Parallelism in Coordination as an Instance of Syntactic Priming: Evidence from Corpus-based Modeling

Amit Dubey and **Patrick Sturt** and **Frank Keller**

Human Communication Research Centre, Universities of Edinburgh and Glasgow

2 Buccleuch Place, Edinburgh EH8 9LW, UK

{adubey,sturt,keller}@inf.ed.ac.uk

Abstract

Experimental research in psycholinguistics has demonstrated a parallelism effect in coordination: speakers are faster at processing the second conjunct of a coordinate structure if it has the same internal structure as the first conjunct. We show that this phenomenon can be explained by the prevalence of parallel structures in corpus data. We demonstrate that parallelism is not limited to coordination, but also applies to arbitrary syntactic configurations, and even to documents. This indicates that the parallelism effect is an instance of a general syntactic priming mechanism in human language processing.

1 Introduction

Experimental work in psycholinguistics has provided evidence for the so-called parallelism preference effect: speakers processes coordinated structures more quickly when the two conjuncts have the same internal syntactic structure. The processing advantage for parallel structures has been demonstrated for a range coordinate constructions, including NP coordination (Frazier et al., 2000), sentence coordination (Frazier et al., 1984), and gapping and ellipsis (Carlson, 2002; Mauner et al., 1995).

The parallelism preference in NP coordination can be illustrated using Frazier et al.'s (2000) Experiment 3, which recorded subjects' eye-movements while they read sentences like (1):

(1) a. Terry wrote a long novel and <u>a short poem</u> during her sabbatical.
     b. Terry wrote a novel and <u>a short poem</u> during her sabbatical

Total reading times for the underlined region were faster in (1-a), where *short poem* is coordinated with a syntactically parallel noun phrase (*a long novel*), compared to (1-b), where it is coordinated with a syntactically non-parallel phrase.

These results raise an important question that the present paper tries to answer through corpus-based modeling studies: what is the mechanism underlying the parallelism preference? One hypothesis is that the effect is caused by low-level processes such as *syntactic priming*, i.e., the tendency to repeat syntactic structures (e.g., Bock, 1986). Priming is a very general mechanism that can affect a wide range of linguistic units, including words, constituents, and semantic concepts. If the parallelism effect is an instance of syntactic priming, then we expect it to apply to a wide range of syntactic construction, and both within and between sentences. Previous work has demonstrated priming effects in corpora (Gries, 2005; Szmrecsanyi, 2005); however, these results are limited to instances of priming that involve a choice between two structural alternatives (e.g., dative alternation). In order to study the parallelism effect, we need to model priming as general syntactic repetition (independent of the structural choices available). This is what the present paper attempts.

Frazier and Clifton (2001) propose an alternative account of the parallelism effect in terms of a *copying mechanism*. Unlike priming, this mechanism is highly specialized and only applies to coordinate structures: if the second conjunct is encountered, then instead of building new structure, the language processor simply copies the structure of the first conjunct; this explains why a speed-up is observed if the second conjunct is parallel to the first one. If the copying account is correct, then we would expect parallelism effects to be restricted to coordinate structures and would not apply in other contexts.

In the present paper, we present corpus evidence that allows us to distinguish between these two competing explanations. Our investigation will proceed as follows: we first establish that there is evidence

for a parallelism effect in corpus data (Section 3). This is a crucial prerequisite for our wider investigation: previous work has only dealt with parallelism in comprehension, hence we need to establish that parallelism is also present in production data, such as corpus data. We then investigate whether the parallelism effect is restricted to coordination, or whether it also applies also arbitrary syntactic configurations. We also test if parallelism can be found for larger segments of text, including, in the limit, the whole document (Section 4). Then we investigate parallelism in dialog, testing the psycholinguistic prediction that parallelism in dialog occurs between speakers (Section 5). In the next section, we discuss a number of methodological issues and explain the way we measure parallelism in corpus data.

