

Implicit Learning and Syntactic Persistence: Surprisal and Cumulativity

T. Florian Jaeger (fjaeger@bcs.rochester.edu)

Department of Brain and Cognitive Science, Meliora Hall
Rochester, NY 14627 USA

Neal Snider (snider@stanford.edu)

Department of Linguistics, Building 460
Stanford, CA 94305 USA

Abstract

Speakers often repeat syntactic structures that have recently been used by them or their interlocutors. Such syntactic persistence has variously been attributed to social causes (mimicking to be well-liked), to an effort to reduce processing costs, or as a side effect of either transient activation of previously processed structure or implicit learning. We present new evidence from syntactic persistence in conversational speech consistent with the view that syntactic persistence is an epiphenomenon. We propose that syntactic persistence is linked to the updating and maintenance of probabilistic syntactic knowledge. We find that syntactic persistence is **SURPRISAL-SENSITIVE** and **CUMULATIVE**: Syntactic persistence is stronger the less probable the prime and it is sufficiently long-lived to be cumulative.

Keywords: priming; surprisal; cumulativity; learning

Introduction

Syntactic persistence (Bock, 1986; Pickering & Branigan, 1998) is the tendency for speakers to repeat a syntactic structure that they have processed previously. In the following example (1) from the Switchboard corpus (Godfrey, Holliman, & McDaniel, 1992), the speaker produces the double object ditransitive structure, or NPNP (*give [a country] [money]*), as opposed to the prepositional object structure, or NPPP (*give [money] [to a country]*), after having produced an NPNP structure with another verb previously in the conversation:

- (1) “... I don’t feel we should **loan [them] [money]**... I wish our leaders were really seeking the Lord on these things, and if we feel led to **give [a country] [money]** to help them, fine”

Syntactic persistence has been well documented for different modalities, different languages, different types of structures. Considerably less is known about the underlying causes though several competing accounts exist. For example, Brennan and Clark (1996) proposed that syntactic persistence is due to conceptual pacts that interlocutors use to maintain common ground: interlocutors come to use similar syntax for certain meanings just as they come to use the same descriptions for novel objects. Conceptual pact accounts of persistence assume some sort of intention to mimic interlocutors’ behavior, e.g. as a sign of social alignment in that producers come to use the same syntax as their interlocutors to show social solidarity just as much as they use the same gestures and body language (Krauss & Fussell, 1996). According to other accounts, persistence is a byproduct of other processes, either ‘transient activation’ (Branigan, Pickering, Stewart, &

McLean, 2000; Pickering & Garrod, 2004) or ‘implicit learning’ (Bock & Griffin, 2000; Bock, Dell, Chang, & Onishi, in press). According to transient activation accounts, accessing a structure during production or comprehension activates it and that activation persists for a short time. Transient activation makes it more likely that the same structure is used again on the next opportunity. According to ‘implicit learning’ accounts, processing of a structure inevitably leads to implicitly learning of the processed linguistic representations and the amount of learning determines the probability of reusing that same structure later on (Bock & Griffin, 2000; Bock et al., in press; Ferreira, 2003b).

Transient activation and implicit learning accounts are often pitched against each other on the basis that they are taken to predict different time courses for the decay of syntactic persistence effects (Bock & Griffin, 2000; Bock et al., in press; Hartsuiker, Bernolet, Schoonbaert, Speybroeck, & Vanderelst, submitted). The transient activation account is taken to predict that syntactic persistence is short-lived. That is, the increase in the posteriori probability of using a structure given that it has just been processed should rapidly decrease over time. The implicit learning account, on the other hand, is taken to predict that persistence should be long lived. Hence repetition probability should not decrease for larger prime-target distances, or decrease only very slowly.

