

# Statistical Structures in Artificial languages Prime Relative Clause Attachment Biases in English

Felix Hao Wang<sup>a</sup>, Mythili Menon<sup>b</sup>, Elsi Kaiser<sup>b</sup> {wang970, mythilim, emkaiser} @usc.edu

<sup>a</sup>Department of Psychology, <sup>b</sup>Department of Linguistics, University of Southern California, Los Angeles, CA, 90089

## Abstract

The phenomenon of syntactic priming is well studied in the literature, but the mechanisms behind it are still under debate. In this study, we trained English-speaking participants in artificial language sequences with dependencies that are either adjacent or non-adjacent. The participants then wrote completions to relative clause (RC) fragments. We found that participants who learn non-adjacent dependencies in the artificial language, exhibit a bias to write high-attachment (non-adjacent) continuations for RCs, when compared to participants in a control condition who exhibit low-attachment (adjacent) biases in RCs. The implications for theories of syntactic priming and its relations to implicit learning are discussed.

**Keywords:** implicit learning; syntactic priming; relative clause attachment bias; non-adjacent dependencies

## Introduction

Although the phenomenon of syntactic priming has been very well studied in the literature, the exact processes behind priming are still unclear. We build on insights from research in child language development and adult sentence processing regarding the representation of abstract dependencies in language and other cognitive domains (e.g. Gomez, 2002; Scheepers, Sturt, Martin, Myachykov, Teevan & Vizkupova, 2011). We explore whether an abstract relation represented through word-level statistical regularities in an artificial language can prime the attachment biases of relative clauses. Specifically, we explore the question whether adjacent and non-adjacent structures derived from statistics can prime the low versus high attachment preferences during the production of English relative clauses.

### Structural priming

Structural priming refers to the observation that people are more likely to reuse syntactic structures that they have already used (e.g. Bock, 1986; Bock, 1989). Researchers have demonstrated structural priming with different syntactic structures, including verb phrase structures (Pickering & Branigan, 1998) and relative clause attachment (Mitchell, Cuetos, Corley & Brysbaert, 1995; Scheepers, 2003). In the classic paradigm (Bock, 1986), participants read sentences and were asked to describe semantically unrelated pictures. The question is whether their structural choices in the descriptions are influenced by the structure of the sentences they had previously read. For example, if participants have read a sentence like “The teacher sent the girl a letter” (recipient *girl* mentioned before direct object

*letter*) and then are asked to describe a picture where a soccer player is giving a ball to a boy, they are more likely to say *The soccer player gave the boy a ball* rather than *The soccer player gave a ball to the boy*. In general, reading or producing sentences of one type can prime the production and comprehension of sentences with the same structure. Structural priming occurs even if the lexical items in the prime sentence and the target sentence are different, and thus cannot be attributed to lexical repetition (e.g., Corley & Scheepers, 2002; Pickering & Branigan, 1998; Branigan, Pickering & Cleland, 2000).

There are two theoretical accounts for structural priming: the Lingering Activation account and the Implicit Learning account. The “lingering activation” account (Pickering & Branigan, 1998) suggests that people are using the same structures repeatedly because the activation for the structures lingers in the language production system. This account predicts that when structures leave activations lingering, these activations are more likely to be reused. Furthermore, since there is also lingering activation of lexical items, priming with the same lexical item with the same syntactic structure will yield stronger priming (“lexical boost”). Crucially, because activation – of lexical items as well as syntactic structures – is assumed to decay over time, this account predicts priming effects should diminish relatively rapidly with time. The “implicit learning account” (Bock & Griffin, 2000; Chang, Dell & Bock, 2006) suggest that syntactic priming does not require (and usually does not involve) any explicit awareness on the part of the participants that they are reusing structures. In the classic structural priming paradigm, exposure to the structures of interest is covert in that participants are not aware that certain sentences are primes while others are fillers. As a whole, these observations suggest that adapting to the structures in question is not a conscious choice, i.e. priming is an implicit process.

Hartsuiker, Bernolet, Schoonbaert, Speybroeck & Vanderelst (2008) suggest that both the “implicit learning” and “lingering activation” accounts are partially right. Hartsuiker et al. monitored the timing course of syntactic priming, using stimuli triggering priming from the same verb or different verbs. Between trials, the timing between the exposure to a structure (prime) and the production of the target sentence was varied. Hartsuiker et al. discovered that there is indeed a larger priming effect when both prime and target use the same verb (“lexical boost”). Furthermore, this effect decays with time, consistent with the predictions of the “lingering activation” account. Given that our design did not use any English verbs to achieve the priming effect, our

results validates the implicit learning account for the phenomenon when syntactic structures get reused without having the same lexical items. However, it does not speak to the lingering activation account. We agree with Hartsuiker et al. that these two accounts are not mutually exclusive.

