

Exemplars in Syntax: Evidence from Priming in Corpora

Neal Snider

Department of Linguistics
Stanford University
Stanford, CA 94305
snider@stanford.edu

Abstract

Two studies are presented that show that exemplar frequency and similarity interacts with syntactic priming. In naturally-occurring spoken corpus data from the dative alternation and relative clause attachment, the strength of the priming effect is shown to depend on the exemplar frequency of the prime structure. Target exemplars that are more frequent are also more likely to occur. Also, using a precise definition of exemplars and exemplar similarity, more similar prime and target exemplars are shown to be more likely to prime. This is evidence for an exemplar model of the mental representation of syntactic knowledge, because the exemplar hypothesis predicts that exemplar similarity and the frequency of occurrence of exemplars will affect language processing.

1 Introduction

Exemplar models of language hold that entire strings of words are stored with their lexical and even phonetic content. Most of the evidence so far for exemplar models has come from phonology and morphology (Bybee, 2000; Bybee and McClelland, 2005; Krott et al., 2006), with the work on syntax consisting primarily of computational models (Bod, 2006), although Hay and Bresnan (2006) is an exception. The studies presented here provide empirical evidence for an exemplar model of syntactic representation by demonstrating that exemplar frequency and similarity affects syntactic priming.

Syntactic priming (Bock, 1986; Bock and Griffin, 2000) is the tendency for speakers to produce a syntactic construction if they have recently

been exposed to that construction. Syntactic priming is interesting because it can be used to probe the nature of the mental representation of linguistic structures. As Branigan and Pickering argue (1995), priming results can be understood as a tendency to repeat structures (produce a 'target') that has been heard or comprehended (the 'prime') when the prime and target are similar along some dimension. If speakers' behavior is sensitive to this similarity, then that similarity must arise from the two structures having the same cognitive representation of that dimension. Thus, by experimentally finding the dimensions of similarity between structures that cause priming, one can determine the nature of the mental representations of those structures. Applying this argument to the exemplar model of syntactic representation, exemplar models provide other dimensions of cognitive similarity for priming, and these can be tested by calculating exemplar frequency and similarity. If two exemplars are similar along their dimensions of representation, then one should be more likely to prime the other. The key prediction of exemplar models is that the exemplars must be compared along dimensions of very fine-grained detail. If such a similarity metric is found to influence language production, then that is a strong argument for the exemplar hypothesis. Similarly, the frequency of those exemplars should affect their likelihood of priming and production. Thus, if a speaker's knowledge of syntax consists of stored fragments of language, then syntactic production, and syntactic priming, should be influenced by the frequency and similarity of these string tokens. Models of representation that do not include exemplar storage do not predict this interaction of priming with exemplar frequency and similarity.

In order to investigate whether priming is af-

ected by exemplar frequency and similarity, two corpus studies of English will be presented. These studies show priming is very sensitive to frequency and similarity effects. This is important empirical evidence that the mental representation of syntactic knowledge consists of exemplars. First, I will discuss a background study that was the first to show how syntactic priming is affected by the frequency of the prime, using data from the ditransitive alternation. Next, I present a new study showing frequency effects on priming of relative clause attachment. In a second study, I show that a precise metric of exemplar similarity, the nearest-neighbor distance, predicts the likelihood of priming. Finally, I conclude and discuss future work.

2 Background

Turning first to frequency effects on priming, the prediction of the exemplar model for the particular direction of the interaction between priming and frequency is not obvious. For the target, the direction of the effect of exemplar frequency is clear, more frequent exemplars should be more likely to be produced. Therefore, a positive effect on target production is predicted for target frequency. However, the direction of the effect of prime frequency is more interesting. It has been noted (Bock, 1986) that syntactic priming is sensitive to the frequency of the prime construction, with infrequent constructions priming more. This makes sense if priming is implicit learning (Bock and Griffin, 2000; Jaeger and Snider, 2007), which predicts that less frequent events lead to more activation (and hence more learning). This predicts a negative effect of exemplar frequency on priming strength.

