Title
Grammatical Bracketing Determines Learning of Non-adjacent Dependencies

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 38(0)

Authors
Wang, Felix Hao
Zevin, Jason
Mintz, Toby

Publication Date
2016

Peer reviewed

Grammatical Bracketing Determines Learning of Non-adjacent Dependencies

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

Grammatical dependencies often involve elements that are not adjacent. However, most experiments in which non-adjacent dependencies are learned bracketed the dependent material with pauses, which is not how dependencies appear in natural language. Here we report successful learning of embedded NAD without pause bracketing. Instead, we induce learners to compute structure in an artificial language by entraining them through processing English sentences. We also found that learning becomes difficult when grammatical entrainment causes learners to compute boundaries that are misaligned with NAD structures. In sum, we demonstrated that grammatical entrainment can induce boundaries that can carry over to reveal structures in novel language materials, and this effect can be used to induce learning of non-adjacent dependencies.

Keywords: non-adjacent dependency learning; grammatical entrainment

Introduction

Due to the hierarchical organization of the syntax of natural languages, lexical items (and morphemes) that are syntactically related are not always linearly adjacent. Thus, to acquire the specifics of the hierarchical grammar, learners must be able to track dependencies of linguistic items that are both linearly adjacent and non-adjacent. For example, given the dependency between the singular subject *child* and the agreeing inflected verb *runs*, the subject and the verb are adjacent in *the child runs*, and non-adjacent in *the child always runs*. Therefore, an important question in language acquisition is how learners acquire adjacent and non-adjacent grammatical dependencies, as they could help learners understand how their language is structured. There has been considerable interest in investigating learning mechanisms that could detect these dependencies in linear sequences within spoken utterances. These mechanisms could be useful for discovering syntactic structures when children start to acquire their first language, and facilitate building syntactic parses later in life. For example, many studies employing artificial and natural languages have investigated how language learners acquire non-adjacent dependencies (e.g., Gómez, 2002; Peña, Bonatti, Nespor & Mehler, 2002; Newport & Aslin, 2004; Romberg & Saffran, 2013), and how early in the acquisition process such dependencies are detected (Santelmann & Jusczyk, 1998; Gómez, 2002).

Whereas most studies on adjacent dependencies have found success in a variety of learning paradigms, the studies on non-adjacent dependency learning to date have only found evidence in limited situations, with some reporting success in learning and others reporting failure. One consistent 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; Romberg & Saffran, 2013). 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.

But in natural languages, non-adjacent dependencies are not restricted to positions of silences (e.g., utterance boundaries), and are often embedded in longer sequences. Learning the dependency relations of a natural language generally require learning non-adjacent dependencies of items that may not always occur at silence boundaries. This may pose a problem for the learning theories mentioned above. One possibility is that the non-adjacent dependency learning with spoken language critically requires the presence of pauses as a prosodic cue. According to this interpretation (see Peña et al. 2002, for a discussion), successful learning requires an interaction between prosody and syntactic analysis. However, it is also possible that the edges that the learning mechanism requires are broader in scope, including boundaries in representations rather than directly in the signal (pauses being the latter). For example, edges or boundaries that are represented as the result of a syntactic parse might play a similar role to pauses. Under this interpretation, prosody per se might not be critical for learning non-adjacent dependencies. But rather, the availability of any kind of boundary, in the signal or computed, could facilitate the detection of non-adjacent dependencies when the boundaries are in close alignment with them.

In order to distinguish these hypotheses, we sought a way to induce edges and bracketing via syntactic analysis without resorting to pauses. To this end, we made use of a recently published phenomenon described as grammatical entrainment. When sentences with the same syntactic structures are presented repeatedly in a cyclical pattern, MEG recordings identified cortical regions which track the syntactic structures and compute syntactic boundaries when structures repeat (Ding, Melloni, Zhang, Tian & Poeppel, 2015). One of the proposed mechanisms is that the repeated presentation of the same syntactic structure makes it possible to predict the syntax of the sentence that comes next.

