Processing Parallel Structure: Evidence from Eye-Tracking and a Computational Model

Jens Apel (j.apel@sms.ed.ac.uk)
Department of Psychology, University of Edinburgh
7 George Square, Edinburgh, EH8 9JZ, UK

Pia Knoeferle (knoferle@coli.uni-sb.de)
Department of Computational Linguistics
Saarland University, 66041 Saarbrücken, Germany

Matthew W. Crocker (crocker@coli.uni-sb.de)
Department of Computational Linguistics
Saarland University, 66041 Saarbrücken, Germany

Abstract

The parallelism effect in human parsing is a phenomenon in which the second constituent of a coordinate structure is processed faster when it parallels the first constituent in comparison with when it does not parallel the first constituent. The main aim of this paper is to investigate whether the parallelism effect, which was first discovered in ambiguous coordinate structures, also occurs in non-coordinate constructions, and in structurally unambiguous sentences. The motivation for investigating these two issues was to determine the extent of, and the mechanisms underlying parallelism effects, with the goal of informing the development of a computational model. Two eye-tracking studies of German showed that parallelism effects are obtained in unambiguous sentences, and strongly suggest that the effect is restricted to coordinate structures. Additionally, the parallelism effect is shown to be sensitive to the fine-grained parts-of-speech of the parallel constituents. Based on these findings, we present a computational system that is able to model our effects, and is consistent with other prevailing results.

Introduction

The Parallelism Effect

The parallelism effect in human parsing is a phenomenon in which the second constituent of a coordinate structure is processed faster when it parallels the first constituent in comparison with when it does not parallel the first constituent. In the first study of this phenomena, Frazier, Taft, Roeper, & Clifton (1984) conducted a series of five self-paced reading experiments which examined the parallelism effect for a range of clause types. The conjoined clauses of their sentences were either parallel or non-parallel. For example, in Sentence (1) the first and second clause are both active and thus, parallel, whereas in Sentence (2) the two coordinated clauses are non-parallel: the first clause is passive, and the second clause active.

(1) The tall gangster hit John and the short thug hit Sam.

(2) John was hit by the tall gangster and the short thug hit Sam.

The most interesting finding in this study was that the parallelism effect (i.e. faster processing of *the short thug hit Sam* in (1) than in (2)) was detected in all tested sentence constructions and seems to be very stable among different syntactic structures.

Frazier, Munn, & Clifton (2000) further investigated the scope of the parallelism effect by testing whether it also occurs in non-coordinate structures. If the effect is not restricted to structures within coordination, then a priming effect that occurs in both coordinate and non-coordinate constructions might be the cause of the facilitation of the second constituent. Frazier et al. (2000) tested this question by first replicating the parallelism effect in sentences containing coordinated noun phrases (NP). In a second study they then restructured the sentences such that the NPs were in subject and object position. In the parallel condition both NPs consisted of a determiner, an adjective and a noun. In the non-parallel condition the first NP consisted of a determiner and a noun and the second NP of determiner, adjective and noun.

Frazier et al. (2000) found parallelism effects in the sentences containing a coordinate structure. However, the results from the second experiment revealed that the noun phrase in object position was not faster processed when its syntactic structure paralleled that of the noun phrase in subject position. Frazier et al. (2000) concluded that the parallelism effect does not affect structures (e.g. NPs) that are not coordinated with each other. To explain their findings, Frazier et al. (2000) argue that parallel structures are most predictable in coordinated environments, and suggest that the parallelism effect might be an instance of this predictability. That is, facilitation of the second constituent might be processed more quickly when its structure can be predicted by the first conjunct.

However, reading times in the first experiment (coordinated NPs) were measured by using eye-tracking, whereas self-paced reading was used in the second experiment (non-coordinated NPs). As a result of the differences in methods, it is difficult to directly compare findings from the two experiments. Furthermore, with self-paced, moving window presentation, the reader is prevented from looking back at earlier regions once they have pressed a button to move to the next part of the sentence. It may be that important re-reading times, which would be observed using eye-tracking, were therefore crucially missing from the reported total times.