Attempts to distinguish transient activation and implicit learning accounts on the basis of evidence from persistence decay have led to mixed results. Most experimental studies have found that there is no effect of prime-target distance (Bock & Griffin, 2000; Bock et al., in press), while corpus studies have mostly found persistence to decay logarithmically with distance (Gries, 2005; Szmrecsányi, 2005; Reitter, Moore, & Keller, 2006, but see Jaeger, 2006b). We return to this apparent empirical conflict in the general discussion. Independent of the empirical facts, it is unclear how slow or fast a decay would be a decisive piece of evidence against either of the theories. It is also unclear what prediction conceptual pact accounts of syntactic persistence make about the time course of persistence decay. In this paper, we look at two other properties of syntactic persistence that distinguish between competing accounts – **SURPRISAL-SENSITIVITY** and **CUMULATIVITY**.

We explore an approach that we consider related to, but conceptually slightly different from, implicit learning accounts. We propose that language users, or rather the system underlying language processing, implicitly maintain and update probabilistic distributions over linguistic structures. In-

deed, there is much evidence that language production and comprehension are sensitive to distributions of structures at all levels of linguistic representation (for a recent overview, see Jurafsky, 2003). While such probability-sensitive processing may be intuitive from an information theoretic perspective, because access to probabilistic information allows provably optimal (i.e. efficient and robust) communication in the face of uncertainty, it raises the question how language users keep track of the distributions of a huge number of structures. Not only does the probability of a structure vary depending on context, the set of relevant cues to a structure's probability, too, may vary across contexts (e.g. due to other unobserved cues). It seems therefore unlikely that language users learn a cue's interpretation once and for all during language acquisition. Instead, the language processing system must be sufficiently flexible to accommodate short-term changes in probability distributions and in the interpretation of specific cues. We propose that the language processing system is set up in such a way that, whenever an instance of a structure (e.g. an NPNP structure) is processed, it is seen as a piece of evidence that affects the structure's probability distribution (e.g. the distribution of NPNP vs. NPPP structures) given the set of cues associated with that instance. In other words, maintenance of probability distributions is assumed to be an inherent part of the language processing system. The phenomenon of syntactic persistence is then taken to be a mere correlate of the maintenance of syntactic probability distributions, where less probable syntactic structures, if observed, lead to a bigger change in the probability distribution, which in turn leads to an increased probability of reusing the same structure. This would link syntactic persistence to recent work on perceptual persistence (Huber & O'Reilly, 2003) and skill maintenance (Huber, Shiffrin, Lyle, & Ruys, 2001), where less probable events lead to greater activation boosts. This approach leads us to formulate the hypothesis of SURPRISAL-SENSITIVE PERSISTENCE: more surprising structures are predicted to prime more strongly (i.e. to lead to a bigger increase in the probability of repetition) than less surprising structures. We use the term 'surprisal' in its information theoretic sense referring to the log inverse of the probability, $surprisal(X) = \frac{1}{\log(probability(X))}$.

Surprisal-sensitive persistence would be compatible with transient activation and implicit learning accounts but not with accounts that attribute syntactic persistence solely to conceptual pact (Brennan & Clark, 1996, a.o.). Initial support for surprisal-sensitive persistence comes from the observation that less frequent structures tend to prime more strongly (Bock, 1986; Ferreira, 2003b, among many others). For example, in an experiment on active/passive choice, Bock (1986) found syntactic persistence for the infrequent passive structures (only 19% of the productions in her experiment). No syntactic persistence, on the other hand, was observed for the much more frequent active structures. While such inverse frequency effects on prime strength are encouraging, they are insufficient as a test of surprisal-sensitive persistence. The hypothesis of surprisal-sensitive persistence makes a stronger prediction: for instances of a syntactic structure, the likelihood that this syntactic structure will be reused should be correlated with its surprisal. Study 1 tests this prediction using a database of actives and passives from spontaneous speech.

