [CP [SpecCP [AdvP [Adv' [Adv Vorne] [PP]]]] [C' [C V_k] [IP [VP [SpecVP DP] [V t_k]]]]]

   Input 2: am

   Parse 2: [CP [SpecCP [AdvP [Adv' [Adv Vorne] [PP [P' [P' [P an] [X_i + (de)m]] [DP [D' [D t_i] [NP [N' [N]]]]]]]]]] [C' [C V_k] [IP [VP [SpecVP DP] [V t_k]]]]]

   Input 3: Bug

   Parse 3: [CP [SpecCP [AdvP [Adv' [Adv Vorne] [PP [P' [P' [P an] [X_i + (de)m]] [DP [D' [D t_i] [NP [N' [N Bug]]]]]]]]]] [C' [C V_k] [IP [VP [SpecVP DP] [V t_k]]]]]

   Input 4: sah

   Parse 4: [CP [SpecCP [AdvP [Adv' [Adv Vorne] [PP [P' [P' [P an] [X_i + (de)m]] [DP [D' [D t_i] [NP [N' [N Bug]]]]]]]]]] [C' [C sah_k] [IP [VP [SpecVP DP] [V' DP]] [V t_k]]]]

   Input 5: man

   Parse 5: [CP [SpecCP [AdvP [Adv' [Adv Vorne] [PP [P' [P' [P an] [X_i + (de)m]] [DP [D' [D t_i] [NP [N' [N Bug]]]]]]]]]] [C' [C sah_k] [IP [VP [SpecVP [DP [D' [D Ø] [NP [N' [N man]]]]]] [V' DP]] [V t_k]]]]

   Input 6: eine

   Parse 6: [CP [SpecCP [AdvP [Adv' [Adv Vorne] [PP [P' [P' [P an] [X_i + (de)m]] [DP [D' [D t_i] [NP [N' [N Bug]]]]]]]]]] [C' [C sah_k] [IP [VP [SpecVP [DP [D' [D Ø] [NP [N' [N man]]]]]] [V' [DP [D' [D eine] [NP [N' [N]]]]]]] [V t_k]]]]

   Input 7: prächtige

   Parse 7: [CP [SpecCP [AdvP [Adv' [Adv Vorne] [PP [P' [P' [P an] [X_i + (de)m]] [DP [D' [D t_i] [NP [N' [N Bug]]]]]]]]]] [C' [C sah_k] [IP [VP [SpecVP [DP [D' [D Ø] [NP [N' [N man]]]]]] [V' [DP [D' [D eine] [NP [N' [AP [A' [A prächtige]]] [N]]]]]]] [V t_k]]]]

   Input 8: Galionsfigur

   Parse 8: [CP [SpecCP [AdvP [Adv' [Adv Vorne] [PP [P' [P' [P an] [X_i + (de)m]] [DP [D' [D t_i] [NP [N' [N Bug]]]]]]]]]] [C' [C sah_k] [IP [VP [SpecVP [DP [D' [D Ø] [NP [N' [N man]]]]]] [V' [DP [D' [D eine] [NP [N' [AP [A' [A prächtige]]] [N Galionsfigur]]]]]]] [V t_k]]]]

6. **Sentence:** Torsten beobachtete gestern eine Maus, die Efeu fraß.

   Input 1: Torsten

   Parse 1: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Torsten]]]]]] [C' [C t$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V t$_j$]] [I t$_j$]]]]]

   Input 2: beobachtete

   Parse 2: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Torsten]]]]]] [C' [C beobachtete$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V' [DP] [V t$_j$]]] [I t$_j$]]]]]

   Input 3: gestern

   Parse 3: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Torsten]]]]]] [C' [C beobachtete$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V' [AdvP [Adv' [Adv gestern]]] [V' [DP] [V t$_j$]]]] [I t$_j$]]]]]

   Input 4: eine

   Parse 4: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Torsten]]]]]] [C' [C beobachtete$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V' [AdvP [Adv' [Adv gestern]]] [V' [DP [D' [D eine] [NP [N' [N]]]]] [V t$_j$]]]] [I t$_j$]]]]]

   Input 5: Maus

   Parse 5: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Torsten]]]]]] [C' [C beobachtete$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V' [AdvP [Adv' [Adv gestern]]] [V' [DP [D' [D eine] [NP [N' [N Maus]]]]] [V t$_j$]]]] [I t$_j$]]]]]

   Input 6: die

   Parse 6: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Torsten]]]]]] [C' [C beobachtete$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V' [AdvP [Adv' [Adv gestern]]] [V' [DP [D' [D eine] [NP$_k$ [N' [N Maus$_k$] [CP [SpecCP [DP$_l$ [D' [D Ø] [NP [N' [N die$_k$]]]]]] [C' [C t$_m$] [IP [SpecIP t$_l$] [I' [VP [SpecVP t$_l$] [V t$_m$]] [I t$_m$]]]]]]]] [V t$_j$]]]] [I t$_j$]]]]]

   Input 7: Efeu

   Parse 7: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Torsten]]]]]] [C' [C beobachtete$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V' [AdvP [Adv' [Adv gestern]]] [V' [DP [D' [D eine] [NP$_k$ [N' [N Maus$_k$] [CP [SpecCP [DP$_l$ [D' [D Ø] [NP [N' [N die$_k$]]]]]] [C' [C t$_m$] [IP [SpecIP t$_l$] [I' [VP [SpecVP t$_l$] [V' [DP [D' [D Ø] [NP [N' [N Efeu]]]]] [V t$_m$]]] [I t$_m$]]]]]]]] [V t$_j$]]]] [I t$_j$]]]]]

Input 8: fraß

Parse 8: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Torsten]]]]]] [C' [C beobachtete$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V' [AdvP [Adv' [Adv gestern]]] [V' [DP [D' [D eine] [NP$_k$ [N' [N Maus$_k$] [CP [SpecCP [DP$_l$ [D' [D Ø] [NP [N' [N die$_k$]]]]]] [C' [C t$_m$] [IP [SpecIP t$_l$] [I' [VP [SpecVP t$_l$] [V' [DP [D' [D Ø] [NP [N' [N Efeu]]]]] [V t$_m$]]] [I fraß$_m$]]]]]]]] [V t$_j$]]]] [I t$_j$]]]]]

7. **Sentence:** Der schüchterne kleine Gnom mied die Nähe der Elfen.

Input 1: Der

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [N' [N]]]]]] [C' [C V$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [t$_j$]]]]]]

Input 2: schüchterne

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [AP [A' [A schüchterne]]] [N' [N]]]]]] [C' [C V$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [t$_j$]]]]]]

Input 3: kleine

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [AP [A' [A schüchterne]]] [N' [AP [A' [A kleine]]] [N' [N]]]]]]] [C' [C V$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [t$_j$]]]]]]

Input 4: Gnom

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [AP [A' [A schüchterne]]] [N' [AP [A' [A kleine]]] [N' [N Gnom]]]]]]] [C' [C V$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [t$_j$]]]]]]

Input 5: mied

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [AP [A' [A schüchterne]]] [N' [AP [A' [A kleine]]] [N' [N Gnom]]]]]]] [C' [C mied$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V' [DP] [V t$_j$]]] [I [t$_j$]]]]]]

Input 6: die

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [AP [A' [A schüchterne]]] [N' [AP [A' [A kleine]]] [N' [N Gnom]]]]]]] [C' [C mied$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D die] [NP [N' [N]]]]] [V t$_j$]]] [I [t$_j$]]]]]]

Input 7: Nähe

Parse 7: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [AP [A' [A schüchterne]]] [N' [AP [A' [A kleine]]] [N' [N Gnom]]]]]]] [C' [C mied$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D die] [NP [N' [N Nähe] [DP]]]]] [V t$_j$]]] [I [t$_j$]]]]]]

Input 8: der

Parse 8: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [AP [A' [A schüchterne]]] [N' [AP [A' [A kleine]]] [N' [N Gnom]]]]]]] [C' [C mied$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D die] [NP [N' [N Nähe] [DP [D' [D der] [NP [N' [N]]]]]]]]] [V t$_j$]]] [I [t$_j$]]]]]]

Input 9: Elfen

Parse 9: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [AP [A' [A schüchterne]]] [N' [AP [A' [A kleine]]] [N' [N Gnom]]]]]]] [C' [C mied$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D die] [NP [N' [N Nähe] [DP [D' [D der] [NP[N' [N Elfen]]]]]]]]] [V t$_j$]]] [I [t$_j$]]]]]]

8. **Sentence:** Claudia hatte ihr Fahrrad auf der Straße stehen gelassen.

Input 1: Claudia

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Claudia]]]]]] [C' [C t$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V t$_j$]] [I t$_j$]]]]]

Input 2: hatte

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Claudia]]]]]] [C' [C hatte$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V' [DP] [V t$_j$]] [V t$_j$]] [I t$_j$]]]]]

Input 3: ihr

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Claudia]]]]]] [C' [C hatte$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V' [DP [D' [D ihr] [NP [N' [N]]]]] [V V [V t$_j$]]]] [I t$_j$]]]]]

Input 4: Fahrrad

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Claudia]]]]]] [C' [C hatte$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V' [DP [D' [D ihr] [NP [N' [N Fahrrad]]]]] [V V [V t$_j$]]]] [I t$_j$]]]]]

Input 5: auf

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Claudia]]]]]] [C' [C hatte$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V' [DP [D' [D ihr] [NP [N' [N Fahrrad]]]]] [V' [PP [P' [P auf] [DP]]] [V V [V t$_j$]]]]] [I t$_j$]]]]]

Input 6: der

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Claudia]]]]] [C' [C hatte$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V' [DP [D' [D ihr] [NP [N' [N Fahrrad]]]]] [V' [PP [P' [P auf] [DP [D' [D der] [NP [N' [N]]]]]]] [V V [V t$_j$]]]]] [I t$_j$]]]]]

Input 7: Straße

Parse 7: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Claudia]]]]] [C' [C hatte$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V' [DP [D' [D ihr] [NP [N' [N Fahrrad]]]]] [V' [PP [P' [P auf] [DP [D' [D der] [NP [N' [N Strasse]]]]]]] [V V [V t$_j$]]]]] [I t$_j$]]]]]

Input 8: stehen

Parse 8: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Claudia]]]]] [C' [C hatte$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V' [DP [D' [D ihr] [NP [N' [N Fahrrad]]]]] [V' [PP [P' [P auf] [DP [D' [D der] [NP [N' [N Strasse]]]]]]] [V [V stehen] [V t$_j$]]]]] [I t$_j$]]]]]

Input 9: gelassen

Parse 9: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Claudia]]]]] [C' [C hatte$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP t$_i$] [V' [DP [D' [D ihr] [NP [N' [N Fahrrad]]]]] [V' [PP [P' [P auf] [DP [D' [D der] [NP [N' [N Strasse]]]]]]] [V [V [V stehen] [V gelassen]] [V t$_j$]]]]] [I t$_j$]]]]]

9. **Sentence:** Wir hätten schon vor einer Stunde wissen sollen, ob ihr kommt.

