Measuring Syntactic Priming in Dialogue Corpora

Christian Pietsch, Armin Buch, Stefan Kopp and Jan de Ruiter

1. Introduction

1.1. What is syntactic priming, and why do we care whether it exists?

1.1.1. Priming

Priming, as studied by psychologists, can be defined as the effect that one occurrence of a stimulus (the prime) influences the processing of a subsequent-stimulus (the target) (paraphrased after Tulving and Schacter, 1990, p. 301). Stimuli can have external (other-priming) or internal (self-priming) origin. The influence on processing can be inhibitory or facilitative. Inhibitory effects have seldom been reported, and we will not discuss them here. The mention of subsequent implies a temporal distance between prime and target which can be measured in different ways, e.g. time or intervening items. In this study, we measure distance in seconds.

1.1.2. Linguistic priming

Priming effects have been assumed (and observed) for several traditional levels of linguistic description, including lexis and syntax. Potentially, these effects offer a way to investigate linguistic units (and their status as such): Building blocks of a linguistic theory that claims cognitive adequacy can be expected to exhibit priming.

We concentrate on lexical priming and syntactic priming\(^1\). Lexical priming can be observed more directly, and therefore it is far better attested than

---

1. Terminological note: We choose to use the relatively established term “syntactic priming” although some (Bock, 1986; Szmrecsanyi, 2005) argue that “syntactic persistence” would be more adequate. Another term, “structural priming”, is sometimes used interchangeably but could also be used outside syntax and even outside linguistics for entirely different phenomena (Pickering and Ferreira, 2008, pp. 427–428).
syntactic priming (see section 1.1.3). An apparent syntactic similarity might be explained solely by a sharing of similar lexical items. This would then not be a genuine case of syntactic priming, but of lexical priming (an insight which might be more obvious for those working within a lexicalist framework, where each lexical head determines the constructions it occurs in).

1.1.3. Syntactic priming

Syntactic priming has been defined succinctly and informally as “the tendency to reuse syntactic constructions” (Gries, 2005), or more generally “as the proposal that processing a particular syntactic structure within a sentence affects the processing of the same (or a related) syntactic structure within a subsequently presented sentence” (Branigan et al., 1995).

In a landmark experiment published in 1986 (Bock, 1986), syntactic production–production priming (a form of syntactic self-priming) was first established and immediately interpreted as evidence for the autonomy of syntax and against the functionalist view of syntax (as held e.g. by Bates and MacWhinney, 1982) which considers syntactic knowledge only a derivative of semantic or superficial properties of utterances. Since then, many psycholinguistic theories of speech production have attempted to take syntactic (self-)priming into account (Pickering and Ferreira, 2008). More recently, syntactic priming results have also been used to judge the comparative merits of different grammar formalisms (Reitter et al., 2006a).

In contrast, evidence for syntactic other-priming has emerged only since 1998 (Potter and Lombardi, 1998; Branigan et al., 2000), and mainly from lab experiments. This evidence suggests that comprehension and production processes operate on the same (or closely connected) representations. It has also been used to argue for the assumption that priming drives dialogue (Pickering and Garrod, 2004). However, the corpus studies on syntactic priming we are aware of paint an inconsistent picture, especially with respect to comprehension–production priming. We mention some of them in section 1.2 below.

As for the cause of priming effects, there are two prominent explanations. Many believe that priming effects are caused by transient activation spreading in the neural networks of the brain (Branigan et al., 2000; Pickering and Garrod, 2004; Jaeger and Snider, 2008). Others see the cause of priming effects in implicit learning (Bock and Griffin, 2000; Chang et al., 2006). The spreading-activation account corresponds to short-lived priming effects, whereas effects of implicit learning are longer-lived. Recent experimental evidence seems to suggest that lexical priming is short-lived while syntactic priming is persistent (Hartsuiker et al., 2008). Alternative explanations for syntactic priming include...
social mimicry (Balcetis and Dale, 2005), and the tendency to reduce process-
ing costs (discussed by Smith and Wheeldon, 2000, p. 127).

In the dialogue corpora we use, the longest time span over which priming
effects could be observed is the length of a single dialogue. Therefore we do not
have access to long-term effects, and will hence be investigating the short-term
effects of syntactic priming.

1.1.4. Motivation
Our own interest in syntactic priming is driven by the following long-term
research questions:

(1) Can syntactic priming be used to identify the cognitively adequate units
of syntactic structure?

These units, known as syntactic exemplars (Bybee, 2006)\(^2\), \(^3\), are conjectured to
span one or more words. As we outlined above, it is desirable for a theory to use
empirically verified building blocks. Results could also be useful for computer-
implemented natural language processing (Dubey et al., 2006).

(2) Is there syntactic priming in natural dialogue?

