

# Using Statistics to Learn Words and Grammatical Categories: How High Frequency Words Assist Language Acquisition

**Rebecca L. A. Frost (r.frost1@lancaster.ac.uk)**

Department of Psychology, Lancaster University,  
Bailrigg, Lancaster, LA1 4YF, UK

**Padraic Monaghan (p.monaghan@lancaster.ac.uk)**

Department of Psychology, Lancaster University,  
Bailrigg, Lancaster, LA1 4YF, UK

**Morten H. Christiansen (christiansen@cornell.edu)**

Department of Psychology, Cornell University,  
Ithaca, NY, 14853-7601, USA

## Abstract

Recent studies suggest that high-frequency words may benefit speech segmentation (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005) and grammatical categorisation (Monaghan, Christiansen, & Chater, 2007). To date, these tasks have been examined separately, but not together. We familiarised adults with continuous speech comprising repetitions of *target words*, and compared learning to a language in which targets appeared alongside high-frequency *marker words*. Marker words reliably preceded targets, and distinguished them into two otherwise unidentifiable categories. Participants completed a 2AFC *segmentation* test, and a similarity judgement *categorisation* test. We tested transfer to a word-picture mapping task, where words from each category were used either consistently or inconsistently to label actions/objects. Participants segmented the speech successfully, but only demonstrated effective categorisation when speech contained high-frequency marker words. The advantage of marker words extended to the early stages of the transfer task. Findings indicate the same high-frequency words may assist speech segmentation and grammatical categorisation.

**Keywords:** statistical language learning, speech segmentation, grammatical categorisation.

## Introduction

Learners must master (at least) two critical tasks prior to reaching linguistic proficiency; identifying individual words from speech, and discovering that these words belong to different grammatical categories. As speech contains no perfectly reliable acoustic cues to word boundaries (Aslin, Woodward, LaMendola, & Bever, 1996) or grammatical category membership (Monaghan, Christiansen, & Chater, 2007), learners must draw upon additional sources of information to accomplish these tasks.

It is well documented that learners can perform powerful computations on the statistical information contained in speech, which can help them to infer word boundaries (Saffran, Aslin, & Newport, 1996). Language learners have been shown to exploit transitional statistics to help identify word boundaries in both artificial (Aslin, Saffran, & Newport, 1998; Saffran et al., 1996; Saffran, Newport,

Aslin, Tunick, & Barrueco, 1997) and natural languages (Pelucchi, Hay, & Saffran, 2009), and can do so from infancy onward, before knowing the meaning of a single word in the language (Saffran et al., 1996; Teinonen, Fellman, Naatanen, Alku, & Huotilainen, 2009). Furthermore, similar statistical information can help learners develop rule-like linguistic regularities (Gerken, 2010; Gómez, 2002; Lany & Gómez, 2008; Lany, Gómez, & Gerken, 2007), even while they are learning to segment speech (Frost & Monaghan, 2016).

Bearing in mind learners' aptitude for exploiting statistics, it follows that items appearing more frequently than others in speech might prove helpful for learning. Indeed, recent research has suggested that language acquisition may benefit from the presence of high-frequency words, with a variety of studies demonstrating that these may be advantageous for speech segmentation in particular (Altwater-Mackensen & Mani, 2013; Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Mersad, & Nazzi, 2012).

One benefit of frequently occurring words for language acquisition lies in their ability to provide learners with helpful information about the boundaries of words that surround them in speech. In a recent study, Mersad and Nazzi (2012) demonstrated that 8-month-old French-learning infants could identify unfamiliar words from speech when they appeared in a stream containing the familiar word "maman" (mommy in French) but not when they appeared with non-word *māma*. Similarly, Bortfeld et al., (2005) demonstrated that infants were better able to identify new words when they appeared in speech next to high frequency, words such as their own name and the word 'mommy', but not when they appeared alongside an unfamiliar item.

Monaghan and Christiansen (2010) examined the possibility that highly frequent words may assist with natural language acquisition through the PUDDLE model of speech segmentation, which they applied to natural language corpora of child-directed speech. Findings indicated that the model was able to quickly extract the high

frequency words in the speech input, which were then used for accurately segmenting the rest of the speech.