We reasoned that such entrainment of structural processing might carry over to an artificial language that is presented immediately following cyclic presentation of structurally similar English sentences presented as in Ding et al. (2015). If this process induces syntactic edges in a predictable, cyclic way, it could potentially allow the same analysis to be applied and transferred to new linguistic materials that are presented also in the same frequency (words-per-second) and phase as the English material. That is, perhaps repeated structure building of familiar material (i.e., English) can be used to guide upcoming parsing of novel material (i.e., and artificial language). To this end, we placed novel artificial language after repetitions of the same English syntactic structure across a variety of English sentences, matched in word length (4 words). At the very least, we expected that the English sentences would entrain listeners to segment into 4-word sequences, effectively inducing virtual boundaries in the novel artificial language. Specifically, we predicted that when English grammatical dependencies are repeated in phase with the artificial language dependencies, they would facilitate detection of the artificial non-adjacent dependencies (Experiment 1). We further predicted that when the phase relationship between English and artificial language was mismatched, the facilitation effect would disappear (Experiment 2).

Experiment 1

Methods

Participants. Twenty-four undergraduate students at University of Southern California were recruited from Psychology Department subject pool. Half of them participated in each counterbalancing condition. The sample size was based on previous studies (Newport & Aslin 2004).

Stimuli. We recorded speech from a native English speaker and digitized the recording at a rate of 44.1 kHz. We recorded 2 types of words: English words and novel words. For English words, we recorded 5 names (Brian, John, Kate, Nate, Clair), 5 monosyllabic verbs in 3rd person singular form (turns, keeps, puts, lets, has), 5 pronouns (these, those, this, that, it) and 5 adverbs (down, on, up, off, in). For the non-adjacent dependency, we used 12 novel words: 3 at position 1 (pel, tink, blit), 3 at position 2 (swech, voy, rud), 3 at position 3 (dap, tood, wesh) and 3 at position 4 (ghire, jub, tiv).

After all the words were recorded in list intonation in a random order, we spliced the words from the recording. Each word by itself from the recording lasted 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 intelligibility and such that when words are concatenated in a continuous stream, they would occur at 3Hz. This was not intended to be a manipulation, and is certainly not the same as the design from Peña, et al., 2002 because the pauses do not pre-segment phrases into four-syllable chunks that line up with the dependency. The pauses are after every word, making it uninformative of the dependency structure.

Design and procedure. There were three blocks of training phase and testing phase, with each testing phase following a training phase. The training materials and testing materials were divided into 3 equal proportions for the 3 blocks.

Training phase. To create the training stream, we mixed English sentences and artificial sentences together. A total of 432 novel word sentences and 858 English sentences were randomly concatenated together in the following fashion. Each English sentence was created by randomly picking a name, a verb, a pronoun, and an adverb, in that order. As such, each sentence consisted of 4 syllables (with the exception of sentences containing the word Brian), and lasted 1.33 seconds. Since words were randomly selected and constrained only by position, there was no statistical dependency on the level of words.

Each novel sentence was a concatenation of 4 novel words, 1 each from choices of 3 for each position, as specified in the Stimuli section. We represent this pattern as YAXB, with one letter for each class of novel word. The second position word predicted the fourth position word (YA_iXB_i), so the fourth word is predictable from the second word. All the other words (at positions 1, 2 and 3) cannot be predicted. Given that there are 3 different words for each non-dependent position (1, 2, & 3), there were 27 possible different quadruplet artificial sentences.

The training stream was made by concatenating alternating English and artificial language sentences in the following way. We concatenated 5, 6 or 7 English sentences together to create the entrainment effect. After the English sentences, three artificial language sentences were presented, followed by more English sentences (see Figure 1 for a demonstration). There are no additional pauses between English sentences, between novel words of artificial language, or the boundary between English words or novel words.

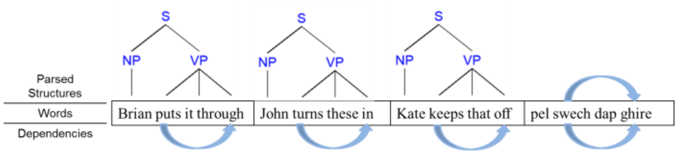

Figure 1. The design of language materials in the training phase of Experiment 1. The English sentence and the artificial dependencies are in phase. If the third sentence is [pel *swech* dap *ghire*] and the following sentence is [blit rud dap tiv], learners may learn the dependency between *swech* and *ghire* (as well as rud & tiv).

