Learning Non-Adjacent Dependencies in Continuous Presentation of an Artificial Language

Felix Hao Wang (wang970@usc.edu)[a], Jason Zevin (zevin@usc.edu)[a,b,c], Toby Mintz (tmintz@usc.edu)[a,b,c]

[a]Department of Psychology, University of Southern California, 3620 McClintock Ave, Los Angeles, CA, 90089
[b]Department of Linguistics, University of Southern California, Los Angeles, CA, 90089
[c]Program in Neuroscience, University of Southern California, Los Angeles, CA, 90089

Abstract

Many grammatical dependencies in natural language involve elements that are not adjacent, such as between the subject and verb in *the child always runs*. To date, most experiments showing evidence of learning non-adjacent dependencies have used artificial languages in which the to-be-learned dependencies are presented in isolation by presenting the minimal sequences that contain the dependent elements. However, dependencies in natural language are not typically isolated in this way. In this study we exposed learners to non-adjacent dependencies in long sequences of words. We accelerated the speed of presentation and learners showed evidence for learning of non-adjacent dependencies. The previous pause-based positional mechanisms for learning of non-adjacent dependency are challenged.

Keywords: implicit learning; non-adjacent dependencies

Introduction

Sentences in natural languages contain grammatical dependencies, such as those that arise from agreement marking between the sentence subject and the verb. Sometimes these dependencies hold between adjacent words (or morphemes), and sometimes the dependencies are non-adjacent. For example, the dependency between the singular subject *child* and the agreeing inflected verb *runs* in *the child runs* is an adjacent dependency, whereas in, *the child always runs* it is non-adjacent. These dependencies are expressed by hierarchical syntactic structures in formal syntactic grammars. However, there has been considerable interest in investigating learning mechanisms that could detect these dependencies in linear sequences within spoken utterances. Such mechanisms could be useful for discovering syntactic structure in children acquiring a language, and could also aid proficient language users in building syntactic parses. For example, there have been a number of studies using artificial and natural languages that have investigated how language learners acquire non-adjacent dependencies (e.g., Gómez, 2002; Newport & Aslin, 2004; Peña, Bonatti, Nespor & Mehler, 2002; Romberg & Saffran, 2013; Pacton & Perruchet 2008), and how early in the acquisition process such dependencies are detected (Gómez, 2002; Gómez & Maye, 2005; Santelmann & Jusczyk, 1998).

While most studies on adjacent dependency learning report success, the same cannot be said for learning of non-adjacent dependencies. The studies to date have found evidence of non-adjacent dependency learning only in limited situations, with some studies reporting success in learning and others reporting failure. Interestingly, a characteristic of experiments that showed successful learning is that the minimal sequences that contained a dependency were presented as discrete chunks. In other words, the chunks were surrounded by silences, and the edges of such a chunk consisted of the (non-adjacent) dependent elements. For example, studies that have probed non-adjacent dependency learning between words in artificial languages typically have used trigrams in which the dependent words were at the trigram edges, and subjects were presented the trigrams one at a time, with silence intervening between presentations (Gómez, 2002; Gómez & Maye, 2005; Gómez, Bootzin & Nadel 2006; Romberg & Saffran, 2013). With the one trigram at a time design, the words immediately before and after the silences are salient for learning the dependencies given that they make up the dependency. Similarly, in experiments investigating non-adjacent dependencies between syllables in syllable sequences, learning occurred only when brief pauses were introduced before (and after) each syllable trigram (Peña et al.,2002). When syllables were concatenated continuously, participants showed no learning (see also Newport & Aslin, 2004). In the studies just discussed, the fact that subjects' success in learning non-adjacent dependencies was correlated with whether the trigrams containing the dependency were pre-segmented suggests that the chunked presentation might have played an important role in learning. One reason in which pre-segmenting the material in this way could be helpful is that it places one or both dependent elements in an edge position. Indeed, Endress, Nespor & Mehler (2009) argued that edges are privileged in the kind of position-related computations they afford, and placement at edges could be an important constraint for learning non-adjacent dependencies.

However, non-adjacent dependencies in natural language are not restricted to edge positions, and are often embedded in longer sequences. Thus, learning the dependency relations of a natural language may require learning non-adjacent dependencies of items that may not always occur at boundaries marked by silences. Given the apparent difficulty in detecting non-adjacent dependencies of continuous sequences of syllables (Newport & Aslin, 2004;

Peña et al., 2002; Gebhart, Newport, & Aslin 2009), the experiments presented here were designed to assess how detection and learning of word-level non-adjacent dependencies fairs when the critical sequences are embedded in longer sequences, such that the dependent items are not at edges.