2 Adaptation

Psycholinguistic studies have shown that priming affects both speech production (Bock, 1986) and comprehension (Branigan et al., 2005). The importance of comprehension priming has also been noted by the speech recognition community (Kuhn and de Mori, 1990), who use so-called caching language models to improve the performance of speech comprehension software. The concept of caching language models is quite simple: a cache of recently seen words is maintained, and the probability of words in the cache is higher than those outside the cache.

While the performance of caching language models is judged by their success in improving speech recognition accuracy, it is also possible to use an abstract measure to diagnose their efficacy more closely. Church (2000) introduces such a diagnostic for lexical priming: adaptation probabilities. Adaptation probabilities provide a method to separate the general problem of priming from a particular implementation (i.e., caching models). They measure the amount of priming that occurs for a given construction, and therefore provide an upper limit for the performance of models such as caching models.

Adaptation is based upon three concepts. First is the *prior*, which serves as a baseline. The prior measures the probability of a word appearing, ignoring the presence or absence of a prime. Second is the *positive adaptation*, which is the probability of a word appearing given that it has been primed. Third is the *negative adaptation*, the probability of a word

appearing given it has not been primed.

In Church's case, the prior and adaptation probabilities are estimated as follows. If a corpus is divided into individual documents, then each document is then split in half. We refer to the halves as the prime set (or prime half) and the target set (or target half).[1] We measure how frequently a document half contains a particular word. For each word w, there are four combinations of the prime and target halves containing the word. This gives us four frequencies to measure, which are summarized in the following table:

$f_{w_{p,t}}$	$f_{w_{\bar{p},t}}$
$f_{w_{p,\bar{t}}}$	$f_{w_{\bar{p},\bar{t}}}$

These frequencies represent:

$$
\begin{aligned}
f_{w_{p,t}} &= \text{\# of times } w \text{ occurs in prime set} \\
& \quad \text{and target set} \\
f_{w_{\bar{p},t}} &= \text{\# of times } w \text{ occurs in target set} \\
& \quad \text{but not prime set} \\
f_{w_{p,\bar{t}}} &= \text{\# of times } w \text{ occurs in prime set} \\
& \quad \text{but not target set} \\
f_{w_{\bar{p},\bar{t}}} &= \text{\# of times } w \text{ does not occur in either} \\
& \quad \text{target set or prime set}
\end{aligned}
$$

In addition, let N represent the sum of these four frequencies. From the frequencies, we may formally define the prior, positive adaptation and negative adaptation:

(1) **Prior** $\quad P_{\text{prior}}(w) = \dfrac{f_{w_{p,t}} + f_{w_{\bar{p},t}}}{N}$

(2) **Positive Adaptation** $\quad P_+(w) = \dfrac{f_{w_{p,t}}}{f_{w_{p,t}} + f_{w_{p,\bar{t}}}}$

(3) **Negative Adaptation** $\quad P_-(w) = \dfrac{f_{w_{\bar{p},t}}}{f_{w_{\bar{p},t}} + f_{w_{\bar{p},\bar{t}}}}$

In the case of lexical priming, Church observes that $P_+ \gg P_{\text{prior}} > P_-$. In fact, even in cases when P_{prior} quite small, P_+ may be higher than 0.8. Intuitively, a positive adaptation which is higher than the prior entails that a word is likely to reappear in the target set given that it has already appeared in the prime set. We intend to show that adaptation probabilities provide evidence that syntactic constructions behave

---

[1] Our terminology differs from that of Church, who uses 'history' to describe the first half, and 'test' to describe the second. Our terms avoid the ambiguity of the phrase 'test set' and coincide with the common usage in the psycholinguistic literature.

similarity to lexical priming, showing positive adaptation P_+ greater than the prior. As P_- must become smaller than P_{prior} whenever P_+ is larger than P_{prior}, we only report the positive adaptation P_+ and the prior P_{prior}.

While Church's technique was developed with speech recognition in mind, we will show that it is useful for investigating psycholinguistic phenomenon. However, the connection between cognitive phenomenon and engineering approaches go in both directions: it is possible that syntactic parsers could be improved using a model of syntactic priming, just as speech recognition has been improved using models of lexical priming.