Input 1: Wir

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Wir]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 2: hätten

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Wir]]]]]] [C' [C [V hätten$_j$]] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [V] [V' [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 3: schon

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Wir]]]]]] [C' [C [V hätten$_j$]] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [AdvP [Adv' [Adv schon]]] [V' [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 4: vor

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Wir]]]]]] [C' [C [V hätten$_j$]] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [AdvP [Adv' [Adv schon]]] [V' [PP [P vor] [NP]] [V' [V t$_j$]]]]] [I [V t$_j$]]]]]]

Input 5: einer

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Wir]]]]]] [C' [C [V hätten$_j$]] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [AdvP [Adv' [Adv schon]]] [V' [PP [P vor] [DP [D' [D einer] [NP [N' [N]]]]]] [V' [V t$_j$]]]]] [I [V t$_j$]]]]]]

Input 6: Stunde

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Wir]]]]]] [C' [C [V hätten$_j$]] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [AdvP [Adv' [Adv schon]]] [V' [PP [P vor] [DP [D' [D einer] [NP [N' [N Stunde]]]]]] [V' [V t$_j$]]]]] [I [V t$_j$]]]]]]

Input 7: wissen

Parse 7: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Wir]]]]]] [C' [C [V hätten$_j$]] [IP [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [AdvP [Adv' [Adv schon]]] [V' [PP [P vor] [DP [D' [D einer] [NP [N' [N Stunde]]]]]] [V' [V t$_j$] [V' [V' [CP [t]] [V' [V wissen] [V t$_j$]]]]]]]] [I [V t$_j$]]]] [CP [SpecCP] [C' [C] [IP [SpecIP [DP$_k$]] [I' [VP [SpecVP [t$_k$]] [V' [V t$_l$]]] [I [t$_l$]]]]]]]]

Input 8: sollen

Parse 8: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Wir]]]]]] [C' [C [V hätten$_j$]] [IP [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [AdvP [Adv' [Adv schon]]] [V' [PP [P vor] [DP [D' [D einer] [NP [N' [N Stunde]]]]]] [V' [V t$_j$] [V' [V' [CP [t]] [V [V [V wissen] [V sollen]] [V t$_j$]]]]]]]] [I t$_j$]]] [CP [SpecCP] [C' [C] [IP [SpecIP [DP$_k$]] [I' [VP [SpecVP [t$_k$]] [V' [V t$_l$]]] [I [t$_l$]]]]]]]]

Input 9: ob

Parse 9: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Wir]]]]]] [C' [C [V hätten$_j$]] [IP [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [AdvP [Adv' [Adv schon]]] [V' [PP [P vor] [DP [D' [D einer] [NP [N' [N Stunde]]]]]] [V' [V t$_j$] [V' [V' [CP [t]] [V [V [V wissen] [V sollen]] [V t$_j$]]]]]]]] [I t$_j$]]] [CP [SpecCP] [C' [C ob] [IP [SpecIP [DP$_k$]] [I' [VP [SpecVP [t$_k$]] [V' [V t$_l$]]] [I [t$_l$]]]]]]]]

Input 10: ihr

Parse 10: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Wir]]]]]] [C' [C [V hätten$_j$]] [IP [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [AdvP [Adv' [Adv schon]]] [V' [PP [P vor] [DP [D' [D einer] [NP [N' [N Stunde]]]]]] [V' [V t$_j$] [V' [V' [CP [t]] [V [V [V wissen] [V sollen]] [V t$_j$]]]]]]]] [I t$_j$]]] [CP [SpecCP] [C' [C ob] [IP [SpecIP [DP$_k$ [D' [D Ø][NP [N' [N ihr]]]]]] [I' [VP [SpecVP [t$_k$]] [V' [V t$_l$]]] [I [t$_l$]]]]]]]]

Input 11: kommt

[Parse11:] [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Wir]]]]]] [C' [C [V hätten$_j$]] [IP [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [AdvP [Adv' [Adv schon]]] [V' [PP [P vor] [DP [D' [D einer] [NP [N' [N Stunde]]]]]] [V' [V t$_j$] [V' [V' [CP [t]] [V [V [V wissen] [V sollen]] [V t$_j$]]]]]]]] [I t$_j$]]] [CP [SpecCP] [C' [C ob] [IP [SpecIP [DP$_k$ [D' [D Ø][NP [N' [N ihr]]]]]] [I' [VP [SpecVP [t$_k$]] [V' [V kommt$_l$]]] [I [t$_l$]]]]]]]]

10. **Sentence:** Die Eltern konnten ihre Kinder im Garten raufen hören.

Input 1: Die

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Die] [NP]]] [C' [C V$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [t$_j$]]]]]]]

Input 2: Eltern

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Eltern]]]]] [C' [C V$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [t$_j$]]]]]]]

Input 3: konnten

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Eltern]]]]] [C' [C konnten$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V' [V] [V t$_j$]]] [I [t$_j$]]]]]]]

Input 4: ihre

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Eltern]]]]] [C' [C konnten$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D ihre] [NP [N' [N]]]]] [V' [V] [V t$_j$]]]] [I [t$_j$]]]]]]]

Input 5: Kinder

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Eltern]]]]] [C' [C konnten$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D ihre] [NP [N' [N Kinder]]]]] [V' [V] [V t$_j$]]]] [I [t$_j$]]]]]]]

Input 6: im

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Eltern]]]]] [C' [C konnten$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D ihre] [NP [N' [N Kinder]]]]] [V' [PP [P' [P [P in] [X$_k$ +(de)m]] [DP [D' [D t$_k$] [NP [N' [N]]]]]] [V' [V] [V t$_j$]]]]] [I [t$_j$]]]]]]]

Input 7: Garten

Parse 7: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Eltern]]]]] [C' [C konnten$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D ihre] [NP [N' [N Kinder]]]]] [V' [PP [P' [P [P in] [X$_k$ +(de)m]] [DP [D' [D t$_k$] [NP [N' [N Garten]]]]]] [V' [V] [V t$_j$]]]]] [I [t$_j$]]]]]]]

Input 8: raufen

Parse 8: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Eltern]]]]] [C' [C konnten$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V' [VP [V' [DP [D' [D ihre] [NP [N' [N Kinder]]]]] [V' [PP [P' [P [P in] [X$_k$ +(de)m]] [DP [D' [D t$_k$] [NP [N' [N Garten]]]]]] [V raufen]]]] [V [V] [V t$_j$]]]] [I [t$_j$]]]]]]]

Input 9: hören

Parse 9: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Eltern]]]]] [C' [C konnten$_j$] [IP [SpecIP t$_i$] [I' [VP [SpecVP [t$_i$]] [V' [VP [V' [DP [D' [D ihre] [NP [N' [N Kinder]]]]] [V' [PP [P' [P [P in] [X$_k$ +(de)m]] [DP [D' [D t$_k$] [NP [N' [N Garten]]]]]] [V raufen]]]] [V [V hören] [V t$_j$]]]] [I [t$_j$]]]]]]]

11. **Sentence:** Wegen ihrer Diät hatte die Gräfin leider keine Austern nehmen dürfen

Input 1: Wegen

Parse 1: [CP [SpecCP [PP [P Wegen] [DP]]] [C' [C] [IP [VP [SpecVP [DP]] [V' [V ti]]]]]]

Input 2: ihrer

Parse 2: [CP [SpecCP [PP [P Wegen] [DP [D' [D ihrer]] [NP [N' [N]]]]]] [C' [C] [IP [VP [SpecVP [DP]] [V' [V ti]]]]]]

Input 3: Diät

Parse 3: [CP [SpecCP [PP [P Wegen] [DP [D' [D ihrer]] [NP [N' [N Diät]]]]]] [C' [C] [IP [VP [SpecVP [DP]] [V' [V ti]]]]]]

Input 4: hatte

Parse 4: [CP [SpecCP [PP [P Wegen] [DP [D' [D ihrer]] [NP[N' [N Diät]]]]]] [C' [C hatte] [IP [VP [SpecVP [DP]] [V' [V] [V ti]]]]]]

Input 5: die

Parse 5: [CP [SpecCP [PP [P Wegen] [DP [D' [D ihrer]] [NP[N' [N Diät]]]]]] [C' [C hatte] [IP [VP [SpecVP [DP [D' [D die] [NP [N' [N]]]]]] [V' [V] [V ti]]]]]]

Input 6: Gräfin

Parse 6: [CP [SpecCP [PP [P Wegen] [DP [D' [D ihrer]] [NP[N' [N Diät]]]]]] [C' [C hatte] [IP [VP [SpecVP [DP [D' [D die] [NP [N' [N Gräfin]]]]]] [V' [V] [V ti]]]]]]

Input 7: leider

Parse 7: [CP [SpecCP [PP [P Wegen] [DP [D' [D ihrer]] [NP[N' [N Diät]]]]]] [C' [C hatte] [IP [VP [SpecVP [DP [D' [D die] [NP [N' [N Gräfin]]]]]] [V' [AdvP [Adv' [Adv leider]]] [V' [V] [V ti]]]]]]]

Input 8: keine

Parse 8: [CP [SpecCP [PP [P Wegen] [DP [D' [D ihrer]] [NP[N' [N Diät]]]]]] [C' [C hatte] [IP [VP [SpecVP [DP [D' [D die] [NP [N' [N Gräfin]]]]]] [V' [AdvP [Adv' [Adv leider]] [DP [D' [D keine] [NP [N' [N]]]]]] [V' [V] [V ti]]]]]]]

Input 9: Austern

Parse 9: [CP [SpecCP [PP [P Wegen] [DP [D' [D ihrer]] [NP[N' [N Diät]]]]]] [C' [C hatte] [IP [VP [SpecVP [DP [D' [D die] [NP [N' [N Gräfin]]]]]] [V' [AdvP [Adv' [Adv leider]] [DP [D' [D keine] [NP [N' [N Austern]]]]]] [V' [V] [V ti]]]]]]]

Input 10: nehmen

Parse 10: [CP [SpecCP [PP [P Wegen] [DP [D' [D ihrer]] [NP[N' [N Diät]]]]]] [C' [C hatte] [IP [VP [SpecVP [DP [D' [D die] [NP [N' [N Gräfin]]]]]] [V' [AdvP [Adv' [Adv leider]] [DP [D' [D keine] [NP [N' [N Austern]]]]]] [V' [V nehmen] [V ti]]]]]]]

Input 11: dürfen

[Parse11:] [CP [SpecCP [PP [P Wegen] [DP [D' [D ihrer]] [NP[N' [N Diät]]]]]] [C' [C hatte] [IP [VP [SpecVP [DP [D' [D die] [NP [N' [N Gräfin]]]]]] [V' [AdvP [Adv' [Adv leider]] [DP [D' [D keine] [NP [N' [N Austern]]]]]] [V' [V nehmen V dürfen] [V ti]]]]]]]

12. **Sentence:** Man sollte nie Geschirr mit einem dreckigen Lappen spülen

müssen.

Input 1: Man

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Man]]]]]] [C' [C V$_j$] [IP [SpecIP [t_i]] [I' [VP [SpecVP [t_i]] [V t_j]] [I t_j]]]]]

Input 2: sollte

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Man]]]]]] [C' [C sollte$_j$] [IP [SpecIP [t_i]] [I' [VP [SpecVP [t_i]] [V' [VP [Spec PRO$_i$] [V' [V' [V]]]] [V t_j]]] [I t_j]]]]]

Input 3: nie

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Man]]]]]] [C' [C sollte$_j$] [IP [SpecIP [t_i]] [I' [VP [SpecVP [t_i]] [V' [NegP nie] [V' [VP [Spec PRO$_i$] [V' [V' [V]]]] [V t_j]]]] [I t_j]]]]]

Input 4: Geschirr

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Man]]]]]] [C' [C sollte$_j$] [IP [SpecIP [t_i]] [I' [VP [SpecVP [t_i]] [V' [NegP nie] [V' [VP [Spec PRO$_i$] [V' [V' [DP [D' [D Ø] [NP [N' [N Geschirr]]]]] [V' [V]]]]] [V t_j]]]] [I t_j]]]]]

Input 5: mit

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Man]]]]]] [C' [C sollte$_j$] [IP [SpecIP [t_i]] [I' [VP [SpecVP [t_i]] [V' [NegP nie] [V' [VP [Spec PRO$_i$] [V' [V' [DP [D' [D ;] [NP [N' [N Geschirr]]]]] [V' [PP [P' [P mit] [DP]]] [V]]]]] [V t_j]]]] [I t_j]]]]]

Input 6: einem

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Man]]]]]] [C' [C sollte$_j$] [IP [SpecIP [t_i]] [I' [VP [SpecVP [t_i]] [V' [NegP nie] [V' [VP [Spec PRO$_i$] [V' [V' [DP [D' [D ;] [NP [N' [N Geschirr]]]]] [V' [PP [P' [P mit] [DP [D' [D einem] [NP [N' [N]]]]]]] [V]]]]] [V t_j]]]] [I t_j]]]]]

Input 7: dreckigen

Parse 7: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Man]]]]]] [C' [C sollte$_j$] [IP [SpecIP [t_i]] [I' [VP [SpecVP [t_i]] [V' [NegP nie] [V' [VP [Spec PRO$_i$] [V' [V' [DP [D' [D ;] [NP [N' [N Geschirr]]]]] [V' [PP [P' [P mit] [DP [D' [D einem] [NP [N' [AP [A' [A dreckigen]]] [N]]]]]]] [V]]]]] [V t_j]]]] [I t_j]]]]]

Input 8: Lappen

Parse 8: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Man]]]]]] [C' [C sollte$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [NegP nie] [V' [VP [Spec PRO$_i$] [V' [V' [DP [D' [D ;] [NP [N' [N Geschirr]]]]] [V' [PP [P' [P mit] [DP [D' [D einem] [NP [N' [AP [A' [A dreckigen]]] [N Lappen]]]]]]] [V]]]]] [V t$_j$]]]] [I t$_j$]]]]]

Input 9: spülen

Parse 9: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Man]]]]]] [C' [C sollte$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [NegP nie] [V' [VP [Spec PRO$_i$] [V' [V' [DP [D' [D ;] [NP [N' [N Geschirr]]]]] [V' [PP [P' [P mit] [DP [D' [D einem] [NP [N' [AP [A' [A dreckigen]]] [N Lappen]]]]]]] [V spülen]]]]] [V t$_j$]]]] [I t$_j$]]]]]

Input 10: müssen

Parse 10: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Man]]]]]] [C' [C sollte$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [NegP nie] [V' [VP [Spec PRO$_i$] [V' [VP [Spec PRO$_i$] [V' [V' [DP [D' [D ;] [NP [N' [N Geschirr]]]]] [V' [PP [P' [P mit] [DP [D' [D einem] [NP [N' [AP [A' [A dreckigen]]] [N Lappen]]]]]]] [V spülen]]]]] [V müssen]]] [V t$_j$]]]] [I t$_j$]]]]]

13. **Sentence:** Kinder essen Quark am liebsten mit Früchten.