In this study, we present non-adjacent dependencies with words concatenated together without pauses. Similar previous attempts without pauses with syllables (Newport & Aslin, 2004; Peña et al, 2002) have all reported failure to learn. Instead of CV syllables in previous studies, our study used recorded monosyllabic words with a presentation rate close to the normal speech rate (3Hz). This temporal characteristic is significantly different from previous experiments with words, when individual words were presented every 0.75 seconds (Gomez, 2002; Romberg & Saffran, 2013). We believe that this faster rate may facilitate learning in a number of ways, for a number of reasons. For one, previous theories suggested that speech processing generally occurs at the theta rate (for a review, see Kiebel, Daunizeau & Friston, 2008). For another, faster presentation may expand short memory capacity (Frensch & Miner, 1994). Moreover, it has been suggested that presenting auditory material rapidly may aid auditory statistical learning (Emberson, Conway & Christiansen 2011). Thus, presenting auditory materials rapidly arguably presents the best chance for people to learn non-adjacent dependencies in speech.

We recently described the effect of presenting English sentences for entraining grammatical boundaries to aid learning non-adjacent dependencies (Wang, Zevin & Mintz, under review). We found that non-adjacent dependency is learnable with English bracketing the boundaries of the dependency. However, whether non-adjacent dependency is learnable without English is unknown, especially under the current learning conditions, where no pauses are inserted to indicate where the dependency boundaries are. The variability at the intermediate position of the dependency has also been theorized to influence the learnability of non-adjacent dependency, where dependency with low variability is generally hard to learn (Gomez, 2002). In the current paper, we employed low variability (n=3) in the intermediate position. To summarize, we used no pauses to indicate dependency boundaries, and low variability in intermediate position of the dependency, both of which has been theorized to exacerbate the learning problem.

However, we found that the fast presentation rate is enough to yield learning in all three experiments, and that makes this the first demonstration of learning of non-adjacent dependencies at the syllable/word level. Given all of the failures to learn in the literature, we present the first success demonstration of learning word level non-adjacent dependencies (Experiment 1). We consequently replicated the finding with similarly designs (Experiment 2 & 3).

Experiment 1

Methods

Participants. Thirty-eight USC undergraduates were recruited. Half of them participated in each counterbalancing condition.

Stimuli. We recorded speech from a native English speaker and digitized the recording at a rate of 44.1 kHz. We recorded 9 novel words to form the non-adjacent dependency: 3 at position 1 (rud, swech, voy), 3 at position 2 (dap, wesh, tood) and 3 at position 3 (tiv, ghire, jub).

After all the words were recorded in list intonation, we spliced the words from the recording. Each word by itself from the recording lasts between 300ms to 737ms, and we used the lengthen function in Praat (Boersma, 2001) to shorten all the words into approximately 250ms. An additional 83ms of silence was added to the end of each word to increase its intelligibility. Thus, words occurred at a rate of 3Hz.

Design and Procedure. The experiment consisted of three blocks, each with a training phase and a number of testing trials. Each learning trial consisted of listening to materials passively. Each learning trial contained 144 non-adjacent dependency triplets. Given the word presentation rate of 3Hz, each sentence lasted 1 second, and each learning trial lasted 2.4 minutes. There were no extra pauses between any novel words of artificial language. The testing section consisted of a set of 18 question trials. Each question trial involved participants giving a familiarity rating after hearing a three word sequence. Half of the 18 questions play a triplet from the language with the correct dependency, and the other half from the counterbalancing condition, with order of presentation randomized each time.

Each novel sentence was a concatenation of 3 novel words, 1 each from choices of 3 for each position, as specified in the Stimuli section. We denote the words making up the dependencies with A and B, and the words filling up other positions with X (and Y). The pattern we tested in Experiment 1 can thus be represented as AXB. All the possible combinations occurred for A_iXB_i where the first position word predicted the third position word. As such, there were 3 AB pairs and 3 X words, which made 9 possible different artificial sentences. A counterbalancing condition was created such that the ungrammatical strings that occurred in the test are grammatical in the training sequence in the counterbalancing condition, similar to Gomez (2002). That is to say, where A_iXB_i is grammatical in one condition and A_iXB_j is ungrammatical, the reverse is true for the other condition, and both conditions use the same test items.