We can consider how this might work in practise by taking the sentence *youeatthecheeseetyoudrinkthewine*. A learner hearing this sentence could recognise high frequency words *you* and *the*, and use these to discern information about the words that surround them in speech. Specifically, in this instance recognising the word *you* would help the learner to identify the way in which the preceding (yet) and succeeding (eat, drink) words end and begin, respectively.

In addition to helping with speech segmentation, highly frequent words may also help inform the formation of grammatical categories (Valian & Coulson, 1988): in the example given above, high frequency word *you* reliably precedes verbs, while *the* reliably precedes nouns, providing valuable information about the word's grammatical role as well as its identity. Interestingly Monaghan and Christiansen (2010) noted that there was substantial overlap between words that were useful for segmentation, and words that were useful for identifying grammatical categories in previous studies of child-directed speech (Monaghan, Christiansen, & Chater, 2007).

In the present study, we examined the way that the same high frequency words affected both of these tasks at the same time. Specifically, we tested whether presence of high frequency words helps learners to identify words that appear alongside them in speech (speech segmentation), and we examined the way that learners used the same high frequency words to learn that words belong to different categories (categorisation). In a subsequent test, we assessed the extent to which learners' pre-existing category knowledge affected their ability to use the language in a word picture mapping task that required them to use the language in a way that was either consistent or inconsistent with their training. We hypothesised that high frequency marker words would assist with speech segmentation, while also constraining the language by contributing to formation of grammatical categories. Further, we varied the ratio of marker words to category words in the study to determine whether variation in the marker words was helpful (Onnis, Monaghan, Christiansen, & Chater, 2004) or an impediment (Valian & Coulson, 1988) to learning.

## Method

### Participants

The experiment was completed by 72 adults (18 males, 54 females), all of whom were students at Lancaster University, with a mean age of 20.39 years (range = 18-48 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.

### Design

The experiment used a between subjects design, with three conditions of training type which varied the number of marker words per category from 0 to 2; *Marker0*, *Marker1*,

and *Marker2*. Participants were randomly allocated to one of these conditions, with 24 participants receiving each type of training.

## Materials

### Stimuli

Speech stimuli were created using the Festival speech synthesiser (Black, Taylor, & Caley, 1990). The language contained 20 monosyllabic items (*no, ro, fo, to, li, gi, ni, ka, ma, sa, za, fe, te, re, de, ve, mu, zu, pu, bu*), which were combined pseudo-randomly to create eight bisyllabic *target words* (e.g., *samu, noli, nifo, fede, tero, buza, kato, mave*), and four monosyllabic *marker words* (e.g., *fo, pu, gi, re*), which preceded items in speech. Phonemes used for target and marker words contained both plosive and continuant sounds. There was no repetition of vowel sounds within target words. Each target word lasted approximately 500ms, and each marker word lasted approximately 250ms.

The eight target words were arbitrarily split into two equal categories (*A* and *B*, with four words in each). Category membership was denoted only by the co-occurrence with marker words in speech: in *Marker1*, one marker word reliably preceded words from each category; in *Marker2*, two words reliably preceded words from each category (so, in *Marker2* markers appeared alongside targets half as often as in *Marker1*). Marker words preceded category words to reflect use of function words in English, though we acknowledge that information marking grammatical category membership can also occur after word stems. The speech stream for *Marker0* contained target words only, so participants in this condition received no information regarding category membership, therefore we did not expect them to demonstrate such knowledge at test.

Eight transitions between words were omitted from the *Marker0* condition (e.g., word4 never succeeded word1), and these omitted transitions were later used for the segmentation test stimuli. This was to ensure that all non-words had not occurred during training in any language condition.

To control for possible preferences for certain syllables in certain positions, and preferences for particular dependencies between syllables not due to the statistical structure of the sequences, six versions of the language were generated by randomly assigning syllables to positions within words and marker words (Onnis, Monaghan, Richmond & Chater, 2005). These were counterbalanced across each of the three conditions.

### Training

A continuous stream of synthetic speech was created using the Festival speech synthesiser (Black et al., 1990) by concatenating target words and marker words (see Table 1). For *Marker0*, the speech stream comprised target words only, and lasted approximately 280 seconds. For *Marker1*, the speech stream comprised target words plus two marker words, and for *Marker2*, speech comprised target words plus four marker words. Streams for *Marker1* and *Marker2* both lasted approximately 420 seconds. The marker words for

each of these conditions were used equally in the speech stream in all instances.

