Jaeger, T. F. and N. Snider. (2007). Implicit learning and syntactic persistence: Surprisal and cumulativity. In L. Wolter and J. Thorson (Eds.), *University of Rochester Working Papers in the Language Sciences, 3(1),* 26-44.

Implicit Learning and Syntactic Persistence: Surprisal and Cumulativity[1]

T. Florian Jaeger
Brain and Cognitive Sciences, University of Rochester

Neal Snider
Department of Linguistics, Stanford University

Syntactic persistence (Bock, 1986; Pickering and 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 et al., 1992), the speaker produces the double object ditransitive structure, or NPNP (1-a), as opposed to the prepositional object structure, or NPPP (1-b), after having produced an NPNP 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"
  a. **(NPNP)** give [a country] [money]
  b. **(NPPP)** give [money] [to a country]

In other words, syntactic persistence in production refers to the phenomenon that a structure's posteriori probability of occurring is increased – compared to its a priori probability of occurring – after another instance of the same structure. As such, syntactic persistence is an empirical fact (Bock, 1986; Pickering and Branigan, 1998, inter alia). Relatively little is known, however, about the *underlying mechanisms* that *cause* syntactic persistence. In this paper, we ask why syntactic persistence exists. We aim to distinguish between two prominent theories about the underlying causes of persistence.

According to one view, persistence is a byproduct of 'transient activation' (Branigan et al., 2000; Pickering and Garrod, 2004), where the activation of the representations that have been accessed to produce or comprehend a structure persists for a short time, and so the representation can be used again on the next relevant opportunity. According to another view, syntactic persistence is due to 'implicit learning' (Bock and Griffin, 2000): when language users process a structure, they automatically and implicitly learn something about that structure and the amount of learning determines the probability of reusing that same structure later on.

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 and Griffin, 2000; Bock et al., in press; Hartsuiker et al., 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

[1]Both authors made equal contributions.

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 and Griffin, 2000; Bock et al., in press), while corpus studies have mostly found persistence to decay logarithmically with distance (Gries, 2005; Szmrecsányi, 2005; Reitter et al., 2006, but see Jaeger, 2006b,a). We will get back to this apparent empirical conflict below, where we also discuss how recent work by Hartsuiker and colleagues (submitted) may resolve it. Independent of the empirical facts, however, there is another problem with attempts to distinguish the two hypotheses solely on the basis of the time course of persistence decay. It is unclear how slow or fast a decay would be compatible with either of the theories. In this paper, we look at two other potential properties of syntactic persistence that distinguish between the transient activation and the implicit learning account – SURPRISAL-SENSITIVITY and CUMULATIVITY.

If syntactic persistence is a consequence of implicit learning, this raises the question of what it is that is being learned. It could be that it is the processing of the structure itself that is being learned (Bock and Griffin, 2000). For example, it could be that implicit learning helps to keep alive low-frequency structures (Ferreira, 2003b), structures that are rarely ever produced. Proponents of this view have referred to error-driven learning (Bock and Griffin, 2000; Bock et al., in press; Ferreira, 2003b), where a greater chance of an error leads to more learning (and hence, presumably, increased likelihood of repetition). A related, but conceptually slightly different view would be that it is the probabilistic distribution of structures that is being learned (or rather: maintained). In this view, each structure (e.g. an NPNP structure) can be seen as a piece of evidence that affects the overall probabilistic distribution (e.g. the distribution of an NPNP vs. NPPP structure). This would link syntactic persistence to recent work on perceptual persistence (Huber and O'Reilly, 2003) and skill maintenance (Huber et al., 2001). In either view, less expected prime structures are predicted to prime more (i.e. to lead to a bigger increase in the probability of repetition) than more expected prime structures. We refer to this hypothesis as surprisal-sensitive persistence, where we use the term 'surprisal' in its information theoretic sense referring to the log inverse of the probability, $surprisal(X) = \frac{1}{probability(X)}$. Initial support for surprisal-sensitive persistence comes from the observation that less frequent structures tend to prime more strongly (Bock, 1986; Ferreira, 2003b). For example, Bock (1986) showed syntactic persistence for passives, which are overall rather infrequent in her production experiments (19% or 32 out of 170 active/passive structures were passives). The much more frequent active structure on the other hand were not found to prime.

So far, this so called INVERSE FREQUENCY EFFECT has only been informally observed (see Ferreira, 2003b), but it is compatible with the hypothesis that syntactic persistence is sensitive to surprisal. The first two studies in this paper investigate whether surprisal affects the case-by-case strength of syntactic persistence. If less expected – and hence more surprising – structures prime more strongly, by which we mean that they are associated with a bigger increase in the probability of repetition, this would support an implicit learning account. Study 1 addresses this question using a database of ditransitives, as in (1) above. Study 2 uses a database of actives and passives.

Study 2 also addresses our second question, whether syntactic persistence is cumulative. If syntactic persistence is caused by language users tracking the distribution of structures, as suggested by the implicit learning account, we might expect syntactic persistence beyond the most recent preceding structure. While it would be possible that only the most recent structure affects language users' assumptions over the distribution of that structure, it is easy to see why a system that gives (weighted) access to several preceding primes may be advantageous (under the assumption that implicit learning updates probability distributions). Considering several preceding structures would give language users access to a better estimate of the probability of a structure *given the current context*. In short, if persistence effects turn out to be cumulative, this would support an implicit learning account of syntactic persistence. Evidence in support of cumulativity comes from recent experiments by Kaschak and colleagues (Kaschak et al., 2006; Kaschak and 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 cannot 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. In addition to Study 2, we present two further studies on cumulativity in spontaneous speech. Study 3 and Study 4 investigate *that*-omission in complement and relative clauses, respectively. We end with a brief discussion of the implications for work on syntactic priming and some questions for future work.