3 Experiment 1: Parallelism in Coordination

In this section, we investigate the use of Church's adaptation metrics to measure the effect of syntactic parallelism in coordinated constructions. For the sake of comparison, we restrict our study to several constructions used in Frazier et al. (2000). All of these constructions occur in NPs with two coordinate sisters, i.e., constructions such as NP$_1$ CC NP$_2$, where CC represents a coordinator such as *and*.

3.1 Method

The application of the adaptation metric is straightforward: we pick NP$_1$ as the prime set and NP$_2$ as the target set. Instead of measuring the frequency of lexical elements, we measure the frequency of the following syntactic constructions:

SBAR An NP with a relative clause, i.e., NP → NP SBAR.

PP An NP with a PP modifier, i.e., NP → NP PP.

NN An NP with a single noun, i.e., NP → NN.

DT NN An NP with a determiner and a noun, i.e., NP → DT NN.

DT JJ NN An NP with a determiner, an adjective and a noun, i.e., NP → DT JJ NN.

Parameter estimation is accomplished by iterating through the corpus for applications of the rule NP → NP CC NP. From each rule application, we create a list of prime-target pairs. We then estimate adaptation probabilities for each construction, by counting the number of prime-target pairs in which the
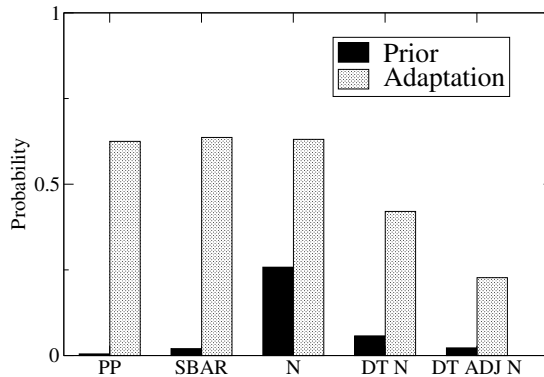

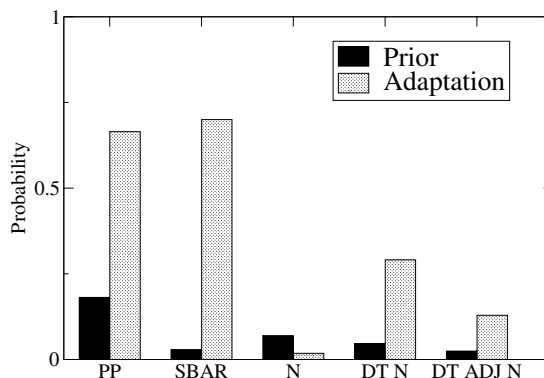Figure 1: Adaptation within coordinate structures in the Brown corpus

Figure 2: Adaptation within coordinate structures in the WSJ corpus

construction does or does not occur. This is done similarly to the document half case described above. There are four frequencies of interest, but now they refer to the frequency that a particular construction (rather than a word) either occurs or does not occur in the prime and target set.

To ensure results were general across genres, we used all three parts of the English Penn Treebank: the Wall Street Journal (WSJ), the balanced Brown corpus of written text (Brown) and the Switchboard corpus of spontaneous dialog. In each case, we use the entire corpus.