Input 1: Kinder

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Kinder]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 2: essen

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Kinder]]]]]] [C' [C essen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 3: Quark

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Kinder]]]]]] [C' [C essen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D Ø] [NP [N' [N Quark$_k$]]]]] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 4: am

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Kinder]]]]]] [C' [C essen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D Ø] [NP [N' [N Quark$_k$]]]]] [V' [PP [P' [P [P an] [X$_l$ +(de)m]] [DP [D' [D t$_l$] [NP [N' [N]]]]]]] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 5: liebsten

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Kinder]]]]] [C' [C essen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D Ø] [NP [N' [N Quark$_k$]]]]] [V' [PP [P' [P [P an] [X$_l$ +(de)m]] [DP [D' [D t$_l$] [NP [AP [A' [A liebsten]]] [N' [N]]]]]] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 6: mit

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Kinder]]]]] [C' [C essen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D Ø] [NP [N' [N Quark$_k$]]]]] [V' [PP [P' [P [P an] [X$_l$ +(de)m]] [DP [D' [D t$_l$] [NP [AP [A' [A liebsten]]] [N' [N Ø]]]]]] [V' [PP [P' [P mit] [DP]]] [V t$_j$]]]]] [I [V t$_j$]]]]]]

Input 7: Früchten

Parse 7: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Kinder]]]]] [C' [C essen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D Ø] [NP [N' [N Quark$_k$]]]]] [V' [PP [P' [P [P an] [X$_l$ +(de)m]] [DP [D' [D t$_l$] [NP [AP [A' [A liebsten]]] [N' [N Ø]]]]]] [V' [PP [P' [P mit] [DP [D' [D Ø] [NP [N' [N Früchten]]]]]] [V t$_j$]]]]] [I [V t$_j$]]]]]]

14. **Sentence:** Die Frauen in den Andendörfern weben Stoff noch auf traditionellen Webstühlen.

Input 1: Die

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 2: Frauen

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Frauen]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 3: in

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Frauen] [PP [P' [P in] [DP]]]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 4: den

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Frauen] [PP [P' [P in] [DP [D' [D den] [NP [N' [N]]]]]]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 5: Andendörfern

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Frauen] [PP [P' [P in] [DP [D' [D den] [NP [N' [N Andendörfern]]]]]]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 6: weben

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Frauen] [PP [P' [P in] [DP [D' [D den] [NP [N' [N Andendörfern]]]]]]]]]]] [C' [C weben$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 7: Stoff

Parse 7: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Frauen] [PP [P' [P in] [DP [D' [D den] [NP [N' [N Andendörfern]]]]]]]]]]] [C' [C weben$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [DØ] [NP [N' [N Stoff]]]]] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 8: noch

Parse 8: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Frauen] [PP [P' [P in] [DP [D' [D den] [NP [N' [N Andendörfern]]]]]]]]]]] [C' [C weben$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [DØ] [NP [N' [N Stoff]]]]] [V' [AdvP [Adv' [Adv noch]]] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 9: auf

Parse 9: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Frauen] [PP [P' [P in] [DP [D' [D den] [NP [N' [N Andendörfern]]]]]]]]]]] [C' [C weben$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [DØ] [NP [N' [N Stoff]]]]] [V' [AdvP [Adv' [Adv noch]]] [V' [PP [P' [P auf] [DP]]] [V t$_j$]]]]] [I [V t$_j$]]]]]]

Input 10: traditionellen

Parse 10: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Frauen] [PP [P' [P in] [DP [D' [D den] [NP [N' [N Andendörfern]]]]]]]]]]] [C' [C weben$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [DØ] [NP [N' [N Stoff]]]]] [V' [AdvP [Adv' [Adv noch]]] [V' [PP [P' [P auf] [DP [D' [DØ] [NP [N' [AP [A' [A traditionellen]]] [N]]]]]] [V t$_j$]]]]] [I [V t$_j$]]]]]]

Input 11: Webstühlen

[Parse11:] [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Frauen] [PP [P' [P in] [DP [D' [D den] [NP [N' [N Andendörfern]]]]]]]]]]] [C' [C weben$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [DØ] [NP [N' [N Stoff]]]]] [V' [AdvP [Adv' [Adv noch]]] [V' [PP [P' [P auf] [DP [D' [DØ] [NP [N' [AP [A' [A traditionellen]]] [N Webstühlen]]]]]] [V t$_j$]]]]] [I [V t$_j$]]]]]]

15. **Sentence:** In den Fässern gären Beize und Lauge.

    Input 1: In

    Parse 1: [CP [SpecCP [PP$_i$ [P' [P In] [DP]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [DP]] [V' [t$_i$] [V t$_j$]]] [I [V t$_j$]]]]]]

    Input 2: den

    Parse 2: [CP [SpecCP [PP$_i$ [P' [P In] [DP [D' [D den] [NP [N' [N]]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [DP]] [V' [t$_i$] [V t$_j$]]] [I [V t$_j$]]]]]]

    Input 3: Fässern

    Parse 3: [CP [SpecCP [PP$_i$ [P' [P In] [DP [D' [D den] [NP [N' [N Fässern]]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [DP]] [V' [t$_i$] [V t$_j$]]] [I [V t$_j$]]]]]]

    Input 4: gären

    Parse 4: [CP [SpecCP [PP$_i$ [P' [P In] [DP [D' [D den] [NP [N' [N Fässern]]]]]]] [C' [C gären$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [DP]] [V' [t$_i$] [V t$_j$]]] [I [V t$_j$]]]]]]

    Input 5: Beize

    Parse 5: [CP [SpecCP [PP$_i$ [P' [P In] [DP [D' [D den] [NP [N' [N Fässern]]]]]]] [C' [C gären$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [DP [D' [D Ø] [NP [N' [N Beize] [N' [Coord'] [NP]]]]]]] [V' [t$_i$] [V t$_j$]]] [I [V t$_j$]]]]]]

    Input 6: und

    Parse 6: [CP [SpecCP [PP$_i$ [P' [P In] [DP [D' [D den] [NP [N' [N Fässern]]]]]]] [C' [C gären$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [DP [D' [D Ø] [NP [N' [N Beize] [N' [Coord' und] [NP]]]]]]] [V' [t$_i$] [V t$_j$]]] [I [V t$_j$]]]]]]

    Input 7: Lauge

    Parse 7: [CP [SpecCP [PP$_i$ [P' [P In] [DP [D' [D den] [NP [N' [N Fässern]]]]]]] [C' [C gären$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [DP [D' [D Ø] [NP [N' [N Beize] [N' [Coord' und] [NP [N' [N Lauge]]]]]]]]] [V' [t$_i$] [V t$_j$]]] [I [V t$_j$]]]]]]

16. **Sentence:** Die Förster küren Ahorn zum Baum des Jahres.

    Input 1: Die

    Parse 1: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 2: Förster

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Förster]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 3: küren

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Förster]]]]]] [C' [C küren$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 4: Ahorn

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Förster]]]]]] [C' [C küren$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D Ø] [NP [N' [N Ahorn]]]]] [V' [PP] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 5: zum

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Förster]]]]]] [C' [C küren$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D Ø] [NP [N' [N Ahorn]]]]] [V' [PP [P' [P zum] [DP]]] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 6: Baum

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Förster]]]]]] [C' [C küren$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D Ø] [NP [N' [N Ahorn]]]]] [V' [PP [P' [P zum] [DP [D’ [D Ø] [NP [N’ [N Baum] [DP]]]]]]] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 7: des

Parse 7: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Förster]]]]]] [C' [C küren$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D Ø] [NP [N' [N Ahorn]]]]] [V' [PP [P' [P zum] [DP [D’ [D Ø] [NP [N’ [N Baum] [DP [D’ [D des] [NP[N’ [N]]]]]]]]]]] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 8: Jahres

Parse 8: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Förster]]]]]] [C' [C küren$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D Ø] [NP [N' [N Ahorn]]]]] [V' [PP [P' [P zum] [DP [D’ [D Ø] [NP [N’ [N Baum] [DP [D’ [D des] [NP[N’ [N Jahres]]]]]]]]]]] [V t$_j$]]]] [I [V t$_j$]]]]]]

17. **Sentence:** Das Pferd ist seinem Reiter auf den Fuß getreten.

Input 1: Das

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Das] [NP [N' [N]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 2: Pferd

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Das] [NP [N' [N Pferd]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 3: ist

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Das] [NP [N' [N Pferd]]]]]] [C' [C ist$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [AP [A' [A]]] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 4: seinem

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Das] [NP [N' [N Pferd]]]]]] [C' [C ist$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP$_k$ [D' [D seinem] [NP [N' [N]]]]] [VP [SpecVP [t$_i$]] [V' [t$_k$] [V' [AP [A' [A]]] [V t$_j$]]]]] [I [V t$_j$]]]]]]

Input 5: Reiter

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Das] [NP [N' [N Pferd]]]]]] [C' [C ist$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP$_k$ [D' [D seinem] [NP [N' [N Reiter]]]]] [VP [SpecVP [t$_i$]] [V' [t$_k$] [V' [AP [A' [A]]] [V t$_j$]]]]] [I [V t$_j$]]]]]]

Input 6: auf

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Das] [NP [N' [N Pferd]]]]]] [C' [C ist$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP$_k$ [D' [D seinem] [NP [N' [N Reiter]]]]] [VP [SpecVP [t$_i$]] [V' [t$_k$] [V' [PP [P' [P auf] [DP]] [V' [V] [V t$_j$]]]]]] [I [V t$_j$]]]]]]]

Input 7: den

Parse 7: [CP [SpecCP [DP$_i$ [D' [D Das] [NP [N' [N Pferd]]]]]] [C' [C ist$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP$_k$ [D' [D seinem] [NP [N' [N Reiter]]]]] [VP [SpecVP [t$_i$]] [V' [t$_k$] [V' [PP [P' [P auf] [DP [D' [D den] [NP [N' [N]]]]]] [V' [V] [V t$_j$]]]]]] [I [V t$_j$]]]]]]]

Input 8: Fuß

Parse 8: [CP [SpecCP [DP$_i$ [D' [D Das] [NP [N' [N Pferd]]]]]] [C' [C ist$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP$_k$ [D' [D seinem] [NP [N' [N Reiter]]]]] [VP [SpecVP [t$_i$]] [V' [t$_k$] [V' [PP [P' [P auf] [DP [D' [D den] [NP [N' [N Fuß]]]]]] [V' [V] [V t$_j$]]]]]] [I [V t$_j$]]]]]]

Input 9: getreten

Parse 9: [CP [SpecCP [DP$_i$ [D' [D Das] [NP [N' [N Pferd]]]]]] [C' [C ist$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP$_k$ [D' [D seinem] [NP [N' [N Reiter]]]]] [VP [SpecVP [t$_i$]] [V' [t$_k$] [V' [PP [P' [P auf] [DP [D' [D den] [NP [N' [N Fuß]]]]]]] [V' [V getreten] [V t$_j$]]]]]] [I [V t$_j$]]]]]]

18. **Sentence:** Sarah hat ihrem Opa ein Bild gemalt.

    Input 1: Sarah

    Parse 1: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Sarah]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

    Input 2: hat

    Parse 2: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Sarah]]]]]] [C' [C hat$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [V] [V t$_j$]]] [I [V t$_j$]]]]]]

    Input 3: ihrem

    Parse 3: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Sarah]]]]]] [C' [C hat$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP$_k$ [D' [D ihrem] [NP [N' [N]]]]] [VP [SpecVP [t$_i$]] [V' [t$_k$] [V' [V] [V t$_j$]]]]] [I [V t$_j$]]]]]]

    Input 4: Opa

    Parse 4: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Sarah]]]]]] [C' [C hat$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP$_k$ [D' [D ihrem] [NP [N' [N Opa]]]]] [VP [SpecVP [t$_i$]] [V' [t$_k$] [V' [V] [V t$_j$]]]]] [I [V t$_j$]]]]]]

    Input 5: ein

    Parse 5: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Sarah]]]]]] [C' [C hat$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP$_k$ [D' [D ihrem] [NP [N' [N Opa]]]]] [VP [SpecVP [t$_i$]] [V' [t$_k$] [V' [DP [D' [D ein] [NP [N' [N]]]]] [V' [V] [V t$_j$]]]]]] [I [V t$_j$]]]]]]

    Input 6: Bild

    Parse 6: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Sarah]]]]]] [C' [C hat$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP$_k$ [D' [D ihrem] [NP [N' [N Opa]]]]] [VP [SpecVP [t$_i$]] [V' [t$_k$] [V' [DP [D' [D ein] [NP [N' [N Bild]]]]] [V' [V] [V t$_j$]]]]]] [I [V t$_j$]]]]]]

    Input 7: gemalt

    Parse 7: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Sarah]]]]]] [C' [C hat$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP$_k$ [D' [D ihrem] [NP [N' [N Opa]]]]] [VP [SpecVP [t$_i$]] [V' [t$_k$] [V' [DP [D' [D ein] [NP [N' [N Bild]]]]] [V' [V gemalt] [V t$_j$]]]]]] [I [V t$_j$]]]]]]

19. **Sentence:** Der Franzose gewann gegen den Belgier.