### **Artificial language paradigm**

To study syntactic priming without the influence from lexical items, we decided to use an artificial language paradigm. Artificial language can provide adjacent and non-adjacent dependencies, which are structural and can be learned.

In order to demonstrate implicit learning, researchers have used the artificial language paradigm (e.g., Saffran, Newport & Aslin, 1996; Gomez, 2002). In an artificial language, people can learn statistical patterns in nonsense words (e.g. *voy glaik fex, choon glaik jub*). Participants can readily learn these structures, in the absence of any semantic information. Constructed with nonsense words, these linguistic materials only convey distributional patterns. Furthermore, statistical learning goes beyond the specific items (words or syllables) being learned (Thiessen, Kronstein & Hufnagle, 2013; Mintz, Wang & Li, 2014). According to Thiessen et al. 2013, statistical learning initially gathers statistics about the input presented to learner and uses this information to learn and infer patterns. Under this view, the representations that learners generate from artificial language input have been argued to be abstract and structural.

### **Adjacency and non-adjacency**

Two key concepts relevant for structural representations are the notions of adjacency and non-adjacency. Starting with the seminal study of Saffran, Newport & Aslin (1996), there is a large body of work showing that adults and children can learn adjacent relations from continuous streams of syllables. More recent artificial language work (Gomez 2002; Maye & Gomez, 2005) demonstrated that adults and children are also able to learn non-adjacent dependencies between words.

### **Relative clauses**

The notions of adjacency and non-adjacency are also relevant in the domain of syntax, for example in the representation of relative clauses. In English sentences with the structure *NP1 of NP2 who* (e.g., *Jessica visited the doctors of the supermodel who*), the following relative clause completions (eg. *who lived in Los Angeles*) can potentially attach to either one of the NPs. In high attachment completions, the following relative clause attaches back to the higher NP1 (eg. *the doctors lived in Los Angeles*). In the low attachment completions, the relative clause attaches back to the lower NP2 (eg. *the supermodel lived in Los Angeles*).

In our experiment, participants are asked to complete sentence fragments ending in ‘who’. Thus, they can complete the sentence fragment modifying either the immediately adjacent noun (NP2), or the non-adjacent noun

(NP1). If there are more high-attachment sentences produced, we call this a high-attachment bias, and vice versa. The high attachment completions are instances of non-adjacency whereas the low attachment completions are adjacency. In English, the default preference for attachment completions is low attachment, i.e. participants tend to attach ambiguous relative clauses to the lower NP (e.g. Cuetos and Mitchell, 1988).

As pointed out by Scheepers (2003), the distinction between high and low in relative clause attachment bias has to do with syntactic sequencing. The syntactic rules used to generate these representations are the same, and the only exception is that in low attachment, the relative clause is modifying the noun immediately preceding it, whereas in high attachment, the relative clause is modifying the noun non-adjacently preceding it. In our opinion, this provides a striking analogy to artificial language dependencies, because artificial language provides combinatorial properties where words are corresponding to other words, according to some combinatorial pattern. The only potential issue is the grain size of sequencing (word vs. phrase level). However, there is previous research suggesting that grain size may not matter to a large extent (Melinger & Dobel, 2005).

### **Aims of this work**

In this study, we test whether structural representations arising from distributional information can prime relative clause completions. If relative clause attachment biases come from representations that are completely different from distributional dependencies, exposure to any artificial language with only distributional properties will not result in any changes in the completion of relative clauses. On the other hand, if relative clause attachment biases come from representations that are shared with sequential representations from an artificial language, the relative clause bias is predicted to change as a result of learning structures that are different from the default.

To this end, we primed participants with an artificial language which conveyed structures that are consistent with our prediction. In this experiment, we will test this hypothesis with English.

## **Experiment**

In this experiment, we explore the effect of statistical structures from an artificial language on participants’ completions of ambiguous relative clauses fragments. Our experiment has a learning phase and a testing phase.

### **Methods**

#### **Participants**

A total of 50 adult native English speakers participated. Given the four conditions described below, there were 20 participants in the critical non-adjacent dependency condition, 10 in the control condition, and 10 in each of the two adjacent dependency conditions.