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 Gómez (2002). There were three $A_i\_B_i$ frames (i.e., $A_i\_B_i$, where i=1-3). The two counterbalancing languages were created by taking three pairs of A_B for the three frames in one training language ($A_1\_B_1$, $A_2\_B_2$, $A_3\_B_3$), and three different pairs for the other training language ($A_1\_B_2$, $A_2\_B_3$, $A_3\_B_1$). This design allowed the use of the same set of test items for the two counterbalancing training languages, where the set included both frames from the two training languages.

Participants listened to the sound stream passively through the headphones while the screen was blank. The listening part lasted about 9 minutes per block.

Test phase. Immediately after each training block, we showed instructions for the test phase 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 per block, half of which were from the correct dependency, and the other half from incorrect ones. The sequence of presenting the test trials was randomized for each participant.

Participants initiated each test trial. Per trial, we played an artificial language sentence, and asked the participant to indicate whether some sequences are from the previous section that they have heard. A scale 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 choices, and the trial ended and next trial began.

Results

We coded the scale of "Definitely", "Maybe", "Not Sure", "Maybe Not" and "Definitely Not" into numeric values of 1 through 5. To compare ratings statistically, we ran mixed effect linear regressions with the data. In the regression, ratings were the dependent variable, where the interaction and main effects of item type (correct vs. incorrect) and block number (1 through 3) was entered in the fixed effect with subject as the random effect. The interaction was found to be significant (chi2(2) = 8.37, p=0.0152). For this interaction, the regression revealed that the difference score between correct items and incorrect items in block 2 was not significantly different from block 1 (β=0.019, z=0.13, p=0.895) but block 3 is (β=-0.343, z=-2.44, p= 0.015). We conclude that the difference of ratings for correct vs. incorrect items was found to be significant. A plot of the data and comparisons can be found in Figure 2.

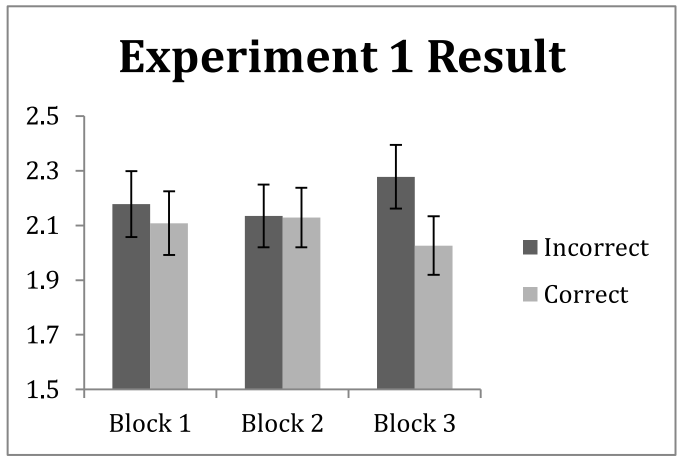

Figure 2. Ratings of test items by block for correct and incorrect items. Smaller rating indicate acceptance of the item. Error bars represent 95% confidence intervals around the mean.

Discussion

In Experiment 1, we introduced a window of analysis from English that participants can use to break into the artificial language. Given the syntactic structures in English, the bracketing of artificial language can be induced such that structural analysis can be applied to artificial language sequences. Applying the boundary every 4 words into the artificial language, we get the chunks [Kate keeps that off] [YA_iXB_i] [YA_jXB_j], which facilitates dependency learning.

It's worth characterizing the way we constructed the training material that produced grammatical entrainment. The English sentences we used did not contain any acoustic information regarding its structure unless the knowledge of English is applied. The construction was such that all the English sentences had constant transitional probability between each word, and the words between each sentences, given that the English sentences were made from a random Markov process. That is, each English word just predicts a word from a set of the next words with equal probability. If no knowledge of English is present, words come randomly without informative landmarks. Without English grammar, listening to the English words would not result in any kind of boundaries that can be used for processing of later artificial language. The fact that our participants were native speakers of English made grammatical entrainment possible, which facilitated the detection of non-adjacent dependencies.

