

Priming Effects in Combinatory Categorical Grammar

David Reitter

School of Informatics
University of Edinburgh
2 Buccleuch Place
Edinburgh EH8 9LW, UK
dreitter@inf.ed.ac.uk

Julia Hockenmaier

Inst. for Res. in Cognitive Science
University of Pennsylvania
3401 Walnut Street
Philadelphia PA 19104, USA
juliahr@cis.upenn.edu

Frank Keller

School of Informatics
University of Edinburgh
2 Buccleuch Place
Edinburgh EH8 9LW, UK
keller@inf.ed.ac.uk

Abstract

This paper presents a corpus-based account of structural priming in human sentence processing, focusing on the role that syntactic representations play in such an account. We estimate the strength of structural priming effects from a corpus of spontaneous spoken dialogue, annotated syntactically with Combinatory Categorical Grammar (CCG) derivations. This methodology allows us to test a range of predictions that CCG makes about priming. In particular, we present evidence for priming between lexical and syntactic categories encoding partially satisfied sub-categorization frames, and we show that priming effects exist both for incremental and normal-form CCG derivations.

1 Introduction

In psycholinguistics, *priming* refers to the fact that speakers prefer to reuse recently encountered linguistic material. Priming effects typically manifest themselves in shorter processing times or higher usage frequencies for reused material compared to non-reused material. These effects are attested both in language comprehension and in language production. *Structural priming* occurs when a speaker repeats a syntactic decision, and has been demonstrated in numerous experiments over the past two decades (e.g., Bock, 1986; Branigan et al., 2000). These experimental findings show that subjects are more likely to choose, e.g., a passive voice construction if they have previously comprehended or produced such a construction.

Recent studies have used syntactically annotated corpora to investigate structural priming. The results have demonstrated the existence of priming effects in corpus data: they occur for specific syntactic constructions (Gries, 2005; Szm-

recsanyi, 2005), consistent with the experimental literature, but also generalize to syntactic rules across the board, which repeated more often than expected by chance (Reitter et al., 2006b; Dubey et al., 2006). In the present paper, we build on this corpus-based approach to priming, but focus on the role of the underlying *syntactic representations*. In particular, we use priming to evaluate claims resulting from a particular syntactic theory, which is a way of testing the representational assumptions it makes.

Using priming effects to inform syntactic theory is a novel idea; previous corpus-based priming studies have simply worked with uncontroversial classes of constructions (e.g., passive/active). The contribution of this paper is to overcome this limitation by defining a computational model of priming with a clear interface to a particular syntactic framework. The general assumption we make is that priming is a phenomenon relating to grammatical constituents – these constituents determine the syntactic choices whose repetition can lead to priming. Crucially, grammatical frameworks differ in the grammatical constituents they assume, and therefore predict different sets of priming effects.

We require the following ingredients to pursue this approach: a syntactic theory that identifies a set of constituents, a corpus of linguistic data annotated according to that syntactic theory, and a statistical model that estimates the strength of priming based on a set of external factors. We can then derive predictions for the influence of these factors from the syntactic theory, and test them using the statistical model. In this paper, we use regression models to quantify structural priming effects and to verify predictions made by Combinatory Categorical Grammar (CCG, Steedman (2000)), a syntactic framework that has the theoretical potential to elegantly explain some of the phenomena discovered in priming experiments.

CCG is distinguished from most other grammatical theories by the fact that its rules are *type-dependent*, rather than structure-dependent like classical transformations. Such rules adhere strictly to the constituent condition on rules, i.e., they apply to and yield constituents. Moreover, the syntactic types that determine the applicability of rules in derivations are transparent to (i.e., are determined, though not necessarily uniquely, by) the semantic types that they are associated with. As a consequence, syntactic types are more expressive and more numerous than standard parts of speech: there are around 500 highly frequent CCG types, against the standard 50 or so Penn Treebank POS tags. As we will see below, these properties allow CCG to discard a number of traditional assumptions concerning surface constituency. They also allow us to make a number of testable predictions concerning priming effects, most importantly (a) that priming effects are type-driven and independent of derivation, and, as a corollary; (b) that lexical and derived constituents of the same type can prime each other. These effects are not expected under more traditional views of priming as structure-dependent.

This paper is organized as follows: Section 2 explains the relationship between structural priming and CCG, which leads to a set of specific predictions, detailed in Section 3. Sections 4 and 5 present the methodology employed to test these predictions, describing the corpus data and the statistical analysis used. Section 6 then presents the results of three experiments that deal with priming of lexical vs. phrasal categories, priming in incremental vs. normal form derivations, and frequency effects in priming. Section 7 provides a discussion of the implications of these findings.