## Stimuli

First, we describe the stimuli used in the training phase of the experiment. In the training phase, we used artificial words, similar to the stimuli used in Gomez (2002)'s non-adjacent dependency experiment. A female American English speaker read and recorded these nonsense words in a sound isolated room. The speaker pronounced the stimuli one word at a time. We digitally spliced the recordings into individual word files that began at the onset of each word. Word files generated from this procedure are all shorter than 0.8 seconds, and silences were added to make each word files 0.8 seconds long. This allowed us to concatenate word files into sentences with words occurring every 0.8 seconds. Between each artificial sentence, there was also a 0.8 second pause in between, to signal the start and the end of each 'sentence'.

Similar to Gomez (2002), each sentence is made of 3 words, which differ in terms of their distributional properties, between the 4 conditions of the experiment. We used monosyllabic words (for eg. *voy, nud, choon, glaik, blit, ghire, ghen, sowch, dess, fex, dap, jub*). In the non-adjacent dependency condition ( $A_iXC_i$ ), words at the beginning and the end always co-occurred. Three different pairs of words co-occured as A words or C words, while a total of 6 different words were used as X words at the intermediate position. Thus, we had a total of 18 unique trigrams. The correspondence between A words and C words were counterbalanced between subjects, such that the wrong correspondence in one condition is correct in the other condition, and vice versa. In the adjacent dependency condition, the X words were moved to the front ( $XA_iC_i$  condition) or to the back ( $A_iC_iX$  condition), such that the dependency is adjacent. In the control condition, 18 unique word trigrams were created such that there were same numbers of adjacent or non-adjacent dependencies in these trigrams.

We also created sentence fragments for participants to complete in the testing phase. There were 2 kinds of sentence fragments: targets and fillers. All target sentence fragments are similar to example (1), where NP1 *the doctors* and NP2 *the supermodel* are connected by the preposition 'of' and are followed by the relative pronoun 'who'. (Targets were constructed using stimuli used by Rohde, Levy & Kehler, 2011. We made sure that none of our verbs had strong implicit causality biases, using Hartshorne and Snedeker (2012).

- (1) John met [the doctors]<sub>NP1</sub> of [the supermodel]<sub>NP2</sub>  
[who invented a vaccine]<sub>RC</sub>.

The participants' task was to write a completion for the sentence. We analyzed the completions for whether the relative clause modifies NP1 (e.g. *the doctors who invented a vaccine*) or NP2 (e.g. *the supermodel invented a vaccine*). Relative clauses that modify NP1 are called high attachments and relative clauses that modify NP2 are called low attachments.

In targets, the subject of the sentence was always a proper name (equal numbers of male and female names). The two

NPs were definite animate nouns, preceded by the definite article. The NPs were controlled for number. Half of the sentences had NP1-singular and NP2-plural (e.g. *the doctor of the supermodels*) and the other had the opposite configuration (e.g. *the doctors of the supermodel*). This facilitates coding because number marking on the verb usually disambiguates (e.g. *...was happy* vs. *...were happy*). All verbs in the target fragments (e.g. *counted*) were non-implicit causality (non-IC verbs), chosen in order to avoid verb semantic bias. Fillers were non-ambiguous English sentence fragments of similar length. Each participant completed the same 18 target sentences and 18 filler sentences. The fillers do not involve relative clauses, and are comprised of a range of sentence types. They are open to a range of reasonable continuations, and do not follow a particular structure.

## Design and Procedure

There are two phases to the experiment, the training phase and the testing phase. During the training phase, participants listened to sequences in the artificial language and in the test phase, participants either answered an artificial language question, or completed a sentence fragment.

The training phase consisted of a simple artificial language learning task. In this phase, participants listened to an artificial language according to the condition that they were in. To briefly reiterate, there were 4 between-subject conditions: the  $A_iXC_i$  condition (Non-adjacent dependency condition), the  $A_iC_iX$  and  $XA_iC_i$  conditions (adjacent dependency condition), and the control condition where about equal numbers of adjacent and non-adjacent dependencies exist in the 18 trigrams used. In between trials, participants were also asked the question "What was the last word you heard?" with 2 words to choose from. Participants then pressed a key to indicate their choice. This question was presented every few minutes, in order to keep them alert during this task. The training phase lasted about 20 minutes.