Therefore, in total, we report 30 probabilities: the prior and positive adaptation for each of the five constructions in each of the three corpora. The primary objective is to observe the difference between the prior and positive adaptation for a given construction in a particular corpus. Therefore, we also perform a χ^2 test to determine if the difference between these two probabilities are statistically significant.
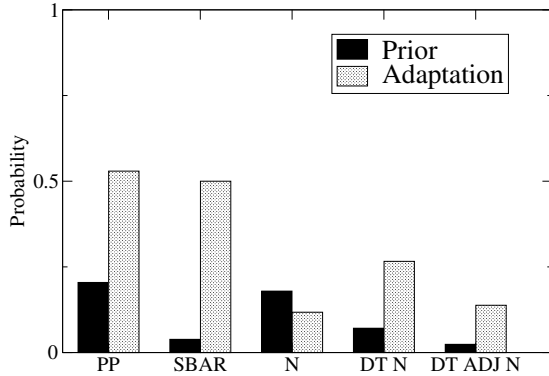
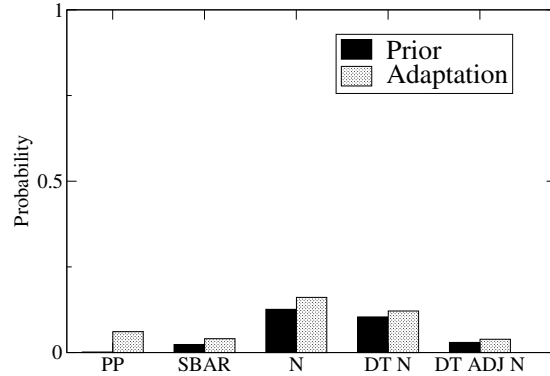Figure 3: Adaptation within coordinate structures in the Switchboard corpus

Figure 4: Adaptation within sentences in the Brown corpus

Figure 5: Adaptation within sentences in the WSJ corpus

3.2 Results

The results are shown in Figure 1 for the Brown corpus, Figure 2 for the WSJ and Figure 3 for Switchboard. Each figure shows the prior and positive adaptation for all five constructions: relative clauses (SBAR) a PP modifier (PP), a single common noun (N), a determiner and noun (DT N), and a determiner adjective and noun (DT ADJ N). Only in the case of a single common noun in the WSJ and Switchboard corpora is the prior probability higher than the positive adaptation. In all other cases, the probability of the given construction is more likely to occur in NP_2 given that it has occurred in NP_1. According to the χ^2 tests, all differences between priors and positive adaptations were significant at the 0.01 level. The size of the data sets means that even small differences in probability are statistically significant. All differences reported in the remainder of this paper are statistically significant; we omit the details of individual χ^2 tests.

3.3 Discussion

The main conclusion we draw is that the parallelism effect in corpora mirrors the ones found experimentally by Frazier et al. (2000), if we assume higher probabilities are correlated with easier human processing. This conclusion is important, as the experiments of Frazier et al. (2000) only provided evidence for parallelism in *comprehension* data. Corpus data, however, are *production* data, which means that the our findings are first ones to demonstrate parallelism effects in production.

The question of the relationship between comprehension and production data is an interesting one.
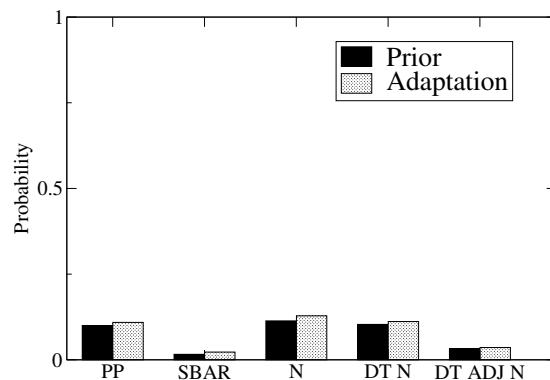
We can expect that production data, such as corpus data, are generated by speakers through a process that involves self-monitoring. Written texts (such as the WSJ and Brown) involve proofreading and editing, i.e., explicit comprehension processes. Even the data in a spontaneous speech corpus such as Swtichboard can be expected to involve a certain amount of self-monitoring (speakers listen to themselves and correct themselves if necessary). It follows that it is not entirely unexpected that similar effects can be found in both comprehension and production data.

