# Testing the limits of non-adjacent dependency learning: Statistical segmentation and generalization across domains

**Rebecca L. A. Frost (rebecca.frost@mpi.nl)**
Language Development Department, Max Planck Institute for Psycholinguistics, Nijmegen, NL

**Erin S. Isbilen (esi6@cornell.edu)**
Department of Psychology, Cornell University, Ithaca, NY, USA

**Morten H. Christiansen (christiansen@cornell.edu)**
Department of Psychology, Cornell University, Ithaca, NY, USA

**Padraic Monaghan (p.j.monaghan@uva.nl)**
Department of English Language and Culture, University of Amsterdam, NL;
Department of Psychology, Lancaster University, UK

**Abstract**

Achieving linguistic proficiency requires identifying words from speech, and discovering the constraints that govern the way those words are used. In a recent study of non-adjacent dependency learning, Frost and Monaghan (2016) demonstrated that learners may perform these tasks together, using similar statistical processes — contrary to prior suggestions. However, in their study, non-adjacent dependencies were marked by phonological cues (plosive-continuant-plosive structure), which may have influenced learning. Here, we test the necessity of these cues by comparing learning across three conditions; *fixed phonology*, which contains these cues, *varied phonology*, which omits them, and *shape*s, which uses visual shape sequences to assess the generality of statistical processing for these tasks. Participants segmented the sequences and generalized the structure in both auditory conditions, but learning was best when phonological cues were present. Learning was around chance on both tasks for the visual shapes group, indicating statistical processing may critically differ across domains.

**Keywords:** statistical learning; speech segmentation; generalization, language learning; non-adjacent dependencies; implicit learning

## Background

Learners must master a number of critical tasks in order to reach linguistic proficiency, including learning how to segment individual words from speech, and learning to identify the constraints that govern the way those words are structured and used. Learners are remarkably adept at these tasks, thanks in part to the myriad cues that speech contains that may assist learning. One such cue is the statistics that describe co-occurrences of items in speech; for instance, the co-occurrence of syllables provides a helpful cue to what constitutes possible words, while information about how those words are used in combination helps learners to discern how the language operates. The ability to detect and draw on this distributional information - *statistical learning* - is suggested to play a key role in language acquisition, for both segmenting speech and for learning about grammatical structure (e.g., Conway, Bauernschmidt, Huang, & Pisoni, 2010; Frost, Monaghan, & Christiansen, 2019; Redington & Chater, 1997).

Since word- and structure-learning appear to have distinct requirements, it is unsurprising that the nature of the (statistical) processes that underlie these tasks has been subject to substantial debate (e.g., Peña, Bonatti, Nespor, & Mehler, 2002; Perruchet, Tyler, Galland, & Peereman, 2004). Central to these discussions have been questions concerning the types of computations required to discover word-like and rule-like items in speech, and learners' capacity to do so by computing over co-occurrence statistics.

These issues have been extensively tested using a classic artificial language learning paradigm (Peña et al., 2002), which examines learners' ability to acquire linguistic structure that is defined in terms of non-adjacent dependencies (i.e., an AxC structure, where A and C are syllables that reliably co-occur, regardless of which x syllable intervenes). AxC languages are used to jointly assess learners' capacity for statistical word and structure learning, since they contain novel words that learners must discover (AxC strings), in addition to structural regularities within those words (A-C relationships).

Initial studies using this paradigm suggested that learners perform statistical computations on the non-adjacent dependencies to segment the speech into individual AxC strings (or *words*), but perform more abstract computations on those words in order to learn about their structure - and perhaps do so only when speech segmentation has been resolved (typically by inserting pauses between words in the training stream).