Training phase. At the start of the experiment subjects heard the following instructions:

"In this study, you will be presented with rapid succession of made-up words. Press Space to start listening."

Participants listened to the sound stream passively while the screen was blank during the training phase.

Testing phase. Immediately after a training phase, we showed the instructions for the testing section on the screen. The instruction made it clear that participants would hear word sequences and make judgments about them. There were a total of 18 test trials during each testing section, half of which were from the correct dependency, and the other half from the counterbalancing condition (i.e., ungrammatical). The sequence of presenting the test trials was randomized for each participant.

Participants initiated each test trial. Per trial, participants clicked on a button to play an artificial language sentence, and a question followed asking the participant to indicate whether some sequences are from the previous section that they have heard. A scale with radio buttons showed up after playing the sentence and participants were asked to answer the question "Do you think that you heard this sequence in the previous section?" There were five possible items to choose from, "Definitely", "Maybe", "Not Sure", "Maybe Not", "Definitely Not". Participants could click on any of the radio buttons to make their choice, and this trial ended and the next trial began.

Results

For each question in the testing section, participants rated their familiarity for a given test sequence. We coded the scale of "Definitely", "Maybe", "Not Sure", "Maybe Not" and "Definitely Not" into numeric values of 1 through 5 (Definitely = 1). This allowed us to compared ratings for grammatical items vs. the ungrammatical items.

Next, we examined the means and standard error of the ratings. We show the rating information of grammatical and ungrammatical items by block in Figure 1. To compare ratings statistically, we ran mixed effect linear regressions with the data. In the regression, ratings were compared with test item (correct vs. incorrect) as the fixed effect, and subject as the random effect. We found that participants were able to learn the non-adjacent dependency in general (β= -0.160, p<0.001).

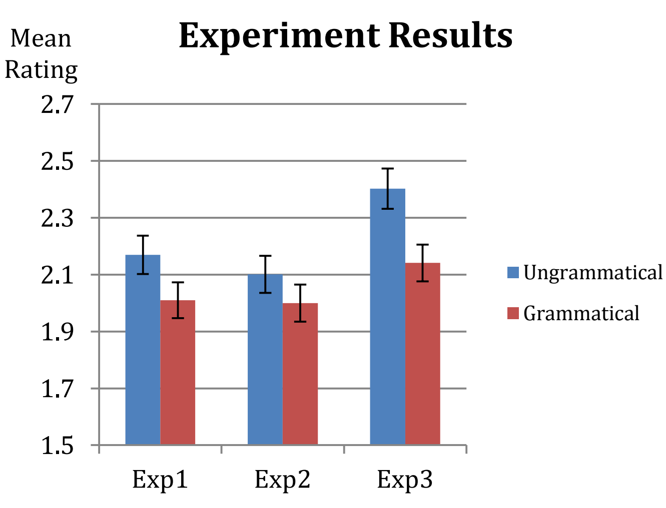

Figure 1. Rating data from all three experiments. The mean and the 95% confidence interval were plotted for each item type (grammatical/ungrammatical). The bar graph indicated that the grammatical items were judged to be more likely to be in the language than the ungrammatical items. All three experiments showed significant learning.

Discussion

Experiment 1 demonstrated that the non-adjacent dependency with one intermediate item can be learned efficiently as long as the words are presented in quick succession. However, natural languages rarely have non-adjacent dependencies stack one after other. In Experiment 2, we explore learning when there is a word between the dependencies, with the pattern of YAXB. Success in Experiment 2 should also be considered as a conceptual replication of Experiment 1.

Experiment 2

Experiment 2 tests non-adjacent dependency learning with the pattern is YAXB, where A & B words formed the dependency. Whereas Experiment 1 presented triplets with the dependency continuously, Experiment 2 has the triplet portion (AXB) separated from the next dependency by a word (Y). The choice of Y words is random with respect to other parts of the artificial language.

Methods

Participants. Thirty-eight USC undergraduates participated in Experiment 2. These participants have not participated in other experiments reported here. Half of the participants were in each counterbalancing condition.

Stimuli. We used the stimuli in Experiment 1. We used 12 novel words to form the non-adjacent dependency: 3 at position 1 (blit, pel, tink), 3 at position 2 (rud, swech, voy), 3 at position 3 (dap, wesh, tood) and 3 at position 4 (tiv, ghire, jub). There are four positions in Experiment 2 because the pattern is YAXB where A & B formed the dependency.