1. Ditransitive alternation (Study 1)

The first study tests the effect of prime surprisal in the ditransitive alternation as in (1) above. We use an existing database of ditransitives from spontaneous speech (Bresnan et al., 2007). We extend the best currently available model of the choice between NPPP and NPNP (Bresnan et al., 2007) to contain additional controls and a measure of prime surprisal (here conditioned on the prime verb's subcategorization bias; see below).

1.1 Data

The database compiled by Bresnan and colleagues (2007) contains all 1,108 ditransitives with preceding primes from the full Switchboard corpus (about 2 million words, Godfrey et al., 1992).

1.2 Method

Bresnan and colleagues tested the effects of many factors on alternation choice, including the role of syntactic persistence. They found that speakers are more likely to produce an NPPP structure like (1-b) (the target) if the most recent ditransitive structure (the prime) was an NPPP structure (Recchia, 2007).

We used a multiple logistic regression model in our analysis, containing all factors Bresnan and colleagues found significant (for a formal introduction to logistic regression, see Agresti 2002; for an informal introduction, see Jaeger 2007; see also below). We excluded the standard syntactic persistence factor included in the model of Bresnan and colleagues. Instead, we split the data set into two parts: one data set with all NPPP primes, and one with all NPNP primes. This was done because we are interested in the effect of the prime's surprisal and this effect could differ depending on the prime structure. The NPPP-prime data set had 249 tokens (160 NPNP targets, and 90 NPPP targets), while the NPNP-prime data set had 947 (736 NPNP and 167 NPPP).

Here we estimate the prime's surprisal as the conditional probability of the NPPP structure given the verb, or verb NPPP bias.[2] As an example, take a verb like *cost*, which is highly biased towards NPNP:

(2) a. "A hard disk drive would cost several thousand dollars to the consumer..."[3]
b. "...inaccurate credit information could cost the consumer tens-of-thousands of dollars..."[4]

The verb *cost* is very rarely used in the NPPP structure, but as (2-a) attests, it does occur. The prediction of the surprisal hypothesis is that a verb occurring in a structure that it is biased against (like (2-a)) will make that structure more likely than if the verb were one that is biased towards that structure. We used subcategorization frequencies from the database in Roland and colleagues (in press) to estimate the surprisal associated with a prime given its verb's NPPP bias. This database includes subcategorization biases calculated based on spoken and written corpora from both British and American English. We used estimates of the subcategorization bias based on the Switchboard corpus (spoken American English) and the spoken portions of the British National Corpus (BNC). Other estimates were also considered, but the estimates based on all spoken corpora turned out to be best.[5].

Subcategorization biases of ditransitive verbs can be defined in several ways, two of which are considered here. One is the probability of the verb appearing in the NPPP subcategorization frame versus the NPNP subcategorization frame (which we call ALTERNATION VERB BIAS), another is the probability of the verb appearing in the NPPP subcategorization frame versus all possible subcategorization frames of the verb (which we call OVERALL VERB BIAS). Both definitions were considered in the analyses. If the relevant surprisal of a structure is its surprisal *given the*

[2]As we discuss at the end of this paper, there are other ways in which a prime can be surprising.

[3]Google,4-19-07,www.dtidata.com/resourcecenter/2007/02/08/

[4]Google,4-19-07,www.ftc.gov/os/comments/creditscorefees/000010.pdf

[5]For many ditransitive verbs, the Switchboard alone does not contain a sufficient number of occurrences to get reliable estimates, but when we restricted our database to only those cases for which the prime verb was observed at least five times in the corpus, we obtain the same results as reported here.

intended meaning, we might expect surprisal conditioned on the alternation verb bias to be the more adequate measure, since the overall verb bias is more likely to be influenced by occurrences where the verb has a completely different sense than in the ditransitive construction.

We included three further control factors in our models. The distance in words between the prime and target verbs was included because some theories of syntactic persistence predict decay of persistence effects over time (e.g. Pickering and Garrod, 2004), while others do not (e.g. Bock and Griffin, 2000). We also included a factor coding whether the prime and target verb lemmas were identical. Prime-target verb lemma identity was included as a control because Pickering and Branigan (1998) showed that persistence is stronger when the prime and target verbs are the same. Finally, the *target* verb's NPPP bias (also taken from Roland et al., in press) was included as a control since it captures a lot of the variation. Again, we consider both an estimate based on the alternation verb bias and an estimate based on the overall verb bias.

Separate logistic regression analysis with all of the above factors were conducted for the NPPP-prime data set and the NPNP-prime data set. To avoid overfitting of the model to our sample, we used 40-fold cross-validation and removed all factors whose removal from the model was not at all significant (Ps > 0.7 associated with χ^2 over data log-likelihood ratios). In 40-fold cross validation, model comparison is performed 40 times using 39/40 of the data to train the model and the remaining 1/40 for testing. This is done 40 times so that each 1/40 of the data set is used as a test set once.