    Input 1: Der

    Parse 1: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [N' [N]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 2: Franzose

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [N' [N Franzose]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 3: gewann

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [N' [N Franzose]]]]]] [C' [C gewann$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 4: gegen

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [N' [N Franzose]]]]]] [C' [C gewann$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [PP [P' [P gegen] [DP]]] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 5: den

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [N' [N Franzose]]]]]] [C' [C gewann$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [PP [P' [P gegen] [DP [D' [D den] [NP [N' [N]]]]]]

Input 6: Belgier

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [N' [N Franzose]]]]]] [C' [C gewann$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [PP [P' [P gegen] [DP [D' [D den] [NP [N' [N Belgier]]]]]]] [V t$_j$]]] [I [V t$_j$]]]]]]

20. **Sentence:** Der alte Kapitän goß stets ein wenig Rum in seinen Tee.

Input 1: Der

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [N' [N]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 2: alte

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [N' [AP [A' [A alte]]] [N]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]]

Input 3: Kapitän

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [N' [AP [A' [A alte]]] [N Kapitän]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 4: goß

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Der] [NP [N' [AP [A' [A alte]]] [N Kapitän]]]]]] [C' [C goß$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 5: stets

Parse 5: [CP [SpecCP [DP_i [D' [D Der] [NP [N' [AP [A' [A alte]]] [N Kapitän]]]]]] [C' [C goß$_j$] [IP [SpecIP [t_i]] [I' [VP [AdvP [Adv' [Adv stets]]] [VP [SpecVP [t_i]] [V' [DP] [V t_j]]]] [I [V t_j]]]]]]

Input 6: ein

Parse 6: [CP [SpecCP [DP_i [D' [D Der] [NP [N' [AP [A' [A alte]]] [N Kapitän]]]]]] [C' [C goß$_j$] [IP [SpecIP [t_i]] [I' [VP [AdvP [Adv' [Adv stets]]] [VP [SpecVP [t_i]] [V' [DP [D' [D ein] [NP [N' [N]]]]] [V' [DP] [V t_j]]]]] [I [V t_j]]]]]]

Input 7: wenig

Parse 7: [CP [SpecCP [DP_i [D' [D Der] [NP [N' [AP [A' [A alte]]] [N Kapitän]]]]]] [C' [C goß$_j$] [IP [SpecIP [t_i]] [I' [VP [AdvP [Adv' [Adv stets]]] [VP [SpecVP [t_i]] [V' [DP [D' [D ein] [NP [N' [AP [A' [A wenig]]] [N]]]]] [V' [PP] [V t_j]]]]] [I [V t_j]]]]]]

Input 8: Rum

Parse 8: [CP [SpecCP [DP_i [D' [D Der] [NP [N' [AP [A' [A alte]]] [N Kapitän]]]]]] [C' [C goß$_j$] [IP [SpecIP [t_i]] [I' [VP [AdvP [Adv' [Adv stets]]] [VP [SpecVP [t_i]] [V' [DP [D' [D ein] [NP [N' [AP [A' [A wenig]]] [N Rum]]]]] [V' [PP] [V t_j]]]]] [I [V t_j]]]]]]

Input 9: in

Parse 9: [CP [SpecCP [DP_i [D' [D Der] [NP [N' [AP [A' [A alte]]] [N Kapitän]]]]]] [C' [C goß$_j$] [IP [SpecIP [t_i]] [I' [VP [AdvP [Adv' [Adv stets]]] [VP [SpecVP [t_i]] [V' [DP [D' [D ein] [NP [N' [AP [A' [A wenig]]] [N Rum]]]]] [V' [PP [P' [P in] [DP]]] [V t_j]]]]] [I [V t_j]]]]]]

Input 10: seinen

Parse 10: [CP [SpecCP [DP_i [D' [D Der] [NP [N' [AP [A' [A alte]]] [N Kapitän]]]]]] [C' [C goß$_j$] [IP [SpecIP [t_i]] [I' [VP [AdvP [Adv' [Adv stets]]] [VP [SpecVP [t_i]] [V' [DP [D' [D ein] [NP [N' [AP [A' [A wenig]]] [N Rum]]]]] [V' [PP [P' [P in] [DP [D' [D seinen] [NP [N' [N]]]]]]] [V t_j]]]]] [I [V t_j]]]]]]

Input 11: Tee

[Parse11:] [CP [SpecCP [DP_i [D' [D Der] [NP [N' [AP [A' [A alte]]] [N Kapitain]]]]]] [C' [C goß$_j$] [IP [SpecIP [t_i]] [I' [VP [AdvP [Adv' [Adv stets]]] [VP [SpecVP [t_i]] [V' [DP [D' [D ein] [NP [N' [AP [A' [A wenig]]] [N Rum]]]]] [V' [PP [P' [P in] [DP [D' [D seinen] [NP [N' [N Tee]]]]]]] [V t_j]]]]] [I [V t_j]]]]]]

21. **Sentence:** Mäuse und Ratten nagen gerne an Stromkabeln.

Input 1: Mäuse

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Mäuse]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 2: und

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Mäuse] [COORD und] [NP [N' [N]]]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 3: Ratten

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Mäuse] [COORD und] [NP [N' [N Ratten]]]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 4: nagen

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Mäuse] [COORD und] [NP [N' [N Ratten]]]]]]]] [C' [C nagen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [V t$_j$]]] [I [V t$_j$]]]]]]

Input 5: gerne

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Mäuse] [COORD und] [NP [N' [N Ratten]]]]]]]] [C' [C nagen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [AdvP [Adv' [Adv gerne]]] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 6: an

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Mäuse] [COORD und] [NP [N' [N Ratten]]]]]]]] [C' [C nagen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [AdvP [Adv' [Adv gerne]]] [V' [PP [P' [P an] [DP]]] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 7: Stromkabeln

Parse 7: [CP [Spec [DP$_i$ [D' [D Ø] [NP [NP [N' [N Mäuse]]] [COORD und] [NP Ratten]]]]] [C' [C nagen$_j$] [IP [Spec t$_i$] [I' [VP [Spec t$_i$] [V' [AdvP [Adv' [Adv gerne]]] [V' [PP [P' [P an] [DP [D' [D Ø] [NP [N' [N Stromkabeln]]]]]]] [V t$_j$]]]] [I t$_j$]]]]]]

22. **Sentence:** Tierärzte impfen keine Kaninchen gegen Tollwut.

Input 1: Tierärzte

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Tierärze]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [V t$_j$]]] [I [V t$_j$]]]]]]

Input 2: impfen

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Tierärze]]]]] [C' [C impfen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 3: keine

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Tierärze]]]]] [C' [C impfen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D keine] [NP [N' [N]]]]] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 4: Kaninchen

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Tierärze]]]]] [C' [C impfen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D keine] [NP [N' [N Kaninchen]]]]] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 5: gegen

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Tierärze]]]]] [C' [C impfen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D keine] [NP [N' [N Kaninchen]]]]] [V' [PP [P' [P gegen] [DP]]] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 6: Tollwut

Parse 6: [CP [Spec [DP$_i$ [D' [D Ø] [NP [N' [N Tierärzte]]]]]] [C' [C impfen$_j$] [IP [Spec t$_i$] [I' [VP [Spec t$_i$] [V' [V' [DP [D' [D keine] [NP [N' [N Kaninchen]]]]] [V' [PP [P' [P gegen] [DP [D' [D Ø] [NP [N' [N Tollwut]]]]]]] [V t$_j$]]]]] [I t$_j$]]]]]

23. **Sentence:** Die Hunde der Wächter bellen beim geringsten Anlaß.

Input 1: Die

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 2: Hunde

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Hunde]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 3: der

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Hunde] [DP [D' [D der] [NP [N' [N]]]]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]]

Input 4: Wächter

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Hunde] [DP [D' [D der] [NP [N' [N Wächter]]]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 5: bellen

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Hunde] [DP [D' [D der] [NP [N' [N Wächter]]]]]]]] [C' [C bellen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 6: beim

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Hunde] [DP [D' [D der] [NP [N' [N Wächter]]]]]]]] [C' [C bellen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [PP [P' [P [P bei] [X$_k$ +(de)m]] [DP [D' [D t$_k$] [NP [N' [N]]]]]] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 7: geringsten

Parse 7: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Hunde] [DP [D' [D der] [NP [N' [N Wächter]]]]]]]] [C' [C bellen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [PP [P' [P [P bei] [X$_k$ +(de)m]] [DP [D' [D t$_k$] [NP [N' [AdjP [Adj' [Adj geringsten]]] [N]]]]] [V t$_j$]]] [I [V t$_j$]]]]]]]

Input 8: Anlaß

Parse 8: [CP [Spec [DP$_i$ [D' [D die] [NP [N' [N Hunde] [DP [D' [D der] [NP [N' [N Wächter]]]]]]]]] [C' [C bellen$_j$] [IP [Spec t$_i$] [I' [VP [Spec t$_i$] [V' [PP [P' [P [P bei] [X$_k$ +(de)m]] [DP [D' [D t$_k$] [NP [N' [AdjP [Adj' [Adj geringsten]]] [N Anlass]]]]]] [V t$_j$]]] [I t$_j$]]]]]

24. **Sentence:** Affen kraulen sich oft stundenlang das Fell.

Input 1: Affen

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Affen]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 2: kraulen

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Affen]]]]]] [C' [C kraulen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP] [VP [SpecVP [t$_i$]] [V' [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 3: sich

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Affen]]]]]] [C' [C kraulen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP [D' [D Ø] [NP [N' [N sich]]]]] [VP [SpecVP [t$_i$]] [V' [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 4: oft

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Affen]]]]] [C' [C kraulen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP [D' [D Ø] [NP [N' [N sich]]]]] [VP [AdvP [Adv' [Adv oft]]] [VP [SpecVP [t$_i$]] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 5: stundenlang

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Affen]]]]] [C' [C kraulen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP [D' [D Ø] [NP [N' [N sich]]]]] [VP [AdvP [Adv' [Adv oft]]] [VP [SpecVP [t$_i$]] [V' [AdvP [Adv' [Adv stundenlang]]] [V t$_j$]]]]] [I [V t$_j$]]]]]]

Input 6: das

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Affen]]]]] [C' [C kraulen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP [D' [D Ø] [NP [N' [N sich]]]]] [VP [AdvP [Adv' [Adv oft]]] [VP [SpecVP [t$_i$]] [V' [AdvP [Adv' [Adv stundenlang]]] [V' [DP [D' [D das] [NP [N' [N]]]]] [V t$_j$]]]]]] [I [V t$_j$]]]]]]

Input 7: Fell

Parse 7: [CP [SpecCP [DP$_i$ [D' [D Ø] [NP [N' [N Affen]]]]] [C' [C kraulen$_j$] [IP [SpecIP [t$_i$]] [I' [VP [DP [D' [D Ø] [NP [N' [N sich]]]]] [VP [AdvP [Adv' [Adv oft]]] [VP [SpecVP [t$_i$]] [V' [AdvP [Adv' [Adv stundenlang]]] [V' [DP [D' [D das] [NP [N' [N Fell]]]]] [V t$_j$]]]]]] [I [V t$_j$]]]]]]

25. **Sentence:** Die Bäume in den Wäldern speichern sehr viel Wasser.