The test phase immediately followed the training phase. Before the test phase started, we reminded participants of the two types of tasks: questions about the artificial language and sentence fragments for them to complete. Half of all the test trials were questions about the artificial language (36 trials), and the other half were sentence completions (36 trials). The artificial language portion consisted of trigrams that are composed in the same fashion as in the training phase. Three words at a pace of 0.8 second per word were presented to the participant, and then a question appeared on the screen: "Did you hear this in the training phase or not?" Across the testing session, there were 36 test items, half of which are targets (trigrams from the artificial language), and the other half foils (trigrams not in the artificial language). The foils in the  $AXC$ ,  $ACX$ ,  $XAC$  conditions are such that the correspondence in terms of the dependency is incorrect ( $A_iXC_k$ ,  $k \neq i$ ). The foils in the control conditions are reversed strings from the training

trigrams. For sentence completion trials, half of the sentence fragments were target and half were foils.

The trials were block pseudo-randomized in the following way. The two types of trials that were critical were the relative clause target sentence completions and artificial language questions that are from the language. The relative clause target sentence completions are always preceded from an artificial language item from the language, that is, participants are supposed to answer, “Yes” to the artificial language question. We mixed these trials with all the other trials in a randomized order within 3 blocks. The testing phase lasted between 15 to 30 minutes, and the whole experiment was done under an hour.

### Coding

We coded only the target sentences. The coding of the sentences resulted in three types: high attachment (HA), low attachment (LA), and ambiguous. For the logistic regression model, HA was coded as 1 and LA was coded as 0, and ambiguous was treated as missing. Coding was done with mostly syntactic considerations, given that the two NPs in our sentences are different in terms of number, so the verb from the continuation in the relative clause shows overt morphological agreement with the NPs. If verb number did not disambiguate (e.g. *went, asked*), semantic cues were used to decide high attachment (e.g. *Emily worked with the mother of the children who just got tenure*) from low attachment (*Chris counted the fans of the singer who just finished the encore*). If both verb marking and semantic cues were unclear, the sentence was coded as ambiguous.

The continuations were double coded by two native English speakers, who exhibited >99% agreement. (The remaining <1% of the items were resolved by discussion).

### Training Phase Results

In all conditions, participants were able to correctly endorse correct items in the artificial language and rejected foils above chance. For each of the four conditions, we ran a mixed-effects logistic regression, with respect to participants’ responses in the testing phase (Table 1). The responses included both the target artificial language items and the foil items. In the regression, subjects were specified as random effect with no fixed effects. This way, the coefficient of the intercept indicates a comparison with chance (Jaeger, 2008), and we report the co-efficient ( $\beta$ ) with the associated z and p-values.

Table 1. Artificial language learning test phase results compared to chance

Condition	$\beta$	Z	p-value
AXC	0.82	2.76	0.006 **
ACX	2.14	3.78	<0.001 ***
XAC	0.281	0.99	0.32
Control	0.637	4.94	<0.001 ***

### Priming results: RC Completion patterns

Now, we turn to the priming results. As mentioned, the outcome is a binary response for high/low attachment, with others were coded as missing. The sentences coded as missing included tiny proportions of continuations where the participant entered continuations that are syntactically incorrect, as well as those that are semantically completely ambiguous with regard to attachment. These sentences were missing at random, and a Goodness of Fit test showed that there are no in-between condition differences in terms of the amount of data missing ( $p=0.09$ ). In a mixed-effects logistic regression model that predicts the proportion of completion being high attachment (coded as 1) vs. low attachment (coded as 0), the artificial language learning condition was specified as the fixed effects while holding subjects and items as random effects. The general model fit was indicated by the Wald chi-squared test, which yields a p-value smaller than 0.001.

Our main results show that when comparing between conditions, we find that the non-adjacent dependency condition is significantly different from the control condition, while the adjacent dependency conditions are not. In other words, as can be seen in Figure 1 (below), we see that although in the control condition, the participants are biased to produce more low attachment completions (55%), in the non-adjacent dependency condition ( $A_iXC_i$ ) the number of low attachment completions is significantly lower (37%). In the adjacency conditions ( $A_iC_iX$ ,  $XA_iC_i$ ), the low attachment completions are not significantly different from the control condition (52 % for  $A_iC_iX$ , 48% for  $XA_iC_i$ ). Moreover, in the non-adjacency condition, there is an overall bias for high attachment completions (42 %). These results are in Table 2, and we plot out the proportions of sentence completions in each artificial language condition in Figure 1.