2 Background

2.1 Structural Priming

Previous studies of structural priming (Bock, 1986; Branigan et al., 2000) have made few theoretical assumptions about syntax, regardless of whether the studies were based on planned experiments or corpora. They leverage the fact that alternations such as *He gave Raquel the car keys* vs. *He gave the car keys to Raquel* are nearly equivalent in semantics, but differ in their syntactic structure (double object vs. prepositional object). In such experiments, subjects are first exposed to a *prime*, i.e., they have to comprehend or produce

either the double object or the prepositional object structure. In the subsequent trial, the *target*, they are free to produce or comprehend either of the two structures, but they tend to prefer the one that has been primed. In corpus studies, the frequencies of the alternative constructions can be compared in a similar fashion (Gries, 2005; Szmeccsanyi, 2005).

Reitter et al. (2006b) present a different method to examine priming effects in the general case. Rather than selecting specific syntactic alternations, general syntactic units are identified. This method detects syntactic repetition in corpora and correlates its probability with the distance between prime and target, where at great distance, any repetition can be attributed to chance. The size of the priming effect is then estimated as the difference between the repetition probability close to the prime and far away from the prime. This is a way of factoring out chance repetition (which is required if we do not deal with syntactic alternations). By relying on syntactic units, the priming model includes implicit assumptions about the particular syntactic framework used to annotate the corpus under investigation.

2.2 Priming and Lexicalized Grammar

Previous work has demonstrated that priming effects on different linguistic levels are not independent (Pickering and Branigan, 1998). Lexical repetition makes repetition on the syntactic level more likely. For instance, suppose we have two verbal phrases (prime, target) produced only a few seconds apart. Priming means that the target is more likely to assume the same syntactic form (e.g., a passive) as the prime. Furthermore, if the head verbs in prime and target are identical, experiments have demonstrated a stronger priming effect. This effect seems to indicate that lexical and syntactic representations in the grammar share the same information (e.g., subcategorization information), and therefore these representations can prime each other.

Consequently, we treat subcategorization as coterminous with syntactic type, rather than as a feature exclusively associated with lexemes. Such types determine the context of a lexeme or phrase, and are determined by derivation. Such an analysis is exactly what categorial grammars suggest. The rich set of syntactic types that categories afford may be just sufficient to describe all and only

the units that can show priming effects during syntactic processing. That is to say that syntactic priming is categorial *type-priming*, rather than structural priming.

Consistent with this view, Pickering and Branigan (1998) assume that morphosyntactic features such as tense, aspect or number are represented independently from combinatorial properties which specify the contextual requirements of a lexical item. Property groups are represented centrally and shared between lexicon entries, so that they may – separately – prime each other. For example, the pre-nominal adjective *red* in *the red book* primes other pre-nominal adjectives, but not a post-nominal relative clause (*the book that's red*) (Cleland and Pickering, 2003; Scheepers, 2003).

However, if a lexical item can prime a phrasal constituent of the same type, and vice versa, then a type-driven grammar formalism like CCG can provide a simple account of the effect, because lexical and derived syntactic types have the same combinatory potential, which is completely specified by the type, whereas in structure-driven theories, this information is only implicitly given in the derivational process.

2.3 Combinatory Categorical Grammar

CCG (Steedman, 2000) is a mildly context-sensitive, lexicalized grammar formalism with a transparent syntax-semantics interface and a flexible constituent structure that is of particular interest to psycholinguistics, since it allows the construction of incremental derivations. CCG has also enjoyed the interest of the NLP community, with high-accuracy wide-coverage parsers (Clark and Curran, 2004; Hockenmaier and Steedman, 2002) and generators¹ available (White and Baldrige, 2003).

Words are associated with lexical categories which specify their subcategorization behaviour, eg. $((S[dc] \backslash NP) / NP) / NP$ is the lexical category for (tensed) ditransitive verbs in English such as *gives* or *send*, which expect two NP objects to their right, and one NP subject to their left. Complex categories X/Y or $X \backslash Y$ are functors which yield a constituent with category X , if they are applied to a constituent with category Y to their right ($/Y$) or to their left ($\backslash Y$).

