Modeling the Noun Phrase versus Sentence Coordination Ambiguity in Dutch: Evidence from Surprisal Theory

Harm Brouwer
University of Groningen
Groningen, the Netherlands
harm.brouwer@rug.nl

Hartmut Fitz
University of Groningen
Groningen, the Netherlands
h.fitz@rug.nl

John C. J. Hoeks
University of Groningen
Groningen, the Netherlands
j.c.j.hoeks@rug.nl

Abstract

This paper investigates whether surprisal theory can account for differential processing difficulty in the NP-/S-coordination ambiguity in Dutch. Surprisal is estimated using a Probabilistic Context-Free Grammar (PCFG), which is induced from an automatically annotated corpus. We find that our lexicalized surprisal model can account for the reading time data from a classic experiment on this ambiguity by Frazier (1987). We argue that syntactic and lexical probabilities, as specified in a PCFG, are sufficient to account for what is commonly referred to as an NP-coordination preference.

1 Introduction

Language comprehension is incremental in that meaning is continuously assigned to utterances as they are encountered word-by-word (Altmann and Kamide, 1999). Not all words, however, are equally easy to process. A word's processing difficulty is affected by, for instance, its frequency or its effect on the syntactic and semantic interpretation of a sentence. A recent theory of sentence processing, surprisal theory (Hale, 2001; Levy, 2008), combines several of these aspects into one single concept, namely the *surprisal* of a word. A word's surprisal is proportional to its expectancy, i.e., the extent to which that word is expected (or predicted). The processing difficulty a word causes during comprehension is argued to be related linearly to its surprisal; the higher the surprisal value of a word, the more difficult it is to process.

In this paper we investigate whether surprisal theory can account for the processing difficulty involved in sentences containing the noun phrase (NP) versus sentence (S) coordination ambiguity. The sentences in (1), from a self-paced reading experiment by Frazier (1987), exemplify this ambiguity:

(1) a. Piet kuste Marie en / **haar zusje** / ook
Piet kissed Marie and / her sister / too
[1,222ms; NP-coordination]

b. Piet kuste Marie en / **haar zusje** / lachte
Piet kissed Marie and / her sister / laughed
[1,596ms; S-coordination]

Both sentences are temporarily ambiguous in the boldface region. Sentence (1-a) is disambiguated as an NP-coordination by the sentence-final adverb ook. Sentence (1-b), on the other hand, is disambiguated as an S-coordination by the sentence-final verb lachte. Frazier found that the verb lachte in sentence (1-b) takes longer to process (1,596 ms) than the adverb ook (1,222 ms) in (1-a).

Frazier (1987) explained these findings by assuming that the human language processor adheres to the so-called *minimal attachment principle*. According to this principle, the sentence processor projects the simplest syntactic structure which is compatible with the material read at any point in time. NP-coordination is syntactically simpler than S-coordination in that it requires less phrasal nodes to be projected. Hence, the processor is biased towards NP- over S-coordination. Processing costs are incurred when this initial preference has to be revised in the disambiguating region, as in sentence (1-b), resulting in longer reading times. Hoeks et al. (2006) have shown that the NP-coordination preference can be reduced, but not entirely eliminated, when poor thematic fit between the verb and a potential object make an NP-coordination less likely (e.g., *Jasper sanded the board and the carpenter laughed*). We argue here that this residual preference for NP-coordination can be explained in terms of syntactic and lexical expectation within the framework of surprisal theory. In contrast to the minimal attachment principle, surprisal theory does not pos-

Proceedings of the 2010 Workshop on Cognitive Modeling and Computational Linguistics, ACL 2010, pages 72–80,
Uppsala, Sweden, 15 July 2010.

tulate specific kinds of syntactic representations or rely on a metric of syntactic complexity to predict processing behavior.

This paper is organized as follows. In section 2 below, we briefly sketch basic surprisal theory. Then we describe how we induced a grammar from a large annotated Dutch corpus and how surprisal was estimated from this grammar (section 3). In section 4, we describe Frazier's experiment on the NP-/S-coordination ambiguity in more detail, and present our surprisal-based simulations of this data. We conclude with a discussion of our results in section 5.