Furthermore, the sentences including a coordinate structure in Frazier et al.'s (2000) study contained a local ambiguity, despite the intention of examining parallelism in unambiguous sentences. Specifically, the second noun phrase of the coordinate structure (as e.g. in *Hilda noticed a strange man and a tall woman...*) could also be attached as the subject of a coordinated sentence resulting in sentences like *Hilda noticed a strange man and a tall woman entered the house*.

In this paper we present two eye-tracking studies of German, to clarify (a) whether or not parallelism is limited to coordinate constructions, (b) whether it can be observed in completely unambiguous sentences, and (c) the level of granularity in lexico-syntactic structure which is required for constituents to be considered parallel. By using eye-tracking for both studies, we are able to more directly compare our findings, and ensure that any possible null-effect in the no-coordination study could not be an artifact of the self-paced reading method used. Based on the evidence from the two experiments presented here, as well as the proposals of Frazier and colleagues, we develop a computational model in which the likelihood that the parser initially interprets a coordinate structure as being parallel is higher than interpreting it as non-parallel.

Experiment 1

Experiment 1 was conducted in order to investigate the parallelism effect in fully unambiguous coordinate structures. In addition to contrasting noun phrases containing an adjective or no noun modifier at all, we added a third condition in which the noun modifier consists of a participle. This enabled us to investigate whether the parallelism effect still occurs when the number of words in the noun phrase stays the same and only the parts of speech (POS) of the noun modifiers differ.

Method

Participants Thirty-six participants from Saarland University took part in Experiment 1. The average age of participants was 23.9 (range 19-50). All participants were native speakers of German and showed no visual impairments. Each participant was paid 7,50 Euro for participation.

Materials Twenty-four sentences were used as experimental items. Examples of the sentences are illustrated below:

(3) **Der Esel** und **der melkende Bauer** sind vor dem Gewitter geflüchtet.
The donkey and the milking farmer fled from the thunderstorm.

(4) **Der dämliche Esel** und **der melkende Bauer** sind vor dem Gewitter geflüchtet.
The dim-witted donkey and the milking farmer fled...

(5) **Der stampfende Esel** und **der melkende Bauer** sind vor dem Gewitter geflüchtet.
The stomping donkey and the milking farmer fled...

The second noun phrase of the coordinate structure always consisted of a determiner, a participle and a noun. The first constituent differed in the type of noun modifier. In the first condition (Sentence 3), no noun modifier was present (Empty-Part) and thus, the constituents were non-parallel concerning number of words and consequently also differed in the parts of speech of the words. In the second condition (Sentence 4) the noun modifier consisted of an adjective (Adj-Part). Thus, the constituents differed only in the part of speech of the noun modifier. In the third condition (Sentence 5) the noun modifier was a participle (Part-Part) and the two constituents of the coordinate structure were parallel concerning number of words and the part of speech of the noun modifier.

The length of the corresponding words (i.e. determiner, noun modifier and noun) within the coordinated noun phrases varied by at most two characters in the three conditions of an item (mean lengths in characters: adjective: 9.11, participle first NP: 9.59, participle second NP: 9.63, first NP: 7.67, second NP: 7.89). Furthermore, the corresponding words in the three conditions of an item possessed the same number of syllables, and had the same stress pattern. The frequency of the word pairs was also kept similar (i.e. the difference of frequency is not more than 0.3 log frequency for the corresponding words in the three conditions of an item; mean frequency obtained from the celex database: adjective: 0.02, participle first NP: 0.01, participle second NP: 0.02, first NP: 0.45, second NP: 0.54). To ensure that the plausibility of the sentences in the three different conditions was similar, a pre-test was conducted. Twenty-one participants rated the 24 sentences on a scale between 7 (very plausible) and 1 (very implausible). A one-way repeated measure analysis of variance (ANOVA; 1 factor, 3 levels), by subjects (F_1) and by items (F_2), was conducted yielding no significant difference between the three conditions [$F_1(2,19)=0.20$, $p > .10$, and $F_2(2,22)=0.23$, $p > .10$].