A recent study by Frost and Monaghan (2016) expanded on this work, aiming to shed further light on two key questions about how word- and structure-learning unfold in language acquisition: whether these tasks occur sequentially or simultaneously, and whether they may actually utilize similar statistical computations – contrary to prior suggestions. In their study, participants were able to draw on the non-adjacent dependencies to segment continuous speech into words, *and* to learn about the non-adjacent dependency structure that those words contained, possibly simultaneously (though further work is required to conclusively establish the time-course of learning for these tasks). The key difference between this and earlier work on this phenomenon was a slight methodological change which addressed a possible confound in the previous measure of generalization. Specifically, prior generalization tasks typically required learners to indicate a preference for 'rule words' over part-words, with rule words comprising a trained dependency, intervened by an onset/coda from another dependency (e.g., $A_1A_2C_1$ or $A_1C_2C_1$). While such comparisons do permit assessment of preference for the overall structure, they require learners to use trained A and C items flexibly in a way that deviates from their knowledge of syllable position, which may affect performance. Indeed, using amended test items (trained dependencies with entirely novel intervening items), Frost and Monaghan (2016) demonstrated that adults can segment statistical nonadjacent dependencies and generalize them to novel grammatically consistent instances in the absence of additional information, such as pauses between words (see Isbilen, Frost, Monaghan, & Christiansen, 2018, for a replication of this effect).

This finding was contrary to prior suggestions that these tasks are fundamentally computationally distinct (e.g., Peña et al., 2002), and provides crucial evidence to suggest that learners may draw on the same type of statistical processing mechanisms for both of these tasks, and they may do so at the same time during language learning.

However, one possibility that cannot be overlooked is that learning in this study was not just driven by computations over transitional probabilities; learning may have been assisted by the phonological properties of the language. In line with Peña et al.'s (2002) landmark study, Frost and Monaghan (2016) employed an artificial language that contained both statistical dependencies between elements, and phonological structure, which aligned with the non-adjacency structure such that A and C syllables contained plosives, whereas intervening x syllables contained continuants.

Prior research has noted that the pattern of phonological information in artificial languages can significantly benefit learning, and phonological similarity between related elements has been found to support learning of non-adjacent dependencies in particular. For instance, in a series of experiments with a similar paradigm, Newport and Aslin, (2004) demonstrated that learning nonadjacent dependencies between syllables was remarkably difficult to accomplish in the absence of phonological cues (though the difficulty there may also have been due to additional factors, including learnability of the language - i.e., the number of dependencies, and the number of intervening items, which has been shown to impact learning - together with the relative complexity of some of the tests). Similarly, in Gomez and Gerken (1999), dependency learning was supported by phonological distinctions between A/C items and x items, where A and C were bisyllabic, and x were monosyllabic. Yet, research has also suggested that this phonological information should not be essential for learning to take place (Onnis, Monaghan, Christiansen, & Chater, 2004). Further research is therefore required to assess the extent to which this phonological information guided learning in Frost and Monaghan's (2016) study, to determine whether learners can indeed discover words and structures together, from distributional information alone.

In the present paper, we replicate Frost and Monaghan (2016), to confirm that participants can compute over non-adjacent dependencies to learn about both words and structure. We also test whether scores on these tasks correlate, to further assess whether these abilities are similar, or distinct. Crucially, we also compare performance for this replication against that for a condition in which participants are trained on the same language but with a more varied phonology (i.e., without phonological cues). Examining the extent to which segmentation and generalization are possible in the absence of these phonological cues will provide critical insights into how learners rely on statistical computations during language acquisition, by removing the possibility that successful performance is due to additional information outside of the syllable distribution.

While manipulating properties of the language allows us to determine how multiple cues interact with statistical learning, it does not inform us about whether that learning is due to domain-specific mechanisms, or whether language learning involves the specific application of general-purpose learning mechanisms (Frost, Monaghan, & Tatstumi, 2017; Siegelman & Frost, 2015). To further explore adults' capacity to compute non-adjacent dependencies, we also assessed whether their ability to do so is unique to language, by extending the paradigm to examine non-adjacent dependency learning from non-linguistic sequences (comprising shapes). This condition will help constrain theorizing on the generality of the mechanisms used for these tasks.