Studies 2 to 4 investigate another potential property of syntactic persistence, namely whether syntactic persistence is cumulative. If syntactic persistence is caused by language users tracking the distribution of structures, as suggested above, we might expect syntactic persistence beyond the most recent structure. Such an effect would be unexpected by transient activation accounts and by accounts that attribute persistence solely to conceptual pacts. While it would be possible that only the most recent structure affects language users' assumptions over the distribution of that structure, a system that gives (weighted) access to several preceding primes may be advantageous (under the assumption spelled out above that language users implicitly maintain probability distributions over syntactic structures). Considering several preceding structures would provide language users with a better estimate of a structure's probability *given the current context*.

Evidence in support of cumulativity comes from recent experiments by Kaschak and colleagues (Kaschak, Loney, & Borreggine, 2006; Kaschak & Borreggine, in press). These studies used a training block before the priming study that contained different distributions of ditransitive structures. Subjects who were made to produce all NPPP in the training block produced more NPPP structures in the priming phase than those who had produced all NPNP in during training. We wished to strengthen these important findings by testing them under natural conditions, so the studies presented here differ from these studies in two ways. First, all our results come from corpora of spontaneous speech. This is important since it shows that our results are unlikely to be the artifact of unnatural distributions that may cause explicit learning rather than implicit, highly automatic learning. Second, our studies differ from the experiments by Kaschak and colleagues in that our data is considerably more heterogeneous. That is, our database contains cases where different prime structures repeat and alternate in many different patterns. We return to these differences when we describe our studies on cumulativity below. Study 2 employs the same database of actives and passives used for Study 1. Studies 3 and 4 use data from *that*-omission in complement and relative clauses, respectively.

Surprisal in the voice alternation (Study 1)

The first study tests the effect of prime surprisal on syntactic persistence in the the active vs. passive voice alternation.

Corpus and data We first extracted all passive structures from the manually parsed Penn Treebank (release 3, approximately, 800,000 words, see Marcus, Santorini, Marcinkiewicz, & Taylor, 1999) portion of the Switchboard corpus of telephone conversations. Structures were considered passives if they contained a verb in the passive participle which was the complement of a form of *be* or *get*, and the participle was coded as having a missing complement NP (object or oblique). Next, we extracted all passivizable actives, which were defined as verbs with a complement NP (object or oblique; verbs that do not participate in the voice alternation (e.g. *suppose*), despite being coded as passives in the Treebank were excluded). The data set contains uses of actives and passives inside both main and subordinate clauses. We then excluded all tokens without a preceding prime, where a prime is defined as the last use of a passivizable verb in the conversation by either speaker (e.g. a token would not

have a prime if it was the first passivizable verb in a conversation). We also excluded tokens for which the prime or target verb occurred less than 10 times in the entire data set (including occurrences without a prime), because the surprisal estimates for these cases would be too unreliable. This left 30,798 prime-target pairs, of which 1,791 targets are passive.

Surprisal estimates We estimate the prime's surprisal as the conditional probability of the passive structure given the verb (henceforth prime verb bias), by counting the number of times each verb lemma occurred in each structure in the full active/passive data set (not just those tokens with primes).

Analysis We used a mixed logit model (Breslow & Clayton, 1993) to test whether prime surprisal affects speakers' choice between active and passive structures. Mixed logit models can be thought of as an extension of logistic regression that allows modeling of random subject effects. Here and in all following studies, we report the coefficient for each independent variable and its levels of significance. Coefficients in logit models are given in log-odds (the space in which logistic models are fitted to the data), but we also report odds ($= e^{\log\text{-odds}}$). Odds range from 0 (for proportions of 0) to positive infinity (for proportions of 1), with proportions of 0.5 corresponding to odds of 1. Odds are a multiplicative scale, so we talk about an x -fold increase or decrease in odds between conditions. Below we model the (log-)odds that the target structure is a passive. So, significant positive coefficients indicate increased odds that the target structure will be a passive and negative coefficients indicate increase odds of an active structure.