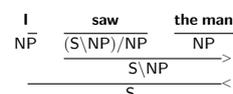
Constituents are combined via a small set of combinatory rule schemata:

Forward Application: $X/Y \ Y \Rightarrow X$

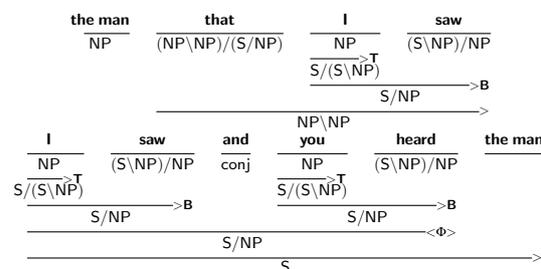
¹<http://opennlp.sourceforge.net/>

Backward Application: $Y \ X \backslash Y \Rightarrow X$
Forward Composition: $X/Y \ Y/Z \Rightarrow B \ X/Z$
Backward Composition: $Y \ Z \ X \backslash Y \Rightarrow B \ X \backslash Z$
Backw. Crossed Composition: $Y/Z \ X \backslash Y \Rightarrow B \ X/Z$
Forward Type-raising: $X \Rightarrow T \ T/(T \ X)$
Coordination: $X \ conj \ X \Rightarrow \Phi \ X$

Function application is the most basic operation (and used by all variants of categorial grammar):



Composition (**B**) and type-raising (**T**) are necessary for the analysis of long-range dependencies and for incremental derivations. CCG uses the same lexical categories for long-range dependencies that arise eg. in wh-movement or coordination as for local dependencies, and does not require traces:



The combinatory rules of CCG allow multiple, semantically equivalent, syntactic derivations of the same sentence. This *spurious ambiguity* is the result of CCG's flexible constituent structure, which can account for long-range dependencies and coordination (as in the above example), and also for interaction with information structure.

CCG parsers often limit the use of the combinatory rules (in particular: type-raising) to obtain a single right-branching *normal form* derivation (Eisner, 1996) for each possible semantic interpretation. Such normal form derivations only use composition and type-raising where syntactically necessary (eg. in relative clauses).

3 Predictions

3.1 Priming Effects

We expect priming effects to apply to CCG *categories*, which describe the type of a constituent including the arguments it expects. Under our assumption that priming manifests itself as a tendency for repetition, repetition probability should be higher for short distances from a prime (see Section 5.2 for details).

3.2 Terminal and Non-terminal Categories

In categorial grammar, lexical categories specify the subcategorization behavior of their heads, capturing local and non-local arguments, and a small set of rule schemata defines how constituents can be combined.

Phrasal constituents may have the same categories as lexical items. For example, the verb *saw* might have the (lexical) category $(S \setminus NP) / NP$, which allows it to combine with an NP to the right. The resulting constituent for *saw Johanna* would be of category $S \setminus NP$ – a constituent which expects an NP (the subject) to its left, and also the lexical category of an intransitive verb. Similarly, the constituent consisting of a ditransitive verb and its object, *gives the money*, has the same category as *saw*. Under the assumption that priming occurs for these categories, we proceed to test a hypothesis that follows from the fact that categories merely encode *unsatisfied* subcategorized arguments.

Given that a transitive verb has the same category as the constituent formed by a ditransitive verb and its direct object, we would expect that both categories can prime each other, if they are cognitive units. More generally, we would expect that lexical (terminal) and phrasal (non-terminal) categories of the same syntactic type may prime each other. The interaction of such conditions with the priming effect can be quantified in the statistical model.

3.3 Incrementality of Analyses

Type-raising and composition allow derivations that are mostly left-branching, or *incremental*. Adopting a left-to-right processing order for a sentence is important, if the syntactic theory is to make psycholinguistically viable predictions (Niv, 1994; Steedman, 2000).

Pickering et al. (2002) present priming experiments that suggest that, in production, structural dominance and linearization do not take place in different stages. Their argument involves verbal phrases with a shifted prepositional object such as *showed to the mechanic a torn overall*. At a dominance-only level, such phrases are equivalent to non-shifted prepositional constructions (*showed a torn overall to the mechanic*), but the two variants may be differentiated at a linearization stage. Shifted primes do not prime prepositional objects in their canonical position, thus priming must occur at a linearized level, and a separate dominance

level seems unlikely (unless priming is selective). CCG is compatible with one-stage formulations of syntax, as no transformation is assumed and categories encode linearization together with subcategorization.