2 Surprisal Theory

As was mentioned in the introduction, language processing is highly incremental, and proceeds on a more or less word-by-word basis. This suggests that a person's difficulty with processing a sentence can be modeled on a word level as proposed by Attneave (1959). Furthermore, it has recently been suggested that one of the characteristics of the comprehension system that makes it so fast, is its ability to anticipate what a speaker will say next. In other words, the language comprehension system works predictively (Otten et al., 2007; van Berkum et al., 2005). Surprisal theory is a model of differential processing difficulty which accommodates both these properties of the comprehension system, incremental processing and word prediction (Hale, 2001; Levy, 2008). In this theory, the processing difficulty of a sentence is a function of word processing difficulty. A word's difficulty is inversely proportional to its expectancy, i.e., the extent to which the word was *expected* or *predicted* in the context in which it occurred. The lower a word's expectancy, the more difficult it is to process. A word's surprisal is linearly related to its difficulty. Consequently, words with lower conditional probabilities (expectancy) lead to higher surprisal than words with higher conditional probabilities.

Surprisal theory is, to some extent, independent of the language model that generates conditional word probabilities. Different models can be used to estimate these probabilities. For all such models, however, a clear distinction can be made between *lexicalized* and *unlexicalized* surprisal. In lexicalized surprisal, the input to the language model is a sequence of words (i.e., a sentence). In unlexicalized surprisal, the input is a sequence of word categories (i.e., part-of-speech tags). While previous studies have used unlexicalized surprisal to predict reading times, evidence for lexicalized surprisal is rather sparse. Smith and Levy (2008) investigated the relation between lexicalized surprisal and reading time data for naturalistic texts. Using a trigram language model, they showed that there was a linear relationship between the two measures. Demberg and Keller (2008) examined whether this relation extended beyond transitional probabilities and found no significant effects. This state of affairs is somewhat unfortunate for surprisal theory since input to the human language processor consists of sequences of words, not part-of-speech tags. In our study we therefore used lexicalized surprisal to investigate whether it can account for reading time data from the NP-/S-coordination ambiguity in Dutch. Lexicalized surprisal furthermore allows us to study how syntactic expectations might be modulated or even reversed by lexical expectations in temporarily ambiguous sentences.

2.1 Probabilistic Context Free Grammars

Both Hale (2001) and Levy (2008) used a Probabilistic Context Free Grammar (PCFG) as a language model in their implementations of surprisal theory. A PCFG consists of a set of rewrite rules which are assigned some probability (Charniak, 1993):

$$\begin{array}{rcll} \text{S} & \rightarrow & \text{NP, VP} & 1.0 \\ \text{NP} & \rightarrow & \text{Det, N} & 0.5 \\ \text{NP} & \rightarrow & \text{NP, VP} & 0.5 \\ \dots & \rightarrow & \dots & \dots \end{array}$$

In this toy grammar, for instance, a noun phrase placeholder can be rewritten to a determiner followed by a noun symbol with probability 0.5. From such a PCFG, the probability of a sentence can be estimated as the product of the probabilities of all the rules used to derive the sentence. If a sentence has multiple derivations, its probability is the sum of the probabilities for each derivation. For our purpose, we also needed to obtain the probability of partial sentences, called *prefix probabilities*. The prefix probability $P(w_1...w_i)$ of a partial sentence $w_1...w_i$ is the sum of the probabilities of all sentences generated by the PCFG which share the initial segment $w_1...w_i$. Hale (2001) pointed out that the ratio of the prefix probabilities $P(w_1 \dots w_i)$ and $P(w_1 \dots w_{i-1})$ equals precisely the conditional probability of word w_i. Given a

PCFG, the difficulty of word w_i can therefore be defined as:

$$\text{difficulty}(w_i) \propto -log_2 \left[\frac{P(w_1 \ldots w_i)}{P(w_1 \ldots w_{i-1})} \right].$$

Surprisal theory requires a probabilistic language model that generates some form of word expectancy. The theory itself, however, is largely neutral with respect to which model is employed. Models other than PCFGs can be used to estimate surprisal. Nederhof et al. (1998), for instance, show that prefix probabilities, and therefore surprisal, can be estimated from Tree Adjoining Grammars. This approach was taken in Demberg and Keller (2009). Other approaches have used trigram models (Smith and Levy, 2008), Simple Recurrent Networks of the Elman type (Frank, 2009), Markov models and Echo-state Networks (Frank and Bod, 2010). This illustrates that surprisal theory is not committed to specific claims about the structural representations that language takes in the human mind. It rather functions as a "causal bottleneck" between the representations of a language model, and expectation-based comprehension difficulty (Levy, 2008). In other words, comprehension difficulty does not critically depend on the structural representations postulated by the language model which is harnessed to generate word expectancy.