1.3 Prediction

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

1.4 Results: NPPP-prime data

Here and in all following studies, we report the coefficient for each independent variable and its levels of significance. Coefficients in logistic regression models are given in log-odds (the space in which logistic models are fitted to the data). For categorical factors, significant positive coefficients mean that a correct answer is more likely in the tested level of the variable than in the other level. For example, if the coefficient of Pronominal Theme for ditransitives is positive, then having a theme that is a pronoun makes the NPPP structure more likely. Negative coefficients mean the opposite. If the coefficient of Animate Recipient is negative, then having a recipient that is animate makes the NPPP structure less likely. For continuous factors, significant positive coefficients indicate how much more likely a correct answer is for each 1 unit increase in the level of the variable. For example if the coefficient for Length of Recipient is 0.72, then for each one word increase in the length of the recipient NP, the NPPP structure is $e^{0.72} = 2.1$ times more likely. Below we also report the difference in odds between conditions (as the name suggests, odds are simply $e^{log-odds}$). 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.

In the smaller NPPP-prime data set, not all of the factors significant in the Bresnan et al. (2007) model were still significant. We removed the following factors from the model because their removal did not affect the model at all: person, number, and concreteness of the recipient, as well as animacy, person, number, and concreteness of the theme. The coefficients in log-odds and standard errors associated with the remaining factors are given in the second and third column of Table 1. The corresponding odds coefficients are given the fourth columns. The fifth and sixth columns summarize the Wald's Z statistic, which tests whether the coefficients are significantly different from zero (given the estimated standard error). Finally, the last two columns give the χ^2 over the change in data likelihood ($\Delta_x(\Lambda)$) associated with the removal of the predictor (x) from the final model. The latter test is more robust against collinearity in the model (Agresti, 2002). The χ^2 value, which literally corresponds to the difference in the model's data likelihood without the predictor, can be seen as a measure of the predictors's importance in the model. The Wald-test is included because it implicitly test the directionality of the effect (unlike the χ^2 over the change in data likelihood).

Table 1: Summary of ditransitive analysis

Predictor	*Parameter estimates*			*Wald's test*		$\Delta_x(\Lambda)$*-test*	
(independent variable)	Log-odds	S.E.	Odds	Z	P	χ^2	P
verb class = communication	-4.459	1.354	0.02	-3.3	$\ll 0.001$	10.7	< 0.002
recipient pronominal	-2.251	0.929	0.08	-2.4	< 0.02	6.4	< 0.02
theme pronominal	1.973	0.822	9.2	2.3	< 0.02	7.1	< 0.008
recipient indefinite	2.261	1.237	6.6	1.8	< 0.06	2.9	< 0.09
theme indefinite	-3.412	0.911	0.03	-3.7	$\ll 0.001$	14.6	$\ll 0.001$
recipient inanimate	4.809	1.364	167	3.5	$\ll 0.001$	14.3	$\ll 0.001$
log argument length difference	-2.063	0.576	0.10	-3.5	$\ll 0.001$	14.5	$\ll 0.001$
target verb bias	4.359	1.543	142	2.8	< 0.005	7.3	$\ll 0.001$
prime V = target V	0.806	0.688	2.2	1.1	> 0.3	1.4	> 0.2
prime verb bias	-5.248	2.632	0.008	-1.9	< 0.05	3.9	< 0.05

The first seven rows in Table 1 summarize controls from Bresnan et al. (2007). Rows eight and nine summarize our additional controls. The final row shows that the prime's surprisal based on the alternation verb bias (see above) is a significant predictor of the NPPP structure, with a negative log-odds coefficient $B = -5.248$ ($SE = 2.632$), and a corresponding odds coefficient $e^B = 0.008$. Thus, NPPP primes with verbs that are biased against the NPPP structure make it more likely that the target will be NPPP than NPNP-biased verbs, as shown in Figure 1. If prime surprisal is estimated based on the overall verb bias, it is not significant (not shown above).

All control factors that reach significance have the expected results. We also found that the target's alternation verb bias was highly significant, such that NPPP-biased target verbs are more likely to occur in the NPPP structure. We found no significant effects of prime-target verb identity or prime-target distance.

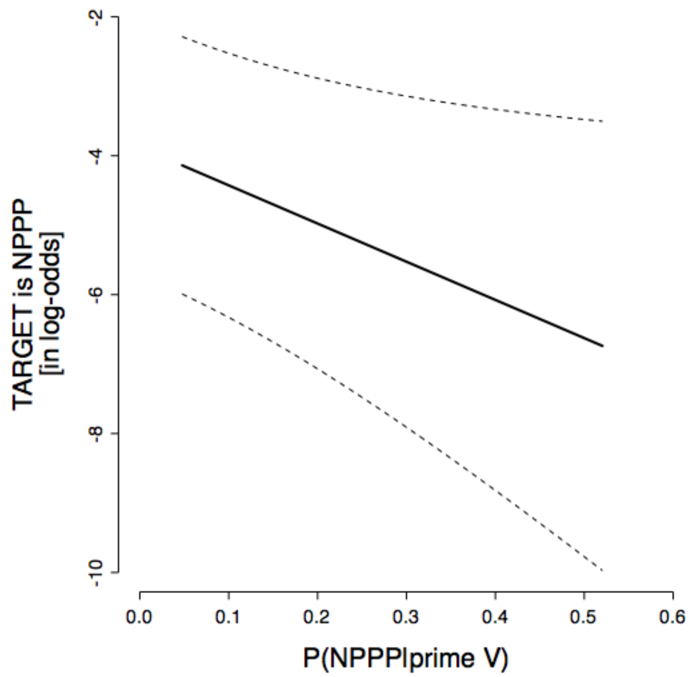

Figure 1: Prime surprisal (based on prime verb's NPPP bias)

1.5 Results: NPNP-prime data

The same study was carried out on the NPNP prime data set, but no effect of prime verb bias was found, although all other control factors have the expected effects.