If the parallelism effect extends to non-ambiguous sentences, we expect faster reading times on the second noun phrase of the coordinate structure when it parallels the first constituent in comparison to when it does not. The experimental design of this study allows us to further determine how a parallel structure is defined. If the second noun phrase is read faster in both the Adj-Part and Part-Part conditions than in the Empty-Part condition, this would suggest that parallelism does not depend on exact parallel parts of speech. If a parallelism effect only occurs in the Part-Part condition, parallel parts of speech are necessary to trigger the parallelism effect.

Design and Procedure One factor [Type of Parallelism (Par-Type)] with three levels (Empty-Part, Adj-Part, Part-Part) was used in the experiment. Par-Type was a within-subject variable, meaning that each participant saw repeated instances of every condition but only one sentence of each item.

The 24 experimental items were mixed with 72 filler items. Yes-no questions on the content were asked for 36 filler items to ensure that participants remained concentrated. The items were pseudo randomized such that all experimental items were separated by at least one filler item. Each participant received an individually randomized list.

An SMI EyeLink head-mounted eye-tracker with a sampling rate of 250 Hz was used to track the eye movements of the participants. This system consists of a headset with two cameras to enable eye-movement recording. Only the dominant eye was recorded.

Data Analysis Fixation duration times and locations were recorded for every word of the sentences. However, we were especially interested in the second conjunct (the second noun phrase) of the coordinate structure since previous studies have found parallelism effects in the second conjunct of coordinate structures. In case we replicate the parallelism effect, faster reading times are expected in this region.

A one-way repeated measure ANOVA analysis was conducted on reading times for all of the regions. The subjects' performance in answering the questions was also analyzed. Participants who deviated more than 2SD from the mean response accuracy across all conditions were excluded from the analysis. The mean of incorrectly answered questions was 34 %. The relatively high error rate presumably resulted from including difficult questions, a choice made to ensure thorough reading. One participant, who answered 44% of the questions wrong, was excluded. Therefore, 35 participants were included in the analysis.

Results and Discussion

As a main observation of this experiment, we replicated the finding from Frazier et al. (2000). We found significant differences between parallel and non-parallel conditions in regression path time (RPD) of the noun in the second noun phrase. RPD is defined as the sum of all fixations beginning with the first fixation in the region and ending when the eye moves rightward to the next region (this includes all regressive fixations). The RPD reading time of the noun in the second constituent of the coordinate structure differed significantly between the conditions $F_1(2, 33) = 3.56$, $p < .05$; $F_2(2, 22) = 3.98$, $p < .05$] [1]. The descriptive data is illustrated in Figure 1.

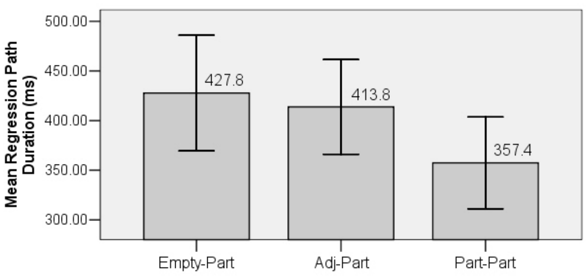

Figure 1: Mean RPD of the Noun of the Second Noun Phrase in Experiment 1.

Post hoc analyses revealed that in both the subjects and items analyses, the RPD fixation durations were significantly shorter in the Part-Part condition (357.8 ms) than in the Empty-Part condition (427.8 ms; p<.05). Furthermore, RPD fixation durations were shorter in the Part-Part condition than in the Adj-Part condition (413.8 ms). However, this difference was only marginally significant in the per subjects analysis and not significant in the per items analysis.

[1]Since the raw data was not normally distributed, the data was transformed using a log_{10} transformation in both Experiments 1 and 2.

Experiment 2

In Experiment 2 we tested whether a parallelism effect can be observed in non-coordinated structures.

Method

Participants Twenty-one further participants from the same population in Experiment 1 took part in Experiment 2. The average age of participants was 29.2 (range 17-51).

Materials Twenty-four sentences were used as experimental items. Examples of the sentences are illustrated below:

(6) **Der Esel** beißt **den melkenden Bauern** ohne jede Vorwarnung.
The donkey bites the milking farmer without any warning.