In the first study to show this negative effect of prime frequency on the likelihood of priming, Jaeger and Snider (Jaeger and Snider, 2007) examined data from the ditransitive alternation, for example:

- (1)a. (NPPP) Yeah, I haven't **given much thought to it**, Im kind of busy raising my kids (SWBD)
 b. (NPNP) Yeah, I haven't **given it much thought**, Im kind of busy raising my kids

They used an existing database of ditransitives from spontaneous speech (Bresnan et al., 2007), and extended the best currently available model of

the choice between NP PP and NP NP (Bresnan et al., 2007) to contain additional controls and what I argue is a partial measure of exemplar frequency (the prime verb's subcategorization bias; see below).

Their data consisted of the database compiled by Bresnan and colleagues (2007)¹, which contains all 1,108 ditransitives with preceding primes from the full Switchboard corpus of approximately 2 million words (Godfrey et al., 1992). Bresnan and colleagues tested the effects of many factors on alternation choice in the ditransitive, including the role of syntactic persistence. They found that speakers are more likely to produce an NPPP structure like (1a) (the target) if the most recent ditransitive structure (the prime) was an NP PP structure (Recchia, 2007).

Similar to the two new studies presented later, Jaeger and Snider analyzed the data using a multiple logistic regression model. To ensure that their priming effects were not artifacts of other features of the local discourse environment, their model contained as controls all factors Bresnan and colleagues found significant in predicting the dative alternation (for a formal introduction to logistic regression, see Agresti (2002); for an informal introduction, see Jaeger (2007); see also below). They excluded the standard syntactic persistence factor used in the models of Bresnan and colleagues, instead, splitting the data set into two parts: one data set with all NP PP primes, and one with all NP NP primes. This was done because the effect of the prime's frequency differs depending on the prime structure. Their study did not assume an exemplar model of representation, but they added a factor of frequency of the head verb in each possible ditransitive alternation, and this can be interpreted as a partial measure of exemplar frequency. The verb bias is defined as the conditional probability of the NP PP structure given the verb. As an example, take a verb like *cost*, which is highly biased towards NP NP:

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

¹The data set is available in the languageR package for the R statistical language

²Google,4-19-07, www.dtidata.com/resourcecenter/2007/02/08/

³Google,4-19-07,www.ftc.gov/os/comments/creditscorefees/000010.pdf

The verb *cost* is very rarely used in the NPPP structure, but as (2a) attests, it does occur. The prediction of the exemplar hypothesis is that the likelihood of a structure being repeated is affected by how likely the prime structure was to have occurred.

Here and in my studies following, I describe 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 NP PP 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. In my later studies, I also report the difference in odds between conditions (as the name suggests, odds are simply $e^{\log\text{-odds}}$). 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 I talk about an x-fold increase or decrease in odds between conditions.

Jaeger and Snider found that the prime's verb bias (exemplar frequency) is a significant predictor of the NPPP structure ($p < 0.05$), with a negative log-odds coefficient. Thus, NPPP primes containing verbs that are biased against the NP PP structure make it more likely that the target will be NPPP than those containing NP PP-biased verbs. They also found a positive effect of target verb bias, or exemplar frequency ($p \ll 0.001$), with a positive log-odds coefficient, so verbs that more frequently occur in the NPPP structure make the NPPP structure more likely. In the NP NP-prime data set, they also found a positive effect of target frequency ($p \ll 0.001$), but no effect of prime frequency.

Jaeger and Snider's study showed that priming is sensitive to the verb biases of both the prime and target. It also shows that the direction of the effect of prime probability is negative, as less likely primes are more likely to be repeated. These effects of verb bias in the prime and target could be naturally explained by an exemplar model of syntactic representation, where the representations that produce the priming effect essentially contain lexical detail about the prime and target constructions. However, one could also argue that these results could be explained by the storage of frequencies in an abstract representation of argument structure, since the effects involve the biases of head verbs in particular argument structures. Thus I undertook the Study 1 to demonstrate the effect of exemplar frequency in priming non-argument constructions.