Again, all the words were approximately 250ms long with an additional 83ms of silence was added to the end of each word. When words are concatenated in a continuous stream, they would occur at a rate of 3Hz.

Design and procedure. The experiment consisted of three blocks, each with a training period followed by a sub-block of 18 question trials. Each learning period consisted of listening to materials passively. Each learning trial contained 144 non-adjacent dependency triplets. Given that words were at 3Hz and each sentence contained 4 words, each sentence took a second and a third. Each learning trial lasted 3.2 minutes. During the testing, each testing section contains 18 questions, half from the language and half from the counterbalancing condition, with order of presentation randomized each time.

Training phase. At the start of the experiment subjects heard the following instructions:
"In this study, you will be presented with rapid succession of made-up words. Press Space to start listening."

Participants listened to the sound stream passively while the screen was blank during the training phase.

Testing phase. Immediately after a training phase, we showed the instructions for the testing section on the screen. The instruction made it clear that participants would hear word sequences and make judgment about the sequences. There were a total of 18 test trials during each testing section, half of which were from the correct dependency, and the other half from the counterbalancing condition. The sequence of presenting the test trials was randomized for each participant.

Participants initiated each test trial. Per trial, participants clicked on a button to play an artificial language sentence, and a question followed asking the participant to indicate whether some sequences were from the previous section that they have heard. A scale with radio buttons showed up after playing the sentence and participants were asked to answer the question "Do you think that you heard this sequence in the previous section?" There were five possible items to choose from, "Definitely", "Maybe", "Not Sure", "Maybe Not", "Definitely Not". Participants could click on any of the radio buttons to make their choice, and this trial ended and the next trial began.

Results

We examined the means and standard error of the ratings (Figure 1). To compare ratings statistically, we ran mixed effect linear regressions with the data. In the regression, ratings were compared with test item (correct vs. incorrect) as the fixed effect, and subject as the random effect. We found that participants were able to learn the non-adjacent dependency in general (β= -0.101, p=0.021).

Experiment 1 & 2 demonstrate that the non-adjacent dependency with one intermediate item can be learned efficiently as long as the words are presented in quick succession.

Experiment 3

Natural languages are not restricted to have only one word in between the dependency (e.g., *the child very rarely runs*). In cases where there is more than one item in between the items that form the dependency, it has been suggested that learning becomes more difficult (Santelmann & Jusczyk, 1998). In Experiment 3, we explore learning when there are two intermediate items between the dependencies, with the pattern of AXYB.

Methods

Participants. Thirty-eight USC undergraduates participated in Experiment 2. These participants have not participated in other experiments reported here. Half of the participants were in each counterbalancing condition.

Stimuli. In Experiment 3, we explore learning when there are two intermediate items between the dependencies, with the pattern of AXYB. We used the stimuli in Experiment 1. We used 12 novel words to form the non-adjacent dependency: 3 at position 1 (rud, swech, voy), 3 at position 2 (blit, pel, tink), 3 at position 3 (dap, wesh, tood) and 3 at position 4 (tiv, ghire, jub).

Again, all the words were approximately 250ms long with an additional 83ms of silence was added to the end of each word. When words are concatenated in a continuous stream, they would occur at a rate of 3Hz.

The experiment consisted of three blocks, each with a training period followed by a sub-block of 18 question trials. Each learning phase consisted of listening to materials passively. Each learning trial contained 216 non-adjacent dependency triplets. Given that words were at 3Hz and each sentence contained 3 words, each sentence took a second. Each trial lasted 3.6 minutes. There were no extra pauses between any novel words of artificial language. During the testing, each testing section contains 18 questions, half from the language and half from the counterbalancing condition, with order of presentation randomized each time.

Training phase. At the start of the experiment subjects heard the following instructions:
"In this study, you will be presented with rapid succession of made-up words. Press Space to start listening."

Participants listened to the sound stream passively while the screen was blank during the training phase.

Testing phase. Immediately after a training phase, we showed the instructions for the testing section on the screen. The instruction made it clear that participants would hear sound sequences and make judgment about the sequences. There were a total of 18 test trials during each testing section, half of which were from the correct dependency, and the other half from the counterbalancing condition. The sequence of presenting the test trials was randomized for each participant.