4 Experiment 2: Parallelism in Documents

The results in the previous section showed that the parallelism effect, which so far had only been demonstrated in comprehension studies, is also attested in corpora, i.e., in production data. In the present experiment, we will investigate the mechanisms underlying the parallelism effect. As discussed in Section 1, there are two possible explana-
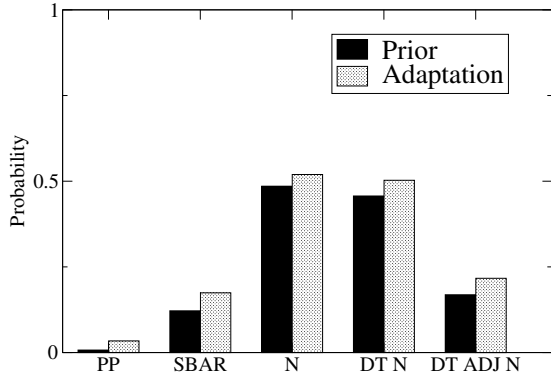
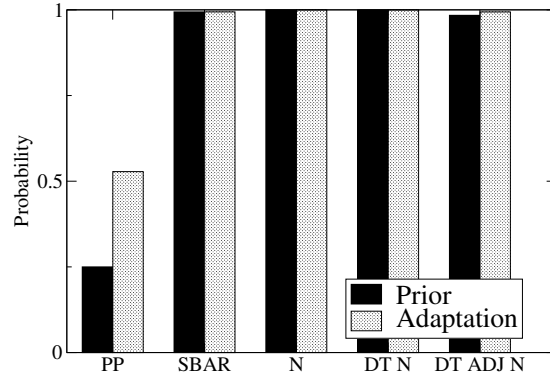Figure 6: Adaptation between sentences in the Brown corpus

Figure 8: Adaptation within documents in the Brown corpus (all items exhibit weak yet statistically significant positive adaptation)
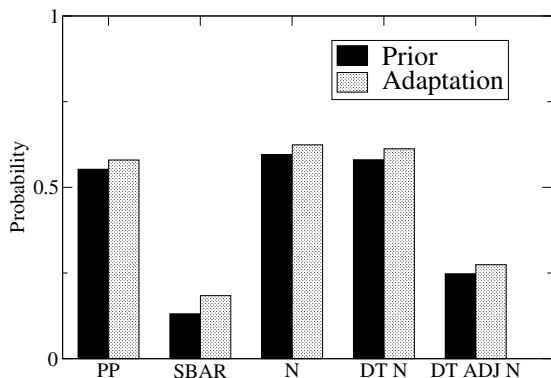

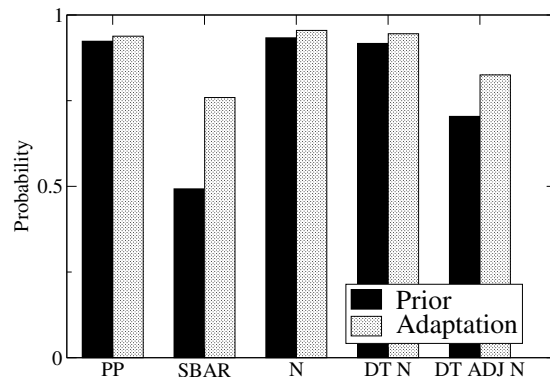Figure 7: Adaptation between sentences in the WSJ corpus

Figure 9: Adaptation within documents in the WSJ corpus

tion for the effect: one in terms of a construction-specific copying mechanism, and one in terms of a generalized syntactic priming mechanism. In the first case, we predict that the parallelism effect is restricted to coordinate structures, while in the second case, we expect that parallelism (a) is independent of coordination, and (b) occurs in the wider discourse, i.e., not only within sentences but also between sentences.

4.1 Method

The method used was the same as in Experiment 1 (see Section 3.1), with the exception that the prime set and the target set are no longer restricted to being the first and second conjunct in a coordinate structure. We investigated three levels of granularity: within sentences, between sentences, and within documents. *Within-sentence parallelism* occurs when the prime NP and the target NP occur within the same sentence, but stand in an ar-

bitrary structural relationship. Coordinate NPs were excluded from this analysis, so as to make sure that any within-sentence parallelism is not confounded coordination parallelism as established in Experiment 1. *Between-sentence parallelism* was measured by regarding as the target the sentence immediately following the prime sentence. In order to investigate *within-document parallelism*, we split the documents into equal-sized halves; then the adaptation probability was computed by regarding the first half as the prime and the second half as the target (this method is the same as Church's method for measuring lexical adaptation).