3 Relative clause attachment height (Study 1)

In this study, I examine exemplar frequency effects on priming in a structure that does not involve an argument relation: relative clauses (RCs), which are not usually considered to be arguments of their head nouns. Specifically, I investigate the priming of RC attachment in multiply-embedded NPs. As was first shown in an experiment by Scheepers (2003), subjects are more likely to complete a sentence with an RC attached "high" (to the first NP in an embedded NP structure like in (3a)) when they have previously processed such a high attachment, than when they had previously processed a "low" attachment (to the second NP).

- (3a. **(HI)** I would definitely consider a try to find [NP₁a school within [NP₂the state]] [RCthat I liked well enough to attend] (SWBD)
- b. **(LO)** I would definitely consider a try to find [NP₁a school within [NP₂the state]] [RCthat I live in]]

3.1 Data

To test for priming and frequency effects in this structure, all the instances of two embedded NPs modified by an RC were extracted from the Penn Treebank Switchboard corpus (Marcus et al., 1994), a corpus of naturalistic, conversational data. There are a total of 610 prime-target pairs, but I excluded those for which there was no prime and for which I did not have a probability estimate

(see below), leaving 272 prime-target pairs (190 LO-attached targets and 82 HI-attached targets).

3.2 Method

I again analyzed the data with multiple logistic regression models, with the HI attachment as the positive response. Because priming of this construction in naturalistic conversation has not been reported previously, a control study was undertaken, to determine if the prime RC attachment height affects the target attachment height. This was indeed the case, for a priming effect was found such that having a preceding HI attachment in the discourse makes a HI attachment much more likely in the next instance ($p < 0.001$).

Having established priming for this construction in a corpus, exemplar frequency was operationalized as the frequency with which each of the two head nouns (*school* and *state* in the example above) occur with an RC, or the probability of the RC given the head noun. This probability was shown to be a predictor of relativizer (*that*) absence in Jaeger (2006), and the probabilities derived in that study were generously provided by Florian Jaeger for use in this one. Jaeger’s study actually examined non-subject extracted relative clauses (NSRCs), so the probability measure I use is actually $P(\text{NSRC} \mid \text{head N})$, which I consider to be an estimate of $P(\text{RC} \mid \text{head N})$. Since the previous study, and other studies in Jaeger and Snider (2007), showed that there is an inverse frequency effect in priming, a prime exemplar frequency factor was added that reflects this.

In order to determine how to calculate exemplar based expectation in a structure such as multiply-embedded NPs, which essentially have two heads to which an RC could attach, one needs to consider what would make a prime unexpected, or *surprising*, in the information theoretic sense of surprisal ($\text{surprisal}(X) = \frac{1}{\text{probability}(X)}$). If the prime is actually observed with a HI attached RC, then that prime would be considered unexpected if the LO NP had a high probability of being modified by an RC. Similarly, if the prime structure had a LO attached RC, then that would be unexpected if the HI NP had a high probability of being modified by an RC. Therefore, the prime exemplar surprisal factor was added as the probability of the **non**-modified noun in the prime to take an RC. Thus if the prime had a HI attachment, the prime probability was the probability of the LO noun to have an RC, and if

the prime had a LO attachment, the prime probability was the **negative** probability of the HI noun to have an RC. The factor is negative for LO attached primes, because these are expected to make the a HI attachment in the target **less** likely. The prime probability factor is defined in equation (1). This factor reflects the prediction that unexpected primes have a stronger effect: if the LO noun of the prime has a high likelihood of being modified by an RC, and the HI noun was actually the one modified by the RC, then this will increase the likelihood that the target will be a HI attachment.

To examine the exemplar frequency effects in the target, I added the probability of the HI and LO target nouns to be modified by an RC. As a final control, I added the distance in utterances between the prime and target.

3.3 Prediction

The exemplar hypothesis predicts that the exemplar frequency of the RC should affect processing. In this case, it predicts that the frequency of the prime RC (the probability that an RC would attach to its head noun in the prime structure) will affect the likelihood of priming. I expect that the probabilistic surprisal of an RC given the non-modified noun in the prime will interact positively with priming. The hypothesis also predicts that the probability of the target nouns to be modified by an RC will affect their likelihood of being modified in the target: the more likely an RC given the HI target noun, the more likely the target will be a HI attachment (a positive coefficient in the model), and the more likely an RC given the LO target noun, the more likely a LO attachment (a negative coefficient).