The use of PCFGs raises some important questions on parallelism in language processing. A prefix probability can be interpreted as a probability distribution over all analyses compatible with a partial sentence. Since partial sentences can sometimes be completed in an indefinite number of ways, it seems both practically and psychologically implausible to implement this distribution as an enumeration over complete structures. Instead, prefix probabilities should be estimated as a by-product of incremental processing, as in Stolcke's (1995) parser (see section 3.2). This approach, however, still leaves open how many analyses are considered in parallel; does the human sentence processor employ full or limited parallelism? Jurafsky (1996) showed that full parallelism becomes more and more unmanageable when the amount of information used for disambiguation increases. Levy, on the other hand, argued that studies of probabilistic parsing reveal that typically a small number of analyses are assigned the majority of probability mass (Roark, 2001). Thus, even when assuming full parallelism, only a small number of 'relevant' analyses would be considered in parallel.

3 Grammar and Parser

3.1 Grammar Induction

In our simulations, we used a PCFG to model the phrase structure of natural language. To induce such a grammar, an annotated corpus was required. We used Alpino (van Noord, 2006)—a robust and wide-coverage dependency parser for Dutch—to automatically generate such a corpus, annotated with phrase structure, for 204.000 sentences, which were randomly extracted from Dutch newspapers. These analyses were then used to induce a PCFG consisting of 650 grammar rules, 89 non-terminals, and 208.133 terminals (lexical items).[1] Moreover, 29 of the 89 nonterminals could result in epsilon productions.

The Alpino parser constructed the phrase structure analyses automatically. Despite Alpino's high accuracy, some analyses might not be entirely correct. Nonetheless, the overall quality of Alpino's analyses is sufficient for corpus studies, and since surprisal theory relies largely on corpus features, we believe the small number of (partially) incorrect analyses should not affect the surprisal estimates computed from our PCFG.

3.2 Earley-Stolcke Parser

To compute prefix probabilities in our model we implemented Stolcke's (1995) probabilistic modification of Earley's (1970) parsing algorithm. An Earley-Stolcke parser is a breadth-first parser. At each point in processing, the parser maintains a collection of *states* that reflect all possible analyses of a partial sentence thus far. A state is a record that keeps track of:

(a) the position up to which a sentence has been processed,

(b) the grammar rule that is applied,

(c) a "dot position" indicating which part of the rule has been processed thus far, and

(d) the leftmost edge of the partial string generated by the rule.

[1] A PCFG can be induced by estimating the relative frequency of each CFG rule $A \to \alpha$:

$$P(A \to \alpha) = \frac{count(A\to\alpha)}{\sum_\beta count(A\to\beta)}.$$

The collection of states is constantly expanded by three operations. First upcoming structural and lexical material is predicted. For all predictions, new states are added with the "dot" placed on the leftmost side of the rule. Then it is determined whether there is a state that predicts the next word in the input sentence. If this is the case, a new state is added with the "dot" placed right to the predicted word. A third operation looks for states with the "dot" rightmost to a grammar rule, and then tries to find states which have the completed state as their leftmost edge. If such states are found, the "dot" in these states is moved to the right of this edge. This step is repeated until no more new states are added. These three operations are cyclically performed until the entire sentence is processed. Our grammar contained 29 non-terminals that could result in epsilon productions. Due to the way epsilon productions are handled within the Earley-Stolcke parser (i.e., by means of "spontaneous dot shifting"), having a large number of epsilon productions leads to a large number of predicted and completed edges. As a consequence, pursuing all possible analyses may become computationally infeasible. To overcome this problem, we modified the Earley-Stolcke parser with a *beam* λ. In prediction and completion, only the λ-number of states with the highest probabilities are added.[2] This constrains the number of states generated by the parser and enforces limited parallelism.