Table 1: Example Speech Streams for Each Condition

	Speech
Marker0	<i>buza-noli-samu-tero-kapu</i>
Marker1	<i>ni-buza-zu-noli-zu-samu-ni-tero-zu-kapu</i>
Marker2	<i>fo-buza-zu-noli-ma-samu-ni-tero-zu-kapu</i>

For all conditions, speech was continuous, with no pauses within or between words. Speech streams had a 5 second fade in and out so that the onset and offset of speech could not be used as a cue to the language structure. Target words were each presented 150 times in all streams, with no immediate repetition.

### Segmentation test

To test segmentation, we created a two-alternative forced-choice task to examine participants' preferences for words versus non-words (see Saffran et al., 1996 for a similar use of words and non-words). Non-words were bisyllabic items that comprised the last syllable of one target word and the first syllable of another. Non-words did not occur during the training stream, as certain syllable combinations were prevented from occurring: for Marker0, they were formed from the omitted word transitions, and for Marker1 and Marker2, non-words did not occur because a marker word intervened. Selecting a word over a non-word on this task would indicate that participants had successfully identified target words in the speech stream.

Eight test pairs were constructed by matching each target word with a corresponding non-word (e.g. /samu/ vs. mu/no). Items in each test pair were separated by a 1s pause. Test pairs were each presented twice, giving 16 test items in total.

### Categorisation test

To test categorisation, we created a similarity-judgement task that contained pairs of target words. Half of all test pairs contained items from the same category (as determined by the marker words that preceded them in speech), with six test pairs containing two A words, and six test pairs containing two B words. Twelve additional mixed test pairs were created (so, one A word and one B word), giving 24 test pairs in total.

### Transfer of category knowledge test

To test transfer, we created a word-picture mapping task which provided a grammatical category distinction onto which the distributionally defined category words could map (see Hay, Pelluchi, Graf Estes, & Saffran, 2011, for a similar experimental design testing transfer of segmented speech into a word meaning acquisition task). For this, we introduced eight images, each printed in black on a white background. Four of these images depicted an action, while the remaining four depicted objects. Each target word was paired with one of eight images, and participants were

required to learn these pairings. Critically, for half of participants, word-picture pairings were *consistent* with the distributionally-defined categories heard during training, such that all A words appeared with actions and all B words appeared with objects. For the remaining participants, pairings were *inconsistent*: two A words and two B words were paired with objects, and two A words and two B word were paired with actions (see Figure 1). Each version of the language contained a different set of pictures, which were selected at random from an array of 8 object and 8 action images taken from The Peabody Picture Vocabulary Test (Dunn & Dunn, 1997).

In all conditions the order of trials on each test was randomised, and correct responses appeared an equal number of times in each position within pairs/arrays.

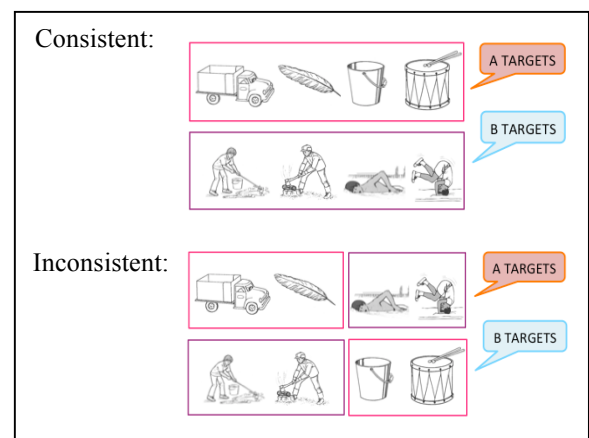


Figure 1. Consistent versus inconsistent word-picture mappings.

### Procedure

Before hearing the familiarisation stream, participants were instructed to pay attention to the language and think about possible words it may contain. Participants were tested immediately after training. All participants received the tasks in the same order: participants completed the segmentation test first, followed by the categorisation test, then the transfer of category knowledge test. Tasks were programmed using EPrime 2.0, with instructions appearing onscreen before each task began.

During the segmentation test, participants were instructed to listen to each test pair then 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.