Below we will present some reasons for being skeptical. In this connection,
we are primarily interested in other-priming, especially comprehension–pro-
duction priming, as this is widely believed to cause syntactic alignment, a key
ingredient of the influential Interactive Alignment Model (IAM) of dialogue
(Pickering and Garrod, 2004).

In a nutshell, the IAM assumes the existence of automatic alignment be-
tween the representations in both participants of a dialogue, and at several
levels of representation, including the lexical and the syntactic one. This is
assumed to lead to situation-model alignment and hence, successful dialogue.
The IAM further postulates that at each level, alignment is caused by a “primi-

\(^2\) The term *exemplar* refers to the actual (and repeated) occurrences of a type.
\(^3\) We would like to list some terms we consider related because a theory of syn-
tax with an adequate notion of “basic unit” could provide a unified treatment for
them: “multi-word expressions” (MWE), “multi-word unit” (MWU), “prefab”,
“word cooccurrence”, “colli-
gation”, “collocation”, “collostruction”, “collexeme”,
“construction”, “idiom”, “phrasal verb”, “formulaic sequences”, “chunks”, “phra-
seological expression”, and more. Ac-
counting for these phenomena is especially
important in theories which try to model performance phenomena as well as “pure”
linguistic competence.
tive and resource-free” priming mechanism, and that alignment at one level “percolates” up to the next level. For obvious reasons, self-priming does not come into consideration as a driver of dialogue.

The study presented here can only address question 1 as it conflates self-priming and other-priming. We will address question 2 in further research.

1.2. Why is syntactic priming controversial?

The existence of syntactic priming is uncontroversial in lab experiments (Pickering and Ferreira, 2008). However, lab experiments cannot provide reliable information about natural dialogue because they are conducted in tightly controlled environments. Experimenters strive to keep as many factors as possible under control so that they can claim a causal relationship between the factor they vary and the outcome. In language experiments, this often leads to situations where subjects are influenced by confederates or instructions to restrict their behaviour because this makes the evaluation easier. This lowers the ecological validity, which is avoided in corpus studies.

In the context of this research, an additional advantage of corpus studies lies in the possibility of investigating not only a few carefully selected constructions but all constructions/exemplars/rules that can be extracted (Reiter et al., 2006b; Reitter, 2008; Healey et al., 2010a,b). Nevertheless, it is still the case that most corpus studies have limited themselves to specific constructions (e. g., Gries, 2005; Szmrecsanyi, 2005; Dubey et al., 2005; Jaeger and Snider, 2008; Howes et al., 2010).

The results of corpus studies on syntactic priming have been inconsistent, especially those which investigate all constructions in a given corpus: Reitter (et al. 2006b; 2008) reported significant effects for self-priming as well as other-priming in the MapTask4 corpus. In the Switchboard5 corpus, he found self-priming but no other-priming when using an utterance-based distance measure. With a time-based distance measure, he did find other-priming. Overall, David Reitter and his colleagues consider their findings compatible with the IAM (which, as we outlined in section 1.1.4, hinges on the existence of other-priming). This is in stark contrast with the results of Patrick Healey and his colleagues (Healey et al., 2010a,b) who found evidence for structural divergence: Structural repetition across adjacent turns in natural dialogue (as documented

in a different set of corpora: the DCPSE\textsuperscript{6} and the BNC\textsuperscript{7}) was below chance level.

2. Methods and data

2.1. Hypothesis

Different linguistic theories propose different structures in the description of natural language. If these structures correspond to mental representations, then they can (and probably should, by the general and subconscious nature of priming) be primed. So finding priming effects for such a structure offers support to the linguistic theory which proposed it. In other words: “[R]epeatable structures are evidence for the units of linguistic cognition” (Reitter, 2008, sec. 1.2).

2.2. Experiments

2.2.1. Preliminaries

Classical priming experiments such as Bock and Griffin (2000) study a single, theory-neutral alternation in controlled experiments. In contrast, we study the distribution of each category in large annotated syntactic corpora (treebanks). Every sub-structure of an annotation is a possible category. We follow Reitter et al. (2006a,b) in studying Combinatory Categorial Grammar (CCG) categories and context-free production rules. In future work, it would be instructive to extend the latter annotation to include subtrees of arbitrary depth as in data-oriented parsing (Bod, 1998).

CCG assumes that there are many equivalent derivations for a given sentence analysis: the same lexical categories, but different modes of combination. Among these, the normal form derivation is the one along the lines of constituent bracketing, which is mostly right-branching for languages like English. The incremental derivation is as left-branching as possible; see Reitter et al. (2006a) for details.

