

**Simultaneous segmentation and generalisation of non-adjacent  
dependencies from continuous speech**

Rebecca L. A. Frost & Padraic Monaghan

Department of Psychology

Lancaster University, Lancaster UK

**Word count: 3881**

**Running Head: Segmentation and generalisation in language learning**

Corresponding author:

Rebecca L. A. Frost

Department of Psychology

Lancaster University

Lancaster

LA1 4YF

UK

E-mail: [r.frost1@lancaster.ac.uk](mailto:r.frost1@lancaster.ac.uk)

Tel: +44 1524 593697

Fax: +44 1524 593744

## **Abstract**

Language learning requires mastering multiple tasks, including segmenting speech to identify words, and learning the syntactic role of these words within sentences. A key question in language acquisition research is the extent to which these tasks are sequential or successive, and consequently whether they may be driven by distinct or similar computations. We explored a classic artificial language learning paradigm, where the language structure is defined in terms of non-adjacent dependencies. We show that participants are able to use the same statistical information at the same time to segment continuous speech to both identify words and to generalise over the structure, when the generalisations were over novel speech that the participants had not previously experienced. We suggest that, in the absence of evidence to the contrary, the most economical explanation for the effects is that speech segmentation and grammatical generalisation are dependent on similar statistical processing mechanisms.

**Key words:** language acquisition, artificial grammar learning, speech segmentation, grammatical processing, statistical learning

## 1. Introduction

In order to achieve linguistic proficiency, language learners must identify words from continuous speech, and work out the relations between those words, in terms of determining grammatical categories and syntactic structures. However, there are no definitive acoustic cues for word boundaries (Aslin, Woodward, LaMendola, & Bever, 1996), nor of grammatical categories of words that can help determine the syntactic dependencies between words (Monaghan, Christiansen, & Chater, 2007). Thus, learning must operate by somehow determining the regularities that are evident within the language, and how these regularities relate to meaning in terms of defining the relations between words and their mapping to intended referents in the environment (Cunillera, Laine, Camara, & Rodriguez-Fornells, 2010; Monaghan & Mattock, 2012).

There are two views about how these learning tasks proceed in language acquisition. One perspective is that similar statistical mechanisms may apply to speech segmentation and to grammatical processing (Perruchet, Tyler, Galland, & Peereman, 2004; Romberg & Saffran, 2010). An alternative view, deriving from classical cognitive psychology approaches to learning (Chomsky, 1957; Pinker, 1997), is that while speech segmentation is likely to depend on processing statistical dependencies, learning grammar relies on rather different algebraic processes that operate between symbolic representations of elements of language. Previous studies of word identification and grammatical processing have tended to be tested by distinct stimuli, and so comparison across tasks is difficult. However, assessing with the same stimuli word identification and abstraction over these sequences for

grammatical processing enables a test of whether processing of these tasks proceeds in tandem or is separated in learning. Though it is not possible to establish for certain whether the same or different processes apply to these tasks, it becomes more challenging to contend that the same statistical process applies to both word identification and grammatical processing if they can be shown to be temporally distinct.

There is good reason to suspect that learning may operate in tandem, because similar sources of information appear to be useful for both segmentation and determining dependencies between words in language acquisition. Monaghan and Christiansen (2010) demonstrated in corpus analyses of child-directed speech that identifying boundaries in speech could usefully rely on determining high-frequency function words that separate other words, forming points of very low transitional probabilities in the speech stream (Ordin & Nespors, 2013). Similarly, they found that these high frequency function words also provide useful markers to the phrase structure of the utterance (Cunillera, Camara, Laine, & Rodriguez-Fornells, 2010), for instance, determiner “the” tends to reliably precede nouns, and pronoun “you” precedes verbs. It is possible that the same sources of information are consulted twice to address these tasks in sequence, but a more economical explanation would be that the same source of information gradually builds up the learners’ understanding of what the words are and how they operate in the grammar of the language.