3.4 Results

The results of my models of priming of RC attachment are in Table 1. 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 in the fourth column. 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

$$\text{prime surprisal} = \begin{cases} P(RC|LO \text{ noun}) & \text{if prime = HI} \\ -P(RC|HI \text{ noun}) & \text{if prime = LO} \end{cases} \quad (1)$$

Table 1: Summary of RC attachment analysis

<i>Predictor</i> (independent variable)	<i>Parameter estimates</i>			<i>Wald's test</i>		$\Delta_x(\Lambda)$ -test	
	Log-odds	S.E.	Odds	Z	P	χ^2	P
prime = HI	0.293	0.449	1.34	0.7	> 0.5	0.4	> 0.5
prime-target distance*prime=HI	-0.814	0.814	0.44	-1.6	> 0.15	2.4	> .12
P(RC LO target N)	-2.298	0.842	0.10	-2.7	< 0.01	8.0	< 0.005
P(RC HI target N)	5.297	1.454	200.33	3.7	< 0.005	14.1	< 0.005
prime surprisal	2.293	0.74	9.90	2.8	< 0.005	8.3	< 0.005

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). As the table shows, a main effect of prime RC surprisal was found ($p < 0.005$) such that less expected primes given the head nouns prime more. This is graphically represented in Figure 1. Interestingly, when the prime probabil-

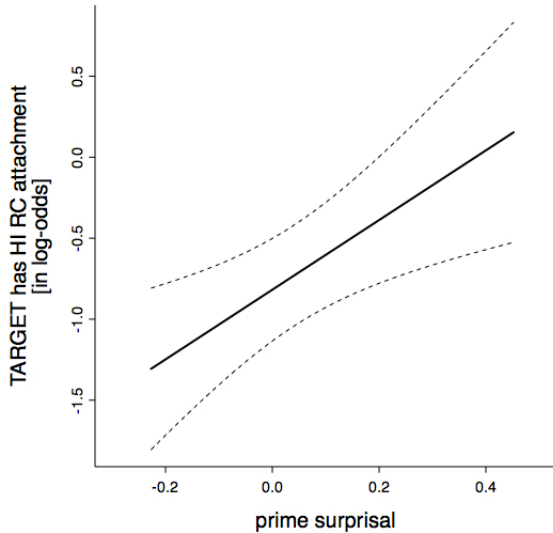


Figure 1: Surprisal of prime RC

ity effect is controlled, there is no main effect of the type of prime structure (HI or LO). Fast backwards model comparison tests showed that the prime surprisal factor explains significantly more of the variance and therefore is the only significant effect. There are also significant effects of the probability of an RC given the the target nouns: if

the HI target noun is frequently modified by an RC, then a HI attachment in the target is more likely ($p < .005$), also if the LO target noun is frequently modified by an RC, then a HI attachment in the target is less likely ($p < .005$). As in the previous study, there was no effect of prime-target distance.

3.5 Discussion

This study shows novel evidence for priming of RC attachments in naturalistic, conversational corpus data. It also provides further evidence for the exemplar hypothesis: the priming effect is sensitive to the exemplar frequencies of both the prime and target. This further strengthens the argument for the exemplar hypothesis because the the study involved a non-argument construction and therefore could not be due to abstract representations of argument structure.

4 Exemplar similarity in ditransitives (Study 2)

The above results argue for an effect of exemplar frequency on priming. It must be acknowledged, however, that the particular results presented here could be explained by other models of representation. For example a lexicalized probabilistic context free grammar, where syntactic knowledge consists of rules and their probabilities given their head word, provides representations that could explain these results. Also, I have not presented a clear model of the exemplars. In this study, I remedy both of these problems by testing exemplar priming effects in a model like that used in k -nearest-neighbor (k -NN) classifiers (Aha et al., 1991; Daelemans and van den Bosch, 2005).

k -NN models classify cases by comparing the

current case with all other observed exemplars (usually a data set derived from a corpus), and picking the best classification based on past experience. An essential part of this model is the distance metric used to determine how similar exemplars are to one another. The algorithm uses this to weight highly the exemplars that are closest to the case to be classified. However, such a distance metric could also be extremely useful in determining whether one structure can prime another. Recall that Branigan and Pickering (1995) argued that priming occurs when the prime and target are similar along some dimension. The k -NN model of exemplar similarity provides a metric for explicitly measuring the similarity of the prime and target. If exemplars are the correct representation, then this measure of exemplar similarity should predict the likelihood of priming.