There are some alternative accounts to our current proposal that listeners transferred syntactic boundaries from English to the artificial language. One possibility is that the length of the English sentences (4 words) encouraged listeners to process the artificial language in smaller chunks, and that was sufficient to ease processing and facilitate detection of non-adjacent dependencies. In other words, the alignment of the chunk boundaries with the non-adjacent patterns may have been irrelevant, and all that was necessary was processing shorter sequences. A related possibility is that presenting the same syntactic structures induced syntactic priming (Bock 1986), whereby the repeated presentation of the same syntactic structure makes it easier to reuse/reactivate structures of the same type. In our case, the fact that we present one syntactic structure (verb + prep. + verb particle) over and over again may have sensitized participants to dependency lengths of three if syntactic priming is operating. Furthermore, given that the boundaries induced from English align with the switch from English to artificial words, there might be boundaries arising from this shift. These boundaries may simply serve as a starting counter for the novel word sequences, from which edge-based computations can be performed. We rule out these alternatives in Experiment 2.

Experiment 2

Methods

Participants. Twenty-four undergraduate students at University of Southern California were recruited from Psychology Department subject pool. Half of them participated in each counterbalancing condition. None of the participants participated in Experiment 1.

Stimuli. We used the same stimuli from Experiment 1.

Design and procedure. Similar to Experiment 1, there were three blocks of training phase and testing phase, with each testing phase following a training phase. The training materials and testing materials were divided into 3 equal proportions for the 3 blocks, and the only difference between Experiment 2 and Experiment 1 is the training materials (described below).

Training phase. The training phase is similar to Experiment 1 except for one key difference in how the English words were ordered. We mixed a total of 432 novel word sentences and 858 English sentences were randomly concatenated together in the same fashion as in Experiment 1, except for the order of the English words. Each English 'sentence' was created by randomly picking a verb, a pronoun, an adverb, and a name, in that order. Given such ordering of English words, parsing of these sentences described above will not start with the beginning to the end of each quadruplets (1-2-3-4); rather, it would restart from the last word of the quadruplet, making use of the name in the current quadruplet, and continuing to the next quadruplet, making a sentence out of the current 4th word and the first 3 from the next (4-1-2-3; see Figure 3 for a demonstration). The purpose of this deliberate phase shift is to change where the syntactic boundaries come in the parsing process, which will become important for the artificial language parsing later. Similar as in Experiment 1, there is no statistical dependency on the level of words, given that any one of the 4 words can appear at its position. Dependency can only be defined grammatically, given listener's knowledge of English and the parsing of the sentences.

The novel sentences were exactly the same as in Experiment 1, and the pattern can be similarly represented as YA_iXB_i. Since the phase of the English sentences was such that the first position and the third position contain the dependency (verb and verb particle), the dependency in the artificial language is out of phase with respect to the English sentences, and instead is aligned with the Y_X structures, which are independent.

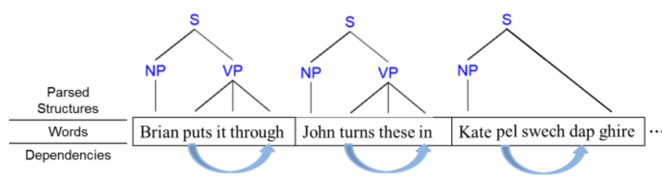

Figure 3. The design of language materials in the training phase of Experiment 2. If the third sentence is [Kate pel *swech* dap] and the following sentence is [*ghire* blit rud dap], learners fail to learn the dependency between *swech* and *ghire*.

Similarly to Experiment 1, 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.

The language stream started with a name, making the first English sentence grammatical. Participants listened to the alternating English and artificial language sentence stream passively through the headphones while the screen was blank.

Test phase. The Test phase is exactly the same as Experiment 1, with 3 blocks, each block containing the same number of correct and incorrect items to be rated.

Results

We coded the scale of "Definitely", "Maybe", "Not Sure", "Maybe Not" and "Definitely Not" into numeric values of 1 through 5. To compare ratings statistically, we ran mixed effect linear regressions with the data. In the regression, ratings were the dependent variable, where the interaction and main effects of item type (correct vs. incorrect) and block number (1 through 3) was entered in the fixed effect with subject as the random intercept. The interaction was found to be not significant (chi2(2) = 2.50, p= 0.287). For this interaction, the regression revealed that the difference score between correct items and incorrect items in block 2 was not significantly different from block 1 (β=-0.157, z= -1.16, p=0.244), nor is in block 3 (β=0.046, z= 0.34, p= 0.732). Dropping the fixed effect of block, I rerun the regression with only item type as the fixed effect with subjects as random effects. Again, there is no evidence for learning, as correct items were not rated differently from the incorrect items (β=-0.078, z=-1.42, p=0.154). We conclude that there is no significant difference in ratings for correct vs. incorrect items. We plot the data in Figure 4.