Thus, in this study we examine whether adults' capacity for segmenting and generalizing non-adjacent dependencies extends to more varied linguistic stimuli, or if it is contingent on a correspondence between distributional and phonological cues to structure. We will also assess whether this capacity is similar or different across modalities. We expect that participants will demonstrate knowledge of words and within-word structure (i.e., non-adjacent dependencies) in both language conditions (Frost & Monaghan, 2016; Onnis et al., 2004), and in the shapes group, in line with the suggestion that statistical learning mechanisms may serve learning broadly across modalities (e.g., Frost et al., 2017). We predict that segmentation and structure learning will benefit from phonological cues, but that these will not be essential for learning (Onnis et al., 2004). Further, we expect that structure learning will be better for linguistic than nonlinguistic input (due to increased experience with learning linguistic structure relative to structured sequences of shapes; Siegelman & Frost, 2015).

# Method

## Participants

90 Cornell University undergraduates (age: $M$ = 19.6 years, range = 18-24 years; 49 females, 41 males) participated for course credit. All participants were native English speakers.

## Design

Participants were randomly allocated to one of three conditions (each N = 30): *fixed phonology*, where AxC sequences contained plosive-continuant-plosive structure (Frost & Monaghan, 2016, Peña et al., 2002), *varied phonology*, which randomized the allocation of plosives and continuants to different positions within words, and *shapes*. These conditions permit comparison of learning from the original training input (fixed phonology) with an amended version containing no reliable phonological cues to word structure (varied phonology), and also a non-linguistic analogue. This will provide critical assessment of whether the pattern of learning demonstrated by Frost and Monaghan (2016) is unique to the properties of the input used in that study, or whether it can be extended to more varied linguistic input, as well as input in a different modality.

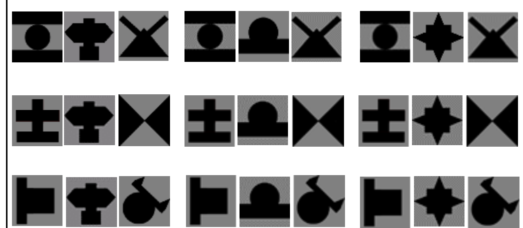
## Stimuli

Speech stimuli were created with Festival speech synthesiser, from a pool of 9 monosyllabic items (*pu, ki, be, du, ta, ga, li, ra, fo*), as used in Peña et al. (2002), and three additional monosyllabic items (*ve, zo, thi*). These additional syllables were reserved for the generalization task for the fixed phonology group in line with prior research (Frost & Monaghan, 2016), but formed part of the general syllable pool for the varied phonology group, to maximise variability. Shape stimuli were created from the Fiser and Aslin (2002) set of novel shapes (novel shapes in black on a grey background).

**Familiarization** Syllables/shapes were concatenated into triadic sequences that followed an AxC structure, with A, x, and C representing an individual syllable/shape. There were three A-C pairings, and three x items that could be used in all pairings ($A_1X_{1–3}C_1$, $A_2X_{1–3}C_2$, and $A_3X_{1–3}C_3$), giving 9 strings in total.

For the fixed phonology condition, syllables were mapped onto words pseudorandomly, such that A and C syllables were plosives, whereas x syllables were continuants, meaning each AxC string had a plosive-continuant plosive structure (e.g., *puraki*). For the varied phonology condition, syllables were randomly allocated to A, x, and C positions, meaning there were no reliable phonological cues that could guide learning. For the shapes condition, shapes were randomly allocated to A, x, and C positions, providing a visual non-linguistic analogue of the varied phonology condition. See Table 1 for example stimuli for each condition.

Table 1: Example stimuli for each condition

| Condition | Triads |
|---|---|
| Fixed Phonology | *puliki, puraki, pufoki*<br>*beliga, beraga, befoga*<br>*talidu, taradu, tafodu* |
| Varied Phonology | *livedu, liradu, likidu*<br>*fovezo, forazo, fokizo*<br>*bevepu, berapu, bekipu,* |
| Shapes |  |

Syllable/shape triplets were concatenated into familiarization streams containing 900 sequences (100 repetitions of each individual AxC sequence), in line with the materials used by Frost and Monaghan (2016). For speech stimuli, this was done using the Festival speech synthesizer (Black et al., 1990), and for shape stimuli this was done using Eprime 2.0. For all conditions, training streams contained no immediate repetition of individual AxC sequences.