Participants initiated each test trial. Per trial, participants clicked on a button to play an artificial language sentence, and a question followed asking the participant to indicate whether some sequences are from the previous section that they have heard. A scale with radio buttons showed up after playing the sentence and participants were asked to answer the question "Do you think that you heard this sequence in the previous section?" There were five possible items to choose from, "Definitely", "Maybe", "Not Sure", "Maybe Not", "Definitely Not". Participants could click on any of the radio buttons to make their choice, and this trial ended and the next trial began.

Results

We examined the means and standard error of the ratings (Figure 1). To compare ratings statistically, we ran mixed effect linear regressions with the data. In the regression, ratings were compared with test item (correct vs. incorrect) as the fixed effect, and subject as the random effect. We found that participants were able to learn the non-adjacent dependency in general (β= -0.261, $p<0.001$).

In sum, we found that robust learning is present even when there are 2 items between the words forming the non-adjacent dependency.

Discussion

As we mentioned, learning non-adjacent dependencies in the lab has been demonstrated in very restricted situations. There are a variety of reasons for this. For the most part, past literature suggested (Newport & Aslin, 2004; Peña et al., 2002) that pauses are critical to the learning of non-adjacent dependencies. Our design does not contain pauses, which makes our study the first we know that showed success of learning non-adjacent dependencies without resorting to pauses. There have been studies of non-adjacent dependencies with auditory artificial language where the non-adjacent dependency is embedded in which dependent items sometimes occur at edges enables the detection of non-adjacent patterns (Mintz et al., 2014; Reeder, Newport & Aslin, 2013; Wang & Mintz, under review). In the cases where successful learning of non-adjacent dependencies has been reported, at least one edge (beginning or ending) is marked with pauses (Mintz et al., 2014; Reeder, Newport & Aslin, 2013). When both edges are not marked with pauses, learning failed (Wang & Mintz, under review). It is possible that having exposure to elements at edge positions facilitated, detecting non-adjacent dependencies at least initially. However, natural languages contain non-adjacent dependencies at non-edge positions, thus making it difficult to evaluate learning theories that requires the presence of pauses.

Why would the presentation rate make a difference? There are a number of possibilities. The auditory system for speech perception may be tuned towards a particular frequency (Kiebel, Daunizeau & Friston 2008), so efficient speech processing may play a role. This line of explanation is along the lines of modality-specific statistical learning theories (Emberson, Conway, & Christiansen 2011). They argued for the central role of modality-specific processing by contrasting the opposite influence of changing presentation rate in visual and auditory statistical learning. These theories hold promising directions for understanding the modality-specific statistical learning mechanisms, but they are also vague regarding why specific kind of statistical learning (in this case, non-adjacent dependency learning) would benefit from fast presentation. We leave these questions for future research.

Success in learning non-adjacent dependencies without pauses point to the possibility the non-adjacent dependency learning mechanisms with spoken language do not critically require the presence of pauses as a prosodic cue, contrary to previous theories (see Peña et al. 2002, for a discussion). Peña et al. 2002 argued that successful learning requires a prosodic analysis whereby boundaries and positional information is obtained before non-adjacent dependencies are learnable. Across all the previous studies that report success on non-adjacent dependency learning with spoken artificial language (Peña et al. 2002; Newport & Aslin, 2004), this has been the case. The current work suggests that pauses are not a necessary condition for learning non-adjacent dependencies. In the absence of explicit pauses, we speculate that the learning mechanism may still have access to virtual boundaries that arise via a distributional analysis that detects simpler repeated patterns. There may be some kind of distributional or syntactic analyses that can make uses of these positional boundaries which in turn may reduce the computational load for calculating dependency relations between non-adjacent items, and induce the detection of higher order dependencies, such as non-adjacent dependency in the current paper. Future work should examine these possibilities. In sum, the position based accounts (Endress et al., 2009) may still apply to the current findings, except that positional information may not come from prosodic processes, but it may be obtained from distributional analysis as well.