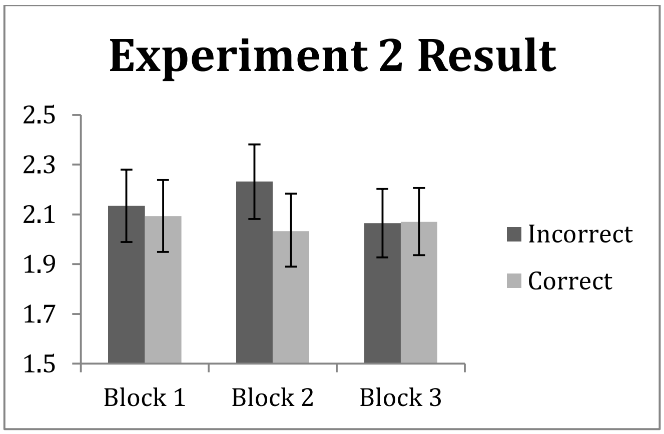

Figure 4. Ratings of test items by block for correct and incorrect items. Smaller rating indicate acceptance of the item. Error bars represent 95% confidence intervals around the mean.

Discussion

In Experiment 2, the phase of the English sentences was mismatched to the phase of the artificial language dependency whereby the non-adjacent dependency was not within each chunk of the segmented material. Participants in Experiment 2 failed to learn the artificial language dependency.

We can eliminate some of the possibilities mentioned in the discussion of Experiment 1. First, similar to Experiment 1, Experiment 2 also used English sentences that were 4 words long. If this helps the artificial language parse into 4-word chunks (as we argue it does, see below), this factor alone is not sufficient to support successful detection. Similarly, a syntactic priming account would not entirely account for the result, given that the same verb to verb particle dependency exists in Experiment 2 as well. We cannot eliminate the potential influence of syntactic priming, given that we presented the same syntactic structure (verb + prep. + verb particle) over and over again, but again, it is not enough to facilitate detection. Lastly, if there were boundaries that were due to changes of English words to artificial words, these boundaries are still present in Experiment 2. Experiment 2 thus suggests that the success of learning in Experiment 1 is not a result merely of computing boundaries in sequences of novel words.

Rather our interpretation is that the English sentences provided edges in specific locations that were sufficiently aligned with the dependencies in the artificial language. If English parsing carried over to parsing the artificial language, even at the coarsest 'sentence' level, the dependencies in Experiment 2 would have been chunked as [Kate YA_iX] [B_iYA_jX]…, and the A_iXB_i dependency would never have been considered, and never be learned as a result. We argue that the reason for failure to learn the dependency is a result of bracketing elements across boundaries. Experiment 2 contrasts with Experiment 1 in that the lack of facilitation effect given the mismatch with phase support the presence of grammatical entrainment, given all other factors (acoustic, transitional probability) are controlled for.

General Discussion

In this paper, we report our first attempt at inducing non-adjacent dependency learning with grammatical bracketing. As we mentioned, non-adjacent dependencies are generally hard to learn for a variety of reasons. For the most part, past literature suggested (Newport & Aslin, 2004; Peña et al., 2002) that pauses are critical to the learning of syllable-level non-adjacent dependencies. Our design does not contain pauses, which makes our study the first we know of that addresses the issue of handling the window of analysis in non-adjacent dependency with language learning without resorting to pauses. We show that this hard problem of learning of non-adjacent dependency can be solved when the non-adjacent dependency is entrained to an English rhythm that provided syntactic edges to the novel artificial language in such a way that the edges line up with the dependencies to be learned. Furthermore, we show that when the syntactic edges slice the novel language into chunks that do not contain the intended dependency, it leads to failure of learning.