Statistical learning has been proposed as the principle by which speech segmentation and learning grammatical structure may be accomplished (Conway,

Bauernschmidt, Huang, & Pisoni, 2010; Lashley, 1951; Redington & Chater, 1997; Rubenstein, 1973). Indeed, transitional probabilities have been found to be effective indicators of word boundaries in both artificial (Aslin, Saffran, & Newport, 1998, Saffran, Aslin & Newport, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997) and natural languages (Pelucchi, Hay, & Saffran, 2009), and can be used to assist learning from infancy onward, even before learners know the meaning of a single word in the language (Saffran et al., 1996; Teinonen, Fellman, Naatanen, Alku, & Huotilainen, 2009). In addition, statistical learning has been shown to be sufficient to account for language learners' acquisition of dependencies between words in sequences (Gerken, 2010; Gómez, 2002; Lany & Gómez, 2008; Lany, Gómez & Gerken, 2007).

If such statistical processing can be demonstrated to be sufficient for word identification and grammar learning, then this weakens the requirement to posit language-specific mechanisms for language acquisition, instead, a simpler domain-general approach to language learning could be assumed, until evidence to the contrary is ascertained (Christiansen & Chater, 2008). In an ingenious set of studies, Peña, Bonatti, Nespor, and Mehler (2002) set out to show this distinction. They focused on learning of non-adjacent dependencies, which are evident in language structure at multiple levels, from orthography (e.g., final *e* changing the pronunciation of the previous vowel, *cap* and *cape*), morpho-syntax (e.g., I go, *he* goes), grammatical categorisation (e.g., high-frequent non-adjacent pairs of words assist grammatical categorisation of the intervening word, "the \_\_ is", "you \_\_ to" (Mintz, 2003; St Clair, Monaghan, & Christiansen, 2010), and hierarchical

grammatical relations (e.g., *the boy the cats chase runs*). As Perruchet et al. (2004) note, if statistical dependencies can be shown to be sufficient for acquiring non-adjacencies then this increases the likelihood of the role of domain general statistical processing in language acquisition.

In Peña et al.'s (2002) study, adults were presented with synthetic speech containing items defined by non-adjacent transitional probabilities (e.g.  $A_1XC_1$ ,  $A_2XC_2$ ), where particular A syllables were always paired with particular C syllables, but the X syllable freely varied over a set of three other syllables. To measure speech segmentation, participants were tested on their ability to identify previously occurring words that were consistent with the non-adjacencies presented in the speech, by assessing preference for words (e.g.  $A_1XC_1$ ) over part-words (e.g.  $XC_1A_2$ ).

Critically, Peña et al. (2002) also used these same stimuli to test the extent to which participants could manipulate non-adjacencies to generalise to new items. This involves going beyond the surface form of the sequences, by abstracting the structure to generalise to these new sequences, and is a key property of grammatical processing (Marcus, Vijayan, Rao, & Vishton, 1999). After the same training, they tested a different set of participants on their preference for "rule-words", constructed by moving an A or a C syllable from elsewhere in the speech stream (e.g., placing  $A_2$  within the  $A_1C_1$  non-adjacency:  $A_1A_2C_1$ ), in comparison with part-words. Participants were not able to generalise. However, when the segmentation task was solved for participants, by placing a 25 ms gap between the syllable triples during training, participants did generalise to the rule-words. Peña et al. (2002) thus suggested that although adults are capable of using statistics to identify words from

a continuous speech stream, they may then apply separate computations that do not depend on learning statistical dependencies between particular elements of the language, to generalise the structure to consistent forms. They suggest that this can occur only once the task of identifying the words in the stimuli has been solved (Chomsky, 1957; Endress & Bonatti, 2007; Marchetto & Bonatti, in press; Marcus et al., 1999; Miller & Chomsky, 1963).

The interpretation of these results has been hotly debated, but previously the focus of disagreement has been on whether non-adjacencies were learned at all, or rather whether participants instead remembered particular items from the speech (Perruchet et al., 2004), or whether participants learned only the general position of syllables in the sequences rather than the dependencies between them (Endress & Bonatti, 2007; Endress & Mehler, 2009; Mueller, Bahlmann, & Friederici, 2008, 2010; Perruchet et al., 2004). However, there has been substantially less focus on the extent to which segmentation and generalisation of structure co-occur, or are temporally distinct processes.