For the categorisation test, participants were instructed to listen to each test pair, then rate how similar the role of the items was in the familiarisation stream. Participants were required to respond on a computer keyboard using a 6-point Likert-scale, with 1 being extremely different, and 6 being very similar. If participants have formed categories based on the co-occurrence of target words and markers, then pairs of

items taken from the same category should receive higher similarity ratings than mixed pairs.

For the transfer of category knowledge task, participants heard a target word and saw its corresponding image onscreen for 2 seconds, presented in randomised order. After all eight word-picture pairs were presented, participants heard each target word in isolation and had to select the corresponding image from an array containing all 8 images, responding via keypress, with a number between 1 and 8. This exposure-test sequence was then repeated 3 additional times (so 4 times total). If prior category knowledge was influencing performance on this task, then participants should find it easier to use words from each of the categories consistently (i.e., all A words labelling objects) than inconsistently (i.e. some A words labelling objects, but some A words labelling actions). Thus, an effect involving consistency on this task would indicate transfer.

All speech was presented at a comfortable volume, through closed-cup headphones.

## Results

### Segmentation

One-sample t-tests were performed on the proportion of correct choices for the segmentation test to compare performance to chance. Performance was significantly above chance for Marker0 ( $M = .613$ ,  $SE = .053$ ,  $t(23) = 2.098$ ,  $p = .047$ ), for Marker1 ( $M = .68$ ,  $SE = .034$ ,  $t(23) = 5.571$ ,  $p < .001$ ), and also for Marker2 ( $M = .66$ ,  $SE = .023$ ,  $t(23) = 6.788$ ,  $p < .001$ ) conditions, indicating that all participants were able to extract individual words from the speech stream (see Figure 2).

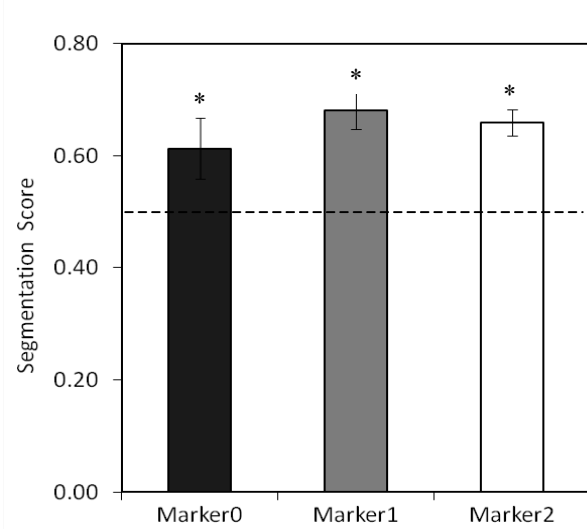


Figure 2: Mean segmentation scores (proportion correct), with standard error. Asterisks indicate above chance performance.

An ANOVA with condition and language version was performed on the segmentation data, with language version included to control for variation across the randomised

assignments of phonemes to syllables (this factor was not further analysed in the results). There was no significant effect of condition,  $F(2, 54) = 1.032$ ,  $p = .363$ , with participants in all conditions performing at a statistically similar level.

### Categorisation

A repeated-measures ANOVA was performed on data from the categorisation test, with test-pair type (same or different category), condition (Marker0, Marker1, Marker2), and language version as factors. There was no significant effect of test-pair type,  $F(1, 54) = .968$ ,  $p = .330$ , and there was no significant effect of condition,  $F(2, 54) = 1.327$ ,  $p = .274$ . However, there was a significant interaction between test-pair type and condition,  $F(2, 54) = 3.316$ ,  $p = .044$ ,  $\eta_p^2 = .109$ , which was led by performance in the Marker1 group, who demonstrated a significant difference between ratings for test pairs containing items from the same ( $M = 3.882$ ,  $SE = .179$ ) versus different categories ( $M = 3.656$ ,  $SE = .170$ ,  $t(23) = 2.085$ ,  $p = .048$ ).