CCG assumes that the processor may produce syntactically different, but semantically equivalent derivations.² So, while neither the incremental analysis we generate, nor the normal-form, represent one single correct derivation, they are two extremes of a ‘spectrum’ of derivations. We hypothesize that priming effects predicted on the basis of incremental CCG analyses will be as strong as those predicted on the basis of their normal-form equivalents.

4 Corpus Data

4.1 The Switchboard Corpus

The Switchboard (Marcus et al., 1994) corpus contains transcriptions of spoken, spontaneous conversation annotated with phrase-structure trees. Dialogues were recorded over the telephone among randomly paired North American speakers, who were just given a general topic to talk about. 80,000 utterances of the corpus have been annotated with syntactic structure. This portion, included in the Penn Treebank, has been time-aligned (per word) in the Paraphrase project (Carletta et al., 2004).

Using the same regression technique as employed here, Reitter et al. (2006b) found a marked structural priming effect for Penn-Treebank style phrase structure rules in Switchboard.

4.2 Disfluencies

Speech is often disfluent, and speech repairs are known to repeat large portions of the preceding context (Johnson and Charniak, 2004). The original Switchboard transcripts contains these disfluencies (marked up as EDITED):

```
( ( S >>>(EDITED
  (RM (-DFL- \bs [] )
  (EDITED
    (RM (-DFL- \bs [] )
    (CC And)
    ( , , )
    (IP (-DFL- \bs +) ) )
  (CC and)
  ( , , )
  (RS (-DFL- \bs [] ) )
  (IP (-DFL- \bs +) ) ) <<<
```

²Selectional criteria such as information structure and intonation allow to distinguish between semantically different analyses.

```

(CC and)
>>>(RS (-DFL- \bs ]) )<<<
(NP-SBJ (PRP I) )
(VP (VBP guess)
  (SBAR (-NONE- 0)
    (S (NP-SBJ (DT that) )
      (VP (BES 's)
        (SBAR-NOM-PRD
          (WHNP-1 (WP what) )
          (S (NP-SBJ (PRP I) )
            (ADVP (RB really) )
            (VP (VBP like)
              (NP (-NONE- *T*-1) ))))))))
  (. .) (-DFL- E_S) ))

```

It is unclear to what extent these repetitions are due to priming rather than simple correction. In disfluent utterances, we therefore eliminate reparanda and only keep repairs (the portions marked with $\langle \dots \rangle$ are removed). Hesitations (uh, etc.), and utterances with unfinished constituents are also ignored.

4.3 Translating Switchboard to CCG

Since the Switchboard annotation is almost identical to the one of the Penn Treebank, we use a similar translation algorithm to Hockenmaier and Steedman (2005). We identify heads, arguments and adjuncts, binarize the trees, and assign categories in a recursive top-down fashion. Nonlocal dependencies that arise through wh-movement and right node raising ($*T*$ and $*RNR*$ traces) are captured in the resulting derivation. Figure 1 (left) shows the rightmost normal form CCG derivation we obtain for the above tree. We then transform this normal form derivation into the most incremental (i.e., left-branching) derivation possible, as shown in Figure 1 (right).

This transformation is done by a top-down recursive procedure, which changes each tree of depth two into an equivalent left-branching analysis if the combinatory rules allow it. This procedure is run until no further transformation can be executed. The lexical categories of both derivations are identical.

5 Statistical Analysis

5.1 Priming of Categories

CCG assumes a minimal set of combinatory rule schemata. Much more than in those rules, syntactic decisions are evident from the *categories* that occur in the derivation.

Given the categories for each utterance, we can identify their repeated use. A certain amount of repetition will obviously be coincidental. But

structural priming predicts that a target category will occur more frequently closer to a potential prime of the same category. Therefore, we can correlate the probability of repetition with the distance between prime and target. Generalized Linear Mixed Effects Models (GLMMs, see next section) allow us to evaluate and quantify this correlation.

Every syntactic category is counted as a potential prime and (almost always) as a target for priming. Because interlocutors tend to stick to a topic during a conversation for some time, we exclude cases of syntactic repetition that are a results of the repetition of a whole phrase.

Previous work points out that priming is sensitive to frequency (Scheepers (2003) for high/low relative clause attachments, (Reitter et al., 2006a) for phrase structure rules). Highly frequent items do not receive (as much) priming. We include the logarithm of the raw frequency of the syntactic category in Switchboard (LNFREQ) to approximate the effect that frequency has on accessibility of the category.