The analyses were conducted using the Wall Street Journal and the Brown portion of the Penn Treebank. The document boundary was taken to be the file boundary in these corpora. The Switchboard corpus is a dialog corpus, and therefore needs to be treated differently: turns between speakers rather

than sentences should be level of analysis. We will investigate this separately in Experiment 3 below.

4.2 Results

The results for the within-sentence analysis are graphed in Figures 4 and 5 for the Brown and WSJ corpus, respectively. We find that there is a parallelism effect in both corpora, for all the NP types investigated. Figures 6–9 show that the same is true also for the between-sentence and within-document analysis: parallelism effects are obtained for all NP types and for both corpora, even it the parallel structures occur in different sentences or in different document halves. (The within-document probabilities for the Brown corpus (in Figure 8) are close to one in most cases; the differences between the prior and adaptation are nevertheless significant.)

In general, note that the parallelism effects uncovered in this experiment are smaller than the effect demonstrated in Experiment 1: The differences between the prior probabilities and the adaptation probabilities (while significant) are markedly smaller than those uncovered for parallelism in coordinate structure.[2]

4.3 Discussion

This experiment demonstrated that the parallelism effect is not restricted to coordinate structures. Rather, we found that it holds across the board: for NPs that occur in the same sentence (and are not part of a coordinate structure), for NPs that occur in adjacent sentences, and for NPs that occur in different document halves. The between-sentence effect has been demonstrated in a more restricted from by Gries (2005) and Szmrecsanyi (2005), who investigate priming in corpora for cases of structural choice (e.g., between a dative object and a PP object for verbs like *give*). The present results extend this finding to arbitrary NPs, both within and between sentences.

The fact that parallelism is a pervasive phenomenon, rather than being limited to coordinate structures, strongly suggests that it is an instance of a general syntactic priming mechanism, which has been an established feature of accounts of the human sentence production system for a while (e.g., Bock,

---

[2]The differences between the priors and adaptation probabilities are also much smaller than noted by Church (2000). The probabilities of the rules we investigate have a higher marginal probability than the lexical items of interest to Church.

1986). This runs counter to the claims made by Frazier et al. (2000) and Frazier and Clifton (2001), who have argued that parallelism only occurs in coordinate structures, and should be accounted for using a specialized copying mechanism. (It is important to bear in mind, however, that Frazier et al. only make explicit claims about comprehension, not about production.)

However, we also found that parallelism effects are clearly strongest in coordinate structures (compare the differences between prior and adaptation in Figures 1–3 with those in Figures 4–9). This could explain why Frazier et al.'s (2000) experiments failed to find a significant parallelism effect in non-coordinated structures: the effect is simply too week to detect (especially using the self-paced reading paradigm they employed).

5 Experiment 3: Parallelism in Spontaneous Dialog

Experiment 1 showed that parallelism effects can be found not only in written corpora, but also in the Switchboard corpus of spontaneous dialog. We did not include Switchboard in our analysis in Experiment 2, as this corpus has a different structure from the two text corpora we investigated: it is organized in terms of turns between two speakers. Here, we exploit this property and conduct a further experiment in which we compare parallelism effects between speakers and within speakers.