One methodological note is regarding the measure we took, which is a rating scale of confidence. This is different from how artificial language is assessed in the past literature, which involves a variant of this question, “Have you heard this sentence in the language before?”, “Is this sentence in the language?”, etc, requiring a yes/no answer from participants. There are a number of problems with this approach, most of which involves the interpretation of the phrase “in the language”. What does it mean to a naïve participant that a novel sentence is in a novel language? Does it mean that the sentence literally heard? Or does it mean that it follows some kind of a rule? Regardless, given any interpretation of “in the language”, participants also need to decide on the criterion when a phrase is “close enough” to be in the language. Many artificial language studies from our labs suggest that participants may simply answer yes to all questions, because it is not clear to them what the experimenter is asking (for similar results, see Gómez, 2002). In light of these findings, we used a rating scale instead of collecting yes/no responses. Rating that making subjects make explicit judgments, we asked participants to report their confidence level that a phrase has been heard. This measure, degrees of certainty, does not require any commitment to any type of meta-linguistic knowledge of knowing what it means to be “in a language”, but rather, assesses familiarity with a phrase. Making use of this measure has yielded much success with multiple artificial language/statistical learning studies from our lab already.

Getting back to our study, we wish to emphasize the role of the timing. There are other word level non-adjacent dependency studies (Gómez, 2002; Romberg & Saffran, 2013), but the timing profile is different. In those experiments, utterances were concatenated words, such that there was around 0.8 s between word onsets. This is a relatively slow rate of speech, which can be considered unnatural as far as speech perception is concerned in terms of its timing characteristics. It is conceivable that this mode of presentation makes detecting patterns of non-adjacent elements more difficult because they are not temporally close.

Lastly, existing theories (Gomez, 2002, among others) suggest that the dependency is hard to detect without highly variable middle elements. This is different from our design in important ways. In our design, the variability of the middle elements (n=3) is very low according to Gomez 2002, making the dependency hard to learn. We show that this hard problem of learning of non-adjacent dependency can be solved when the non-adjacent dependency is presented at a typical speech rate. It remains possible that the variability issue is important when the presentation rate of speech is slow, but at least with fast presentation rate, low variability does not seem to lead to failure to learn. Future work is needed to examine whether increase variability will make learning more robust.

In sum, we have shown that temporally controlled word-level non-adjacent dependency is learnable without pauses. We propose that learning about distributional analysis may be best obtained the learning material is presented at the optimal rate is critical, and the importance of the speech rate may outweigh constraints previously proposed, such as presence of pauses and variability in the middle element.

References

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International* (2001): 341-345.

Chemla, E., Mintz, T. H., Bernal, S., & Christophe, A. (2009). Categorizing words using 'frequent frames': what cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental Science*, 12(3), 396-406.

Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *The Quarterly Journal of Experimental Psychology*, 64(5), 1021-1040.

Endress, A.D., Nespor, M. & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences*, 13(8), 348-353.

Frensch, P. A., & Miner, C. S. (1994). Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory & Cognition*, 22(1), 95-110.

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing Structures in Midstream: Learning Along the Statistical Garden Path. *Cognitive Science*, *33*(6), 1087–1116. http://doi.org/10.1111/j.1551-6709.2009.01041.x

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13(5), 431-436.

Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, 7, 183-206.

Gómez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychological Science*, 17(8), 670-674.

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS computational biology*, 4(11).

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.

Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(1), 80.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298(5593), 604-607.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive psychology*, 48(2), 127-162.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive psychology*, 66(1), 30-54.

Romberg, A. R., & Saffran, J. R. (2013). All together now: Concurrent learning of multiple structures in an artificial language. *Cognitive Science*, *37*(7), 1290-1320.

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition*, 69(2), 105-134.

Shi, R., & Melançon, A. (2010). Syntactic Categorization in French-Learning Infants. *Infancy*, 15(5), 517-533.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*, 552-564.

Van den Bos, E., & Christiansen, M.H. (2009). Sensitivity to non-adjacent dependencies embedded in sequences of symbols. *Proceedings of the 31st Annual Meeting of the Cognitive Science Society.*

Vuong, L.C., Meyer, A.S. & Christiansen, M.H. (in press). Concurrent statistical learning of adjacent and non-adjacent dependencies. *Language Learning*.

Wang, H., Höhle, B., Ketrez, N. F., Küntay, A. C., & Mintz, T. H. (2011). Cross-linguistic Distributional Analyses with Frequent Frames: The Cases of German and Turkish. In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the 35th annual Boston University Conference on Language Development* (pp. 628-640). Somerville, MA: Cascadilla Press.

Wang, F. H., & Mintz, T. H. (under review). Learning Non-Adjacent Dependencies Embedded in Sentences of an Artificial Language: When Learning Breaks Down.