Table 2. Result of logistic regression for priming

Cond.	Manipulation	$\beta$	Z	p-value
AXC	Non-Adjacent	0.951	4.38	<0.001***
ACX	Adjacent	0.229	0.89	0.372
XAC	Adjacent	0.25	1.00	0.318

Sorted by artificial language manipulation type for Experiment 1, as compared to the control condition.

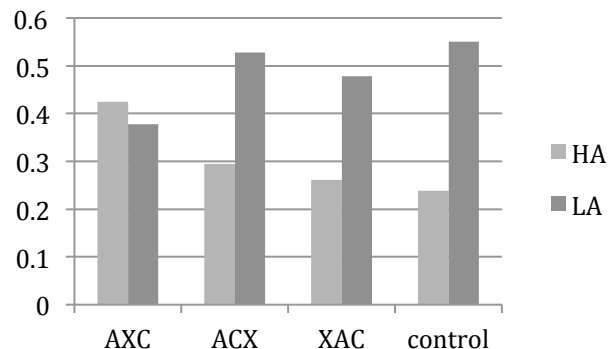


Figure 1. Proportions of sentence completions in each artificial language condition

Since the non-adjacent dependency condition is changing participants' performance, we ran more tests to examine the patterns of data in this condition closely. Two further analyses investigate the specific relationship between participants' item-level judgment about artificial language tests and their tendencies to complete a target sentence with a high attachment continuation. Table 3 (below) details the numbers of yes/no responses for the artificial language item immediately before the completion of the target sentence completions.

Table 3. Completing high-attachment RC and answering correctly to the AXC question preceding it.

Response	High attachment	Low attachment
Yes (correctly)	105	96
No	40	48

Table 3 shows this non-existent relationship in the non-adjacent condition. Fisher's exact yields  $p=0.798$ , ns.

Figure 2 demonstrate the relationship at the participant level, correlating general performance on artificial language tasks and the proportion of high attachment completions. Both of these analyses show no apparent relations between the two. We come back to this point in the discussion.

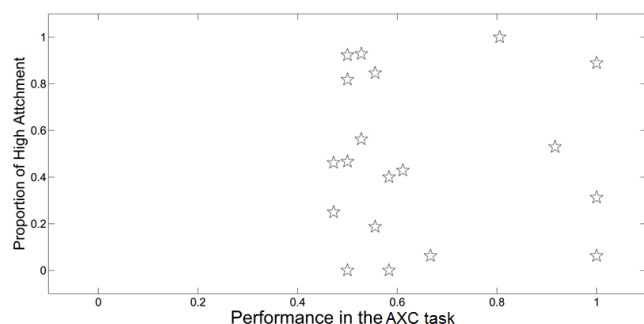


Figure 2. Participant level artificial language task performance and high attachment completion proportions per subject.

The correlation between (i) correctly remembering specific word-level ( $A_i$  with  $C_i$ ) correspondences and (ii) a bias towards producing non-adjacent dependencies is weak ( $-0.053$ ) and non-significant ( $p=0.82$ ).

### Summary

In this experiment, we studied attachment biases in the completion of relative clause fragments in English. We observe a main effect of artificial language condition. Specifically, the non-adjacent dependency condition changes the attachment bias significantly. On average, participants produced 18.6% more high attachment relative clause completions in the non-adjacent artificial language

condition than in the control condition. We find this result that participants reuse the structure from the artificial language to the participants' native language very similar to the phenomenon of syntactic priming. This result has interesting theoretical implications, as we discuss in the next section.

## General Discussion

We conducted an experiment where we provided structures for people to learn in an artificial language task, and we tested how these structures change the biases in relative clause attachment in natural language. In doing so, we provide the first demonstration that implicit learning of structures changes the bias in relative clause attachment, providing empirical evidence for the link between the two processes, structural learning and sentence production.

We used the artificial grammar learning paradigm in this study to induce implicit learning. Different language learning tasks require different kinds of learning mechanisms and it is important to choose the right task to induce implicit learning. Unlike learning the meaning of lexical items, which is dependent on the explicit learning system (Trueswell et al, 2013; Wang & Mintz, in revision), learning grammars from an artificial language stream uses an implicit learning process (e.g., Ullman 2004). For these reasons, we chose the artificial language learning task that yield abstract structural representations.