In addition to prime verb bias, we included three controls in the model: *target verb bias* (estimated the same way as prime verb bias) to capture the lexical bias of the target verb; *prime-target distance* (the distance between prime and target in utterances) to test for prime decay; *verb-identity* (whether the prime and target verb lemma were identical). The model also contained a random effect of speaker to control for speaker differences in the rate of passivization. We report the results for those tokens with passive primes and those with active primes separately.¹

Prediction The learning-surprisal hypothesis predicts that increasing prime verb bias towards passives should make the passive structure less likely in the target.

Results: Passive primes There was a significant effect of prime verb bias ($p < 0.003$), with a negative log-odds coefficient $B = -1.205$ ($SE = 0.404$), corresponding odds coefficient $e^B = 0.30$. Active-biased prime verbs appearing in the passive make the target more likely to be a passive than passive-biased prime verbs. There were also significant effects of target bias ($p \ll 0.001$), with a positive log-odds coefficient $B = 1.212$ ($SE = 0.025$), corresponding odds coefficient $e^B = 3.35$, and prime-target verb identity ($p \ll 0.001$), with a positive log-odds coefficient $B = 1.754$ ($SE = 0.259$), corresponding odds coefficient $e^B = 5.75$. In this data set, there was a significant effect of distance ($p < 0.01$), with a negative log-odds coefficient $B = -0.060$ ($SE = 0.026$), corresponding odds coefficient $e^B = 0.94$.

¹Collinearity was removed from all reported models using residuals to avoid spurious results due to inflated standard errors.

Results: Active primes The same study was carried out on the active prime data set, but no effect of prime verb bias was found. All controls had the expected effects.

Discussion Syntactic persistence effects on passives seem to be sensitive to the prime's surprisal. At the same time, no such surprisal-sensitivity is observed for actives. That is, we have not found an effect of surprisal in the more frequent structures. This may be due to the overall much weaker syntactic persistence effects for the more frequent active structure (cf. the inverse frequency effect, e.g. Ferreira, 2003b). The results presented here are preliminary in nature and we refer to Snider (2008) for further developments.

We also found a distance effect. The probability of a passive after a passive prime has been processed decreases by 6% for each intervenes utterance between prime and target. This does not mean, however, that the persistence effect is short-lived. In the next set of studies, we show that the effect of many primes is cumulative across the conversation: the more primes comprehended or produced, the more likely the prime structure is to be produced later.

Cumulativity in the voice alternation (Study 2)

Recent work by Kaschak and colleagues (Kaschak et al., 2006; Kaschak & Borreggine, in press) has provided evidence that syntactic persistence effects are cumulative. The purpose of the current study was two-fold: to test whether cumulative persistence is also found in spontaneous conversational speech and to further investigate the nature of cumulative persistence. We test whether the *number* of preceding active and passive primes correlates with speakers' choices in the voice alternation. That is, unlike Kaschak and colleagues, we use a gradient measure of cumulativity.

Cumulativity measure Cumulativity was measured by counting the number of primes of each structure previously comprehended or produced by the speaker in the conversation (excluding the most recent prime). We examined cumulative effects along both of the aforementioned dimensions: the effect of each type of prime structure (active, passive), and the processing modality (comprehension, production).

Analysis Four cumulativity variables were added to the mixed logit model from Study 1 (with the same controls). Two variables encode within-speaker persistence: the number of passives produced previously by the target speaker (the speaker who uttered the target sentence), and the number of actives produced previously by the target speaker. The other two new variables encode across-speaker persistence: the number of passives and the number of actives comprehended previously by the target speaker.

Prediction If language users are sensitive to the distribution of recently processed structures, we should be able to observe effects of primes beyond the most recent prime. More generally, implicit learning accounts of syntactic persistence predict that persistence effects are long-lasting. Hence, persistence effects should be observable beyond the most recent prime. Given previous results from production experiments (Bock et al., in press, a.o.), within-speaker persistence is expected to be stronger than across-speaker persistence.