The Peña et al. (2002) rule-word generalisation stimuli were constructed by moving an A or a C syllable to a new position in the sequence. An advantage of this is that the frequencies of individual syllables were controlled across the target and the part-word stimuli in forced choice tests, so any observed preferences must then be due to syllable co-occurrences, either of adjacent or non-adjacent elements in speech. However, this design may have made generalisation performance harder to detect because it requires not only generalisation of the non-adjacency but also unlearning of the dependency relations for the moved syllable. For instance, the

moved-syllable test of Peña et al.'s (2002) study would be analogous to training participants on “the boy the cats chase runs” and “the girl the dog nuzzles smiles”, and then testing whether they can flexibly apply the non-adjacency to “the boy smiles runs”. Participants may reject these items because they are not able to generalise the non-adjacent structure, or because they fail to accept a violation of relational structure. The observed importance of the pause between syllable triples may then be required not to solve the segmentation task, but rather to increase the salience of syllables with regard to their position (Endress & Mehler, 2009; Perruchet et al., 2004), thus providing an additional cue to relative positions of elements of the language in the speech.

In the current study we tested whether participants are able to simultaneously segment and generalise structure of a non-adjacent dependency language if new, rather than moved, syllables comprise the sequences to be generalised. A novel syllable intervening between an  $A_i$ - $C_i$  dependency is a stronger test of generalisation, but without interference from previous learning of relative syllable positions. Participants listened to a continuous speech stream, and then completed either a test of segmentation, or of generalisation to rule words containing a moved syllable as in Peña et al. (2002). An additional condition tested generalisation to rule words containing novel syllables. If participants are able to use the same information for segmentation and generalisation simultaneously, but were affected by having to unlearn positional information in Peña et al.'s (2002) test of rule-word generalisation, then we expect learning for the novel syllable rule-words in addition to learning for the segmentation task. However, if segmentation and structural generalisation are



separable processes, then we expect to see a null effect for the novel syllable generalisation task, with similar performance to that seen in Peña et al.'s (2002) original study.

## **2. Method**

### **2.1 Participants**

The experiment was completed by 54 adults (8 males, 46 females) with a mean age of 18.52 years (range = 18-24 years). All participants were native-English-speakers, with no known history of auditory, speech or language disorder. Participants were paid £3.50, or received course credit.

### **2.2 Design**

The experiment used a between participants design with three conditions of test type: *segmentation*, *moved-syllable generalisation*, and *novel-syllable generalisation*. Participants were randomly allocated to one of these conditions, with 18 participants receiving each type of test (segmentation: 2 males, 16 females; moved-syllable generalisation: 4 males, 14 females; novel-syllable generalisation: 2 males, 16 females).

### **2.3 Materials**

#### **2.3.1 Stimuli**

Speech stimuli were created using the Festival speech synthesiser (Black, Taylor, & Caley, 1990) and were based on those used by Peña et al. (2002). The language contained nine monosyllabic items (*pu, ki, be, du, ta, ga, li, ra, fo*) taken from

Peña et al. (2002), which were used to create three non-adjacent pairings with three possible X items which intervened between the dependencies ( $A_1X_{1-3}C_1$ ,  $A_2X_{1-3}C_2$ , and  $A_3X_{1-3}C_3$ ). As in Peña et al. (2002), phonemes used for A and C items contained plosive sounds (*pu, ki, be, du, ta, ga*), while X items contained continuants (*li, ra, fo*). Similarities in phonological properties of non-adjacent dependent syllables have been shown to support acquisition of those non-adjacencies (Newport & Aslin, 2004), but they are not essential for such learning to occur (Onnis, Monaghan, Christiansen & Chater, 2004). Nevertheless, words from the same grammatical category tend to be coherent with regard to phonological properties (Monaghan et al., 2007), and so this property of the artificial language is consistent with natural language. Each AXC string lasted approximately 700 ms. To control for possible preferences for particular dependencies between syllables not due to the statistical structure of the sequences, 8 versions of the language were generated by randomly assigning syllables to A and C roles. The same X items were used across all versions of the language. These were counterbalanced across the three conditions of the study. There were three additional syllables comprising continuant phonemes (again consistent with a correlation between relational structure and phonology in natural language), which were reserved for testing generalisation to novel items (*ve, zo, thi*).

### **2.3.2 Training**

A 10.5-min-long continuous stream of synthetic speech was created using the Festival speech synthesiser (Black et al., 1990) by concatenating AXC words in the language. No  $A_iC_i$  dependency was immediately repeated. Speech streams had a 5s

fade in and out so that onset and offset of the speech could not be used as a cue to the language structure.