5.2 Generalized Linear Mixed Effects Regression

We use generalized linear mixed effects regression models (GLMM, Venables and Ripley (2002)) to predict a response for a number of given categorical (‘factor’) or continuous (‘predictor’) explanatory variables (features). Our data is made up of instances of repetition examples and non-repetition examples from the corpus. For each target instance of a syntactic category c occurring in a derivation and spanning a constituent that begins at time t , we look back for possible instances of constituents with the same category (the prime) in a time frame of $[t - d - 0.5; t - d + 0.5]$ seconds. If such instances can be found, we have a positive example of repetition. Otherwise, c is included as a data point with a negative outcome. We do so for a range of different distances d , commonly $1 \leq d \leq 15$ seconds.³ For each data point, we include the logarithm of the distance d between *priming period* and *target* as an explanatory variable LNDIST. (See Reitter et al. (2006b) for a worked example.)

In order to eliminate cases of lexical repetition of a phrase, e.g., names or lexicalized noun

³This approach uses a number of data points per target, looking backwards for primes. The opposite way – looking forwards for targets – would make similar predictions.

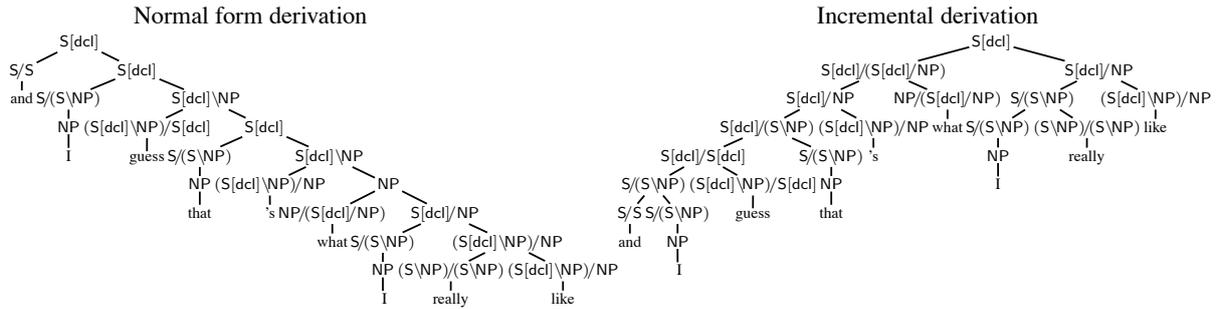


Figure 1: Two derivations (normal form: left), incremental: right) for the sentence fragment *and I guess that's what I really like* from Switchboard.

phrases, which we consider topic-dependent or instances of lexical priming, we only collect syntactic repetitions with at least one differing word.

Without syntactic priming, we would assume that there is no correlation between the probability that a data point is positive (repetition occurs) and distance d . With priming, we would expect that the probability is inversely proportional to d . Our model uses $\ln d$ as predictor LNDIST, since memory effects usually decay exponentially.

The regression model fitted is then simply a choice of coefficients β_i , among them one for each explanatory variable i . β_i expresses the contribution of i to the probability of the outcome event, that is, in our case, successful priming. The coefficient of interest is the one for the time correlation, i.e. $\beta_{\ln Dist}$. It specifies the strength of decay of repetition probability over time. If no other variables are present, a model estimates the repetition probability for a data point i as

$$\hat{p}_i = \beta_0 + \beta_{\ln Dist} \ln DIST_i$$

Priming is present if the estimated parameter is negative, i.e. the repetition probability decreases with increasing distance between prime and target.

Other explanatory variables, such as ROLE, which indicates whether priming occurs within a speaker (production-production priming, PP) or in between speakers (comprehension-production priming, CP), receive an interaction coefficient that adds linearly to $\beta_{\ln Dist}$. Additional interaction variables are included depending on the experimental question.⁴

⁴Lastly, we identify the target utterance in a random factor in our model, grouping the several measurements (15 for the different distances from each target) as *repeated measurements*, since they depend on the same target category occurrence and are partially inter-dependent.

From the data produced, we include all cases of repetition and an equal number of randomly sampled non-repetition cases.⁵

6 Experiments

6.1 Experiment 1: Priming in Incremental and Normal-form Derivations

Hypothesis CCG assumes a multiplicity of semantically equivalent derivations with different syntactic constituent structures. Here, we investigate whether two of these, the normal-form and the most incremental derivation, differ in the strength with which syntactic priming occurs.

Method A joint model was built containing repetition data from both types of derivations. Since we are only interested in cases where the two derivations differ, we excluded all constituents where a string of words was analyzed as a constituent in both derivations. This produced a data set where the two derivations could be contrasted.