4 NP-/S-coordination ambiguities

4.1 Frazier's experiment

Our aim was to determine to what extent lexicalized surprisal theory can account for reading time data for the NP-/S-coordination ambiguity in Dutch. This type of ambiguity was investigated by Frazier (1987) using a self-paced reading experiment. The sentences in (2) are part of Frazier's materials. Sentence (2-a) and (2-b) exemplify an NP-/S-coordination ambiguity. The sentences are identical and temporarily ambiguous up to the NP haar zusje (her sister). In (2-a) this NP is followed by the adverb ook, and therefore disambiguated to be part of an NP-coordination; Marie and haar zusje are conjoined. In (2-b), on other hand, the same NP is followed by the verb lachte, and therefore disambiguated as the subject of a conjoined sentence; Piet kuste Marie and haar zusje lachte are conjoined.

(2) a. Piet kuste Marie en **haar zusje** ook
Pete kissed Marie and her sister too
(Ambiguous; NP-coordination)

b. Piet kuste Marie en **haar zusje** lachte
Pete kissed Marie and her sister laughed
(Ambiguous; S-coordination)

c. Annie zag **haar zusje** ook
Annie saw her sister too
(Unambiguous; NP-control)

d. Annie zag dat **haar zusje** lachte
Annie saw that her sister laughed
(Unambiguous; S-control)

Sentence (2-c) and (2-d) functioned as unambiguous controls. These sentences are identical up to the verb zag. In (2-c), the verb is followed by the single NP haar zusje, and subsequently the adverb ook. The adverb eliminates the possibility of an NP-coordination. In (2-d), on the other hand, the same verb is followed by the complementizer dat, indicating that the clause her sister laughed is a subordinate clause (the complementizer is obligatory in Dutch).

Frazier constructed twelve sets consisting of four of such sentences each. The 48 sentences were divided into three frames. The first frame included all the material up to the critical NP haar zusje in (2). The second frame contained only the critical NP itself, and the third frame contained all the material that followed this NP.

40 native Dutch speakers participated in the experiment. Reading times for the final frames were collected using a self-paced reading task. Figure 1 depicts the mean reading times for each of the four conditions.

Frazier found a significant interaction between Type of Coordination (NP- versus S-coordination) and Ambiguity (ambiguous versus control) indicating that the effect of disambiguation was larger for S-coordinations (ambiguous: 1596 ms; control: 1141 ms) than for NP-coordinations (ambiguous: 1222 ms; control: 1082 ms).

4.2 Simulations

We simulated Frazier's experiment in our model. Since one set of sentences contained a word that was not covered by our lexicon (set 11; "Lorraine"), we used only eleven of the twelve sets of test items from her study. The remaining 44 sentences were successfully analyzed. In our first

[2] A similar approach was used in Roark (2001) and Frank (2009).

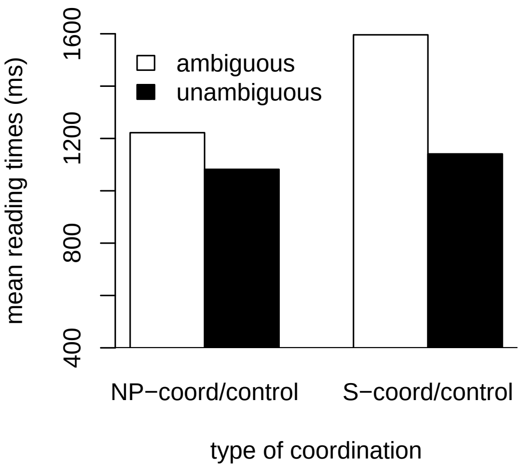

Figure 1: Reading time data for the NP-/S-coordination ambiguity (Frazier, 1987).

simulation we fixed a beam of $\lambda = 16$. Figure 2 depicts surprisal values in the sentence-final frame as estimated by our model. When final frames contained multiple words, we averaged the surprisal values for these words. As Figure 2 shows,