We use the same data as Reitter et al. (2006a,b): The Switchboard corpus (Godfrey and Holliman, 1997) augmented with timing information (Calhoun


\textsuperscript{7} BNC: British National Corpus (Burnard, 2007).
et al., 2010), and annotated with either context-free rule expansions (sw-CFG) (Marcus et al., 1999) or CCG categories (sw-CCG-I and sw-CCG-N for incremental and normal form derivations, respectively (Hockenmaier and Steedman, 2007); and the MapTask corpus (Anderson et al., 1991) with CFG annotation (mp-CFG). We also look at lexical priming (sw-words).

Switchboard is a corpus of telephone conversations on loosely defined topics, whereas MapTask (as the name suggests) contains dialogues in which an instructor has to communicate a path on a map to a follower in a cooperative task.

2.2.2. A simple measure of priming

Newman et al. (2009) directly measure syntactic priming as a reduced reaction time in a brain region related to syntactic processing, facilitated by lexical priming of verbs. This effect is not particular to single syntactic categories or words, so priming as mental activation of representations cannot be measured directly (yet). In corpus studies, we observe the distribution of a category. The null hypothesis is a random distribution, described as a Poisson process. For this, the (temporal) distances between adjacent occurrences are exponentially distributed \( p(x) = \lambda_0 e^{-\lambda_0 x} \), where \( \lambda_0 \) equals the frequency of the category within the corpus. \( p(x) \) is the expected frequency of seeing the next instance of the category at exactly distance \( x \). At distance 0 this is the frequency of the category itself, and integrated over all distances (ad infinitum) it is 1: In an infinite corpus, the category will surely be instantiated at some point.

We compare this expected distribution of distances to the actual distribution. These also look like an exponential distribution (see Figure 1), yet with more short distances than expected from a random distribution. The gaps without any occurrence of the category now are longer. We fit an exponential curve with decay parameter \( \lambda \) to the actual distribution. We interpret the ratio \( r = \lambda / \lambda_0 \) as the strength of priming: The more the fitted parameter deviates from the expected one, the more skewed is the distribution. This parameter can be obtained for every single category, or as an aggregate over all categories.

2.2.3. Single Categories

Figure 1 shows the estimated density function, a random distribution (dotted line), and the fitted, much steeper exponential (dashed line), for the expansion \( VP \rightarrow VB S \).

Across all corpora or their annotations, estimated parameters \( \lambda \) are always larger than \( \lambda_0 \). Rare categories show more priming with \( r \) up to 2.3, and close to 1 for the very common expansion \( S \rightarrow NP VP \) (0.34 occurrences per second). Exponential decay fits well, with standard deviation around 0.005.
The exponential decay supports the suggestion that priming is an effect of (short-term) memory. While frequent categories have generally less room for skewed distributions, there is still something more to be explained about the effect of frequency.

2.2.4. Corpus averages

Measuring the overall priming in a corpus allows to compare several settings: different linguistic frameworks (CCG vs. CFG), spoken vs. written language, conversational (Switchboard) vs. task-oriented (MapTask). We normalize all categories for frequency (so that $\lambda_0 = 1$) and take the average.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>decay parameter $\lambda$</th>
<th>standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>sw-CFG</td>
<td>1.1589</td>
<td>0.0044</td>
</tr>
<tr>
<td>sw-CCG-I</td>
<td>1.0523</td>
<td>0.0054</td>
</tr>
<tr>
<td>sw-CCG-N</td>
<td>1.0364</td>
<td>0.0051</td>
</tr>
<tr>
<td>mp-CFG</td>
<td>1.4666</td>
<td>0.0049</td>
</tr>
<tr>
<td>sw-words</td>
<td>1.2521</td>
<td>0.0113</td>
</tr>
</tbody>
</table>

Standard errors are low, we have thus a good estimate of the actual distribution of distances. Yet Figure 2 suggests an even more extreme distribution. This
might be a result of cumulating activation: short distances trigger more short distances.

We see strong lexical priming (1.25). Task-oriented dialogue outranks conversational dialogue (Reitter et al., 2006b; Pickering and Garrod, 2004). CCG annotation shows comparably little priming. Results by Reitter et al. (2006a) stated that it is significant, but that the difference is not.

2.3. Results

We have devised a notably simple priming measure.8 A single parameter λ per category (or per corpus) suffices, modeling the distribution of distances. Experiments show it to be larger than its expected value, which is the category’s frequency. The effect appears to be larger for rare categories. Interpreting the fitted λ as a frequency is somewhat paradoxical: Primed categories seem more frequent than they actually are.

8. Our measure is simple in comparison to Reitter’s approach which relies on sophisticated statistical modelling using generalized linear mixed models (GLMM) for a logistic regression with random effects.
So far we have viewed categories as mutually exclusive. This does not take into account priming of similar categories. Adding pairwise similarities to the model could improve it. A simple lexical example is stemming or lemmatization: A word also primes all inflected forms. In the syntactic domain, one might consider a measure of similarity between constructions.