A factor DERIVATION in the model indicates whether the repetition occurred in a normal-form (NF) or an incremental derivation (INC).

Results Significant and substantial priming is present in both types of derivations, for both PP and CP priming. There is no significant difference in priming strength between normal-form and incremental derivations ($\beta_{\ln Dist:NF} = 0.008, p = 0.95$). The logarithm of the raw category frequency is negatively correlated with the priming strength ($\beta_{\ln Dist:\ln Freq} = 0.151, p < 0.0001$). Note that a negative coefficient for LNDIST indicates

⁵We trained our models using Penalized Quasi-Likelihood (Venables and Ripley, 2002). This technique works best if data is balanced, i.e. we avoid having very rare positive examples in the data. Experiment 2 was conducted on a subset of the data.

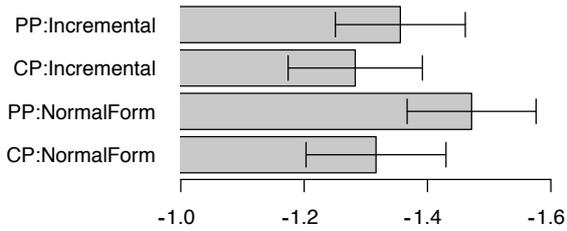


Figure 2: Decay effect sizes in Experiment 1 for combinations of comprehension-production or production-production priming and in incremental or normal-form derivations. Error bars show (non-simultaneous) 95% confidence intervals.

decay. The lower this coefficient, the more decay, hence priming).

If there was no priming of categories for incrementally formed constituents, we would expect to see a large effect of DERIVATION. In the contrary, we see no effect at a high p , where the regression method used is demonstrably powerful enough to detect even small changes in the priming effect. We conclude that there is no detectable difference in priming between the two derivation types. In Fig. 2, we give the estimated priming effect sizes for the four conditions.⁶

The result is compatible with CCG’s separation of derivation structure and the type of the result of derivation. It is not the derivation structure that primes, but rather the type of the result. It is also compatible with the possibility of a non-traditional constituent structure (such as the incremental analysis), even though it is clear that neither incremental nor normal-form derivations necessarily represent the ideal analysis.

The category sets occurring in both derivation variants was largely disjunct, making testing for actual overlap between different derivations impossible.

6.2 Experiment 2: Priming between Lexical and Phrasal Categories

Hypothesis Since CCG categories simply encode unsatisfied subcategorization constraints, constituents which are very different from a traditional linguistic perspective can receive the same category. This is, perhaps, most evident in phrasal

⁶Note that Figures 2 and 3 stem from nested models that estimate the effect of LNDIST within the four/eight conditions. Confidence intervals will be larger, as fewer data-points are available than when the overall effect of a single factor is compared.

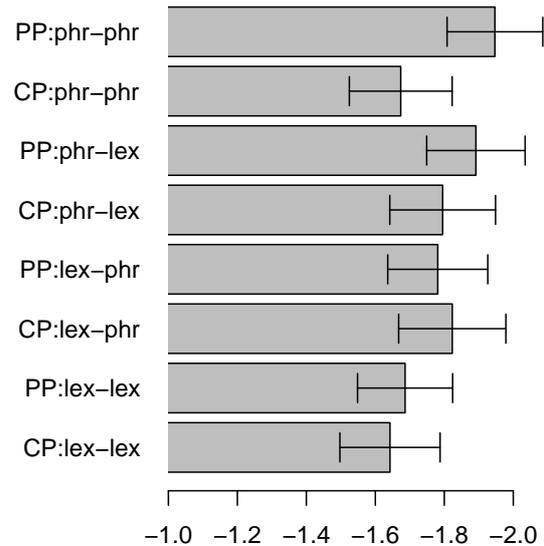


Figure 3: Decay effect sizes in Experiment 2, for combinations of comprehension-production or production-production priming and lexical or phrasal primes and targets, e.g. the third bar denotes the decay in repetition probability of a phrasal category as prime and a lexical one as target, where prime and target occurred in utterances by the same speaker. Error bars show (non-simultaneous) 95% confidence intervals.

and lexical categories (where, e.g., an intransitive verb is indistinguishable from a verb phrase).

Bock and Loebell (1990)’s experiments suggest that priming effects are independent of the subcategorization frame. There, an active voice sentence primed a passive voice one with the same phrase structure, but a different subcategorization. If we find priming from lexical to phrasal categories, then our model demonstrates priming of subcategorization frames.