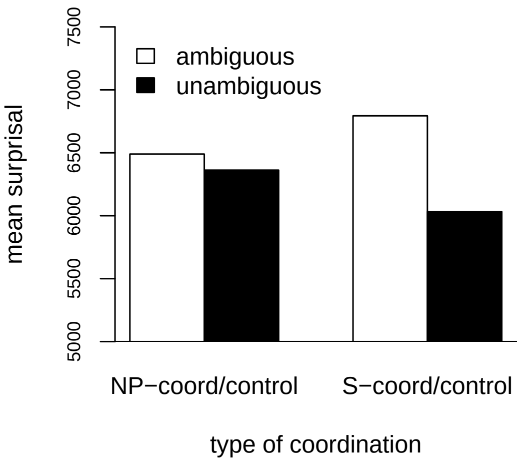

Figure 2: Mean surprisal values for the final frame in the model ($\lambda = 16$).

our model successfully replicated the effects reported in Frazier (1987): In both types of coordinations there was a difference in mean surprisal between the ambiguous sentences and the controls, but in the S-coordinations this effect was larger than in the sentences with NP-coordination. Statistical analyses confirmed our findings. An ANOVA on surprisal values per item revealed an interaction between Type of Coordination (NP- vs. S-coordination) and Ambiguity (ambiguous vs. control), which was marginally significant ($p = 0.06$), most probably due to the small number of

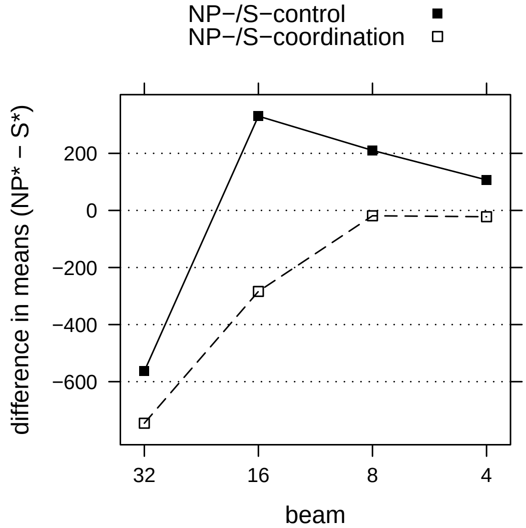

Figure 3: Differences between NP versus S surprisal for different beam sizes (λs).

items (i.e., 11) available for this statistical test (recall that the test in the original experiment was based on 40 participants). Follow-up analyses revealed that the difference between S-coordination and S-control was significant ($p < 0.05$), whereas the difference between NP-coordination and NP-control was not ($p = 0.527$).

To test the robustness of these findings, we repeated the simulation with different beam sizes (λs) by iteratively halving the beam, starting with $\lambda = 32$. Figure 3 shows the differences in mean surprisal between NP-coordination and S-coordination, and NP-control and S-control. With the beam set to four ($\lambda = 4$), we did not obtain full analyses for all test items. Consequently, two sets of items had to be disregarded (sets 8 and 9). For the remaining items, however, we obtained an NP-coordination preference for all beam sizes. The largest difference occurred for $\lambda = 16$. When the beam was set to $\lambda \leq 8$, the difference stabilized. Taking everything into account, the model with $\lambda = 16$ led to the best overall match with Frazier's reading time data.

As for the interaction, Figure 4 depicts the differences in mean surprisal between NP-coordination and NP-control, and S-coordination and S-control. These results indicate that we robustly replicated the interaction between coordination type and ambiguity. For all beam sizes, S-coordination benefited more from disambiguation than NP-coordination, i.e., the difference in means between S-coordination and S-control was larger

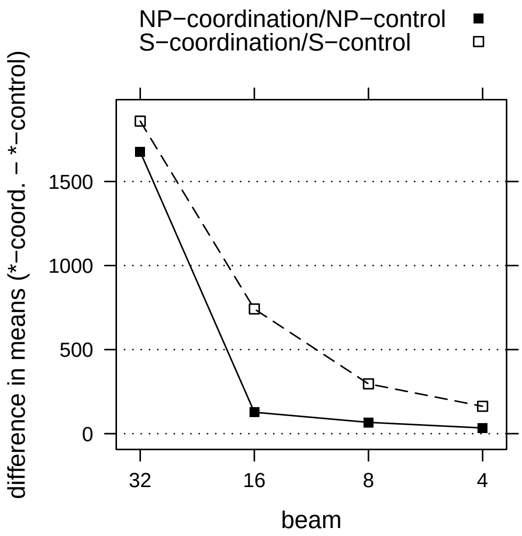

Figure 4: Differences in coordination versus control surprisal for different beam sizes (λs).