1.6 Discussion

The inverse effect of prime surprisal (here conditioned on the prime verb's subcategorization bias) supports implicit learning accounts of syntactic persistence. The lack of an effect in the NPNP-prime data is consistent with the observation that more frequent structures seem to prime less strongly (the inverse frequency effect). Alternatively, there could be less variation with regard to the prime verbs' subcategorization bias in the the NPNP-prime data. We will address this question in future work. Further support for implicit learning accounts may come from the fact that we found no effect of prime-target distance. However, given the small number of ditransitives in our data sets, this null effect may simply reflect a lack of power.

2. Voice alternation (Study 2)

The first study showed the predicted effect of prime verb surprisal in syntactic persistence of the ditransitive alternation. To provide further evidence for the surprisal effect, we collected another data set, this time utilizing the voice alternation - that is, the choice between an active or passive structure.

2.1 Surprisal in the voice alternation

2.1.1 Data

We extracted the voice alternation data set from the manually parsed Penn Treebank (release 3, approximately, 800,000 words, see Marcus et al., 1999) portion of the Switchboard corpus. We first extracted all passives from the Treebank Switchboard. We considered a structure passive if it 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). We included only the tokens that had a prime and for which the verb occurred at least 10 times in the entire data set (including occurrences without a prime), which yields 1,791 passives and 29,007 actives. Verbs with fewer occurrences in the corpus were excluded from the analyses because the surprisal estimates for them are too unreliable. We also excluded verbs that do not participate in the voice alternation (*e.g. suppose*), despite being coded as passives in the Treebank. Finally, we added prime and target verb bias to the data set, which was determined 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).

2.1.2 Method

To test the effect of surprisal in passives, we analyzed the data set with mixed logit models (Breslow and Clayton, 1993; Bates and Sarkar, 2006). Mixed logit models can be thought of as an extension of logistic regression that includes modeling of random subject effects. Inclusion of random effects is necessary to generalize beyond the subjects in the current data (Clark, 1973). Subject effects could be controlled in this study, unlike Study 1, because the data set contained sufficient numbers of examples from each speaker. The interpretation of coefficients in a mixed logit model is the same as in the logistic regression in the previous study.

The dependent variable was structure choice, active or passive. For independent variables, in addition to prime and target verb bias, we added the distance between prime and target in utterances, to test for prime decay. We also added a control factor that indicated whether the prime and target verbs were the same. Finally, as mentioned above, we added a random effect of speaker, to control for the fact that different speakers might have different rates of passivization. As in the first study, the data set was split into those tokens with passive primes and those with active primes.

2.1.3 Prediction

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

2.1.4 Results: Passive-prime data

The results of our models of priming and surprisal in the voice alternation are in Table 2.

Table 2: Summary of passive surprisal analysis

Predictor	*Parameter estimates*			*Wald's test*		$\Delta_x(\Lambda)$*-test*	
(independent variable)	Log-odds	S.E.	Odds	Z	P	χ^2	P
prime V = target V	1.754	0.259	5.75	6.8	$\ll 0.001$	46.8	$\ll 0.001$
target verb pass. bias	1.212	0.025	3.35	47.9	$\ll 0.001$	3489.7	$\ll 0.001$
prime-target distance	-0.060	0.026	0.94	-2.3	< 0.04	6.7	< 0.01
prime verb pass. bias	-1.205	0.404	0.30	-3.0	< 0.002	9.2	< 0.003

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, as shown in Figure 2. 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 $e^B = 0.94$. This can be interpreted as a decrease of 6% in the likelihood of producing a passive when a passive prime has been processed for each utterance that intervenes between prime and target. However, further evidence will be presented in the next section that the persistence of passives is still long-lived – passive persistence accumulates.

2.1.5 Results: Active-prime data

The same study was carried out on the active prime data set, but no effect of prime verb bias was found, although all other control factors have the expected effects.

2.1.6 Discussion

The results of this study extend the results of Study 1: syntactic persistence effects on both NPPP structures and passives seem to be sensitive to the prime's surprisal. At the same time, no such surprisal-sensitivity is observed for NPNP structures and actives. That is, in both studies, 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 structure (cf. the inverse effect

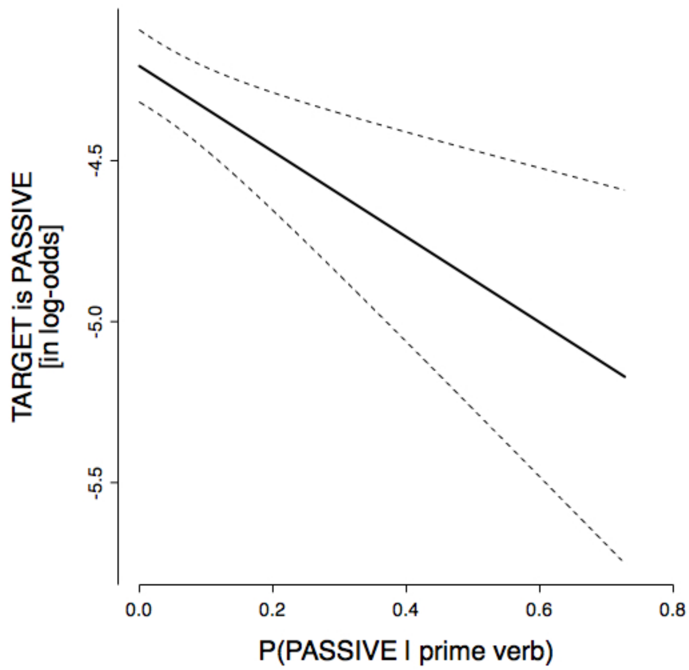

Figure 2: Prime surprisal (based on prime verb's passive bias)

frequency, e.g. Ferreira, 2003b). Another potential source for these null effects is noise in our data sets. The results presented here are somewhat preliminary in nature and we plan to include further controls to our models of the voice alternation.