From a processing point of view, phrasal categories are distinct from lexical ones. Lexical categories are bound to the lemma and thereby linked to the lexicon, while phrasal categories are the result of a structural composition or decomposition process. The latter ones represent temporary states, encoding the syntactic *process*.

Here, we test whether lexical and phrasal categories can prime each other, and if so, contrast the strength of these priming effects.

Method We built a model which allowed lexical and phrasal categories to prime each other. A factor, STRUCTURAL LEVEL was introduced

to distinguish the four cases: priming in between phrasal categories and in between lexical ones, from lexical ones to phrasal ones and from phrasal ones to lexical ones.

Recall that each data point encodes a possibility to repeat a CCG category, referring to a particular instance of a target category at time t and a time span of duration of one second $[t - d - 0.5, t - d + 0.5]$ in which a priming instance of the same category could occur. If it occurred at least once, the data point was counted as a possible example of priming (response variable: true), otherwise it was included as a counter-example (response variable: false). For the target category, its type (lexical or phrasal) was clear. For the category of the prime, we included two data points, one for each type, with a response indicating whether a prime of the category of such a type occurred in the time window. We built separate models for incremental and normal form derivations. Models were fitted to a balanced subset, including all repetitions and a randomly sampled subset of non-repetitions.

Results Both the normal-form and the incremental model show qualitatively the same results. `STRUCTURALLEVEL` has a significant influence on priming strength (`LN DIST`) for the cases where a lexical item serves as prime (e.g., normal-form PP: $\beta_{\text{LnDist:lex-lex}} = 0.261$, $p < 0.0001$; $\beta_{\text{LnDist:lex-phr}} = 0.166$, $p < 0.0001$; $\beta_{\text{LnDist:phr-lex}} = 0.056$, $p < 0.05$; as compared to the baseline $\text{phr} - \text{phr}$. N.B. higher values denote less decay & priming). Phrasal categories prime other phrasal and lexical categories, but there is a lower priming effect to be seen from lexical categories. Figure 3 presents the resulting effect sizes.

Albeit significant, we assume the effect of prime type is attributable to processing differences rather than the strong difference that would indicate that there is no priming of, e.g., lexical subcategorization frames. As the analysis of effect sizes shows, we can see priming from and in between both lexical and phrasal categories.

Additionally, there is no evidence suggesting that, once frequency is taken into account, syntactic processes happening high up in derivation trees show more priming (see Scheepers 2003).

7 Discussion

We can confirm the syntactic priming effect for CCG categories. Priming occurs in incremental as well as in normal-form CCG derivations, and at

different syntactic levels in those derivations: we demonstrated that priming effects persists across syntactic stages, from the lowest one (lexical categories) up to higher ones (phrasal categories). This is what CCG predicts if priming of categories is assumed.

Linguistic data is inherently noisy. Annotations contain errors, and conversions such as the one to CCG may add further error. However, since noise is distributed across the corpus, it is unlikely to affect priming effect strength or its interaction with the factors we used: priming, in this study, is defined as *decay* of repetition probability. We see the lack of control in the collection of a corpus like Switchboard not only as a challenge, but also as an advantage: it means that realistic data is present in the corpus, allowing us to conduct a controlled experiment to validate a claim about a specific theory of competence grammar.

The fact that CCG categories prime could be explained in a model that includes a basic form of subcategorization. All categories, if lexical or phrasal, contain a subcategorization frame, with only those categories present that have yet to be satisfied. Our CCG based models make predictions for experimental studies, e.g., that specific heads with open subcategorization slots (such as transitive verbs) will be primed by phrases that require the same kinds of arguments (such as verbal phrases with a ditransitive verb and an argument).

The models presented take the frequency of the syntactic category into account, reducing noise, especially in the conditions with lower numbers of (positive) repetition examples (e.g., CP and incremental derivations in Experiment 1). Whether there are significant qualitative and quantitative differences of PP and CP priming with respect to choice of derivation type – which would point out processing differences in comprehension vs. production priming – will be a matter of future work.

At this point, we do not explicitly discriminate different syntactic frameworks. Comparing priming effects in a corpus annotated in parallel according to different theories will be a matter of future work.