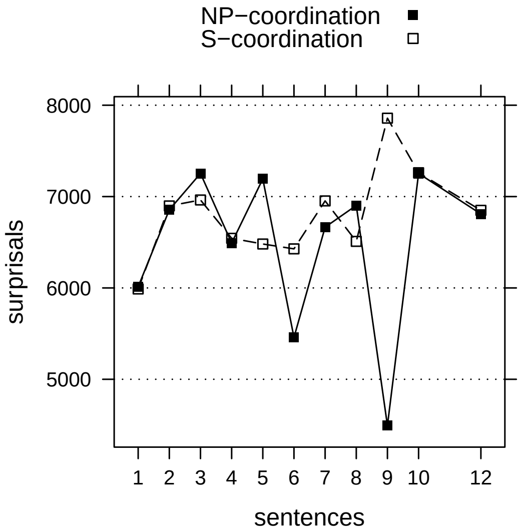

Figure 5: Surprisal per sentence for final frames in the ambiguous condition.

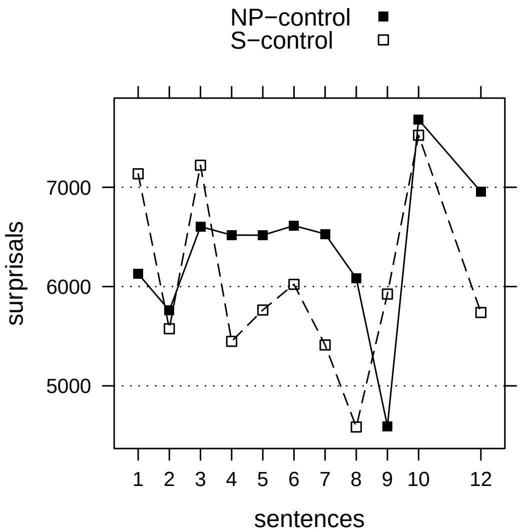

Figure 6: Surprisal per sentence for final frames in the unambiguous condition.

than the difference in means between NP-coordination and NP-control.

In our simulations, we found that surprisal theory can account for reading time data from a classic experiment on the NP-/S-coordination ambiguity in Dutch reported by Frazier (1987). This suggests that the interplay between syntactic and lexical expectancy might be sufficient to explain an NP-coordination preference in human subjects. In the remainder of this section, we analyze our results and explain how this preference arises in the model.

4.3 Model Analysis

To determine what caused the NP-preference in our model, we inspected surprisal differences item-by-item. Whether the NP-coordination preference was syntactic or lexical in nature should be reflected in the grammar. If it was syntactic, NP-coordination would have a higher probability than S-coordination according to our PCFG. If, on the other hand, it was lexical, NP- and S-coordination should be equally probable syntactically. Another possibility, however, is that syntactic and lexical probabilities interacted. If this was the case, we should expect NP-coordinations to lead to lower surprisal values on average only, but not necessarily on every item. Figure 5 shows the estimated surprisal values per sentence-final frame for the ambiguous condition and Figure 6 for the unambiguous condition. Figure 5 indicates that although NP-coordination led to lower surprisal overall (see Figure 2), this was not the case for all tested items. A similar pattern was found for the NP-control versus S-control items in Figure 6. S-controls led to lower surprisal overall, but not for all items. Manual inspection of the grammar revealed a bias towards NP-coordination. A total of 115 PCFG rules concerned coordination ($\approx$ 18% of the entire grammar). As these rules expanded the same grammatical category, their probabilities summed to 1. A rule-by-rule inspection showed that approximately 48% of the probability mass was assigned to rules that dealt with NP-coordinations, 22% to rules that dealt with S-coordinations, and the remaining 30% to rules that dealt with coordination in other structures. In other words, there was a clear preference for NP-coordination in the grammar. Despite this bias, for some tested items the S-coordination received lower surprisal than the NP-coordination (Figure 5). Individual NP-coordination rules might have lower probability than individual S-coordination rules, so the overall preference for NP-coordination in the grammar therefore does not have to be reflected in every test item. Secondly, syntactic probabilities could be modified by lexical probabilities. Suppose for a pair of test items that NP-coordination was syntactically preferred over S-coordination. If the sentence was disambiguated as an NP-coordination by a highly improbable lexical item, and disambiguated as an S-coordination by a highly probable lexical item, surprisal for the NP-coordination might turn out higher than surprisal for the S-coordination. In this way, lexical factors could override the NP-coordination bias in the grammar, leading to a preference for S-coordination in some items.