### 2.3.3 *Testing*

For testing segmentation, a forced choice task tested preference for word compared to part-word comparisons. Part-words were trisyllabic items that occurred in the training speech but straddled word boundaries, comprising the last syllable of one word and the first two syllables of another word ( $C_iA_jX$ ), or the last two syllables of one word and the first syllable of another ( $XC_iA_j$ ). Both types of part-word were created for each of the nine  $AXC$  items. Eighteen test pairs were constructed by matching each part-word with its corresponding word (so, for example, an  $A_1X_2C_1$  item was paired with an  $X_2C_1A_2$  part-word).

For testing moved-syllable generalisation, a forced choice task compared preference for rule-word compared to part-words. Rule words comprised an  $A_iC_i$  non-adjacency containing an  $A$  or a  $C$  item from elsewhere in the speech stream. There were three rule-words for each  $A_iC_i$  dependency. There were 9 test items altogether, five rule-words paired with a  $C_iA_jX$  part-word and four paired with a  $XC_iA_j$  part-word.

For novel-syllable generalisation, nine forced choice tests comprised a rule-word containing one of the three novel syllables (so of the form  $A_iNC_i$ ), where  $N$  indicates the novel syllable, and a novel part-word. Each novel rule-word appeared once in each  $A_iC_i$  dependency. Part-words comprised two syllables that occurred during training in their respective positions, with the same novel syllable that

occurred in the rule-word sequence. The novel syllable could appear in the first, second or third position (so were of the form  $NC_iA_j$ ,  $XNA_i$ , or  $C_iA_jN$ ) with each novel syllable occurring once in each position. Presence of novelty in both rule-word and part-word controlled for the effect of the novel syllable, but still tested for generalisation of the non-adjacent structure.

In all conditions the order of test-pairs was randomized across participants as was the position of the correct response within each pair. Items in each pair were separated by a 1 s pause.

## **2.4 Procedure**

Before hearing the training language, participants were instructed to pay attention to the language and think about possible words it may contain. Participants were then immediately tested on the forced-choice tasks. On each trial, participants listened to a test pair and were instructed to select which item best matched the language they had just heard, responding “1” for the first or “2” for the second sequence on a computer keyboard. Participants listened to the speech through closed-cup headphones in a quiet testing room.

## **3. Results**

Accuracy for each condition is shown in Figure 1. Performance on the segmentation task was significantly higher than chance ( $M = .719$ ,  $SD = .153$ ),  $t(17) = 6.089$ ,  $p < .001$ ,  $d = 1.435$ . Performance on the moved-syllables task did not differ significantly from chance ( $M = .487$ ,  $SD = .246$ ),  $t(17) = -.220$ ,  $p = .828$ . Performance on

the novel-syllable task was significantly higher than chance ( $M = .693$ ,  $SD = .160$ ),  $t(17) = 5.129$ ,  $p < .001$ ,  $d = 1.209$ .

A  $3 \times 8$  univariate ANOVA with task type and randomised language version as factors revealed a significant effect of task type,  $F(2, 30) = 7.896$ ,  $p = .002$ ,  $\eta_p^2 = .345$ . Segmentation and novel-syllable generalisation were not significantly different ( $p = .650$ ), but both were significantly more accurate than moved-syllable generalisation ( $p = .001$ , and  $p = .003$ , respectively). There was no significant effect of language condition,  $F(7, 30) = 1.836$ ,  $p = .117$ ,  $\eta_p^2 = .300$ , or interaction between language condition and test type,  $F < 1$ .

Additional analyses were performed to examine whether there was any learning over testing, and whether any such learning differed between the generalisation conditions, to rule out the possibility that using novel phonemes in the novel-syllable generalisation test made the non-adjacencies more salient in this condition than the moved-syllable or segmentation conditions. For each condition, testing was divided into three blocks of equal number of trials, distinguishing early, middle and late testing trials, and responses were then reanalysed with block as an additional, within-subjects factor.

There was no significant effect of block,  $F(2, 60) = 1.736$ ,  $p = .185$ ,  $\eta_p^2 = .055$ , and the linear contrast for block was not significant,  $F < 1$ . There was no significant interaction between block and condition,  $F < 1$ , which was also not significant in the linear contrast,  $F < 1$ . Thus, there was no evidence of learning during the test trials across conditions, and no differential effect of learning between the conditions. To ensure that there was no learning within individual conditions, we conducted an

ANOVA with block as a within subjects factor for each condition separately. Again, there were no significant main effects or linear contrasts for block in any condition. For segmentation, main effect,  $F(2, 20) = 1.154, p = .335, \eta_p^2 = .103$ , linear contrast,  $F < 1$ ; for moved-syllable generalisation, main effect,  $F(2, 20) = 1.069, p = .362, \eta_p^2 = .097$ , linear contrast,  $F < 1$ ; and for novel-syllable generalisation, main effect and linear contrast both  $F < 1$ .