8 Conclusions

We have discussed an empirical, corpus-based approach to use priming effects in the validation of general syntactic models. The analysis we presented is compatible with the reality of a lexical-

ized, categorial grammar such as CCG as a component of the human sentence processor. CCG is unusual in allowing us to compare different types of derivational analyses within the same grammar framework. Focusing on CCG allowed us to contrast priming under different conditions, while still making a statistical and general statement about the priming effects for *all* syntactic phenomena covered by the grammar.

Acknowledgements

We would like to thank Mark Steedman, Roger Levy, Johanna Moore and three anonymous reviewers for their comments. The authors are grateful for being supported by the following grants: DR by The Edinburgh Stanford Link, JH by NSF ITR grant 0205456, FK by The Leverhulme Trust (grant F/00 159/AL – Syntactic Parallelism).

References

- J. Kathryn Bock. 1986. Syntactic persistence in language production. *Cognitive Psychology*, 18:355–387.
- J. Kathryn Bock and Helga Loebell. 1990. Framing sentences. *Cognition*, 35:1–39.
- Holly P. Branigan, Martin J. Pickering, and Alexandra A. Cleland. 2000. Syntactic co-ordination in dialogue. *Cognition*, 75:B13–25.
- Jean Carletta, S. Dingare, Malvina Nissim, and T. Nikitina. 2004. Using the NITE XML toolkit on the Switchboard corpus to study syntactic choice: a case study. In *Proc. 4th Language Resources and Evaluation Conference*. Lisbon, Portugal.
- Stephen Clark and James R. Curran. 2004. Parsing the WSJ using CCG and log-linear models. In *Proc. of the 42nd Annual Meeting of the Association for Computational Linguistics*. Barcelona, Spain.
- A. A. Cleland and M. J. Pickering. 2003. The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, 49:214–230.
- Amit Dubey, Frank Keller, and Patrick Sturt. 2006. Integrating syntactic priming into an incremental probabilistic parser, with an application to psycholinguistic modeling. In *Proc. of the 21st International Conference on Computational Linguistics and 44th Annual Mtg of the Association for Computational Linguistics*. Sydney, Australia.
- Jason Eisner. 1996. Efficient normal-form parsing for combinatory categorial grammar. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, pages 79–86. Santa Cruz, CA.
- Stefan Th. Gries. 2005. Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research*, 34(4):365–399.
- Julia Hockenmaier and Mark Steedman. 2002. Generative models for statistical parsing with Combinatory Categorical Grammar. In *Proc. 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, PA.
- Julia Hockenmaier and Mark Steedman. 2005. CCGbank: Users’ manual. Technical Report MS-CIS-05-09, Computer and Information Science, University of Pennsylvania.
- Mark Johnson and Eugene Charniak. 2004. A tag-based noisy channel model of speech repairs. In *Proc. 42nd Annual Meeting of the Association for Computational Linguistics*, pages 33–39. Barcelona, Spain.
- M. Marcus, G. Kim, M. Marcinkiewicz, R. MacIntyre, A. Bies, M. Ferguson, K. Katz, and B. Schasberger. 1994. The Penn treebank: Annotating predicate argument structure. In *Proc. ARPA Human Language Technology Workshop*. Plainsboro, NJ.
- Michael Niv. 1994. A psycholinguistically motivated parser for CCG. In *Mtg. of the Association for Computational Linguistics*, pages 125–132.
- Martin J. Pickering and Holly P. Branigan. 1998. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39:633–651.
- Martin J. Pickering, Holly P. Branigan, and Janet F. McLean. 2002. Constituent structure is formulated in one stage. *Journal of Memory and Language*, 46:586–605.
- David Reitter, Frank Keller, and Johanna D. Moore. 2006a. Computational modelling of structural priming in dialogue. In *Proc. Human Language Technology conference - North American chapter of the Association for Computational Linguistics annual mtg*. New York City.
- David Reitter, Johanna D. Moore, and Frank Keller. 2006b. Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In *Proc. 28th Annual Conference of the Cognitive Science Society*.
- Christoph Scheepers. 2003. Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition*, 89:179–205.
- Mark Steedman. 2000. *The Syntactic Process*. MIT Press.
- Benedikt Szmrecsanyi. 2005. Creatures of habit: A corpus-linguistic analysis of persistence in spoken english. *Corpus Linguistics and Linguistic Theory*, 1(1):113–149.
- William N. Venables and Brian D. Ripley. 2002. *Modern Applied Statistics with S. Fourth Edition*. Springer.
- Mike White and Jason Baldridge. 2003. Adapting chart realization to CCG. In *Proc. 9th European Workshop on Natural Language Generation*. Budapest, Hungary.