As for decay of persistence, our evidence has been mixed: we found no significant effect of distance in datives, but did find a distance effect in passives. However, there is still might be evidence that the persistence effect is long lived in the passives if it can be shown to accumulate. 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. This effect can be found in the work of Kaschak and colleagues (Kaschak et al., 2006; Kaschak and Borreggine, in press), although showing cumulative persistence was not the point of the study, so they do not present evidence for a proportional effect of cumulative primes.

2.2 Cumulativity in the voice alternation

2.2.1 Data

We first examined cumulativity in the passive data set. Cumulativity was measured by counting the number of primes of each structure that was comprehended or produced by the speaker in the conversation up to the point of the target (excluding the immediately previous 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).

2.2.2 Method

Using the same mixed logistic model as in the last study, and all the same controls, we added four cumulativity variables. 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 comprehended previously by the target speaker, the number of actives comprehended previously by the target speaker.

For the current study, it was especially crucial to control for speakers' inherent biases towards active or passive structures, since this could potentially confound the two measures of cumulative within-speaker persistence.

2.2.3 Prediction

The learning-cumulativity hypothesis predicts that the more passive primes a speaker has produced or comprehended, the more likely they are to produce a passive structure in the target. Conversely, the more actives produced or comprehended, the less likely a passive target will be produced.

2.2.4 Results

The results of our models of cumulative priming in the voice alternation are shown in Table 3. We found a significant effect of the number of passives produced within speaker ($p \ll 0.001$),

Table 3: Summary of passive cumulativity analysis

Predictor	*Parameter estimates*			*Wald's test*		$\Delta_x(\Lambda)$*-test*	
(independent variable)	Log-odds	S.E.	Odds	Z	P	χ^2	P
prime=*passive*	1.036	0.169	2.82	6.1	$\ll 0.001$	35.1	$\ll 0.001$
prime V = target V * prime=*pass.*	1.757	0.259	5.75	6.8	$\ll 0.001$	46.8	$\ll 0.001$
target verb pass. bias	1.212	0.025	3.35	47.9	$\ll 0.001$	3490.1	$\ll 0.001$
prime verb pass. bias * prime=*pass.*	-1.205	0.404	0.30	-3.0	< 0.002	9.2	< 0.003
prime-target distance * prime=*pass.*	-0.059	0.026	0.94	-2.3	< 0.04	6.7	< 0.01
prec. passives within-speaker	0.107	0.016	1.12	6.6	$\ll 0.001$	36.9	$\ll 0.001$
prec. passives across-speaker	0.055	0.020	1.05	2.8	< 0.02	7.6	< 0.006
prec. actives within-speaker	-0.008	0.002	0.99	-4.1	$\ll 0.001$	16.2	$\ll 0.001$
prec. actives across-speaker	-0.001	0.002	0.99	-0.6	< 0.6	0.3	> 0.5

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. This is consistent with the results of Bock and colleagues (in press), who found that production-to-production persistence is stronger that comprehension-to-production. Also, the significant effect of passives comprehended could not be confounded by a speaker effect. 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 cumulative effect of the number of actives produced.

These effects are illustrated in Figure 3. The y-axis of each of the four graphs represents the log-odds of producing a passive. The x-axis of the top two graphs represents the count of passives, and in the bottom two, it represents the count of actives. Further, in the left two graphs, the x-axis shows the count of structures within-speaker (those produced by the target speaker), while in the right two graphs, the x-axis shows the count of structures across-speaker (those comprehended by the target speaker).

2.2.5 Discussion

The passive data from Study 2 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. However, further tests revealed that we do not find a significant effect of cumulativity for the database of ditransitives used in Study 1. This null effect could be due to a lack in power – the database for Study 1 is much smaller than the database for Study 2.

3. Complementizer and relativizer omission (Study 3 & 4)

Next we look at cumulative persistence in two alternations that differ qualitatively from the syntactic choices in Study 1 and Study 2: *that*-omission in complement clauses and relative clauses. Such syntactic reduction does not involve any word order choices, but rather the choice between a full form (with *that*) and a reduced form (without *that*).

3.1 Data

The data for Study 3 & 4 comes from Jaeger (2006a; 2006b), and consists of 5,701 complement clauses (CCs), as in (3), and 2,071 relative clauses (RCs), as in (4), respectively. The CCs and RCs were extracted from the Switchboard portion of the Penn Treebank (Marcus et al., 1999).