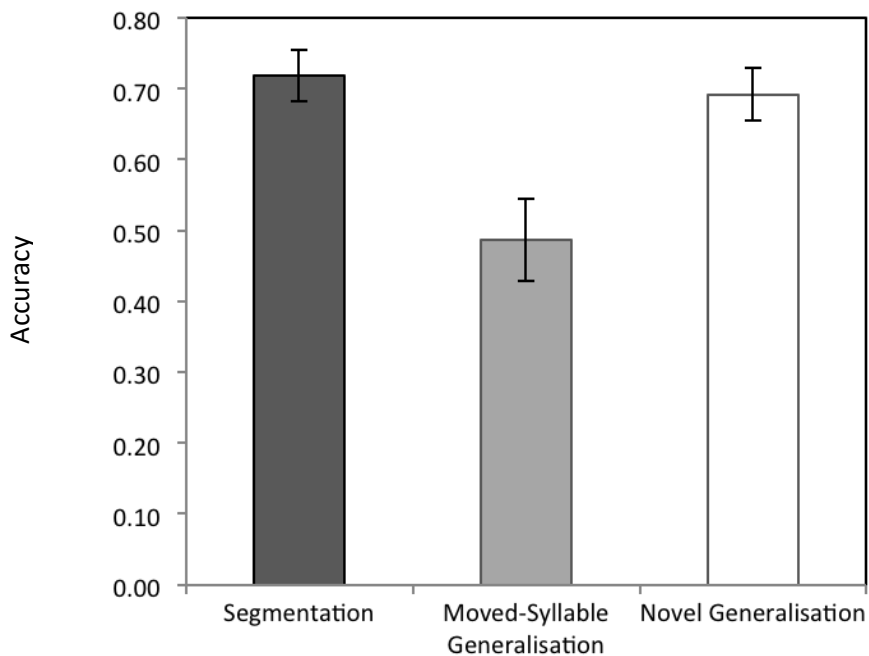


Figure 1. Mean accuracy for each test condition.

Repeated measures ANOVAs were conducted to assess the effect of each type of part-word on each of the tasks individually. Individual analysis was necessary as the novel generalisation task utilised different types of part-words to the segmentation and moved-syllable generalisation tasks. For the segmentation task, there was no significant effect of part-word,  $F < 1$ , with participants performing

equally well on XCA ( $M = .712$ ,  $SD = .181$ ) and CAX ( $M = .723$ ,  $SD = .227$ ) comparisons. There was no significant effect of part-word for the moved-syllable generalisation task,  $F < 1$ , indicating that participants performed equally well on XCA ( $M = .528$ ,  $SD = .256$ ) and CAX ( $M = .456$ ,  $SD = .348$ ) comparisons. There was also no significant effect of part-word for the novel-syllable generalisation task,  $F(2, 34) = 2.013$ ,  $p = .149$ ,  $\eta_p^2 = .106$ , with participants performing similarly on CAN ( $M = .778$ ,  $SD = .256$ ), NCA ( $M = .630$ ,  $SD = .254$ ), and XNA ( $M = .668$ ,  $SD = .230$ ) part-words.

#### **4. Discussion**

We examined adults' ability to learn non-adjacent dependencies from a continuous stream of synthetic speech, to establish whether segmentation and generalisation may occur within the same brief learning period. In accord with prior literature, we found that adults are able to use non-adjacent transitional probabilities to segment a continuous artificial speech stream (Onnis et al., 2005; Peña et al., 2002; Perruchet et al., 2004). As anticipated, performance on the moved-syllable generalisation task did not provide evidence of generalisation, corresponding with the findings of Peña et al. (2002). However, critically, participants were able to generalise non-adjacency structures to sequences that contained novel syllables.