This study allows us to characterize syntactic priming as implicit learning using an experimental approach. Previous work (Chang, Dell & Bock, 2006) used a connectionist model to specify how the process of implicit learning happens. In their model, it is assumed that reading sentences of a particular structure changes the weights over that structure such that the bias for that structure increases. This was in turn used to demonstrate, in production, why syntactic priming occurs. This model provides a computational account of how implicit learning happens and how it influences syntactic behaviors. Our approach provides an empirical validation for this computational account, in that we directly measure the result of implicit learning via assessing outcomes of artificial language learning. Our data provide a causal link between implicit learning and syntactic priming. This can explain the presence of syntactic priming only when the artificial language with the combinatorial properties (non-adjacent) that are different from default biases (adjacent) which lead to preference for non-adjacent attachment in RC completion.

Once the structural representation is learned from the artificial language, the artificial language tests suggest that it does not matter whether participants are aware of the particular dependencies at the lexical level. The canonical way of assessing artificial grammar learning task (asking yes/no questions for one string at a time) requires explicit reflection of whether strings are grammatical or not, which is not the best way to probe implicit representations. Future work should use a more implicit measure of artificial

language learning to observe more subtle effects. In our data at least, we observe a zero correlation between the performance in the ‘explicit’ artificial language task and the sentence completion task. We take this as an indication that the learning of the abstract patterns results in implicit representations (Fiser & Aslin, 2001; Saffran, Newport, Aslin, Tunick & Barrueco, 1997).

We have a few future directions from this work. We are interested in investigating how to assess implicit representations better and find a correlation between implicit learning measures with priming. Also, we are interested in the generalizability of the current finding with regard to language (English, in the present study) to a different language. Preliminary work with Spanish suggests that the priming effect is present for Spanish speakers as well. In the Spanish data, we collected data for second language background, confirming that the priming effect is not a result of sampling bias. Future work with implicit learning processes in domains other than language processing are also underway to assess the domain generality of syntactic priming.

## References

- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive psychology*, 18(2), 355-387.
- Bock, K. (1989). Closed-class immanence in sentence production. *Cognition*, 31, 163–186.
- Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, 129, 177-192.
- Branigan, H.P., Pickering, M.J. and Cleland, A.A. (2000) Syntactic priming in written production: evidence for rapid decay. *Psychonomic Bulletin & Review*. B13-B25.
- Chang, F., Dell, G. S., & Bock, J. K. (2006). Becoming syntactic. *Psychological Review*, 113, 234–272.
- Corley, M and Scheepers, C. 2002. Syntactic Priming in English Sentence Production: Categorical and Latency Evidence from an Internet-Based Study. *Psychonomic Bulletin & Review* 9 (1), 126-131.
- Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, 30(1), 73-105.
- Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12(6), 499-504.
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13(5), 431-436.
- Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, 7(2), 183-206.
- Hartshorne, J. K., & Snedeker, J. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes*, 28(10), 1474-1508.
- Hartsuiker, Bernolet, Schoonbaert, Speybroeck & Vanderelst (2008). Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language* 58 (2), 214-238.
- Jaeger, T. Florian (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, 59, 434-446.
- Melinger, A., & Dobel, C. (2005). Lexically-driven syntactic priming. *Cognition*, 98(1), B11-B20.
- Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: evidence for the use of course-grained (non-lexical) statistical records. *Journal of Psycholinguistic Research*, 24, 469-488.
- Mintz, T. H., Wang, F. H., & Li, J. (2014). Word categorization from distributional information: Frames confer more than the sum of their (Bigram) parts. *Cognitive Psychology*, 75, 1-27.
- Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39(4), 633-651.
- Rohde, H., Levy, R., & Kehler, A. (2011). Anticipating explanations in relative clause processing. *Cognition*, 118(3), 339-358.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928.
- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, 8(2), 101-105.
- Scheepers, C. (2003). Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition*, 89(2), 179-205.
- Scheepers, C., Sturt, P, Martin, C.J, Myachykov, A, Teevan, K, Vizkupova, I. (2011). Structural Priming Across Cognitive Domains: From Simple Arithmetic to Relative-Clause Attachment. *Psychological Science* 22: 1319-1326.
- Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*, 139(4), 792-814.
- Trueswell, J.C., Medina, T.N., Hafri, A. & Gleitman, L.R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, 66 (1), 126-156.
- Wang and Mintz (in revision). Characterizing the Difference Between Learning about Adjacent and Non-adjacent Dependencies.
- Ullman, M.T. (2004). Contributions of Memory Circuits to Language: The declarative/procedural model. *Cognition*, 92 (1), 231-270.