For the fixed phonology and varied phonology conditions, the training stream lasted for 10.5 minutes, and was edited to have a 5-second fade-in and fade-out, to avoid providing cues to word boundaries.

For the shape sequences, presentation of the training stream took 22 minutes overall. For comfort this was split into 3 blocks of 300 sequences, and participants were invited to take short breaks in between blocks if desired. To ensure stimuli were analogous to the linguistic input, sequences were programmed such that shapes were presented sequentially, one by one. Shapes were presented for 225 ms in the centre of the screen, with a 225 ms inter-item interval between all shapes for comfortable viewing (note that since this occurs between all shapes, it does not cue segmentation). Presentation criteria were in line with those used in a comparable study by Frost et al. (2017). Analogous to the 5 second fade-in/-out applied to the speech streams, visual sequences always began and ended mid-triad, to prevent participants receiving any information about sequence boundaries at the start/end of the streams (this is true for the beginning and end of the entire sequence, and also for either side of the scheduled breaks).

To control for the relative ease of learning particular dependencies, for each condition 8 versions of the language were generated and counterbalanced across participants. For the varied phonology and shapes stimuli, these were created by randomly assigning syllables/shapes to A, x and C roles. For the fixed phonology stimuli, these were created by randomly assigning plosives to the A and C roles, while x items were always the same (see Frost & Monaghan, 2016).

**Testing** Learning was assessed using a two-alternative forced-choice (2AFC) test of segmentation and generalization. This contained 18 trials, nine of which assessed segmentation, and nine of which assessed generalization. Segmentation trials contained word versus part-word comparisons, with words being AxC items that occurred in the training stream, and part-words spanning word boundaries such that they comprised the end of one word and the start of another (e.g., xCA, CAx). Generalization trials contained rule-word versus part-word comparisons, where rule-words were trained dependencies but with novel intervening items (e.g., $A_1NC_1$), and part words were structured as before, but with one syllable replaced with a novel syllable (e.g., NCA, CNA, CAN). This was to control for the possibility that participants' responses on these trials were due to novelty alone (see Frost & Monaghan, 2016, for further discussion. Ongoing work by Isbilen, Frost, Monaghan and Christiansen further explores these generalization effects using $A_1N_1C_1$ vs. $A_1N_1C_2$ comparisons).

## Procedure

**Familiarization** Participants were presented with a familiarization stream which comprised either sequences of speech (10.5 minutes), or sequences of shapes (~22 minutes). Participants were instructed to pay attention to the sequences, and the shapes group was instructed to take optional breaks at the designated pauses if required.

**Testing** At test, participants completed a 2AFC task comprising 18 trials; nine segmentation trials (words versus part-word comparisons) and nine generalization trials (rule-words versus part-word comparisons). Presentation of segmentation and generalization trials was randomized. Participants were instructed to carefully listen to/look at each test pair, and indicate which of the two best matched the training stream they had just heard/seen.

# Results and Discussion

## Accuracy Scores

Accuracy scores for each condition are shown in Figure 1. One-sample t-tests (two-tailed) were conducted on the data for each group to compare performance to chance.

For the fixed phonology group, performance was significantly above chance for both the segmentation ($M$ = .709, $SD$ = .245), $t(29) = 4.659$, $p < .001$, $d = .853$ and generalization tasks ($M$ = .661, $SD$ = .173), $t(29) = 5.100$, $p < .001$, $d = .936$, replicating Frost and Monaghan's (2016) demonstration that learners can segment and generalize non-adjacent dependencies from continuous speech. For the varied phonology group, performance was also significantly above chance for both tasks (segmentation: $M$ = .623, $SD$ = .199, $t(29) = 3.391$, $p = .002$, $d = .618$; generalization: $M$ = .594, $SD$ = .217, $t(29) = 2.366$, $p = .025$, $d = .433$), suggesting that acquisition of statistically defined non-adjacent dependencies in this task is not contingent on the phonological properties of the speech input (i.e., phonological similarity between dependent syllables).