Results We found a significant effect of the number of passives produced within speaker ($p \ll 0.001$), with a positive log-odds coefficient $B = 0.107$ ($SE = 0.016$), corresponding odds coefficient $e^B = 1.12$. Thus, the more passives the target speaker has previously produced in the conversation, the more likely they are to produce another passive in the target. It should be noted that this is not confounded by a speaker effect, because we controlled for a random effect of speaker. We also found a significant effect of the number of passives comprehended ($p < 0.006$), with a positive log-odds coefficient $B = 0.055$ ($SE = 0.020$), corresponding odds coefficient $e^B = 1.05$, so the effect size was less than for cumulativity in production. Turning to actives, we found a significant effect of the number of actives produced, such that the more actives the target speaker produced previously, the less likely they are to produce a passive ($p \ll 0.001$), with a negative log-odds coefficient $B = -0.008$ ($SE = 0.002$), corresponding odds coefficient $e^B = 0.992$. The smaller log odds coefficient indicates that effect size of cumulative actives is less than for cumulative passives. This is consistent with the inverse frequency effect: the less frequent structure primes more. We found no significant cumulative effect of the number of actives comprehended.

Discussion The passive data from Study 1 shows that persistence is sufficiently long-lived that it is cumulative. The more prime structures processed, the more likely that prime structure is to be produced later. Our results therefore corroborate recent findings from laboratory experiments (Kaschak et al., 2006; Kaschak & Borreggine, in press) and extend them to conversational speech. Hence, cumulativity of syntactic persistence cannot be reduced to the rather unnatural distributions of structures that participants were exposed to in previous laboratory experiments (Kaschak and colleagues exposed participants to blocks of many trials of the same syntactic structure to show increased persistence effects).

Interestingly, we even found an effect of the cumulative number of active primes (when produced by the target speaker), an effect that did not reach significance if only the most recent prime is considered. As expected, comprehended primes had a weaker cumulative effect compared to produced primes. In combination with the observation that more frequent structures, prime less, weaker across-speaker persistence also provides a likely explanation for the lack of a significant effect of the cumulative comprehended active primes.

Cumulativity in *that*-omission (Studies 3 & 4)

Next we look at cumulative persistence in two alternations that differ qualitatively from the voice alternation: *that*-omission in complement clauses (Study 3) and in relative clauses (Study 4). Such syntactic reduction does not involve word order choices, but rather the choice between a full form (with *that*) and a reduced form (without *that*).

Data The data for Studies 3 and 4 comes from Jaeger (2006a), and consists of 5,701 complement clauses (CCs), as in (2-a), and 2,071 relative clauses (RCs), as in (2-b), respectively, that were extracted from the same telephone conversation corpus as used in Studies 1 and 2. In English, CCs and RCs can be produced in a FULL form with relativizer/complementizer *that* or in a REDUCED form without

that (indicated below by parentheses). Optional *that* in CCs and RCs is affected by syntactic persistence (Ferreira, 2003a; Jaeger, 2006a).

- (2) a. "... and i don't believe [_{CC} (that) any of us would have to purchase any extra vacation days] ..."
- b. "[MUMBLE] the only thing [_{RC} (that) we get paid for] are the aluminum cans ..."