Existing theories posits that perceptual or memory primitives guide aspects of statistical based learning, and more specifically, that edge-based computations are critically required for computing non-adjacent dependencies (Peña et al., 2002). However, this theory left the description of edges rather vague. In the studies cited here on non-adjacent dependencies from Endress and colleagues, the implementation of an edge has been a period of silence. As we mentioned, this points to a possibility that the non-adjacent dependency learning mechanisms critically require the presence of pauses as a prosodic cue, whereby an interaction between prosody and syntactic analysis. However, our data suggest that this may not be the case. One of the advantages of an edge-based computation is that it shortens the sentence length that the computations are performed on. This has great implication for how long-distance dependencies can be detected from a computational perspective. Linguistic dependency can exist between any element within a sentence to any other element the number of dependencies grows factorially with the number of elements within a sentence (between-sentence dependencies also exist in language, which is not considered here). If boundaries are only at utterance boundaries, detection of long-distance dependencies may quickly become intractable as the length of sentence increases (see Wang & Mintz, under review, for a demonstration).

It follows that the learning mechanism for non-adjacent dependencies may require edges. Yet non-adjacent dependencies are not limited to start or end of sentences in natural languages, and as such, the learning mechanism in natural language for effective detection of non-adjacent dependencies may resort to edges beyond utterance boundaries or silences. To solve this problem, we demonstrate that non-adjacent dependencies are learnable when the boundaries in a syntactic sense brackets the continuous speech stream into chunks that contain the dependency. Our study shows that prosody by itself is not critical for learning non-adjacent dependencies. But rather, computing structural boundaries at the beginning and end of chunks that contain the dependency also facilitate their detection. In this sense, the pauses are serving the same function as the syntactic boundaries here, perhaps restricting the window of analysis the detection mechanism operate within. However, characterizing that the learning mechanism requires some kind of prosodic processes such as pauses would be an under-specification.

We can speculate about the role of the English sentences in Experiment 1 further. One possibility is that the English sentence simply sets a pace at every 4 word, bracketing artificial language to 4-word chunks. Potentially, having a chunk close to the size of dependency is enough for learning to happen, though given the low variability in the middle element, learning would occur slowly. Existing theories (Gomez, 2002, among others) suggest that the dependency is hard to detect without highly variable middle elements. In our design, the variability of the middle elements (n=3) is very low according to Gomez 2002, making the dependency hard to learn. Alternatively, processing the English sentences result in a hierarchical parse, such that the verb and the verb particle have a syntactic non-adjacent dependency (e.g., *puts … on*). If this parse transfers to new linguistic material, it would imply that the detection of non-adjacent dependency is facilitated by this narrow window of in-phase pattern matching. Learning is a result of many factors, and future research will separate these possibilities in detail. Although this is an intriguing possibility, from our data we cannot determine if the internal dependency is playing a role, as its alignment to the artificial dependency is confounded with the alignment of the of the sentence boundary and the artificial dependency.

We recognize that various aspects of our design are artificial, especially in terms of the temporal nature of the English and artificial language material. The goal was to use the rhythmic properties as a tool to stimulate syntactic transfer from processing of known structures to novel ones. At one level, this could be viewed as a kind of syntactic bootstrapping, either at a coarse grain level (the sentence level), or fine grain level (phrase level), where prior structures organize the interpretation of novel material.

There are many future directions to this preliminary work. As we just discussed, it is possible that entrainment and transfer involved the internal syntactic dependency. Moreover, it is conceivable the syntactic categories have carried over to the artificial words. There could be multiple reasons this can happen. Given the cyclical nature of the syntactic assignment of English words, the syntactic assignment can potentially carry over. That is to say, different artificial words may become more verb like or pronoun like, depending on its position and the non-adjacent dependency may facilitate syntactic categorization (Mintz, Wang & Li, 2014). Another direction to go is to examine how syntactic bracketing happens in artificial language in general. Does the presence of adjacent and non-adjacent dependencies alone induce bracketing? What kind of mechanism is involved in deducing chunks from statistical information?

In sum, we have argued that correct bracketing is crucial to learn about elements within a chunk. We propose that thinking about bracketing is a useful way of puzzle solving around learning linguistic dependencies, and that having a correct window of analysis is crucial for such purposes.

References

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive psychology*, 18(3), 355-387.

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

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2015). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature neuroscience*.

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

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

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.

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

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

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.

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