(7) **Der dämliche Esel** beißt **den melkenden Bauern** ohne jede Vorwarnung.
The dim-witted donkey bites the milking farmer...

(8) **Der stampfende Esel** beißt **den melkenden Bauern** ohne jede Vorwarnung.
The stomping donkey bites the milking farmer...

The sentences were based on the items of Experiment 1. The first noun phrase from the coordination in Experiment 1 served as the subject. The second noun phrase took the place of the object. The same conditions as in Experiment 1 were used: Empty-Part, Adj-Part and Part-Part.

A plausibility rating test was also conducted for Experiment 2, yielding no significant difference between the three conditions [$F_1(2, 19) = 1.14$, $p > .10$, and F_2(2,22)=1.85, $p > .10$.].

Design and Procedure Design and procedure were the same as in Experiment 1.

Data Analysis The data analysis was the same as in Experiment 1. Response accuracy on the questions was high (15% error rate), and no participant mean deviated more than 2SD from the mean response accuracy.

Results and Discussion

Since the parallelism effect in Experiment 1 did only appear at the noun of the second noun phrase, we also expected the effect to appear on this region (i.e. the noun of the object noun phrase) in Experiment 2 if the parallelism effect extends to non-coordinate structures. Furthermore, as in Experiment 1, we expected an effect in the RPD measure. As a main result, however, we could not find any differences in reading time between the three conditions at this region.

Figure 2 illustrates the RPD-data at the noun of the noun phrase in object position. The main effect was not

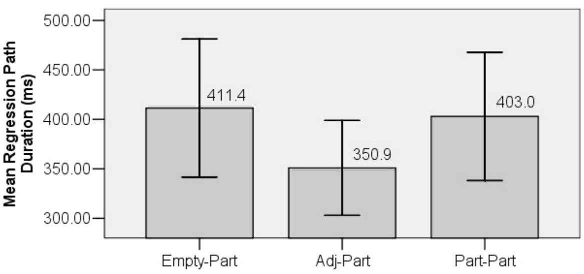

Figure 2: Mean RPD of the Noun of the Second Noun Phrase in Experiment 2.

reliable $[F_1(1,19) = 1.87, p = .18; F_2(2,22) = 1.22, p = .31]$.

In order to test the possibility that the effect might occur earlier or later than in Experiment 1, we examined reading times for regions adjacent to the second noun, and also other measures than RPD. However, we could not find any significant effects in support of a parallelism effect.

Discussion of Results

The two experiments replicated the results from the study by Frazier et al. (2000). A significant parallelism effect was found in Experiment 1 but not in Experiment 2. Thus, the parallelism effect seems to be restricted to coordinate structures and does not come to play when the noun phrases take a subject and object position. A general priming approach as the origin of the parallelism effect can be rejected. It is more likely that a mechanism specific to coordinate structures is responsible for the effect. Furthermore, since in Experiment 1 a parallelism effect was only found in parallel syntactic structures, the parallelism effect seems to depend on parallel parts of speech. Keeping the surface structure of the constituents the same is not sufficient. Finally, since the sentences in Experiment 1 do not exhibit any ambiguity, it can be concluded that the parallelism effect does not only occur in ambiguous sentences but extends to unambiguous structures.

In the next section, we introduce the computational model of the study.

A Computational Model

Several proposals have been made to explain parallelism effects. Frazier and Clifton (2001) postulated a copy α mechanism, which copies the syntactic structure of the first constituent of a coordinate structure to the second constituent and therefore, saves processing costs if the first constituent parallels the second one. This mechanism only applies to coordinate constructions. When the model encounters a conjunction, a pointer is set at the beginning of the first constituent of a coordinate structure. The whole syntactic structure is then copied to the second constituent and if this constituent parallels the first one, no extra cost for building the structure of the second constituent is spent.