Input 1: Die

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 2: Bäume

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Bäume]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 3: in

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Bäume]] [PP [P' [P in] [DP]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 4: den

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Bäume]] [PP [P' [P in] [DP [D' [D den] [NP [N' [N]]]]]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 5: Wäldern

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Bäume]] [PP [P' [P in] [DP [D' [D den] [NP [N' [N Wäldern]]]]]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 6: speichern

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Bäume]] [PP [P' [P in] [DP [D' [D den] [NP [N' [N Wäldern]]]]]]]]]] [C' [C speichern$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 7: sehr

Parse 7: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Bäume]] [PP [P' [P in] [DP [D' [D den] [NP [N' [N Wäldern]]]]]]]]]]] [C' [C speichern$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D Ø] [NP [N' [AdvP [Adv' [Adv sehr] [AP [A' [A]]]]] [N]]]]] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 8: viel

Parse 8: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Bäume]] [PP [P' [P in] [DP [D' [D den] [NP [N' [N Wäldern]]]]]]]]]]] [C' [C speichern$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D Ø] [NP [N' [AdvP [Adv' [Adv sehr] [AP [A' [A viel]]]]] [N]]]]] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 9: Wasser

Parse 9: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N Bäume] [PP [P' [P in] [DP [D' [D den] [NP [N' [N Wäldern]]]]]]]]]]] [C' [C speichern$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D Ø] [NP [N' [AdvP [Adv' [Adv sehr] [AP [A' [A viel]]]]] [N Wasser]]]]] [V t$_j$]]] [I [V t$_j$]]]]]]

26. **Sentence:** Die meisten Leute schummeln beim Spielen gelegentlich.

Input 1: Die

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 2: meisten

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [AP [A' [A meisten]]] [N]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 3: Leute

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [AP [A' [A meisten]]] [N Leute]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 4: schummeln

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [AP [A' [A meisten]]] [N Leute]]]]]] [C' [C schummeln$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 5: beim

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [AP [A' [A meisten]]] [N Leute]]]]]] [C' [C schummeln$_j$] [IP [SpecIP [t$_i$]] [I' [VP [PP$_l$ [P' [P [P bei] [X$_k$ + (de)m]] [DP [D' [D t$_k$] [NP [N' [N]]]]]]] [VP [SpecVP [t$_i$]] [V' [t$_l$] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 6: Spielen

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [AP [A' [A meisten]]] [N Leute]]]]]] [C' [C schummeln$_j$] [IP [SpecIP [t$_i$]] [I' [VP [PP$_l$ [P' [P [P bei] [X$_k$ + (de)m]] [DP [D' [D t$_k$] [NP [N' [N Spielen]]]]]]] [VP [SpecVP [t$_i$]] [V' [t$_l$] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 7: gelegentlich

Parse 7: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [AP [A' [A meisten]]] [N Leute]]]]]] [C' [C schummeln$_j$] [IP [SpecIP [t$_i$]] [I' [VP [PP$_l$ [P' [P [P bei] [X$_k$ + (de)m]] [DP [D' [D t$_k$] [NP [N' [N Spielen]]]]]]] [VP [AdvP [Adv' [Adv gelegentlich]]] [VP [SpecVP [t$_i$]] [V' [t$_l$] [V t$_j$]]]]] [I [V t$_j$]]]]]]

27. **Sentence:** Nach dem Spiel massieren die Therapeuten den Spielern die Beine.

Input 1: Nach

Parse 1: [CP [SpecCP [PP$_k$ [P' [P Nach] [DP]]]] [C' [C V$_j$] [IP [SpecIP [DP$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [t$_k$] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 2: dem

Parse 2: [CP [SpecCP [PP$_k$ [P' [P Nach] [DP [D' [D dem] [NP [N' [N]]]]]]]] [C' [C V$_j$] [IP [SpecIP [DP$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [t$_k$] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 3: Spiel

Parse 3: [CP [SpecCP [PP$_k$ [P' [P Nach] [DP [D' [D dem] [NP [N' [N Spiel]]]]]]]] [C' [C V$_j$] [IP [SpecIP [DP$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [t$_k$] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 4: massieren

Parse 4: [CP [SpecCP [PP$_k$ [P' [P Nach] [DP [D' [D dem] [NP [N' [N Spiel]]]]]]]] [C' [C massieren$_j$] [IP [SpecIP [DP$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [DP] [V' [t$_k$] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 5: die

Parse 5: [CP [SpecCP [PP$_k$ [P' [P Nach] [DP [D' [D dem] [NP [N' [N Spiel]]]]]]]] [C' [C massieren$_j$] [IP [SpecIP [DP$_i$ [D' [D die] [NP [N' [N]]]]]] [I' [VP [SpecVP [t$_i$]] [V' [DP] [V' [t$_k$] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 6: Therapeuten

Parse 6: [CP [SpecCP [PP$_k$ [P' [P Nach] [DP [D' [D dem] [NP [N' [N Spiel]]]]]]]] [C' [C massieren$_j$] [IP [SpecIP [DP$_i$ [D' [D die] [NP [N' [N Therapeuten]]]]]] [I' [VP [SpecVP [t$_i$]] [V' [DP] [V' [t$_k$] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 7: den

Parse 7: [CP [SpecCP [PP$_k$ [P' [P Nach] [DP [D' [D dem] [NP [N' [N Spiel]]]]]]]] [C' [C massieren$_j$] [IP [SpecIP [DP$_i$ [D' [D die] [NP [N' [N Therapeuten]]]]]] [I' [VP [SpecVP [t$_i$]] [V' [DP [D' [D den] [NP]]] [V' [t$_k$] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 8: Spielern

Parse 8: [CP [SpecCP [PP$_k$ [P' [P Nach] [DP [D' [D dem] [NP [N' [N Spiel]]]]]]]] [C' [C massieren$_j$] [IP [SpecIP [DP$_i$ [D' [D die] [NP [N' [N Therapeuten]]]]]] [I' [VP [DP [D' [D den] [NP [N' [N Spielern]]]]] [VP [SpecVP [t$_i$]] [V' [DP] [V' [t$_k$] [V t$_j$]]]]] [I [V t$_j$]]]]]]

Input 9: die

Parse 9: [CP [SpecCP [PP$_k$ [P' [P Nach] [DP [D' [D dem] [NP [N' [N Spiel]]]]]]]] [C' [C massieren$_j$] [IP [SpecIP [DP$_i$ [D' [D die] [NP [N' [N Therapeuten]]]]]] [I' [VP [DP [D' [D den] [NP [N' [N Spielern]]]]] [VP [SpecVP [t$_i$]] [V' [DP [D' [D die] [NP [N' [N]]]]] [V' [t$_k$] [V t$_j$]]]]] [I [V t$_j$]]]]]]

Input 10: Beine

Parse 10: [CP [SpecCP [PP$_i$ [P' [P Nach] [DP [D' [D dem] [NP [N' [N Spiel]]]]]]]] [C' [C massieren$_j$] [IP [SpecIP [DP$_k$ [D' [D die] [NP [N' [N Therapeuten]]]]]] [I' [VP [DP [D' [D den] [NP [N' [N Spielern]]]]] [VP [SpecVP [t$_k$]] [V' [t$_i$] [V' [DP [D' [D die] [NP [N' [N Beine]]]]] [V t$_j$]]]]] [I [V t$_j$]]]]]]

28. **Sentence:** Die beiden Mädchen tuscheln während des Unterrichts.

Input 1: Die

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 2: beiden

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [AP [A' [A beiden]]] [N]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 3: Mädchen

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [AP [A' [A beiden]]] [N Mädchen]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 4: tuscheln

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [AP [A' [A beiden]]] [N Mädchen]]]]]] [C' [C tuscheln$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 5: während

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [AP [A' [A beiden]]] [N Mädchen]]]]]] [C' [C tuscheln$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [PP [P' [P während] [DP]]] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 6: des

Parse 6: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [AP [A' [A beiden]]] [N Mädchen]]]]]] [C' [C tuscheln$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [PP [P' [P während] [DP [D' [D des] [NP [N' [N]]]]]]] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 7: Unterrichts

Parse 7: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [AP [A' [A beiden]]] [N Mädchen]]]]]] [C' [C tuscheln$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [PP [P' [P während] [DP [D' [D des] [NP [N' [N Unterrichts]]]]]]] [V t$_j$]]] [I [V t$_j$]]]]]]

29. **Sentence:** Die Häuser am Horizont flimmern in der Sonne.

Input 1: Die

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Die] [NP [N' [N]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 2: Häuser

Parse 2: [CP [SpecCP [DP_i [D' [D Die] [NP [N' [N Häuser]]]]]] [C' [C V_j] [IP [SpecIP [t_i]] [I' [VP [SpecVP [t_i]] [V t_j]] [I [V t_j]]]]]]

Input 3: am

Parse 3: [CP [SpecCP [DP_i [D' [D Die] [NP [N' [N Häuser] [PP [P' [P' [P an] [X_k + (de)m]] [DP [D' [D t_k] [NP [N' [N]]]]]]]]]]]] [C' [C V_j] [IP [SpecIP [t_i]] [I' [VP [SpecVP [t_i]] [V t_j]] [I [V t_j]]]]]]

Input 4: Horizont

Parse 4: [CP [SpecCP [DP_i [D' [D Die] [NP [N' [N Häuser] [PP [P' [P' [P an] [X_k + (de)m]] [DP [D' [D t_k] [NP [N' [N Horizont]]]]]]]]]]]] [C' [C V_j] [IP [SpecIP [t_i]] [I' [VP [SpecVP [t_i]] [V t_j]] [I [V t_j]]]]]]

Input 5: flimmern

Parse 5: [CP [SpecCP [DP_i [D' [D Die] [NP [N' [N Häuser] [PP [P' [P' [P an] [X_k + (de)m]] [DP [D' [D t_k] [NP [N' [N Horizont]]]]]]]]]]]] [C' [C $flimmern_j$] [IP [SpecIP [t_i]] [I' [VP [SpecVP [t_i]] [V t_j]] [I [V t_j]]]]]]

Input 6: in

Parse 6: [CP [SpecCP [DP_i [D' [D Die] [NP [N' [N Häuser] [PP [P' [P' [P an] [X_k + (de)m]] [DP [D' [D t_k] [NP [N' [N Horizont]]]]]]]]]]]] [C' [C $flimmern_j$] [IP [SpecIP [t_i]] [I' [VP [SpecVP [t_i]] [V' [PP [P' [P in] [DP]]] [V t_j]]] [I [V t_j]]]]]]

Input 7: der

Parse 7: [CP [SpecCP [DP_i [D' [D Die] [NP [N' [N Häuser] [PP [P' [P' [P an] [X_k + (de)m]] [DP [D' [D t_k] [NP [N' [N Horizont]]]]]]]]]]]] [C' [C $flimmern_j$] [IP [SpecIP [t_i]] [I' [VP [SpecVP [t_i]] [V' [PP [P' [P in] [DP [D' [D der] [NP [N' [N]]]]]]] [V t_j]]] [I [V t_j]]]]]]

Input 8: Sonne

Parse 8: [CP [SpecCP [DP_i [D' [D Die] [NP [N' [N Häuser] [PP [P' [P' [P an] [X_k + (de)m]] [DP [D' [D t_k] [NP [N' [N Horizont]]]]]]]]]]]] [C' [C $flimmern_j$] [IP [SpecIP [t_i]] [I' [VP [SpecVP [t_i]] [V' [PP [P' [P in] [DP [D' [D der] [NP [N' [N Sonne]]]]]]] [V t_j]]] [I [V t_j]]]]]]

30. **Sentence:** Den ganzen Tag über konnte man die Raben krächzen hören.