For the shapes group, however, performance on the segmentation task was only marginally above chance ($M$ = .552, $SD$ = .156), $t(29) = 1.827$, $p = .078$, $d = .333$), and performance on the generalization task was at chance level ($M$ = .485, $SD$ = .205), $t(29) = -.410$, $p = .685$, $d = -0.073$) – indicating that adults' ability to segment and generalize sequences using non-adjacent transitional probabilities may not extend to visually presented non-linguistic input.

Segmentation and generalization performance were significantly correlated for the fixed phonology ($r = .385$, $p = .036$) and varied phonology ($r = .625$, $p <. 001$) groups, but not for the shapes group ($r = .281$, $p = .133$).

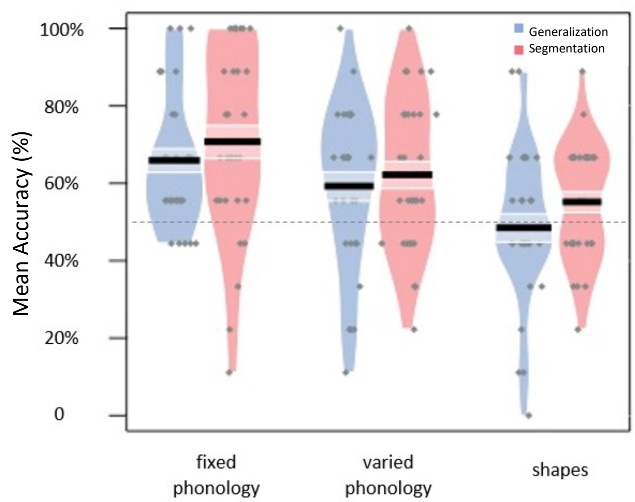



Figure 1. Pirate plot depicting performance on the segmentation and generalization tasks for each condition. Mean scores are shown in black, with standard error in white. The distribution of scores is depicted in red for the segmentation task, and blue for the generalization task, with individual participants' scores in grey. The dashed line indicates chance level.

## Comparing performance across groups

To compare performance across each of these groups, Generalized Linear Mixed Effects (GLMER) analysis was conducted on the data, examining whether segmentation and generalization scores differed according to whether participants were trained on sequences comprising varied or fixed phonology, or shapes. A significant main effect of condition would imply different overall performance across the groups, while a significant main effect of test type would indicate that participants performed differently on the segmentation and generalization tasks overall. An interaction between these variables would tell us that participants' performance on the segmentation and generalization tasks differed as a function of their condition – indicating that adults' capacity for statistical learning on these tasks differs across conditions, and possibly across domains, shedding light on the generality of the possible mechanism(s) that may underlie performance.

GLMER analysis was performed on the data (Baayen, Davidson, & Bates, 2008), modelling the probability (log odds) of response accuracy at test considering variation across participants and materials. The model was built incrementally, with random effects of subjects, particular test-pairs, and language version (to control for variation across the randomized assignments of phonemes to syllables). Random slopes were omitted if the model failed to converge with their inclusion (Barr, Levy, Scheepers, & Tily, 2013).

We then added condition (varied phonology, fixed phonology, and shapes) as a fixed effect, and considered its effect on model fit with likelihood ratio test comparisons. There was a significant effect of condition (model fit improvement over the model containing random effects: $\chi(2)^2 = 7.903$, $p = .019$), with the shapes group performing significantly worse than the fixed phonology group (difference estimate = -.767, $SE = .257$, $z = -2.987$, $p = .003$). The fixed phonology group also outperformed the varied phonology group, however this difference was marginal (difference estimate = -.389, $SE = .217$, $z = -1.788$, $p = .074$). We then added test type (segmentation and generalization), to see whether participants performed differently on each type of task. The effect of test type was marginal (model fit improvement over the model containing random effects: $\chi(2)^2 = 3.144$, $p = .076$) with participants performing better on the segmentation task than the generalization task (difference estimate = .224, $SE = .125$, $z = 1.791$, $p = .073$).