Steiner (2003) proposes an alternative mechanism. The distinctive feature of this model is that the syntactic tree can be three-dimensional. Whenever a coordinate structure is encountered (triggered by a conjunction), the parser jumps back to the beginning of the first constituent and opens a third dimension. The terminal nodes of the second constituent are attached to the existing syntactic structure of the first constituent if the two constituents are parallel. If the coordinate structure is not parallel a new structure has to be built. Steiner (2003) argues that reusing structure saves time, and therefore, a parallel second constituent is processed faster than a non-parallel one.

Neither of these two models is very clear concerning how the first constituent of a parallel coordinate structure is to be identified. In the model by Frazier and Clifton (2001), a pointer is moved to the antecedent constituent to copy the structure of this constituent to the second one. However, Frazier and Clifton (2001) do not make explicit how the beginning of the antecedent constituent is found. The conjunction in sentences like *The man saw the boy and...* could coordinate the noun phrase, the verb phrase or the sentential phrase of the sentence. Thus, the beginning of the antecedent constituent is ambiguously defined. Furthermore, how does the parser recognize that the first clause parallels the second one in order to copy the syntactic structure of the first to the second constituent? Parsing the whole sentence again after a conjunction was found, to detect and analyze the antecedent clause, would probably nullify the advantages of the copy α mechanism. Steiner (2003) integrated a connectionist component with a spreading activation principle into her model in order to solve the problem of comparing the first constituent with the second one to use the same structure when they are parallel.

In the model of the current study, an alternative two-dimensional parsing mechanism is introduced. We further describe a method that is able to deal with the problem of detecting the beginning of the first constituent in a parallel structure. The validity of the model will be evaluated by testing it with sentences from the current study and with sentences from another study by Knoeferle & Crocker (2006).

The Model of the Current Study

The system is divided into two main processing modules. The first module is an incremental serial arc-eager left-corner chart parser. The second module which is termed coordination-module (`C-Module`[2]) is specifically designed to handle coordinate structures. The `Parser` works entirely independent from the `C-Module`.

The structure of the whole system is illustrated in Figure 3. It was implemented using the computational modeling environment COGENT (Cooper, 2002). The `Parser` and the `C-Module` are processing modules and both have reading (indicated by arrows with a triangle) and writing (indicated by normal arrows) capability to various modules such as `Potential Operators`, `Oracle`, `Choice Points`, `Chart` and the `C-Chart`. The `Parser` is only able to read from the `Grammar Rules`

[2]The components of the model are kept in a different font in order to better recognize them.

while the C-Module both reads from and writes to the Grammar Rules. In addition, the Parser alone is able to read from the Lexicon and read from and write to the Edge Labeling. In the next section, the components of the model will be briefly described before explaining the C-Module in more detail.

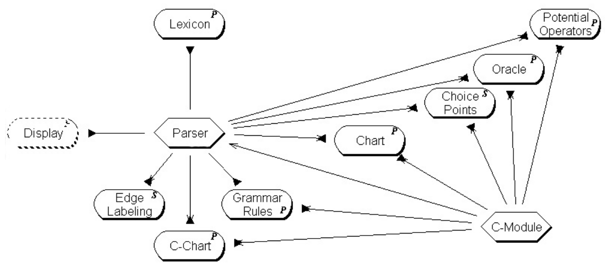

Figure 3: The structure of the system.

In the Lexicon and the Grammar Rules the lexicon and grammar rules are stored. The grammar rules consist of phrase structure rules that each exhibit a resting activation (RA). In the parser, the RAs are all set to a value of 10. To represent a lower frequency of object-verb-subject sentences and to process the sentences from the Knoeferle & Crocker (2006) study (see Evaluation section), the corresponding rule received a lower RA of 2. If the parser processes an ambiguous structure, the grammar rule with the higher RA is more likely to be applied. In the Potential Operators module, all possible parsing routes are stored. Possible parsing routes are: Inserting a new word into the parse tree (1), allocating a POS to a word (2), applying a grammar rule to a POS (3) and merging two edges with each other (4). The numbers (1) - (4) indicate the priority of each of the parsing routes. If more than on route could be applied the one with the higher priority wins the race. This approach is similar to the "race model" proposed by Traxler, Pickering, and Clifton (1998). The remaining routes are stored in the Choice Points module, for potential later access, in case a pursued route fails (see Cooper (2002) for a similar system). The parser uses a so-called oracle, that identifies a wrong parsing route early, without unnecessary backtracking, improving the performance of the parser. In the Chart the chart of the whole parse tree is stored and can be illustrated. The usage of the C-Chart is explained in the following section. Finally the Parser reads the linguistic information (i.e. words from the sentences) word-by-word from the Display.