Input 1: Den

Parse 1: [CP [SpecCP [DP_k [D' [D Den] [NP [N' [N]]]]]] [C' [C V_j] [IP [SpecIP [DP_i]] [I' [VP [SpecVP [t_i]] [V' [t_k] [V t_j]]] [I [V t_j]]]]]]

Input 2: ganzen

Parse 2: [CP [SpecCP [DP$_k$ [D' [D Den] [NP [N' [AP [A' [A ganzen]]] [N]]]]]] [C' [C V$_j$] [IP [SpecIP [DP$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [t$_k$] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 3: Tag

Parse 3: [CP [SpecCP [DP$_k$ [D' [D Den] [NP [N' [AP [A' [A ganzen]]] [N Tag]]]]]] [C' [C V$_j$] [IP [SpecIP [DP$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [t$_k$] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 4: über

Parse 4: [CP [SpecCP [PP$_k$ [P' [DP [D' [D Den] [NP [N' [AP [A' [A ganzen]]] [N Tag]]]]] [P über]]]] [C' [C V$_j$] [IP [SpecIP [DP$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [t$_k$] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 5: konnte

Parse 5: [CP [SpecCP [PP$_k$ [P' [DP [D' [D Den] [NP [N' [AP [A' [A ganzen]]] [N Tag]]]]] [P über]]]] [C' [C konnte$_j$] [IP [SpecIP [DP$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [t$_k$] [V' [V] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 6: man

Parse 6: [CP [SpecCP [PP$_k$ [P' [DP [D' [D Den] [NP [N' [AP [A' [A ganzen]]] [N Tag]]]]] [P über]]]] [C' [C konnte$_j$] [IP [SpecIP [DP$_i$ [D' [D Ø] [NP [N' [N man]]]]]] [I' [VP [SpecVP [t$_i$]] [V' [t$_k$] [V' [V] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 7: die

Parse 7: [CP [SpecCP [PP$_k$ [P' [DP [D' [D Den] [NP [N' [AP [A' [A ganzen]]] [N Tag]]]]] [P über]]]] [C' [C konnte$_j$] [IP [SpecIP [DP$_i$ [D' [D Ø] [NP [N' [N man]]]]]] [I' [VP [SpecVP [t$_i$]] [V' [t$_k$] [V' [DP [D' [D die] [NP [N' [N]]]]] [V' [V] [V t$_j$]]]]] [I [V t$_j$]]]]]]

Input 8: Raben

Parse 8: [CP [SpecCP [PP$_k$ [P' [DP [D' [D Den] [NP [N' [AP [A' [A ganzen]]] [N Tag]]]]] [P über]]]] [C' [C konnte$_j$] [IP [SpecIP [DP$_i$ [D' [D Ø] [NP [N' [N man]]]]]] [I' [VP [SpecVP [t$_i$]] [V' [t$_k$] [V' [DP [D' [D die] [NP [N' [N Raben]]]]] [V' [V] [V t$_j$]]]]] [I [V t$_j$]]]]]]

Input 9: krächzen

Parse 9: [CP [SpecCP [PP$_k$ [P' [DP [D' [D Den] [NP [N' [AP [A' [A ganzen]]] [N Tag]]]]] [P über]]]] [C' [C konnte$_j$] [IP [SpecIP [DP$_i$ [D' [D Ø] [NP [N' [N man]]]]]] [I' [VP [SpecVP [t$_i$]] [V' [t$_k$] [V' [VP [Spec PRO$_i$] [V' [VP [Spec [DP [D' [D die] [NP [N' [N Raben]]]]]] [V krächzen]] [V]]] [V t$_j$]]]] [I t$_j$]]]]]

Input 10: hören

Parse 10: [CP [SpecCP [PP$_k$ [P' [DP [D' [D Den] [NP [N' [AP [A' [A ganzen]]] [N Tag]]]]] [P über]]]] [C' [C konnte$_j$] [IP [SpecIP [DP$_i$ [D' [D Ø] [NP [N' [N man]]]]]] [I' [VP [SpecVP [t$_i$]] [V' t$_k$ [V' [VP [Spec PRO$_i$] [V' [VP [Spec [DP [D' [D die] [NP [N' [N Raben]]]]]] [V krächzen]] [V hören]]] [V t$_j$]]]] [I t$_j$]]]]]

31. **Sentence:** Vor dem Auftritt schminken die Schauspieler sich.

Input 1: Vor

Parse 1: [CP [SpecCP [PP$_k$ [P' [P Vor] [DP]]]] [C' [C V$_j$] [IP [SpecIP [DP$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [t$_k$] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 2: dem

Parse 2: [CP [SpecCP [PP$_k$ [P' [P Vor] [DP [D' [D dem] [NP [N' [N]]]]]]] [C' [C V$_j$] [IP [SpecIP [DP$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [t$_k$] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 3: Auftritt

Parse 3: [CP [SpecCP [PP$_k$ [P' [P Vor] [DP [D' [D dem] [NP [N' [N Auftritt]]]]]]] [C' [C V$_j$] [IP [SpecIP [DP$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [t$_k$] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 4: schminken

Parse 4: [CP [SpecCP [PP$_k$ [P' [P Vor] [DP [D' [D dem] [NP [N' [N Auftritt]]]]]]] [C' [C schminken$_j$] [IP [SpecIP [DP$_i$]] [I' [VP [DP [D' [D]]] [VP [SpecVP [t$_i$]] [V' [t$_k$] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 5: die

Parse 5: [CP [SpecCP [PP$_k$ [P' [P Vor] [DP [D' [D dem] [NP [N' [N Auftritt]]]]]]] [C' [C schminken$_j$] [IP [SpecIP [DP$_i$ [D' [D die] [NP [N' [N]]]]]] [I' [VP [DP [D' [D]]] [VP [SpecVP [t$_i$]] [V' [t$_k$] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 6: Schauspieler

Parse 6: [CP [SpecCP [PP$_k$ [P' [P Vor] [DP [D' [D dem] [NP [N' [N Auftritt]]]]]]] [C' [C schminken$_j$] [IP [SpecIP [DP$_i$ [D' [D die] [NP [N' [N Schauspieler]]]]]] [I' [VP [DP] [VP [SpecVP [t$_i$]] [V' [t$_k$] [V t$_j$]]]] [I [V t$_j$]]]]]]

Input 7: sich

Parse 7: [CP [SpecCP [PP$_k$ [P' [P Vor] [DP [D' [D dem] [NP [N' [N Auftritt]]]]]]] [C' [C schminken$_j$] [IP [SpecIP [DP$_i$ [D' [D die] [NP [N' [N Schauspieler]]]]]] [I' [VP [DP [D' [D Ø] [NP [N' [N sich]]]]] [VP [SpecVP [t$_i$]] [V' [t$_k$] [V t$_j$]]]] [I [V t$_j$]]]]]]

32. **Sentence:** Manche Menschen stottern bei Nervosität.

Input 1: Manche

Parse 1: [CP [SpecCP [DP$_i$ [D' [D Manche] [NP [N' [N]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 2: Menschen

Parse 2: [CP [SpecCP [DP$_i$ [D' [D Manche] [NP [N' [N Menschen]]]]]] [C' [C V$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t $_j$]]]]]]

Input 3: stottern

Parse 3: [CP [SpecCP [DP$_i$ [D' [D Manche] [NP [N' [N Menschen]]]]]] [C' [C stottern$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V t$_j$]] [I [V t$_j$]]]]]]

Input 4: bei

Parse 4: [CP [SpecCP [DP$_i$ [D' [D Manche] [NP [N' [N Menschen]]]]]] [C' [C stottern$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [PP [P' [P bei] [DP]]] [V t$_j$]]] [I [V t$_j$]]]]]]

Input 5: Nervosität

Parse 5: [CP [SpecCP [DP$_i$ [D' [D Manche] [NP [N' [N Menschen]]]]]] [C' [C stottern$_j$] [IP [SpecIP [t$_i$]] [I' [VP [SpecVP [t$_i$]] [V' [PP [P' [P bei] [DP [D' [D Ø] [NP [N' [N Nervosität]]]]]]] [V t$_j$]]] [I [V t$_j$]]]]]]

Appendix B

Stimuli for the experiment in Chapter 3

The appendix lists 24 items in four conditions from the experiment in Chapter 3. For every condition the sentence is followed by a question and the correct—Y (yes), N (No)—answer where applicable.

1. (a) The dedicated firefighter that Henry recommended for the new job cut himself on broken glass on the floor./Did Henry cut himself on broken glass?/N.
   (b) The dedicated firefighter that Linda recommended for the new job cut himself on broken glass on the floor./Did the firefighter recommend Linda for the new job?/N.
   (c) The dedicated firefighter that Linda recommended for the new job cut herself on broken glass on the floor./Did Linda cut herself on broken glass?/N.
   (d) The dedicated firefighter that Henry recommended for the new job cut herself on broken glass on the floor./Did the firefighter recommend Henry for the new job?/N.
2. (a) The head engineer that Peter had visited in the factory convinced himself that the building was safe./Did Peter convince himself about the safety of the building?/N.
   (b) The head engineer that Nancy had visited in the factory convinced himself that the building was safe./Did the engineer visit Nancy in the factory?/N.
   (c) The head engineer that Nancy had visited in the factory convinced herself that the building was safe./Did Nancy convince herself about the safety of the building?/N.

(d) The head engineer that Peter had visited in the factory convinced herself that the building was safe./Did the engineer visit Peter in the factory?/N.

3. (a) The professional electrician that John consulted about the wiring problems blamed himself for the fire./Did John blame himself for the fire?/N.
   (b) The professional electrician that Mary consulted about the wiring problems blamed himself for the fire./Did the electrician blame Mary for the fire?/N.
   (c) The professional electrician that Mary consulted about the wiring problems blamed herself for the fire./Did Mary blame herself for the fire?/N.
   (d) The professional electrician that John consulted about the wiring problems blamed herself for the fire./Did the electrician blame John for the fire?/N.
4. (a) The boat mechanic that Brian found in the newspaper had taught himself everything about boats from a book./Did Brian teach himself everything about boats?/N.
   (b) The boat mechanic that Jane found in the newspaper had taught himself everything about boats from a book./Did Jane teach the mechanic everything about boats?/N.
   (c) The boat mechanic that Jane found in the newspaper had taught herself everything about boats from a book./Did Jane teach herself everything about boats?/N.
   (d) The boat mechanic that Brian found in the newspaper had taught herself everything about boats from a book./Did Brian teach the mechanic everything about boats?/N.
5. (a) The famous senator that Robert volunteered for in the campaign compared himself with previous presidents./Was Robert's senator famous?/Y.
   (b) The famous senator that Melissa volunteered for in the campaign compared himself with previous presidents./Was Melissa's senator famous?/Y.
   (c) The famous senator that Melissa volunteered for in the campaign compared herself with previous presidents./Was Melissa's senator famous?/Y.
   (d) The famous senator that Robert volunteered for in the campaign compared herself with previous presidents./Was Robert's senator famous?/Y.
6. (a) The old rancher that Paul chose as a milk supplier exhausted himself after building a huge barn on the hill./Did the rancher build a barn on the hill?/Y.
   (b) The old rancher that Nina chose as a milk supplier exhausted himself after building a huge barn on the hill./Did the rancher build a barn on the hill?/Y.

   (c) The old rancher that Nina chose as a milk supplier exhausted herself after building a huge barn on the hill./Did the rancher build a barn on the hill?/Y.
   (d) The old rancher that Paul chose as a milk supplier exhausted herself after building a huge barn on the hill./Did the rancher build a barn on the hill?/Y.

7. (a) The truck driver that Frank met in the gas station hated himself for smoking so heavily./Did Frank meet any truck drivers?/Y.
   (b) The truck driver that Stacy met in the gas station hated himself for smoking so heavily./Did Stacy meet any truck drivers?/Y.
   (c) The truck driver that Stacy met in the gas station hated herself for smoking so heavily./Did Stacy meet any truck drivers?/Y.
   (d) The truck driver that Frank met in the gas station hated herself for smoking so heavily./Did Frank meet any truck drivers?/Y.

8. (a) The construction worker that David drove by on the road hit himself on the face by accident./Did the construction worker hit himself?/Y.
   (b) The construction worker that Ashlee drove by on the road hit himself on the face by accident./Did the construction worker hit himself?/Y.
   (c) The construction worker that Ashlee drove by on the road hit herself on the face by accident./Did the construction worker hit herself?/Y.
   (d) The construction worker that David drove by on the road hit herself on the face by accident./Did the construction worker hit herself?/Y.

9. (a) The tough soldier that Fred treated in the military hospital introduced himself to all the nurses.
   (b) The tough soldier that Katy treated in the military hospital introduced himself to all the nurses.
   (c) The tough soldier that Katy treated in the military hospital introduced herself to all the nurses.
   (d) The tough soldier that Fred treated in the military hospital introduced herself to all the nurses.

10. (a) The taxi driver that Andrew crashed into on the road protected himself from being hurt by steering to the side.
   (b) The taxi driver that Amelia crashed into on the road protected himself from being hurt by steering to the side.
   (c) The taxi driver that Amelia crashed into on the road protected herself from being hurt by steering to the side.
   (d) The taxi driver that Andrew crashed into on the road protected herself from being hurt by steering to the side.

11. (a) The rough lumberjack that Kevin saw on the mountain had separated himself from the outside world.

(b) The rough lumberjack that Karen saw on the mountain had separated himself from the outside world.
(c) The rough lumberjack that Karen saw on the mountain had separated herself from the outside world.
(d) The rough lumberjack that Kevin saw on the mountain had separated herself from the outside world.

12. (a) The tidy janitor that Harold insulted in the school simply told himself to ignore the stupid comments.
(b) The tidy janitor that Veronica insulted in the school simply told himself to ignore the stupid comments.
(c) The tidy janitor that Veronica insulted in the school simply told herself to ignore the stupid comments.
(d) The tidy janitor that Harold insulted in the school simply told herself to ignore the stupid comments.

13. (a) The rude receptionist that Melinda spoke with on the phone locked herself out of the office./Did Melinda lock herself out of the office?/N.
(b) The rude receptionist that Melvin spoke with on the phone locked herself out of the office./Did Melvin lock himself out of the office?/N.
(c) The rude receptionist that Melvin spoke with on the phone locked himself out of the office./Did Melvin lock himself out of the office?/N.
(d) The rude receptionist that Melinda spoke with on the phone locked himself out of the office./Did Melinda lock herself out of the office?/N.