The phenomenon of structural repetition between speakers has been discussed in the experimental psycholinguistic literature (see Pickering and Garrod 2004 for a review). According to Pickering and Garrod (2004), the act of engaging in a dialog facilitates the use of similar representations at all linguistic levels, and these representations are shared between speech production and comprehension processes. Thus structural adaptation should be observed in a dialog setting, both within and between speakers. An alternative view is that production and comprehension processes are distinct. Bock and Loebell (1990) suggest that syntactic priming in speech production is due to facilitation of the retrieval and assembly procedures that occur during the formulation of utterances. Bock and Loebell point out that this production-based procedural view predicts a lack of priming between comprehension and production or vice versa, on the assumption that
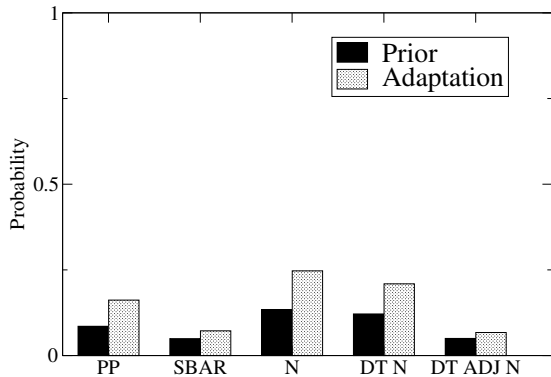
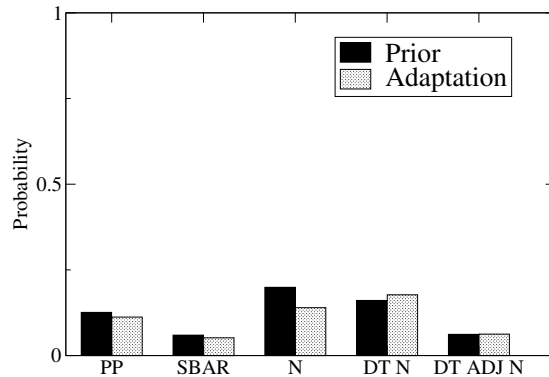Figure 10: Adaptation between speakers in the Switchboard corpus

Figure 11: Adaptation within speakers in the Switchboard corpus

production and parsing use distinct mechanisms. In our terms, it predicts that between-speaker positive adaptation should not be found, because it can only result from priming from comprehension to production, or vice versa. Conversely, the prodedural view outlined by Bock and Loebell predicts that positive adaptation *should* be found within a given speaker's dialog turns, because such adaptation can indeed be the result of the facilitation of production routines within a given speaker.

5.1 Method

We created two sets of prime and target data to test within-speaker and between-speaker adaptation. The prime and target sets were defined in terms of pairs of utterances. To test between-speaker adaptation, we took each adjacent pair of utterances spoken by speaker A and speaker B, in each dialog, and these were treated as prime and target sets respectively. In the within-speaker analysis, the prime and target sets were taken from the dialog turns of only one speaker—we took each adjacent pair of dialog turns uttered by a given speaker, excluding the intervening utterance of the other speaker. The earlier utterance of the pair was treated as the prime, and the later utterance as the target. The remainder of the method was the same as in Experiments 1 and 2 (see Section 3.1).

5.2 Results

The results for the between-speaker and within-speaker adaptation are shown in Figure 10 and Figure 11 for same five phrase types as in the previous experiments.

A positive adaptation effect can be seen in the between-speaker data. For each phrase type, the adaptation probability is greater than the prior. In the within-speaker data, by comparison, the magnitude of the adaptation advantage is greatly decreased, in comparison with Figure 10. Indeed, for most phrase types, the adaptation probability is lower than the prior, i.e., we have a case of negative adaptation.

5.3 Discussion

The results of the two analyses confirm that adaptation can indeed be found between speakers in dialog, supporting the results of experimental work reviewed by Pickering and Garrod (2004). The results do not support the notion that priming is due to the facilitation of production processes within a given speaker, an account which would have predicted adaptation within speakers, but not between speakers.

The lack of clear positive adaptation effects in the within-speaker data is harder to explain—all current theories of priming would predict some effect here. One possibility is that such effects may have been obscured by decay processes: doing a within-speaker analysis entails skipping an intervening turn, in which priming effects were lost. We intend to address these concerns using more elaborate experimental designs in future work.

6 Conclusions

In this paper, we have demonstrated a robust, pervasive effect of parallelism for noun phrases. We found the tendency for structural repetition in two different corpora of written English, and also in a dialog cor-

pus. The effect occurs in a wide range of contexts: within coordinate structures (Experiment 1), within sentences for NPs in an arbitrary structural configuration, between sentences, and within documents (Experiment 2). This strongly indicates that the parallelism effect is an instance of a general processing mechanism, such as syntactic priming (Bock, 1986), rather than specific to coordination, as suggested by (Frazier and Clifton, 2001). However, we also found that the parallelism effect is strongest in co-ordinate structures, which could explain why comprehension experiments so far failed to demonstrate the effect for other structural configurations (Frazier et al., 2000). We leave it to future work to explain *why* adaptation is much stronger in co-ordination: is co-ordination special because of extra constrains (i.e., some kind of expected contrast/comparison between co-ordinate sisters) or because of fewer constraints (i.e., both co-ordinate sisters have a similar grammatical role in the sentence)?