To summarize, the PCFG displayed an overall NP-coordination preference when surprisal was averaged over the test sentences and this result is consistent with the findings of Frazier (1987). The NP-coordination preference, however, was not invariably reflected on an item-by-item basis. Some S-coordinations showed lower surprisal than the corresponding NP-coordinations. This reversal of processing difficulty can be explained in terms of differences in individual rules, and in terms of interactions between syntactic and lexical probabilities. This suggests that specific lexical expectations might have a much stronger effect on disambiguation preferences than supposed by the minimal attachment principle. Unfortunately, Frazier (1987) only reported mean reading times for the two coordination types.[3] It would be interesting to compare the predictions from our surprisal model with human data item-by-item in order to validate the magnitude of lexical effects we found in the model.

5 Discussion

In this paper we have shown that a model of lexicalized surprisal, based on an automatically induced PCFG, can account for the NP-/S-ambiguity reading time data of Frazier (1987). We found these results to be robust for a critical model parameter (beam size), which suggests that syntactic processing in human comprehension might be based on limited parallelism only. Surprisal theory models processing difficulty on a word level. A word's difficulty is related to the expectations the language processor forms, given the structural and lexical material that precedes it. The model showed a clear preference for NP-coordination which suggests that structural and lexical expectations as estimated from a corpus might be sufficient to explain the NP-coordination bias in human sentence processing.

Our account of this bias differs considerably from the original account proposed by Frazier (minimal attachment principle) in a number of ways. Frazier's explanation is based on a metric of syntactic complexity which in turn depends on quite specific syntactic representations of a language's phrase structure. Surprisal theory, on the other hand, is largely neutral with respect to the form syntactic representations take in the human mind.[4] Moreover, differential processing in surprisal-based models does not require the specification of a notion of syntactic complexity. Both these aspects make surprisal theory a parsimonious explanatory framework. The minimal attachment principle postulates that the bias towards NP-coordination is an initial processing primitive. In contrast, the bias in our simulations is a function of the model's input history and linguistic experience from which the grammar is induced. It is further modulated by the immediate context from which upcoming words are predicted during processing. Consequently, the model's preference for one structural type can vary across sentence tokens and even be reversed on occasion. We argued that our grammar showed an overall preference for NP-coordination but this preference was not necessarily reflected on each and every rule that dealt with coordinations. Some S-coordination rules could have higher probability than NP-coordination rules. In addition, syntactic expectations were modified by lexical expectations. Thus, even when NP-coordination was structurally favored over S-coordination, highly unexpected lexical material could lead to more processing difficulty for NP-coordination than for

[3] Thus it was not possible to determine the strength of the correlation between reading times in Frazier's study and surprisal in our model.

[4] This is not to say, of course, that the choice of language model to estimate surprisal is completely irrelevant; different models will yield different degrees of fit, see Frank and Bod (2010).

S-coordination. Surprisal theory allows us to build a formally precise computational model of reading time data which generates testable, quantitative predictions about the differential processing of individual test items. These predictions (Figure 5) indicate that mean reading times for a set of NP-/S-coordination sentences may not be adequate to tap the origin of differential processing difficulty.