Exemplars in k -NN models are defined as sets of features. The distance between exemplar X and exemplar Y , $\Delta(X, Y)$, is the weighted sum of the difference between the two exemplars along each of the dimensions defined by the features:

$$\Delta(X, Y) = \sum_i w_i \delta(x_i, y_i) \quad (2)$$

where w_i is the weight associated with feature i , and $\delta(x_i, y_i)$ is defined as:

$$\delta(x_i, y_i) = \begin{cases} \frac{x_i - y_i}{\max_i - \min_i} & \text{if numeric, otherwise} \\ 0 & \text{if } x_i = y_i \\ 1 & \text{if } x_i \neq y_i \end{cases}$$

If the feature is continuous, then the difference between the two exemplars is just the difference in the values of feature i for the two exemplars, divided by the maximum range of that feature (so that features won't be weighted too highly just because they involve larger numbers). If the feature is categorical, the the difference between the two exemplars is 0 if they have the same value for the feature, and 1 if they have different values for the feature.

Applying such a model to priming, the basic idea is that two exemplars that are more similar, or have a lower NN-distance between them, are more likely to prime. This study tests this prediction in the ditransitive alternation.

4.1 Data

I use the same data as in Bresnan and colleagues (2007) and Jaeger and Snider (2007). For greater

statistical power, I use the full data set of 1,108 ditransitives with preceding primes (as opposed to splitting it in two) from the full Switchboard corpus. Again, the data consists of naturalistic conversation.

4.2 Method

I again analyzed the data with logistic regression models. I added to the ditransitive data set an exemplar NN-distance metric as defined in equation (2). The similarity space in which the distance was calculated is quite fine-grained, consistent with an exemplar model of representation. It has as dimensions all the features used to predict the dative alternation in the original study by Bresnan and colleagues (2007). Thus, in the ditransitive data set, the prime and target exemplars can be compared in terms of the pronominality, animacy, givenness, etc. of the recipient and theme arguments, as well as the verb identity (Bresnan and colleagues modeled this as a random effect). Verb identity between prime and target has been found to correlate positively with priming (Pickering and Branigan, 1998), which is an obvious prediction of the exemplar similarity hypothesis: exemplars with the same verb are clearly more similar.

In order to determine the feature weights used in equation (2). I ran the k -NN classifier TiMBL (Daelemans and van den Bosch, 2005) on the entire ditransitive data set, using the same features as Bresnan and colleagues. Using *Information Gain* weighting, TiMBL classified the ditransitive alternation with 92.8% accuracy, which is comparable to the 96% originally found by Bresnan and colleagues. Using the feature weights determined by TiMBL, I calculated the exemplar NN-distance between each prime and target in the ditransitive data set.

4.3 Prediction

The NN-distance metric can be used to define a similarity metric between the prime and target exemplars in a data set. The prediction is that the lower the distance between the prime and target exemplars, the more likely the prime and target will occur in the same construction.