We then added the interaction between condition and test type, to see whether performance on the tasks differed according to the input participants had received. The interaction was not significant (model fit improvement over the model containing random effects: $\chi(2)^2 = .366$, $p = .833$), suggesting participants performed similarly across each of the conditions. See Table 2 for a summary of the final model.

Table 2: Summary of the GLMER (log odds) for accuracy scores.

| Fixed effects | Estimated coefficient | SE | Wald confidence intervals 2.50% | 97.50% | z | Pr (>\|z\|) |
|---|---|---|---|---|---|---|
| (Intercept) | .7405 | .2082 | .3325 | 1.149 | 3.557 | .0004 |
| Condition: Shapes | -.7658 | .2583 | -1.272 | -.2595 | -2.965 | .003 |
| Condition: Varied Phono | .-.3883 | .2183 | -.8161 | .0395 | -1.779 | .0753 |
| Test_type | .2235 | | -.0211 | .4680 | 1.791 | .07332 |
| **Random effects** | **Variance** | **Std. Dev.** | | | | |
| Subject (Intercept) | .355 | .5958 | | | | |
| Test Pair (Intercept) | .5871 | .773 | | | | |
| Lang_version | .0019 | .0435 | | | | |
| | AIC | BIC | logLik | Deviance | | |
| | 2097.6 | 2140.8 | -1040.8 | 2081.6 | | |

1620 observations, 90 participants, 18 trials. R syntax for the final model is: NAD_DG3 <- glmer (testresponse.ACC ~ condition + test_type + (1|subject) + (1+lang_ver|test_pair), data =NAD_DG, family=binomial, control=glmerControl(optimizer="bobyqa",optCtrl=list(maxfun=100000))).

## General Discussion

Recent evidence for the similarity (and possible simultaneity) of statistical segmentation and generalization has advanced our understanding of the way these processes unfold during language acquisition (see Frost & Monaghan, 2016, and see e.g,, Peña et al., 2002 and Perruchet et al. 2004, for more on the earlier debate about the nature of these tasks). Yet, due to the phonological properties of the training language, it is possible that learning in this recent study was not solely contingent on the statistical regularities contained within the language; learning may have been assisted by the plosive-continuant-plosive structure that AxC sequences adhered to (e.g., Newport & Aslin, 2004).

To explore this possibility, the study at hand examined adults' capacity for non-adjacent dependency learning across three conditions; the first of which used the input from Frost and Monaghan (2016) (see also Peña et al., 2002), which contained the phonological structure described above (termed the *fixed phonology* condition). The second condition omitted these phonological cues, such that AxC sequences had no fixed phonological structure (the *varied phonology* condition). The third condition tested learning from sequences of shapes, to provide a non-linguistic assessment of non-adjacent dependency learning, with a view of considering whether learning was comparable across modalities — perhaps drawing on similar statistical mechanisms. The critical test was whether participants in each group demonstrated learning (i.e., performed above chance), and whether performance in the varied phonology and shapes groups differed significantly from the fixed phonology group.

Participants in both language conditions performed significantly above chance on the segmentation and generalization tasks. This finding replicates the results of Frost and Monaghan (2016), showing that speech segmentation and structural generalization may proceed together during language learning, and can be accomplished from the same distributional statistics (though additional research is required to conclusively establish the precise time-course of learning for these tasks). Further, our results demonstrate that adults' capacity for learning non-adjacent dependencies extends to more phonologically diverse input. However, the difference in overall performance in these conditions was approaching significance, with results indicating that phonological cues were advantageous for learning (evidenced by marginally higher scores for the fixed phonology than the varied phonology group) — in line with Newport and Aslin's (2004) suggestion that such cues were important for learning. Critically though, our data indicate that these cues were not essential (Onnis et al., 2004).

In previous studies of word and structure learning, segmentation and generalization have tended to be tested separately. In the current study, these tasks were completed by all participants (within subjects). We show that the same learners can segment non-adjacencies from speech, and generalize them to new instances (see also Isbilen et al., 2018). In line with previous studies, performance on the segmentation task was higher than that seen for the generalization task (see Isbilen et al., 2018, for a comparable finding), and crucially performance on these tasks was significantly correlated for both language conditions — adding further support to the notion that they may be underpinned by similar mechanisms.