We used a left-corner parser, because in comparison to the top-down and bottom-up parser it performs similar as the human parser concerning memory load while processing embedded sentences (Resnik, 1992). Furthermore, by using the arc-eager version of the left-corner parser we were able to model the preference of most languages to attach a new phrase at the last constructed one (Late-closure principle; Frazier, 1978).

The C-Module consists of four main processing steps. These steps are activated when the parser encounters the coordinating conjunction ("and").

Step 1) Add possible candidates to C-chart
Possible candidates are all phrases that might become the first constituent of a coordinate structure. For example in the sentence *The boy saw the dog and*, possible candidates would be *the dog* (NP), *saw the dog* (VP), *The boy saw the dog* (S).

Step 2) Select a candidate from the C-chart
Since the parser follows the late-closure principle (Frazier, 1978), this would be the lowest attached phrase (*the dog* in the example above).

Step 3) Increase the resting activation of the grammar rule, which built the selected candidate in the C-Chart
This procedure has the effect that when more than one grammar rule could be applied, the probability that the parser chooses the one applied earlier in coordination is increased.[3]

Step 4) IF the coordinate structure was completely explored
THEN delete all entries in the C-chart and set the resting activations of the grammar rules to their initial state
ELSE select another candidate in the C-chart and continue at Step 3.

As outlined above, the basic functionality of the model is that the probability of applying the grammar rule, which was applied for the first constituent of a coordinate structure to the second constituent rises when a conjunction is encountered. Thus, the probability that the parser initially applies the grammar rule which postulates a parallel structure, is higher than a rule postulating a non-parallel structure. This approach is not unlike Frazier et al.'s (2000) proposal that a parallel coordinate structure is more predictable and therefore, faster processed. The problem of the existing approache (Frazier and Clifton, 2001) to select the appropriate syntactic structure of the antecedent phrase in an ambiguous coordinate construction to copy it to the second constituent does not occur in the current approach. The RAs of the grammar rules of all possible phrases are boosted and due to the inherent properties of the arc-eager left-corner parser, the model selects independently from this procedure the lowest-attached phrase.

Evaluation

To evaluate the model, we used the sentences from the two experiments of the present paper, as well as sentences from the Knoeferle & Crocker (2006) study.

Processing steps in COGENT are counted in cycles. To evaluate the system, we directly compared these processing steps with the RPD fixation times of the experiments. Specifically, the RPDs of the whole second

[3] A reviewer correctly pointed out that the resting activation of the grammar rules of all candidates could be raised instead of only the one from the selected candidate. This would not change the results but probably speed up the performance of the parser. In a possible next version of the system this change will be implemented.

constituent are compared with the number of cycles the parser needed to parse this section of the sentences. In order to compare the two values and present them in the same charts, the numbers of cycles were multiplied by 27. This number seems to be a good approximation to align the processing cycles of COGENT with the RPD times of an eye-tracking study. Figure 4 illustrates the results of the evaluation.