(3) a. "… and i don't believe [$_{CC}$ any of us would have to purchase any extra vacation days] …"
b. **(Full CC)** believe [$_{CC}$ that any of us …
c. **(Reduced CC)** believe [$_{CC}$ any of us …

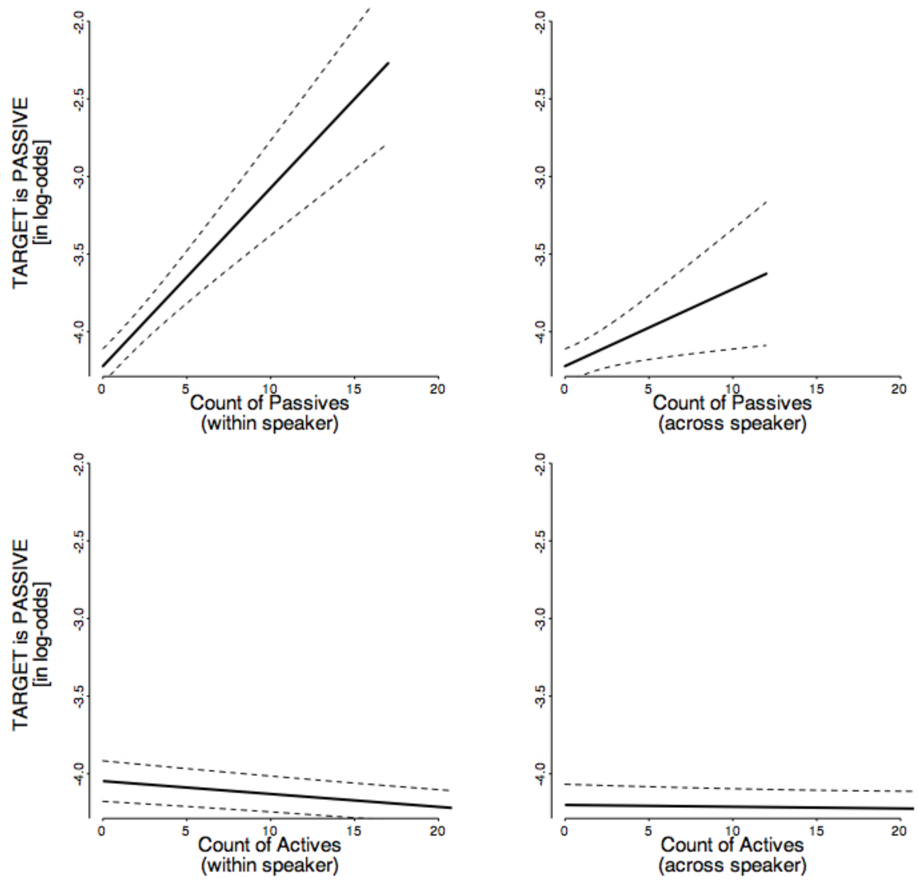

Figure 3: Cumulativity in passives

(4) a. "[MUMBLE] the only thing [$_{RC}$ we get paid for] are the aluminum cans ..."
b. **(Full RC)** the only thing [$_{RC}$ that we ...
c. **(Reduced RC)** the only thing [$_{RC}$ we ...

In English, CCs like (3-a) and RCs like (4-a) can be produced with or without the *that*. Jaeger (2006b) showed that complementizer and relativizer omission are affected by syntactic persistence while controlling for other factors known to affect *that*-omission.

3.2 Method

We used the ordinary logistic regression (as in Study 1) developed in Jaeger (2006a; 2006b) and added measures for the cumulative number of primes, just like we did for the passive data set (here the number of full CC/RCs and reduced CC/RCs, both produced and comprehended). To capture speaker effects, the model was bootstrapped 10,000 times with random replacement of speaker clusters (see Jaeger, 2006a, Ch. 2, mixed effect models with speakers as random variable return the same results). The models reported here do *not* contain prime-target distance as a control factor because none of several prime-target distance measures were found to be significant for complementizer/relativizer omission in Jaeger (2006b).

For brevity's sake we do not report the complete models below (there are over 35 fitted parameters in each of the models). None of the control effects changed qualitatively compared to Jaeger (2006a, Ch 3 & 4).

3.3 Results: Complementizer omission

Table 4 summarizes the effects of the four predictors associated with cumulative persistence after all other controls are accounted for.

Table 4: Summary of CC analysis

Predictor	*Parameter estimates*			*Wald's test*		$\Delta_x(\Lambda)$*-test*	
(independent variable)	Log-odds	S.E.	Odds	Z	P	χ^2	P
prec. full CCs within-speaker	0.214	0.031	1.24	6.9	$\ll 0.001$	47.4	$\ll 0.001$
prec. full CCs across-speaker	0.049	0.040	1.05	1.2	> 0.2	1.5	> 0.2
prec. red. CCs within-speaker	-0.039	0.017	0.96	-2.3	< 0.03	5.3	< 0.03
prec. red. CCs across-speaker	-0.045	0.019	0.96	-2.4	< 0.02	5.7	< 0.02

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 again consistent with an inverse frequency effect, because the full CC is the less frequent structure. The cumulative persistence effects are visualized in Figure 4. We found no effect of prime-target distance.

3.4 Results: Relativizer omission

Table 5 summarizes the effects of the four predictors associated with cumulative persistence after all other controls are accounted for. We again found significant cumulative persistence effects:

Table 5: Summary of RC analysis

Predictor	*Parameter estimates*			*Wald's test*		$\Delta_x(\Lambda)$*-test*	
(independent variable)	Log-odds	S.E.	Odds	Z	P	χ^2	P
prec. full RCs within-speaker	0.170	0.041	1.19	4.1	$\ll 0.001$	17.1	$\ll 0.001$
prec. full RCs across-speaker	0.035	0.045	1.04	0.8	> 0.4	0.6	> 0.4
prec. red. RCs within-speaker	-0.213	0.048	0.81	-4.5	$\ll 0.001$	20.0	$\ll 0.001$
prec. red. RCs across-speaker	-0.010	0.049	1.01	0.2	> 0.8	0.1	> 0.8

the more full RCs that have been produced ($p < 0.001$) or comprehended ($p < 0.001$), the more