Another result (Experiment 3) is that the parallelism effect occurs between speakers in dialog. This finding is compatible with Pickering and Garrod's (2004) interactive alignment model, and strengthens the argument for parallelism as an instance of a general priming mechanism.

Previous experimental work has found parallelism effects, but only in comprehension data. The present work demonstrates that parallelism effects also occur in production data, which raises an interesting question of the relationship between the two data types. It has been hypothesized that the human language processing system is tuned to mirror the probability distributions in its environment, including the probabilities of syntactic structures (Mitchell et al., 1996). If this *tuning hypothesis* is correct, then the parallelism effect in comprehension data can be explained as an adaptation of the human parser to the prevalence of parallel structures in its environment (as approximated by corpus data) that we demonstrated in this paper.

Note that the results in this paper not only have an impact on theoretical issues regarding human sentence processing, but also on engineering problems in natural language processing, e.g., in probabilistic parsing. To avoid sparse data problems, probabilistic parsing models make strong independence assumptions; in particular, they generally assume that sentences are independent of each other. This is partly due to the fact it is difficult to parameterize the many possible dependencies which may occur between adjacent sentences. However, in this paper, we show that structure re-use is one possible way in which the independence assumption is broken. A simple and principled approach to handling structure re-use would be to use adaptation probabilities for probabilistic grammar rules, analogous to cache probabilities used in caching language models (Kuhn and de Mori, 1990). We are currently conducting further experiments to investigate of the effect of syntactic priming on probabilistic parsing.

References

Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18:355–387.

Bock, Kathryn and Helga Loebell. 1990. Framing sentences. *Cognition* 35(1):1–39.

Branigan, Holly P., Marin J. Pickering, and Janet F. McLean. 2005. Priming prepositional-phrase attachment during comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition* 31(3):468–481.

Carlson, Katy. 2002. The effects of parallelism and prosody on the processing of gapping structures. *Language and Speech* 44(1):1–26.

Church, Kenneth W. 2000. Empirical estimates of adaptation: the chance of two Noriegas is closer to $p/2$ than p^2. In *Proceedings of the 17th Conference on Computational Linguistics*. Saarbrücken, Germany, pages 180–186.

Frazier, Lyn, Alan Munn, and Chuck Clifton. 2000. Processing coordinate structures. *Journal of Psycholinguistic Research* 29(4):343–370.

Frazier, Lyn, Lori Taft, Tom Roeper, Charles Clifton, and Kate Ehrlich. 1984. Parallel structure: A source of facilitation in sentence comprehension. *Memory and Cognition* 12(5):421–430.

Frazier, Lynn and Charles Clifton. 2001. Parsing coordinates and ellipsis: Copy α. *Syntax* 4(1):1–22.

Gries, Stefan T. 2005. Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research* 35.

Kuhn, Roland and Renate de Mori. 1990. A cache-based natural language model for speech recognition. *IEEE Transanctions on Pattern Analysis and Machine Intelligence* 12(6):570–583.

Mauner, Gail, Michael K. Tanenhaus, and Greg Carlson. 1995. A note on parallelism effects in processing deep and surface verb-phrase anaphors. *Language and Cognitive Processes* 10:1–12.

Mitchell, Don C., Fernando Cuetos, Martin M. B. Corley, and Marc Brysbaert. 1996. Exposure-based models of human parsing: Evidence for the use of coarse-grained (non-lexical) statistical records. *Journal of Psycholinguistic Research* 24(6):469–488.

Pickering, Martin J. and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27(2):169–225.

Szmrecsanyi, Benedikt. 2005. Creatures of habit: A corpus-linguistic analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1(1):113–149.