The skewness in the distribution of a category (or rather, in the distribution of the pairwise distances of its instances) may be attributed to priming. If the skewness parameter $r$ systematically deviates from 1 across different corpora, then this may be taken as evidence for a mental representation which corresponds to the category and is subject to priming. Thus the categories proposed by linguistic theories can be evaluated for their psycholinguistic validity. This is an approach to inform linguistic theory (about linguistic competence) by performance data (Reitter et al., 2006a).

3. Conclusion

In this corpus study, we have presented a method for measuring syntactic priming based on the decay of repetition probability in a given window of prime–target distance.

While our simple priming measure can be easily used to compare corpora, corpus annotation schemes, and grammar formalisms, it cannot distinguish between self-priming and other-priming, and we have not yet been able to define an absolute baseline for this model.

Before our measure is used for evaluating theories of grammar, a word of caution is in order. In this study, we have interpreted repetitions of syntactic structure as evidence for priming. However, there are certainly other reasons for repeating linguistic constructions, namely the limited range of expressions for talking about certain states of affairs. Coherent texts and dialogues tend to concentrate on certain topics for a while. This is why task-oriented dialogue exhibits more repetition than conversational dialogue. A reliable priming measure should provide means for factoring out semantic and pragmatic aspects – an empirical baseline or control condition.

4. Outlook

The most important next steps for us are to provide (a) methods to measure other-priming separately from self-priming, and (b) to compare these results against an empirical baseline. This would allow us to test theories of dialogue
such as the Interactive Alignment Model (Pickering and Garrod, 2004). Reitter (et al. 2006a; et al. 2006b; 2008) provided a solution for (a) but not for (b).

5. Acknowledgements

This research has been supported by the Center of Excellence 277 “Cognitive Interaction Technology” (CITEC) of the DFG at Bielefeld University. The impetus for this study came from Prof. Gerhard Jäger (who has since moved from Bielefeld University to Tübingen University).

We would like to thank David Reitter and Julia Hockenmaier for kindly providing us with their corpus data, and the anonymous reviewers for reading and commenting on a text that was far from finished.

References

Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry S. Thompson and Regina Weinert

Balcetis, Emily E. and Rick Dale

Bates, E. and B. MacWhinney

Bock, Kathryn

Bock, Kathryn and Zenzi M. Griffin

Bod, Rens

Branigan, Holly, Martin Pickering and Alexandra Cleland
Branigan, Holly, Martin Pickering, Simon Liversedge andrew Stewart and Thomas Urbach

Burnard, Lou

Bybee, Joan

Calhoun, Sasha, Jean Carletta, Jason Brenier, Neil Mayo, Dan Jurafsky, Mark Steedman and David Beaver
2010 The NXT-format Switchboard Corpus: a rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language Resources and Evaluation* 44: 387–419. ISSN 1574-020X.

Chang, Franklin, Gary S. Dell and Kathryn Bock
2006 *Becoming syntactic.* *Psychological Review* 113 (2): 234–272. ISSN 0033-295X.

Dubey, Amit, Frank Keller and Patrick Sturt

Dubey, Amit, Patrick Sturt and Frank Keller

Godfrey, John J. and Edward Holliman

Gries, Stefan Th.

Hartsuiker, R.J., S. Bernolet, S. Schoonbaert, S. Speybroeck and D. Vanderelst
2008 Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language* 58 (2): 214–238. ISSN 0749-596X.

Healey, Patrick G. T., Matthew Purver and Christine Howes
Healey, Patrick G. T., Matthew Purver and Christine Howes  

Hockenmaier, Julia and Mark Steedman  

Howes, Christine, Patrick G. T. Healey and Matthew Purver  

Jaeger, T.F. and N. Snider  

Marcus, Mitchell P., Beatrice Santorini, Mary Ann Marcinkiewicz and Ann Taylor  

Nelson, Gerald, Sean Wallis and Bas Aarts  

Newman, Sharlene D., Kristen Ratliff, Tara Muratore and Thomas Burns Jr.  

Pickering, Martin and Victor Ferreira  

Pickering, Martin and Simon Garrod  

Potter, Mary C. and Linda Lombardi  

Reitter, David  

Reitter, David, Julia Hockenmaier and Frank Keller  

Reitter, David, Johanna D. Moore and Frank Keller  
2006b Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In Proceedings of the 28th Annual Conference of the Cog-
Measuring Syntactic Priming in Dialogue Corpora

Smith, Mark and Linda Wheeldon

Svartvik, Jan, (ed.)

Szmrecsanyi, Benedikt

Tulving, E. and D. L. Schacter