4.4 Results

The results of the ditransitive analysis are in Table 2. The first 14 rows in Table 2 summarize controls from Bresnan *et al* and Snider and Jaeger. The last

Table 2: Summary of ditransitive analysis

<i>Predictor</i> (independent variable)	<i>Parameter estimates</i>			<i>Wald's test</i>		$\Delta_x(\Lambda)$ -test	
	Log-odds	S.E.	Odds	Z	P	χ^2	P
verb class = communication	-2.505	0.631	0.06	-3.9	< 0.001	33.8	\ll 0.001
verb class = future transfer	-0.154	0.812	0.94	-0.2	> 0.8		
verb class = prevent transfer	-4.598	2.409	0.01	-1.9	< 0.1		
verb class = transfer	0.434	0.418	1.22	1.0	> 0.2		
recipient nongiven	2.488	0.462	10.69	5.4	\ll 0.001	25.1	\ll 0.001
theme nongiven	-1.467	0.442	0.22	-3.3	< 0.001	11.4	< 0.001
recipient pronominal	-0.288	0.484	0.73	-0.6	> 0.5	0.4	> 0.5
theme pronominal	1.469	0.449	4.06	3.3	< 0.005	9.9	< 0.005
theme indefinite	-2.146	0.411	0.11	-5.2	\ll 0.001	29.8	\ll 0.001
recipient inanimate	3.481	0.579	28.51	6.0	\ll 0.001	33.0	\ll 0.001
recipient non-local person	0.676	0.390	2.02	1.7	< 0.1	3.3	< 0.1
theme singular	-0.995	0.377	0.43	-2.6	< 0.01	4.9	< 0.03
log argument length difference	-1.342	0.243	0.24	-5.5	\ll 0.001	34.1	\ll 0.001
target verb bias	3.710	0.730	42.34	4.9	\ll 0.001	25.8	\ll 0.001
prime=NPPP	1.665	1.236	11.85	1.4	> 0.15	11.0	< 0.005
prime V = target V * prime=NPPP	-0.024	0.825	0.97	-0.1	< 1	0.03	< 1
NN distance * prime=NPPP	-2.365	0.869	0.05	-2.7	< 0.01	10.8	< 0.005

row (17) contains the result of interest, that exemplar similarity is a significant predictor of priming. The negative coefficient indicates that, if the prime is NP PP, and the NN-distance between the prime and target is small, then the target is more likely to be NP PP. This is graphically represented in Figure 2. This effect is still significant controlling for prime-target verb identity, because row 16 shows that factor is not significant. This indicates that the exemplar similarity effect is not due just to the verb identity effect of Pickering and Branigan (1998). Row 15 shows the priming due to the previous structure being NP PP is not significant as a main effect, but its interactions are significant, as rows 16 and 17 show. Finally, although it is not shown in the table, I tested for the effect of prime surprisal. I found it to be marginally significant ($p < 0.1$). This is probably due to the null effect that Jaeger and Snider found when the prime structure is NP NP.

4.5 Discussion

This study shows that priming is sensitive to exemplar similarity. The target is more likely to occur in same structure as the prime when the NN-distance between the prime and target is small. Further, the exemplar similarity metric used is quite precisely defined and provides a clear model of the exemplar representation. Also, the effect of exemplar

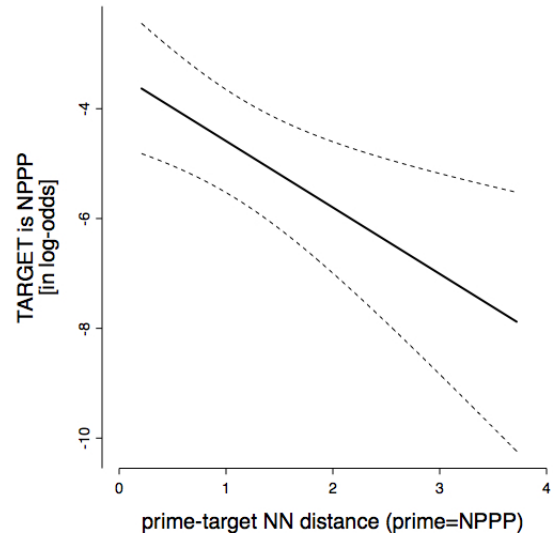


Figure 2: Interaction between exemplar similarity and priming

similarity is not likely to be an artifact of local discourse and syntactic factors, because these were controlled.