14. (a) The enthusiastic cheerleader that Tanya heard in the big stadium gave herself a sore throat./Was the cheerleader too quiet?/N.
(b) The enthusiastic cheerleader that Ted heard in the big stadium gave herself a sore throat./Was the cheerleader too quiet?/N.
(c) The enthusiastic cheerleader that Ted heard in the big stadium gave himself a sore throat./Was the cheerleader too quiet?/N.
(d) The enthusiastic cheerleader that Tanya heard in the big stadium gave himself a sore throat./Was the cheerleader too quiet?/N.

15. (a) The respected beautician that Amy interviewed in the fashion magazine educated herself through years of hard work./Was the beautician interviewed by Amy in the newspaper?/N.
(b) The respected beautician that Arnold interviewed in the fashion magazine educated herself through years of hard work./Was the beautician interviewed by Arnold in the newspaper?/N.
(c) The respected beautician that Arnold interviewed in the fashion magazine educated himself through years of hard work./Was the beautician interviewed by Arnold in the newspaper?/N.

(d) The respected beautician that Amy interviewed in the fashion magazine educated himself through years of hard work./Was the beautician interviewed by Amy in the newspaper?/N.

16. (a) The talkative cosmetician that Janice photographed for the local newspaper poisoned herself by accident./Did Janice poison herself?/N.
    (b) The talkative cosmetician that Phillip photographed for the local newspaper poisoned herself by accident./Did Phillip poison himself?/N.
    (c) The talkative cosmetician that Phillip photographed for the local newspaper poisoned himself by accident./Did Phillip poison himself?/N.
    (d) The talkative cosmetician that Janice photographed for the local newspaper poisoned himself by accident./Did Janice poison herself?/ N.

17. (a) The flight attendant that Wendy troubled on the long flight restrained herself from getting impatient./Did Wendy trouble the flight attendant?/Y.
    (b) The flight attendant that Richard troubled on the long flight restrained herself from getting impatient./Did Richard trouble the flight attendant?/Y.
    (c) The flight attendant that Richard troubled on the long flight restrained himself from getting impatient./Did Richard trouble the flight attendant?/Y.
    (d) The flight attendant that Wendy troubled on the long flight restrained himself from getting impatient./Did Wendy trouble the flight attendant?/Y.

18. (a) The tired hairdresser that Betsy kept at work all day drove herself home at midnight./Did the hairdresser drive herself home at night?/Y.
    (b) The tired hairdresser that Ron kept at work all day drove herself home at midnight./Did the hairdresser drive herself home at night?/Y.
    (c) The tired hairdresser that Ron kept at work all day drove himself home at midnight./Did Ron keep the hairdresser working until very late?/Y.
    (d) The tired hairdresser that Betsy kept at work all day drove himself home at midnight./Did Betsy keep the hairdresser working until very late?/Y.

19. (a) The kindergarten teacher that Helen called for a parent-teacher meeting sent herself an email reminder./Did the kindergarten teacher send herself an email?/Y.
    (b) The kindergarten teacher that Herbert called for a parent-teacher meeting sent herself an email reminder./Did the kindergarten teacher send herself an email?/Y.
    (c) The kindergarten teacher that Herbert called for a parent-teacher meeting sent himself an email reminder./Did Herbert call the kindergarten

teacher?/Y.

(d) The kindergarten teacher that Helen called for a parent-teacher meeting sent himself an email reminder./Did Helen call the kindergarten teacher?/Y.

20. (a) The beauty consultant that Isabelle met on the small island covered herself with ocean mud./Did the beauty consultant cover herself with ocean mud?/Y.
    (b) The beauty consultant that Matthew met on the small island covered herself with ocean mud./Did the beauty consultant cover herself with ocean mud?/Y.
    (c) The beauty consultant that Matthew met on the small island covered himself with ocean mud./Did the beauty consultant cover himself with ocean mud?/Y.
    (d) The beauty consultant that Isabelle met on the small island covered himself with ocean mud./Did the beauty consultant cover himself with ocean mud?/Y.

21. (a) The ballet dancer that Roxanne accidentally pushed to the ground examined herself for any signs of injuries or bruises.
    (b) The ballet dancer that Robby accidentally pushed to the ground examined herself for any signs of injuries or bruises.
    (c) The ballet dancer that Robby accidentally pushed to the ground examined himself for any signs of injuries or bruises.
    (d) The ballet dancer that Roxanne accidentally pushed to the ground examined himself for any signs of injuries or bruises.

22. (a) The eloquent feminist that Victoria collaborated with on a book defined herself as an advocate for women's rights.
    (b) The eloquent feminist that Joshua collaborated with on a book defined herself as an advocate for women's rights.
    (c) The eloquent feminist that Joshua collaborated with on a book defined himself as an advocate for women's rights.
    (d) The eloquent feminist that Victoria collaborated with on a book defined himself as an advocate for women's rights.

23. (a) The popular matchmaker that Elizabeth met in the online chat-room described herself as a servant of Cupid.
    (b) The popular matchmaker that Benjamin met in the online chat-room described herself as a servant of Cupid.
    (c) The popular matchmaker that Benjamin met in the online chat-room described himself as a servant of Cupid.
    (d) The popular matchmaker that Elizabeth met in the online chat-room described himself as a servant of Cupid.

24. (a) The kind caregiver that Amanda hired for the summer vacation reproached herself when the children were ill-behaved.
    (b) The kind caregiver that Johnny hired for the summer vacation reproached herself when the children were ill-behaved.
    (c) The kind caregiver that Johnny hired for the summer vacation reproached himself when the children were ill-behaved.
    (d) The kind caregiver that Amanda hired for the summer vacation reproached himself when the children were ill-behaved.

Appendix C

Predictions of the models for controls and aphasics

Table C.1: The predictions of sentence-picture matching accuracies in percentages across four models.

Sentence Type	Control	M1	M2	M3
Canonical	98.6	97.7	80.8	77.6
Non-canonical	92.5	92.8	57.5	53.1

Table C.2: The predictions of mean response times in milliseconds across four models.

Sentence Type	Response	Control	M1	M2	M3
Canonical	Correct	4688.657	5673.658	4742.642	5724.723
	Incorrect	4703.548	5692.813	4804.913	5787.199
Non-canonical	Correct	4966.801	6034.119	4904.102	5914.505
	Incorrect	4948.329	5931.523	4843.680	5847.385

Table C.3: The predictions of eye movements in terms of the differences in the activation of correct pictures and incorrect pictures across four models.

Sentence Type	Response	Region	Control	M1	M2	M3
Canonical	Correct	NP1	0	0	0	0
		Verb	0.64	0.63	0.63	0.60
		NP2	1.34	1.34	1.32	1.34
		Silence	2.40	2.40	2.05	2.09
Non-canonical	Correct	NP1	0	0	0	0
		Verb	0.63	0.61	0.60	0.62
		NP2	1.36	1.33	1.31	1.35
		Silence	2.33	2.31	1.69	1.68
Canonical	Incorrect	NP1	0	0	0	0
		Verb	−0.24	−0.18	−0.56	−0.52
		NP2	−0.47	−0.37	−1.17	−1.14
		Silence	−0.33	−0.26	−1.03	−1.00
Non-canonical	Incorrect	NP1	0	0	0	0
		Verb	−0.58	−0.50	−0.58	−0.60
		NP2	−1.06	−1.05	−1.26	−1.25
		Silence	−0.67	−0.88	−1.43	−1.44

References

Aho, A. V., & Ullman, J. D. (1972). *The theory of parsing, translation and compiling volume I: Parsing.* Upper Saddle River, NJ: Prentice-Hall.

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518.

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* Oxford, England: Oxford University Press.

Anderson, J. R., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1050.

Avrutin, S. (2006). Weak syntax. In Y. Grodzinsky & K. Amunts (Eds.), *Broca's region* (pp. 49–62). New York: Oxford University Press.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 748–769.

Bates, D., & Sarkar, D. (2007). lme4: Linear mixed-effects models using S4 classes (R package version 0.9975-11) [Computer software].

Beretta, A., & Munn, A. (1998, Dec). Double-agents and trace-deletion in agrammatism. *Brain and Language*, *65*(3), 404–421.

Blackwell, A. W., & Bates, E. (1995). Inducing grammatic profiles in normals: Evidence for the selective vulnerability of morphology under cognitive resource limitation. *Journal of Cognitive Neuroscience*, *7*, 228–257.

Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, *2*(1), 1-12.

Boston, M. F., Hale, J. T., Vasishth, S., & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, *26*(3), 301–349.

Burchert, F., De Bleser, R., & Sonntag, K. (2003, Nov). Does morphology make the difference? agrammatic sentence comprehension in german. *Brain and Language*, *87*(2), 323–342.

Burkhardt, P., Mercedes Piñango, M., & Wong, K. (2003). The role of the anterior left hemisphere in real-time sentence comprehension: Evidence from split intransitivity. *Brain and Language*, *86*(1), 9–22.

Caplan, D. (2001, Feb). The measurement of chance performance in aphasia, with specific reference to the comprehension of semantically reversible passive sentences: a note on issues raised by Caramazza, Capitani, Rey, and Berndt (2001) and Drai, Grodzinsky, and Zurif (2001). *Brain and Language*, *76*(2), 193–201.

Caplan, D. (2006, Aug). Aphasic deficits in syntactic processing. *Cortex*, *42*(6), 797–804.

Caplan, D. (2009). The neural basis of syntactic processing. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (4 ed., pp. 805–816). Cambridge, MA: MIT Press.

Caplan, D., Baker, C., & Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition*, *21*(2), 117–175.

Caplan, D., & Hildebrandt, N. (1988). *Disorders of syntactic comprehension.*

Cambridge: MIT Press.

Caplan, D., & Waters, G. (2003, Feb). On-line syntactic processing in aphasia: studies with auditory moving window presentation. *Brain and Language*, *84*(2), 222–249.

Caplan, D., Waters, G., Dede, G., Michaud, J., & Reddy, A. (2007, May). A study of syntactic processing in aphasia I: Behavioral (psycholinguistic) aspects. *Brain and Language*, *101*(2), 103–150.

Caplan, D., Waters, G., & Hildebrandt, N. (1997). Determinants of sentence comprehension in aphasic patients in sentence-picture matching tasks. *Journal of Speech, Language, and Hearing Research*, *40*, 542–555.

Caplan, D., & Waters, G. S. (1995). Aphasic disorders of syntactic comprehension and working memory capacity. *Cognitive Neuropsychology*, *12*(6), 637–649.

Caramazza, A., Capasso, R., Capitani, E., & Miceli, G. (2005, Jul). Patterns of comprehension performance in agrammatic Broca's aphasia: a test of the Trace Deletion Hypothesis. *Brain and Language*, *94*(1), 43–53.

Caramazza, A., & Zurif, E. B. (1976, Oct). Dissociation of algorithmic and heuristic processes in language comprehension: evidence from aphasia. *Brain and Language*, *3*(4), 572–582.

Chen, Z., Jäger, L., & Vasishth, S. (in press). How structure sensitive is the parser? Evidence from Mandarin Chinese. In B. Stolterfoht & S. Featherston (Eds.), *Empirical approaches to linguistic theory: Studies of meaning and structure.* Berlin: Mouton de Gruyter.

Chomsky, N. (1981). *Lectures on government and binding.* Berlin: Mouton de Gruyter.

Chomsky, N. (1986). *Barriers.* Cambridge, MA: MIT Press.

Choy, J. J., & Thompson, C. K. (2010). Binding in agrammatic aphasia: Processing to comprehension. *Aphasiology*, *24*(5), 551–579.

Clifton, C., Staub, A., & Rayner, K. (1988). Process synchronization, lexical ambiguity resolution, and aphasia. In S. I. Small, G. W. Cottrell, & M. K. Tanenhaus (Eds.), *Lexical ambiguity resolution perspectives from* *psycholinguistics, neuropsychology, and artificial intelligence* (pp. 229–267). Palo Alto, CA: Morgan Kaufmann Publishers, Inc.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. van Gompel (Ed.), *Eye movements: A window on mind and brain* (pp. 341–372). Amsterdam: Elsevier.

Coltheart, M. (2006). Acquired dyslexias and the computational modelling of reading. *Cognitive Neuropsychology*, *23*(1), 96–109.

Cooper, R. M. (1974). The control eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, 84–107.

Crescentini, C., & Stocco, A. (2005). Agrammatism as a failure in the lexical activation process. In B. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum Associates.

Cunnings, I., & Felser, C. (in press). The role of working memory in the processing of reflexives. *Language and Cognitive Processes.*

Dahan, D., & Tanenhaus, M. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*, 453–459.

De Bleser, R., Schwarz, W., & Burchert, F. (2006, Feb). Quantitative neurosyntactic analyses: the final word? *Brain and Language*, *96*(2), 143–146.