It is possible that there were conflicting forces preventing participants from displaying a preference for the moved-syllable rule-words; on one hand, the non-adjacencies were being used for processing the structure, but on the other participants may have been affected by the unfamiliarity of the repositioned A or C

syllable contravening the dependencies within the language. In previous studies, the importance of the pause cue between triples in the language may have been not to solve the segmentation task but rather to provide an additional cue increasing saliency of the positions of individual syllables (Endress & Mehler, 2009), resulting in enhanced learning of syllables in particular positions (Endress & Bonatti, 2007; Perruchet et al., 2004). Without this conflicting information, we have shown that generalisation of the non-adjacencies can be observed in tandem with segmentation, and, furthermore, that this can be accomplished without requiring additional cues to the structure of the language (e.g., Mueller et al., 2008, 2010).

An alternative reason for the observation of novel-syllable generalisation is that the part-word items do not occur in the speech, though for the moved-syllable generalisation task the part-word items did occur, and it is therefore harder to reject them. This is a possibility, but does not affect the overall result that generalisations depending on non-adjacencies are driving the results in the novel-syllable generalisation task. In either case, there was no significant effect of the type of part-word, indicating that a part-word containing a syllable pair that occurred during training (NCA, or CAN) was no harder to reject than a part-word containing no syllable pairs that occurred during training (XNA), making this an unlikely cause for the distinction between the two generalisation conditions.

The current study demonstrates that segmentation and generalisation are not separable behaviourally within the same time period, where such a differentiation had previously been claimed (Peña et al., 2002). Thus, evidence suggesting that there was a distinction between processes for word identification and grammatical



processing is shown to not be supported. There remains a possibility that the tasks are solved simultaneously but in different ways, such that non-adjacencies are utilised for segmentation using statistical learning, but that operations over the structure are still symbolic, applying to abstract generalisations of the relations between elements (Marcus et al., 1999). We suggest that, in the absence of evidence to the contrary, the same class of mechanisms – statistical learning – should be assumed to be sufficient for driving word learning as well as structural generalisation (Aslin & Newport, 2014).

Previous claims of the need for symbolic, algebraic processing for generalisation of sequences rely on a narrow interpretation of statistical mechanisms that permit only computation of dependencies between elements in experienced stimuli (Marcus et al., 1999). However, statistical processing is consistent with learning to generalise, as well as learning precise co-occurrences between experienced elements in sequences (Romberg & Saffran, 2010). The traditions of symbolic and statistical processing in cognitive psychology, and language acquisition research, have undergone substantial convergence (Perruchet & Pacton, 2006). For instance, Redington, Chater, and Finch (1998) demonstrated how statistical processing of clustering can support generalisations as well as learning individual correspondences in grammatical structure. Similarly, French, Addyman, and Mareschal (2011) showed how the same statistical learning mechanism could apply to both speech segmentation studies and studies of implicit learning of rule-based sequences. Our results confirm that such results also occur behaviourally.

The breadth of possible statistical processes that can support speech segmentation, grammatical categorisation, and syntactic processing reduces the requirement to stipulate that language acquisition processes may be domain-specific, rather than applications of powerful general-purpose learning mechanisms (Christiansen & Chater, 2008). However, exactly what statistical mechanism is being applied remains a difficult issue to resolve. In the current experiment, the distinction between word identification and grammatical processing can be understood in terms of learning dependencies between experienced sequences and learning to generalise dependencies to new sequences. However, scaling this distinction up to the dependencies observed in natural language requires explaining how long-distance dependencies between hierarchical structures may be acquired (see Lai & Poletiek, 2011; Lany & Gómez, 2008; Lai & Poletiek, 2011; Onnis, Monaghan, Christiansen, & Chater, 2004; Saffran et al., 1996, for progress in this field). Nevertheless, the study we present here demonstrates that, from the same input and at the same time, participants are able to identify particular sequences as words, and generalise the structure of those sequences. Any qualitative distinctions between the processes involved in these tasks as yet remain to be demonstrated.

## References

- Aslin, R. N., & Newport, E. L. (2014). Distributional language learning: Mechanisms and models of category formation. *Language Learning, 64*, 86-105.
- Aslin, R. N., Saffran, J., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321-324.
- Aslin, R. N., Woodward, J., LaMendola, N., & Bever, T. (1996). Models of word segmentation in fluent maternal speech to infants. In J. Morgan and K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Black, A. W., Taylor, P., & Caley, R. (1990). *The Festival Speech Synthesis System*, available from <http://www.cstr.ed.ac.uk/projects/festival.html>, Centre for Speech Technology Research (CSTR), University of Edinburgh, Edinburgh, UK.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Christiansen, M. H. & Chater, N. (2008). Language as shaped by the brain. *Behavioral Brain Sciences, 31*, 489-558.
- 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.
- Cunillera, T., Camara, E., Laine, M., & Rodriguez-Fornells, A. (2010). Words as anchors: Known words facilitate statistical learning. *Experimental Psychology, 57*, 134-141.