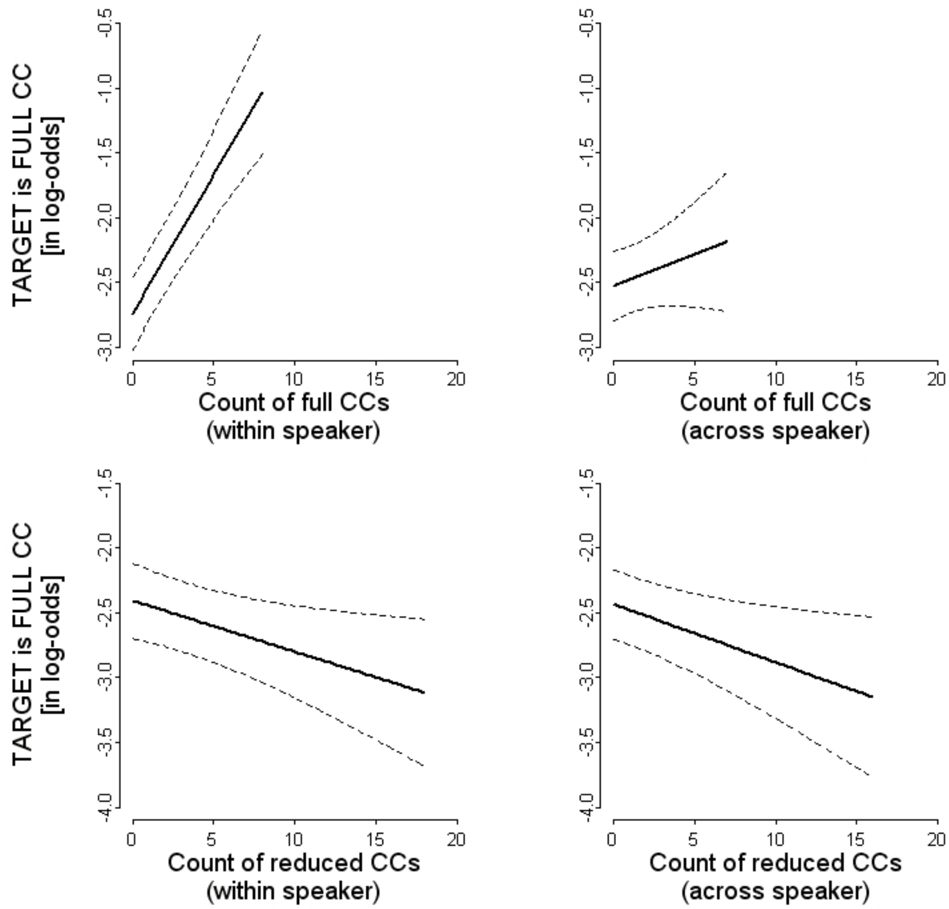

Figure 4: Cumulativity in complementizer omission

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 is also consistent with an inverse frequency effect, because the two structures have approximately the same frequency. The effects are visualized in Figure 5. We found no effect of prime-target distance.

3.5 Discussion

The results from Study 3 & 4 extend the results of Study 2. Syntactic persistence effects on both word order choice and choice in *that*-omission are cumulative. Within speaker persistence seems to be generally stronger than between speaker persistence in our studies. We also found that both for CCs and RCs differences in the strength of syntactic persistence between the full and the