The results for the shapes group followed the same general pattern as those seen in the varied phonology and fixed phonology conditions, with a trend toward higher performance on the segmentation task than the generalization task. However, scores for this group were significantly lower than those seen for the fixed phonology group, with accuracy scores on the segmentation task being only marginally above chance, while performance on the generalization task was at chance level. It is important to note that the shape stimuli differ from the speech stimuli in two key ways: they are both visual and non-linguistic, and therefore differ both in modality and domain. Thus, this pattern of results could be attributed to a number of possible explanations.

One possibility for the difference between the language and the shape task is that there are critical differences in statistical learning across modalities, with tasks being underpinned by different mechanisms (e.g., Conway & Christiansen, 2005). A second possibility is that, for the shapes group, performance could have been negatively affected by participants' relative lack of experience with learning distributionally defined streams containing sequences of visual non-speech input (compared to experience with heard speech) (e.g., Siegelman et al., 2018). Another possibility is that the difference in performance is due to key differences in task demands: in the speech conditions, the presentation of stimuli is such that participants have no choice but to attend (be that actively, or passively). However, in the shapes condition, this is not necessarily the case. Thus, it is possible that the lower scores observed for this group are (at least in part) due to participants attending less to the input during training (and thus, learning less during familiarization). Ongoing replications of this work employing a cover task that maintains participants' attention will help to unpack these possibilities.

To summarise, these data provide further evidence that adults can compute non-adjacent dependencies to discover words and within-word structure from continuous speech. This supports the notion that these tasks may be underpinned by similar statistical processes, and may occur together during language learning. Further, results illustrate that these abilities are not dependent on phonological cues, suggesting that adults' capacity for performing statistical computations over linguistic input is even more powerful than previously suggested.

## Acknowledgments

We thank Dante Dahabreh, Phoebe Ilevbare, Eleni Kohilakis, Farah Mawani, Olivia Wang, Emily Zhang and Sophia Zhang for their help with data collection. ESI was supported by a National Science Foundation Graduate Research Fellowship (#DGE-1650441). PM was supported by the International Centre for Language and Communicative Development (LuCiD) at Lancaster University, funded by the Economic and Social Research Council (United Kingdom; ES/L008955/1).

## References

Baayen, R. H, Davidson, D. J, & Bates D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Black, A. W., Taylor, P., & Caley, R. (1990). *The festival speech synthesis system*. Edinburgh, UK: Centre for Speech Technology Research (CSTR), University of Edinburgh, http://www.cstr.ed.ac.uk/projects/festival.html

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition, 114,* 356-371.

Conway, C. & Christiansen, M.H. (2005). Modality constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory & Cognition, 31,* 24-39.

Frost, R. L. A., & Monaghan P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition, 147*, 70-74.

Frost, R. L. A., Monaghan, P., & Christiansen, M. H. (2019). Mark my words: high frequency marker words impact early stages of language learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*.

Frost, R. L. A., Monaghan P. & Tatsumi, T. (2017). Domain-General Mechanisms for Speech Segmentation: The Role of Duration Information in Language Learning. *Journal of Experimental Psychology: Human Perception and Performance 43*(3), 466-476.

Gómez, R., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. Cognition, 70, 109–135.

Isbilen, E. S., Frost, R. L. A, Monaghan, P., & Christiansen, M. H. (2018). Bridging artificial and natural language learning: Comparing processing- and reflection-based measures of learning. *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*. Madison, WI, USA.

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

Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in non-adjacent dependencies. *Proceedings of the 26th annual conference of the cognitive science society. Mahwah, NJ: Lawrence Erlbaum.*

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

Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning non-adjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General, 133*(4), 573-583).

Redington, M. & Chater, N. (1997). Probabilistic and distributional approaches to language acquisition. *Trends in Cognitive Sciences, 1*(7), 273-281.

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, 177, 198-213.

Siegelman N, & Frost R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. Journal of Memory and Language, 81:105–120.