5 Conclusions and Future Work

In conclusion, these two studies have shown exemplar frequency and similarity interacts with priming. Specifically, the strength of the priming effect depends of the exemplar frequency of the prime

structure. Also, target structures that are more frequent are also more likely to be produced. This is evidence for an exemplar model of the mental representation of syntactic knowledge, because the exemplar hypothesis predicts that the frequency of occurrence of exemplars will affect language processing. Further, priming was shown to be more likely when the prime and target are similar. This is particularly strong evidence for the exemplar hypothesis because a precise model was used to define the exemplars and their similarity metric.

In future work, I will further refine the tests of the exemplar hypothesis. I plan to add new features to the distance metric that represent the degree of similarity between other words in the prime and target structures, for example the head nouns of the recipient and theme arguments. This is a particularly important test of the exemplar hypothesis, because the hypothesis predicts that the representations contain fine details such as the head words involved, in addition to the more general features already in the data set.

6 Acknowledgments

I am especially grateful to Florian Jaeger for advice on analysis and on describing the modeling techniques. I also thank Joan Bresnan for discussions and advice on interpretation. Thanks to Joan Bresnan, Gabe Recchia, and colleagues for providing me with the database of ditransitive prime-target pairs from the Switchboard corpus. Also, thanks to Tom Wasow for very helpful feedback and questions.

References

- Alan Agresti. 2002. *Categorical Data Analysis*. Wiley.
- D.W. Aha, D. Kibler, and M.K. Albert. 1991. Instance-based learning algorithms. *Machine Learning*, 6(1):37–66.
- K. Bock and Z.M. Griffin. 2000. The persistence of structural priming: Transient activation or implicit learning. *Journal of Experimental Psychology: General*, 129(2):177–192.
- J.K. Bock. 1986. Syntactic persistence in language production. *Cognitive Psychology*, 18(3):355–387.
- R. Bod. 2006. Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review: Special Issue on Exemplar-Based Models in Linguistics*, 23.
- H.P. Branigan, M.J. Pickering, S.P. Liversedge, A.J. Stewart, and T.P. Urbach. 1995. Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research*, 24(6):489–506.
- J. Bresnan, A. Cueni, T. Nikitina, and H. Baayen. 2007. Predicting the dative alternation. *Proceedings of the Royal Netherlands Academy of Science Workshop on Foundations of Interpretation*. Amsterdam.
- J. Bybee and J.L. McClelland. 2005. Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, 22:381–410.
- J. Bybee. 2000. The phonology of the lexicon: evidence from lexical diffusion. In M. Barlow and S. Kemmer, editors, *Usage-based Models of Language*, pages 65–85. CSLI, Stanford.
- W. Daelemans and A. van den Bosch. 2005. *Memory-based Language Processing*. Cambridge University Press.
- J.J. Godfrey, E.C. Holliman, and J. McDaniel. 1992. Switchboard: Telephone speech corpus for research and development. In *IEEE ICASSP*, pages 517–520.
- J. Hay and J. Bresnan. 2006. Spoken syntax: The phonetics of giving a hand in new zealand english. *The Linguistic Review: Special Issue on Exemplar-Based Models in Linguistics*, 23.
- T.F. Jaeger and N. Snider. 2007. Implicit learning and syntactic persistence: Surprisal and cumulativity. *University of Rochester Working Papers in the Language Sciences*.
- T. Florian Jaeger. 2006. *Redundancy and Syntactic Reduction in Spontaneous Speech*. Ph.D. thesis, Stanford University.
- T. Florian Jaeger. 2007. Better categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. submitted.
- A. Krott, R. Schreuder, RH Baayen, and WU Dressler. 2006. Analogical effects on linking elements in German. *Language and Cognitive Processes*.
- Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1994. Building a large annotated corpus of english: The penn treebank. *Computational Linguistics*, 19(2):313–330.
- MJ Pickering and HP Branigan. 1998. The Representation of Verbs: Evidence from Syntactic Priming in Language Production. *Journal of Memory and Language*, 39(4):633–651.
- G. Recchia. 2007. Strata: Search tools for richly annotated and time-aligned linguistic data. *Stanford Undergraduate Honors Thesis*.

C. Scheepers. 2003. Syntactic priming of relative clause attachments: persistence of structural configuration in sentence production. *Cognition*, 89(3):179–205.