Analysis We used an ordinary logistic regression model with the same controls as in Jaeger (2006a) in addition to the measures of interest, the cumulative number of primes at the time of the target utterance (the number of full CC/RCs and reduced CC/RCs, both produced and comprehended). For brevity's sake, we do not describe the controls in details here (there were over 35 control parameters in each model). The types of controls are very similar for the two case studies. To give an example, the controls for CCs include, among others:

- Position of CC in matrix clause, and adjacency to CC-embedding verb
- Complexity of CC (log length in words) and of matrix clause (5 parameters)
- Accessibility of CC subject (8 parameters, including pronominality, givenness, parallelism, animacy, frequency)
- Predictability of CC
- Frequency of CC-embedding verb
- Fluency, pauses, and speech rate around CC onset (6 parameters)
- Phonological and prosodic context (3 parameters)
- Speaker Gender

Since none of the control effects included here changed qualitatively compared to the models presented in Jaeger (2006a, Ch 3 & 4), we do not further report them. To correct for speaker effects, the model was bootstrapped 10,000 times with random replacement of speaker clusters (Feng, McLerran, & Grizzle, 1996, mixed effect models with random speaker effects return the same results, Jaeger, 2006a, Ch. 2). The models reported here do *not* contain prime-target distance as a control factor because previous models of the same data have not found any effect of prime-target decay (Jaeger, 2006b), although this null result may be due to the relatively high number of parameters already in the model given the number of data points. The models also do not control for potential effects of prime surprisal. We plan to include these factors in future work.

Results: Complementizer omission We found significant cumulative persistence effects: the more full CCs the target speaker produced, the more likely the target was to be a full CC ($p < 0.001$). There was no significant effect of the number of full CCs comprehended, but there were significant effects of the number of reduced CCs produced ($p < 0.05$) and comprehended ($p < 0.05$). Similar to the passive data, the size of the cumulative persistence effect in full CCs is greater than that in reduced CCs. This result is consistent with an inverse frequency effect, because full CCs are the less frequent structure. Lack of an effect for comprehended full CCs where there is an effect for comprehended reduced CCs is, however, not expected given the inverse frequency effect.

Results: Relativizer omission We found significant cumulative persistence effects: the more full RCs that have been produced ($p < 0.001$) or comprehended ($p < 0.001$), the more likely full RCs are to be produced in the target. Again, the cumulative effect of RCs produced is greater than the effect of those comprehended. We also found significant cumulative effects of the number of reduced RCs produced, such that more reduced RCs produced make a reduced RC more likely in the target. Interestingly, the effect size for full and reduced RCs is approximately the same. This, too, consistent with an inverse frequency effect, because the two structures have approximately the same frequency.

Discussion The results from Studies 3 and 4 extend the results of Study 2 to syntactic reduction. Syntactic persistence effects on both word order choice and choice in *that*-omission are cumulative. We also found that both for CCs and RCs differences in the strength of syntactic persistence reflect an inverse frequency effect. In this context, it is interesting to point out that, in a sentence recall production experiment, Ferreira found that CCs without a complementizer primed *more strongly* (Ferreira, 2003a). This is the opposite of what we found in spontaneous speech and it seems to conflict with an inverse frequency effect. Note, however, that for the structures in the production experiment, CCs without a complementizer were actually *less* frequent (*ibid*) – again the opposite of what we found for spontaneous speech. In other words, both for sentence recall production experiments and for spontaneous speech, we find that the less frequent structure primes more strongly.

Summary and conclusion

In conclusion, we present evidence that syntactic persistence is sensitive to prime surprisal and also that persistence is cumulative. Although the data is still preliminary, the effects are compelling. Recent evidence from a study on syntactic persistence effects on speakers' choice in Dutch ditransitives, provides further evidence that the surprisal of the prime structure given its verb is positively correlated with prime strength (Bernolet & Hartsuiker, 2007). Surprisal-sensitivity and cumulativeness in syntactic persistence are unexpected by conceptual pact accounts of persistence (Brennan & Clark, 1996, e.g.) or accounts that attribute syntactic persistence effects to social alignment, although they do not rule out that conceptual pacts or social alignment have an independent effect on the probability that a speaker will reuse a structure. If confirmed by further studies, surprisal-sensitivity and cumulativeness provide evidence that language users constantly follow the distribution of linguistic structures in their input and update their mental representations of those distributions accordingly. Specifically, the surprisal-sensitivity we observe for syntactic persistence is consistent with our proposal that the language processing system is set up to implicitly maintain probability distributions of linguistic structures. Further, the cumulativeness effects observed could derive from incremental updating, or learning, of the probability distribution associated with the prime construction as each prime is processed. Syntactic persistence may then be an epiphenomenon of this 'maintenance' work. We are the first to agree, however, that the current evidence for this hypothesis is extremely weak and that more direct tests of our proposal are necessary.