- Cunillera, T., Laine, M., Camara, E., & Rodriguez-Fornells, A. (2010). Bridging the gap between speech segmentation and word-to-world mappings: Evidence from an audiovisual statistical learning task. *Journal of Memory and Language*, 63, 295-305.
- Endress, A. D. & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, 105(2), 247-299.
- Endress, A. D. & Mehler, J. (2009). Primitive computations in speech processing. *The Quarterly Journal of Experimental Psychology*. 62(11), 2187-209.
- French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, 118, 614-636.
- Gerken, L. A. (2010). Infants use rational decision criteria for choosing among models of their input. *Cognition*, 115, 362-366.
- Gómez, R. L. (2002). Variability and detection of invariant structure, *Psychological Science*, 13(5) 431-436.
- Lai, J. & Poletiek, F. H. (2011). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition*, 118(2), 265-273.
- Lany, J. & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, 19(12), 1247 - 1252.
- Lany, J. & Gómez, R. L., & Gerken, L. (2007). The role of prior experience in language acquisition, *Cognitive Science*, 31, 481-507.

- Lashley, K. S. (1951). The problem of serial order in behavior. In L.A. Jeffress (Ed.), *Cerebral Mechanisms in Behavior* (pp. 112- 146). New York: Wiley.
- Marchetto, M. & Bonatti, L. L. (2015). Finding words and word structure in artificial speech: The development of infants' sensitivity to morphosyntactic regularities. *Journal of Child Language*, in press. doi:10.1017/S0305000914000452
- Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven month-old infants. *Science*, 283, 77-80.
- Miller, G. & Chomsky, N. (1963). Finitary Models of Language Users. In R. Luce et al. (Eds.), *Handbook of mathematical psychology* (pp286-287). New York: John Wiley.
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90, 91-117.
- Monaghan, P., & Christiansen, M. H. (2010). Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. *Journal of Child Language*, 37, 545-564.
- Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The Phonological Distributional coherence Hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, 55, 259-305.
- Monaghan, P. & Mattock, K. (2012). Integrating constraints for learning word-referent mappings. *Cognition*, 123, 133-143.
- Mueller, J. L., Bahlmann, J., & Friederici, A. D. (2008). The role of pause cues in language learning: The emergence of event-related potentials related to sequence processing. *Journal of Cognitive Neuroscience*, 20(5), 892-905.

- Mueller, J. L., Bahlmann, J., & Friederici, A. D. (2010). Learnability of embedded syntactic structures depends on prosodic cues. *Cognitive Science*, 34(2), 338-349.
- 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 nonadjacent dependencies. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.
- Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in speech processing. *Journal of Memory and Language*, 53, 225-237.
- Ordin, M. & Nespors, M. (2013). Transition probabilities and different levels of prominence in segmentation. *Language Learning*, 63(4), 800-834.
- Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, 10, 233-238.
- Pelucchi, B., Hay, J. F., Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, 80(3), 674-685.
- Peña, M., Bonatti, L., Nespors, 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.

- Pinker, S. (1997). *How the mind works*. New York: Penguin Books.
- Redington, M. & Chater, N. (1997). Probabilistic and distributional approaches to language acquisition. *Trends in Cognitive Sciences*, 1(7), 273-281.
- Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, 22, 425-469.
- Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6), 906-914.
- Rubenstein, H. (1973). Language and probability. In G. A. Miller (Ed.), *Communication, language, and meaning: Psychological perspectives* (pp. 185-195). New York: Basic Books, Inc.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, 8, 101-105.
- St Clair, M. C., Monaghan, P., & Christiansen, M. H. (2010). Learning grammatical categories from distributional cues: What information is useful and when? *Cognition*, 116, 341-360.
- Teinonen, T., Fellman, V., Naatanen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, 10(21).

## **Acknowledgements**

This work was supported by the International Centre for Language and Communicative Development (LuCiD) at Lancaster University, funded by the Economic and Social Research Council (UK) [ES/L008955/1].