Dell, G. S., & Caramazza, A. (2008, Mar). Introduction to special issue on computational modelling in cognitive neuropsychology. *Cognitive Neuropsychology*, *25*(2), 131–135.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997, Oct). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801–838.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *108*(2), 193-

210.

Dick, F., Bates, E., Wulfeck, B., Utman, J. A., Dronkers, N., & Gernsbacher, M. A. (2001, Oct). Language deficits, localization, and grammar: evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychological Review*, *108*(4), 759–788.

Dickey, M. W., Choy, J. J., & Thompson, C. K. (2007, Jan). Real-time comprehension of wh- movement in aphasia: evidence from eyetracking while listening. *Brain and Language*, *100*(1), 1–22.

Dillon, B. (2011). *Structured access in sentence comprehension.* Unpublished doctoral dissertation, University of Maryland.

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2011). *Contrasting interference profiles for agreement and anaphora: Experimental and modeling evidence.* The 24th Annual CUNY Conference on Human Sentence Processing. Stanford, USA. (Available from http://linguistics.stanford.edu/cuny/schedule.php)

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Redfern, B. B., & Jaeger, J. J. (2004, May-Jun). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*(1-2), 145–177.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*, 641-655.

Engbert, R., & Kliegl, R. (2003, Apr). Microsaccades uncover the orientation of covert attention. *Vision Res*, *43*(9), 1035–1045.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). Swift: A dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777– 813.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203.

Frank, S. (2009). Surprisal-based comparison between a symbolic and a connectionist model of sentence processing. In *Proceedings of the 31st*

Annual Conference of the Cognitive Science Society (pp. 1139–1144).

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies.* Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Friederici, A. D., & Kilborn, K. (1989, July). Temporal constraints on language processing: Syntactic priming in Broca's aphasia. *Journal Cognitive Neuroscience*, *1*, 262–272.

Gavarro, A., & Romeu, J. (2010). Agentive versus non-agentive passives in catalan agrammatism. *Procedia Social and Behavioral Sciences*, *6*, 200–201.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* New York: Cambridge University Press.

Geyken, A. (2007). The DWDS corpus: A reference corpus for the German language of the 20th century. In C. Fellbaum (Ed.), *Idioms and Collocations: Corpus-based Linguistic, Lexicographic Studies.* London: Continuum Press.

Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium.* Cambridge, MA: MIT Press.

Gigley, H. M. (1983). HOPEAI and the dynamic process of language behavior. *Cognition & Brain Theory*, *6*(1), 39–88.

Gordon, P., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1411–1423.

Gordon, P., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, *51*(1), 97–114.

Gordon, P., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning,* *Memory, and Cognition*, *32*(6), 1304–1321.

Grasemann, U., Kiran, S., Sandberg, C., & Miikkulainen, R. (2011). Impairment and rehabilitation in bilingual aphasia: A som-based model. In J. Laaksonen & T. Honkela (Eds.), *Proceedings of WSOM11, 8th Workshop on Self-Organizing Maps, LNCS 6731* (pp. 207–217). Espoo, Finland: Springer Verlag.

Grodner, D., & Gibson, E. (2005). Some consequences of the serial nature of linguistic input. *Cognitive Science*, *29*(2), 261–290.

Grodzinsky, Y. (1995, Jul). A restrictive theory of agrammatic comprehension. *Brain and Language*, *50*(1), 27–51.

Grodzinsky, Y. (2000, Feb). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences*, *23*(1), 1–71.

Grodzinsky, Y. (2006, May). The language faculty, Broca's region, and the mirror system. *Cortex*, *42*(4), 464–468.

Grodzinsky, Y., Piñango, M. M., Zurif, E., & Drai, D. (1999, Apr). The critical role of group studies in neuropsychology: Comprehension regularities in Broca's aphasia. *Brain and Language*, *67*(2), 134–147.

Grodzinsky, Y., & Santi, A. (2008, Dec). The battle for Broca's region. *Trends in Cognitive Sciences*, *12*(12), 474–480.

Gutman, R., DeDe, G., Michaud, J., Liu, J. S., & Caplan, D. (2010, May). Rasch models of aphasic performance on syntactic comprehension tests. *Cognitive Neuropsychology*, *27*(3), 230–244.

Haarmann, H. J., Just, M. A., & Carpenter, P. A. (1997). Aphasic sentence comprehension as a resource deficit: A computational approach. *Brain and Language*, *59*(1), 76–120.

Haarmann, H. J., & Kolk, H. H. (1991a). A computer model of the temporal course of agrammatic sentence understanding: The effects of variation in severity and sentence complexity. *Cognitive Science*, *15*(1), 49–87.

Haarmann, H. J., & Kolk, H. H. J. (1991b). Syntactic priming in Broca's aphasics: Evidence for slow activation. *Aphasiology*, *5*(3), 247–263.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model.

In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics* (pp. 159–166). Pittsburgh, PA.

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, *32*, 101–123.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, *30*(4), 643–672.

Hanne, S., Sekerina, I. A., Vasishth, S., Burchert, F., & De Bleser, R. (2011). Chance in agrammatic sentence comprehension: What does it really mean? Evidence from eye movements of german agrammatic aphasic patients. *Aphasiology*, *25*(2), 221–244.

Hickok, G., & Avrutin, S. (1995). Representation, referentiality, and processing in agrammatic comprehension: Two case studies. *Brain and Language*, *50*(1), 10–26.

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*(2), 137-194.

Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics.* MIT Press.

Just, M. A., & Carpenter, P. A. (1992). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354.

Kempen, G., & Vosse, T. (1989). Incremental syntactic tree formation in human sentence processing: A cognitive architecture based on activation decay and simulated annealing. *Connection Science*, *1*(3), 273–290.

Kilborn, K. (1991, Aug). Selective impairment of grammatical morphology due to induced stress in normal listeners: Implications for aphasia. *Brain and Language*, *41*(2), 275–288.

Kliegl, R., Geyken, A., Hanneforth, T., & Würzner, K. (2006). *Corpus matters: A comparison of German DWDS and CELEX lexical and sublexical frequency norms for the prediction of reading fixations.* Unpublished manuscript.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency and predictability effects of words on eye movements in reading. In *Eye movements and information processing during reading* (Vol. 16, p. 262-284). East Sussex: Psychology Press. (Special Issue)

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*, 12-35.

Kolk, H. H., & Van Grunsven, M. M. (1985). Agrammatism as a variable phenomenon. *Cognitive Neuropsychology*, *2*(4), 347–384.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177.

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*, *25*, 93–115.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454.

Logačev, P., & Vasishth, S. (2006). The em package for computing eyetracking measures [Computer software].

Love, T., Swinney, D., Walenski, M., & Zurif, E. (2008, Dec). How left inferior frontal cortex participates in syntactic processing: Evidence from aphasia. *Brain and Language*, *107*(3), 203–219.

Martin, R. C. (2006). The neuropsychology of sentence processing: Where do we stand? *Cognitive Neuropsychology*, *23*(1), 74–95.

Matzke, M., Mai, H., Nager, W., Rüsseler, J., & Münte, T. (2002). The costs of freedom: An ERP-study of non-canonical sentences. *Clinical Neurophysiology*, *113*(6), 844–852.

Mauner, G., Fromkin, V. A., & Cornell, T. L. (1993, Oct). Comprehension and acceptability judgments in agrammatism: Disruptions in the syntax of referential dependency. *Brain and Language*, *45*(3), 340–370.

McElree, B. (1993). The locus of lexical preference effects in sentence comprehension: A time-course analysis. *Journal of Memory and Language*, *32*, 536–571.

McElree, B. (1998). Attended and non-attended states in working memory: Accessing categorized structures. *Journal of Memory and Language*, *38*(2), 225–252.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, *29*(2), 111–123.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General*, *118*(4), 346–373.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory and language memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*(1), 67–91.

Mirman, D., Yee, E., Blumstein, S. E., & Magnuson, J. S. (2011, May). Theories of spoken word recognition deficits in aphasia: Evidence from eye-tracking and computational modeling. *Brain and Language*, *117*(2), 53–68.

Mitchum, C. C., & Berndt, R. (2008). Comprehension and production of sentences. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related neurogenic communication disorders* (5 ed., p. 632-653). Philadelphia: Wolters Kluwer.

Miyake, A., Carpenter, P. A., & Just, M. A. (1994). A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology*, *11*(6), 671–717.

Miyake, A., Carpenter, P. A., & Just, M. A. (1995). Reduced resources and specific impairments in normal and aphasic sentence comprehension. *Cognitive Neuropsychology*, *12*(6), 651–679.

Myung, J.-Y., Blumstein, S. E., & Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: Manipulation knowledge of objects. *Cognition*, *98*(3), 223–243.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251–269.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA, USA: Harvard University Press.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, *18*(1), 5–19.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*(4), 601–626.

Patil, U., Vasishth, S., & Kliegl, R. (2009). Compound effect of probabilistic disambiguation and memory retrievals on sentence processing: Evidence from an eye-tracking corpus. In A. Howes, D. Peebles, & R. Cooper (Eds.), *9th International Conference on Cognitive Modeling — ICCM2009.* Manchester, UK.

Patil, U., Vasishth, S., & Lewis, R. L. (2011). *Early retrieval interference in syntax-guided antecedent search.* (Talk presented at CUNY Conference on Human Sentence Processing)

Penke, M., & Westermann, G. (2006, May). Broca's area and inflectional morphology: Evidence from Broca's aphasia and computer modeling. *Cortex*, *42*(4), 563–576.

Phillips, C., Wagers, M. W., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In J. T.

Runner (Ed.), *Experiments at the interfaces (syntax and semantics, volume 37)* (pp. 147–180). Emerald Group Publishing Limited.

Piñango, M. M. (2000). On the proper generalization for Broca's aphasia comprehension pattern: Why argument movement may not be at the source of Brocas deficit. *Behavioral & Brain Sciences*, *23*, 48–49.

Plaut, D. C. (2002, Oct). Graded modality-specific specialisation in semantics: A computational account of optic aphasia. *Cognitive Neuropsychology*, *19*(7), 603–639.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996, Jan). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115.

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology*, *52*, 1–56.

Prather, P. A., Zurif, E., Love, T., & Brownell, H. (1997, Oct). Speed of lexical activation in nonfluent Broca's aphasia and fluent Wernicke's aphasia. *Brain and Language*, *59*(3), 391–411.

R Development Core Team. (2009). *R: A Language and Environment for Statistical Computing.* Vienna, Austria. (ISBN 3-900051-07-0)

Reichle, E., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model effects of higher level language processing on eye-movements during reading. *Psychonomic Bulletin & Review*, *16*(1), 1–21.

Rogers, T. T., Ralph, M. A. L., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., et al. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, *111*(1), 205–235.

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2006). Processing reflexives and pronouns in picture noun phrase. *Cognitive Science*, *30*(2), 193–241.

Saffran, E. M., Schwartz, M. F., & Linebarger, M. C. (1998, Apr). Semantic influences on thematic role assignment: Evidence from normals and aphasics. *Brain and Language*, *62*(2), 255–297.

Schwartz, M. F., Linebarger, M. C., Saffran, E. M., & Pate, D. S. (1987). Syntactic transparency and sentence interpretation in aphasia. *Language and Cognitive Processes*, *2*(2), 85–113.

Schwartz, M. F., Saffran, E. M., & Marin, O. S. M. (1980). The word order problem in agrammatism : I. comprehension. *Brain and Language*, *10*(2), 249–262.

Stolcke, A. (1995). An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational Linguistics*, *21*, 165–202.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*(3), 542–562.

Sussman, R. S., & Sedivy, J. C. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes*, *18*(2), 143–163.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995, Jun). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Taylor, W. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, *30*, 415–433.

Thompson, C., & Choy, J. (2009). Pronominal resolution and gap filling in agrammatic aphasia: Evidence from eye movements. *Journal of Psycholinguistic Research*, *38*(3), 255–283.

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*(3), 285–316.

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*(3), 247–263.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*, 157–166.

Vasishth, S., Brüssow, S., Lewis, R. L., & Drenhaus, H. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, *32*(4), 685–712.

Vasishth, S., & Lewis, R. L. (2006). Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language*, *82*(4), 767–794.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*, 105–143.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206–237.

Wong, T. J., Cokel, E. T., & Schooler, L. J. (2010). An online database of act-r parameters: Towards a transparent community-based approach to model. In D. D. Salvucci & G. Gunzelmann (Eds.), *Proceedings of the 10th international conference on cognitive modeling* (pp. 282–286). Philadelphia, PA: Drexel University.

Woollams, A. M., Ralph, M. A. L., Plaut, D. C., & Patterson, K. (2007). Sd-squared: On the association between semantic dementia and surface dyslexia. *Psychological Review*, *114*(2), 316–339.

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain and Language*, *108*(1), 40–55.