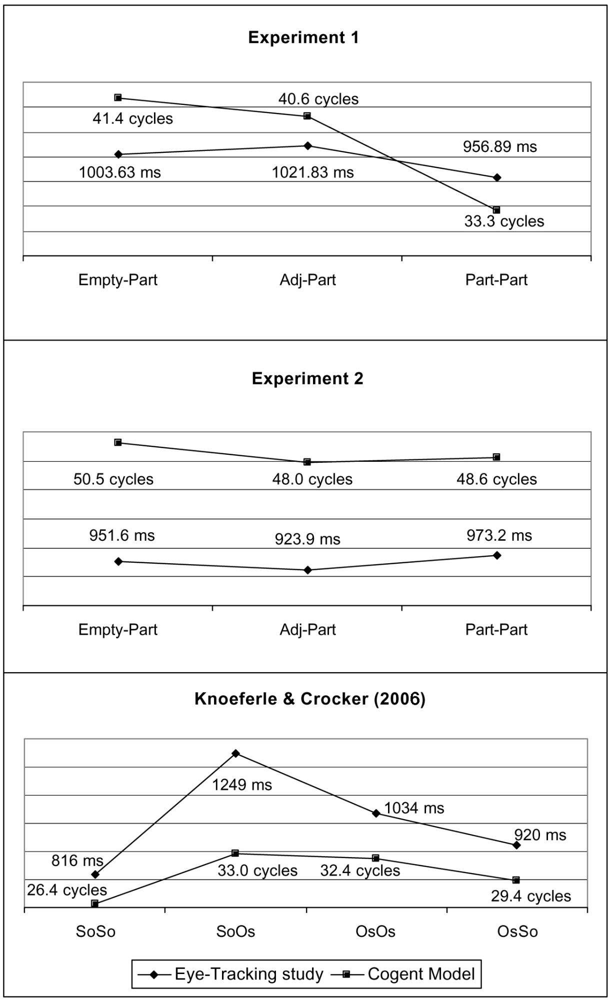

Figure 4: Evaluation of the model using results from Experiment 1 and 2 and the Knoeferle & Crocker (2006) study

As can be seen, the proportions of the cycles and the fixation duration times were approximately the same in all three experiments.

Conclusion

The results of the two experiments conducted in this study confirmed the findings by Frazier et al. (2000). The parallelism effect seems to be restricted to coordinate structures. Our findings strongly suggest that the origin of the parallelism effect is not a general priming effect but that it only applies in coordinate structures. Since we tested sentences that did not exhibit syntactic ambiguities we also conclude that the parallelism effect is not restricted to ambiguous structures. Furthermore, the absence of a parallelism effect between the Empty-Part and the Adj-Part conditions suggests that the parallelism effect depends crucially on the specific parts of speech being parallel in both constituents rather than just a parallel surface structure. Based on these findings, we constructed a computational model, which is able to account for the results found in the experiments of this thesis and the results of experiments conducted by Knoeferle and Crocker (2006). In this model, the parser increases the resting activation of grammar rules which where applied in the first constituent of a coordinate structure. When selecting grammar rules for the second constituent, the rules applied in the first constituent are preferred. In comparison with existing models, our mechanism further eschews the problem of finding the beginning of the antecedent constituent, and qualitatively fits the reading time behavior observed in three experiments.

Acknowledgments

This research was supported by a German Research Foundation grant (CR 135/1-2) awarded to MWC.

References

Cooper, R. P. (2002). *Model ling high-level cognitive processes.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies.*. Doctoral dissertation, University of Connecticut.

Frazier, L., & Clifton, C. (2001). Parsing coordination and ellipsis: Copy α. *Syntax*, *4(1)*, 1–22.

Frazier, L., Munn, A., & Clifton, C. (2000). Processing coordinate structures. *Journal of Psycholinguistic Research*, *29*, 345–370.

Frazier, L., Taft, L., Roeper, T., & Clifton, C. (1984). Parallel structure: A source of facilitation in sentence comprehension. *Memory & Cognition*, *12(5)*, 421–430.

Knoeferle, P., & Crocker, M. (2006). The mechanism(s) underlying parallelism: eye tracking coordinate structures in german. *Proceedings of the CUNY Sentence Processing Conference*, CUNY, New York City.

Resnik, P. (1992). Left-corner parsing and psychological plausibility. *In Proceedings of the 14th Conference on Computational Linguistics - Volume 1*, Nantes, France.

Steiner, I. (2003). Parsing syntactic redundancies in coordinate structures. *Proceedings of EuroCogSci 03 The European Cognitive Science Conference 2003* (pp. 443). Osnabrueck, Germany: Lawrence Erlbaum.

Traxler, M., Pickering, M., & Clifton, J., C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, *39*, 558–592.