At a more general level, we take the current results to provide evidence that persistence is at least in part due to implicit learning. While surprisal-sensitivity is also compatible with transient activation accounts of syntactic persistence, long term cumulative effects are arguably incompatible with transient activation accounts. One may object, however, that our results only show cumulativeness beyond the first prime and not over long periods of time. In this context, it is worth revisiting the issue of prime decay. We examined whether the distance between prime and target structures affects the prime strength. We found a very weak and barely significant effect in only one of our studies, for passives. The lack or relative weakness of persistence decay is consistent with most laboratory experiments (Bock & Griffin, 2000; Bock et al., in press, on ditransitives and passives) and some corpus studies (Jaeger, 2006a, on complementizer and relativizer omission), but conflicts with several other corpus studies (Gries, 2005; Szmrecsányi, 2005). Recent work by Hartsuiker and colleagues (submitted) may provide an explanation for these apparently conflicting results. In a series of production experiments on the Dutch ditransitive alternation, Hartsuiker and colleagues show that syntactic persistence seems to persist over a long time, while the lexical boost that is observed if the prime and the target verb overlap decays rapidly. Since most verbs in our data sets are simply not frequent enough to lead to situations where prime and target share the same head lemma, the fact that we observe little to no decay of syntactic persistence is consistent with the new data from Hartsuiker and colleagues.

While transient activation may account for some of the observed persistence effects (cf. Hartsuiker et al., submitted), longer-lived effects of previously processed structures, such as cumulativeness and lack of decay in the absence of verb identity between prime and target, support implicit learning accounts.

In ongoing studies, we investigate whether syntactic persistence effects on *that*-omission in complement and relative clauses, too, is affected by prime surprisal. The goal of this work is to achieve a better understanding what makes a structure surprising. Recall that surprisal is defined as low probability. So, we can ask "probable given *what*?". For example, a passive structure could be surprising because of its active-biased verb or it could be surprising because of the properties of the argument NPs (e.g. because the patient NP is much shorter than the agent NP – a factor that would bias strongly towards the passive word order). A passive structure could also be surprising because all recent prime structures were actives. Note that sensitivity to the recent distribution of prime structures is not incompatible with the observed cumulativeness of syntactic persistence. Syntactic persistence can be cumulative, and nevertheless the effect of an individual prime may depend on its surprisal given the distribution of recent preceding prime structures. In sum, what surprisal is conditioned on depends on what cues speakers are sensitive to. That is, it is an empirical question what determines the surprisal of a prime. We plan future experiments to address this question. In particular, we plan to investigate the interplay between cumulativeness and prime surprisal: is there evidence that the recent distribution of prime structures determines the effect of a prime, so that prime structures that have been used

overly frequently in the recent discourse prime less? If so, this would suggest an intriguing link to research on perceptual priming, for which it has been observed that short exposures to a stimulus facilitate responses to similar stimuli, whereas longer presentations provide less facilitation, or even inhibit responses to similar stimuli (e.g. Huber & O'Reilly, 2003, 404).

Acknowledgments

We also thank K. Bock, V. Ferreira, D. Huber, R. Levy, and our anonymous reviewers for their very helpful feedback. We thank N. Sierra for proofreading. FJ's work was supported by a post-doctoral fellowship at the Department of Psychology, UCSD (V. Ferreira's NICHD grant R01 HD051030).