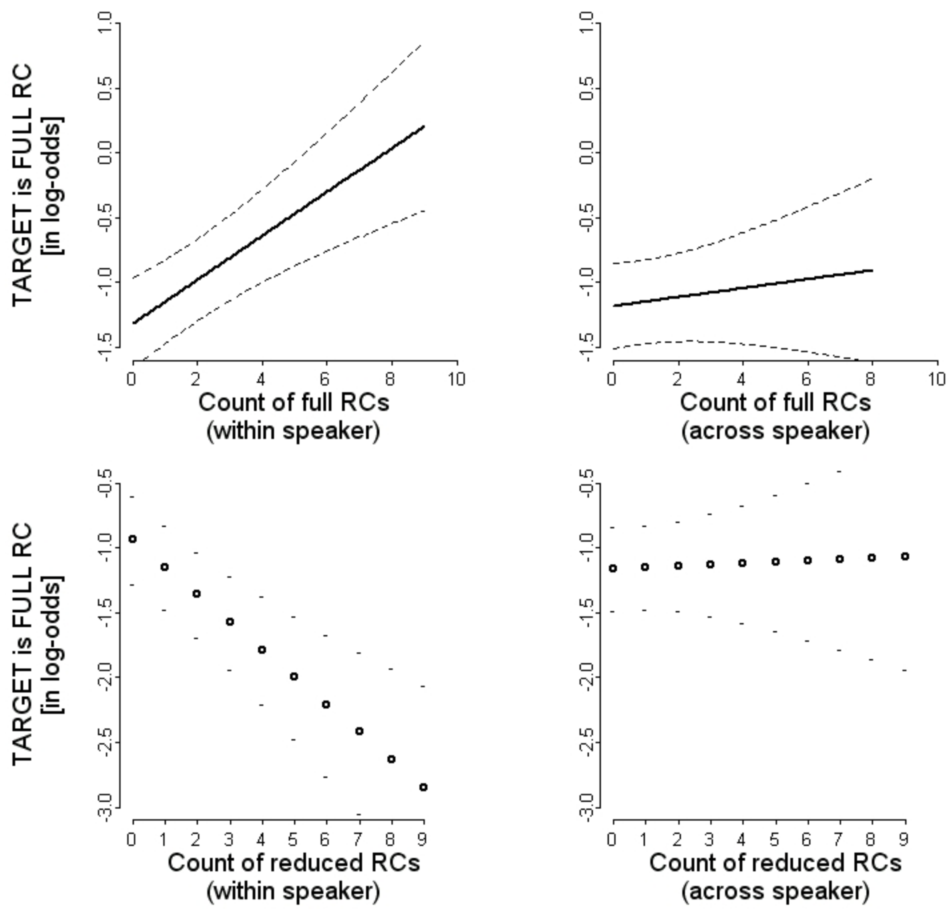

Figure 5: Cumulativity in relativizer omission

reduced variants are consistent with an inverse frequency effect. In this context, it is interesting to point out that, in a sentence recall production experiment, V. 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 in the production experiment, CCs without a complementizer were actually *less* frequent (ibid; see also Ferreira and Dell, 2000) – 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.

4. Concluding remarks

In conclusion, we present several pieces of 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. We believe these effects argue for a type of implicit learning account of syntactic persistence. Just like learning, persistence seems to be error driven, because the persistence effect is stronger the less expected the prime structure is. We found this effect of prime surprisal in two structures, the voice and ditransitive alternations, using data sets from naturally-occurring ecologically valid conversation.

We also showed that persistence is sufficiently long-lived as to be cumulative, as would be expected in an implicit learning account of persistence. We found cumulative persistence in passives, complementizer omission, and relativizer omission. Cumulativity has been observed previously in laboratory experiments (Kaschak et al., 2006; Kaschak and Borreggine, in press). Our results come from natural occurring conversational data, which shows that 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).

Two of our studies also examined whether the distance between prime and target structures affects the prime strength. We found no effect for ditransitives (contrary to Gries, 2005), and only a very weak and barely significant effect for passives. The lack of or relative weakness of persistence decay is consistent with most laboratory experiments (Bock and Griffin, 2000; Bock et al., in press, on ditransitives and passives) and some corpus studies (Jaeger, 2006a, on complementizer and relativizer omisssion), 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.

Taken together the lack of decay (in the absence of prime-target lemma overlap), cumulativity, and surprisal-sensitivity suggest that the underlying mechanism that causes so called syntactic persistence effects is something akin to implicit learning.

In ongoing studies, we investigate whether syntactic persistence effects on *that*-omission in complement and relative clauses, too, is affected by prime surprisal. We are also interested in what makes a structure surprising. Recall that surprisal is defined as low probability. So, we can ask "probable given *what*?". For example, an NPNP structure could be surprising because of the ditransitive verb or it could be surprising because of the properties of the two NPs (e.g. because the second NP is much shorter than the first – a factor that would bias strongly towards the opposite word order). An NPNP structure could also be surprising because all recent prime structures were NPPPs. Note that sensitivity to the recent distribution of prime structures is not incompatible with the observed cumulativity 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 cumulativity 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 and O'Reilly, 2003, 404).

Acknowledgments

We are grateful to Joan Bresnan, Gabe Recchia, and colleagues for providing us with the database of ditransitive prime-target pairs from the Switchboard. We also thank Kay Bock, Vic Ferreira, David Huber, Roger Levy, and audiences at CRL (UCSD), SPLAT (Stanford University), the University of Rochester, and CUNY 2007 for their very helpful feedback and questions (many of which are yet to be addressed). We thank Natalie Sierra for proofreading. FJ's work was supported by a post-doctoral fellowship at the Department of Psychology, UC San Diego (V. Ferreira's NICHD grant R01 HD051030).

References

Agresti, A., 2002. Categorical Data Analysis. Wiley.

Bates, D., Sarkar, D., 2006. lme4: Linear mixed-effects models using S4 classes. R package version 0.995-2.

Bock, J., 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.

Breslow, N. E., Clayton, D. G., 1993. Approximate inference in generalized linear mixed models. Journal of the American Statistical Association 88 (421), 9–24.

Bresnan, J., Cueni, A., Nikitina, T., Baayen, H., 2007. Predicting the dative alternation. Proceedings of the Royal Netherlands Academy of Science Workshop on Foundations of Interpretation. Amsterdam.

Clark, H., 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior 12 (4), 335–359.

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.

Ferreira, V. S., Dell, G. S., 2000. The effect of ambiguity and lexical availability on syntactic and lexical production. Cognitive Psychology 40, 296–340.

Godfrey, J., Holliman, E., McDaniel, J., 1992. Switchboard: Telephone speech corpus for research and development. In: IEEE ICASSP. pp. 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., O'Reilly, R., 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. Ph.D. thesis, Stanford University, Stanford, CA.

Jaeger, T. F., March 2006b. Syntactic persistence in real life (spontaneous speech). In: Proceedings of the 19th Annual CUNY Conference on Sentence Processing. CUNY, New York, NY.

Jaeger, T. F., 2007. Better categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. Submitted.

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).

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.

Recchia, G., 2007. Strata: Search tools for richly annotated and time-aligned linguistic data. Stanford Undergraduate Honors Thesis.

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.

Roland, D., Dick, F., Elman, J. L., in press. Frequency of basic english grammatical structures: A corpus analysis.

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.

UNIVERSITY OF ROCHESTER WORKING PAPERS IN THE LANGUAGE SCIENCES — VOL. 3, NO. 1

Lynsey Wolter and Jill Thorson, Editors
Joyce McDonough, Series Editor