Our results are consistent with the findings of Hoeks et al. (2002), who also found evidence for an NP-coordination preference in a self-paced reading experiment as well as in an eye-tracking experiment. They suggested that NP-coordination might be easier to process because it has a simpler *topic structure* than S-coordination. The former only has one topic, whereas the latter has two. Hoeks et al. (2002) argue that having more than one topic is unexpected. Sentences with more than one topic will therefore cause more processing difficulty. This preference for simple topic-structure that was evident in language comprehension may also be present in language production, and hence in language corpora. Thus, it may very well be the case that the NP-coordination preference that was present in our training corpus may have had a pragmatic origin related to topic-structure. The outcome of our surprisal model is also compatible with the results of Hoeks et al. (2006) who found that thematic information can strongly reduce but not completely eliminate the NP-coordination preference. Surprisal theory is explicitly built on the assumption that multiple sources of information can interact in parallel at any point in time during sentence processing. Accordingly, we suggest here that the residual preference for NP-coordination found in the study of Hoeks et al. (2006) might be explained in terms of syntactic and lexical expectation. And finally, our approach is consistent with a large body of evidence indicating that language comprehension is incremental and makes use of expectation-driven word prediction (Pickering and Garrod, 2007). It remains to be tested whether our model can explain behavioral data from the processing of ambiguities other than the Dutch NP- versus S-coordination case.

References

G. Altmann and Y. Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73:247–264.

F. Attneave. 1959. *Applications of Information Theory to Psychology: A summary of basic concepts, methods, and results*. Holt, Rinehart and Winston.

E. Charniak. 1993. *Statistical Language Learning*. MIT Press.

V. Demberg and F. Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109:193–210.

V. Demberg and F. Keller. 2009. A computational model of prediction in human parsing: Unifying locality and surprisal effects. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, Amsterdam, the Netherlands.

J. Earley. 1970. An efficient context-free parsing algorithm. *Communications of the ACM*, 6:451–455.

S. Frank and R. Bod. 2010. The irrelevance of hierarchical structure to human sentence processing. Unpublished manuscript.

S. Frank. 2009. Surprisal-based comparison between a symbolic and a connectionist model of sentence processing. In *Proceedings of the 31th Annual Conference of the Cognitive Science Society*, pages 1139–1144, Amsterdam, the Netherlands.

L. Frazier. 1987. Syntactic processing: Evidence from Dutch. *Natural Langauge and Linguistic Theory*, 5:519–559.

J. Hale. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the 2nd Conference of the North American Chapter of the Association for Computational Linguistics*, volume 2, pages 159–166.

J. Hoeks, W. Vonk, and H. Schriefers. 2002. Processing coordinated structures in context: The effect of topic-structure on ambiguity resolution. *Journal of Memory and Language*, 46:99–119.

J. Hoeks, P. Hendriks, W. Vonk, C. Brown, and P. Hagoort. 2006. Processing the noun phrase versus sentence coordination ambiguity: Thematic information does not completely eliminate processing difficulty. *The Quarterly Journal of Experimental Psychology*, 59:1581–1599.

D. Jurafsky. 1996. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20:137–147.

R. Levy. 2008. Expectation-based syntactic comprehension. *Cognition*, 106:1126–1177.

M. Nederhof, A. Sarkar, and G. Satta. 1998. Prefix probabilities from stochastic tree adjoining grammar. In *Proceedings of COLING-ACL '98*, pages 953–959, Montreal.

M. Otten, M. Nieuwland, and J. van Berkum. 2007. Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience*.

M. Pickering and S. Garrod. 2007. Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11:105–110.

B. Roark. 2001. Probabilistic top-down parsing and language modeling. *Computational Linguistics*, 27:249–276.

N. Smith and R. Levy. 2008. Optimal processing times in reading: A formal model and empirical investigation. In *Proceedings of the 30th annual conference of the Cognitive Science Society*, pages 595–600, Austin, TX.

A. Stolcke. 1995. An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational linguistics*, 21:165–201.

J. van Berkum, C. Brown, P. Zwitserlood, V. Kooijman, and P. Hagoort. 2005. Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31:443–467.

G. van Noord. 2006. At last parsing is now operational. In *Verbum Ex Machina. Actes de la 13e conférence sur le traitement automatique des langues naturelles*, pages 20–42. Presses universitaires de Louvain.