References

- Bernolet, S., & Hartsuiker, R. J. (2007, August). Verb bias modulates syntactic priming for datives. In *Proceedings of Architectures and Mechanisms of Language Processing (AMLaP)*. Turku, Finland.
- Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18(3), 355-387.
- Bock, K., Dell, G., Chang, F., & Onishi, K. (in press). Persistent Structural Priming from Language Comprehension to Language Production. *Cognition*.
- Bock, K., & Griffin, Z. (2000). The persistence of structural priming: Transient activation or implicit learning. *Journal of Experimental Psychology: General*, 129(2), 177-192.
- Branigan, H., Pickering, M., Stewart, A., & McLean, J. (2000). Syntactic priming in spoken production: Linguistic and temporal interference. *Memory & Cognition*, 28(8), 1297-1302.
- Brennan, S., & Clark, H. (1996). Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482-1493.
- Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421), 9-24.
- Feng, Z., McLerran, D., & Grizzle, J. (1996). A comparison of statistical methods for clustered data analysis with Gaussian error. *Stat in Med*, 15, 1793-1806.
- Ferreira, V. S. (2003a). The persistence of optional complementizer mention: Why saying a "that" is not saying "that" at all. *Journal of Memory and Language*, 48, 379-398.
- Ferreira, V. S. (2003b). The processing basis of syntactic persistence: We repeat what we learn. *Paper presented at the 44th Annual Meeting of the Psychonomic Society, Vancouver, Canada*.
- Godfrey, J., Holliman, E., & McDaniel, J. (1992). Switchboard: Telephone speech corpus for research and development. In *Ieee icassp* (p. 517-520).
- Gries, S. (2005). Syntactic Priming: A Corpus-based Approach. *Journal of Psycholinguistic Research*, 34(4), 365-399.
- Hartsuiker, R., Bernolet, S., Schoonbaert, S., Speybroeck, S., & Vanderelst, D. (submitted). Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language*.
- Huber, D. E., & O'Reilly, R. C. (2003). Persistence and accommodation in short-term priming and other perceptual paradigms: temporal segregation through synaptic depression. *Cognitive Science*, 27(3), 403-430.
- Huber, D. E., Shiffrin, R. M., Lyle, K. B., & Ruys, K. I. (2001). Perception and preference in short-term word priming. *Psychological Review*, 108, 149-182.
- Jaeger, T. F. (2006a). *Redundancy and syntactic reduction in spontaneous speech*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
- Jaeger, T. F. (2006b, March). Syntactic persistence in real life (spontaneous speech). In *Proceedings of the 19th annual cuny conference on sentence processing*. New York, NY.
- Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay, & S. Janned (Eds.), *Probabilistic linguistics*. Cambridge, MA: MIT Press.
- Kaschak, M., & Borreggine, K. (in press). Is long-term structural priming affected by patterns of experience with individual verbs? *Journal of Memory and Language*.
- Kaschak, M., Loney, R., & Borreggine, K. (2006). Recent experience affects the strength of structural priming. *Cognition*, 99(3).
- Krauss, R. M., & Fussell, S. (1996). Social psychological models of interpersonal communication. In E. Higgins & A. Kruglanski (Eds.), *Social psychology: A handbook of basic principles* (pp. 655-701).
- Marcus, M. P., Santorini, B., Marcinkiewicz, M. A., & Taylor, A. (1999). *Treebank-3*.
- Pickering, M., & Branigan, H. (1998). The Representation of Verbs: Evidence from Syntactic Priming in Language Production. *Journal of Memory and Language*, 39(4), 633-651.
- Pickering, M., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02), 169-190.
- Reitter, D., Moore, J., & Keller, F. (2006). Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. In *Proceedings of the 28th annual conference of the cognitive science society*.
- Snider, N. (2008). *An exemplar model of syntactic priming*. Unpublished doctoral dissertation, Stanford University.
- Szmrecsányi, B. M. (2005). Language users as creatures of habit: a corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory*, 1, 113-149.