There was no significant difference between ratings for test-pairs containing items of the same ( $M = 3.451$ ,  $SE = .118$ ) versus different ( $M = 3.556$ ,  $SE = .108$ ) categories for the Marker2 condition ( $t(23) = 1.969$ ,  $p = .061$ ). There was no significant difference between ratings for Marker0 (same:  $M = 3.559$ ,  $SE = .12$ ; different:  $M = 3.525$ ,  $SE = .125$ ,  $t(23) = .339$ ,  $p = .737$ ), which was as expected due to the fact that there were no cues to category membership in this condition (see Figure 3).

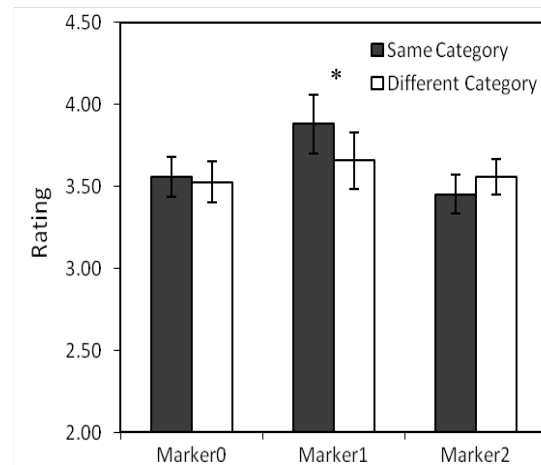


Figure 3: Mean similarity ratings, with standard error. Asterisks indicate significant difference.

### Transfer of category knowledge

One-sample t-tests were performed on word-picture mapping data, collapsed across testing time, to compare performance to chance (taken as 12.5%, in accordance with the number of options available per trial). Performance was significantly above chance for Marker0 ( $M = .73$ ,  $SE = .035$ ,  $t(23) = 17.392$ ,  $p < .001$ ), for Marker1 ( $M = .69$ ,  $SE = .039$ ,  $t(23) = 14.505$ ,  $p < .001$ ), and for Marker2 ( $M = .68$ ,  $SE =$

.045,  $t(23) = 12.470$ ,  $p < .001$ ), indicating that all participants were able to learn the word-picture mappings.

A repeated-measures ANOVA was performed, with condition (Marker0, Marker1, Marker2), consistency (consistent, inconsistent), test time (test1, test2, test3, test4), and language version as factors. There was a significant effect of test time,  $F(3, 108) = 86.266$ ,  $p < .001$ ,  $\eta_p^2 = .706$ , with participants improving steadily across testing phases. There was no significant effect of condition ( $F < 1$ ), and there was no significant interaction between test time and cue condition ( $F < 1$ ). There was no significant effect of consistency ( $F < 1$ ), and there was no significant interaction between consistency and condition,  $F(2, 36) = 1.890$ ,  $p = .166$ .

A univariate ANOVA was performed on the data for the first test to examine whether participants' immediate responses were influenced by their training. There was no significant effect of condition, and there was no significant effect of consistency (both  $F < 1$ ), but the interaction between these two variables approached significance,  $F(2, 36) = 2.554$ ,  $p = .092$ ,  $\eta_p^2 = .124$  (see Figure 4), indicating that prior knowledge about grammatical categories may have influenced early performance. Trends in the means indicate that this interaction was driven by better performance for participants receiving consistent compared with inconsistent labeling in Marker1,  $t(22) = 1.661$ ,  $p = .111$ , and better performance for participants receiving inconsistent compared with consistent labeling in Marker2,  $t(22) = -1.460$ ,  $p = .158$ , however these differences were not significant.

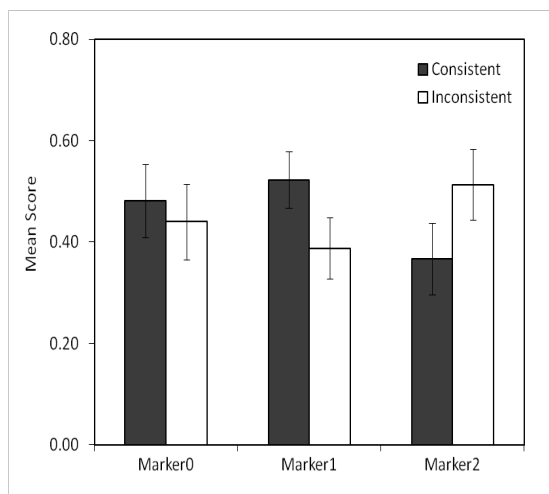


Figure 4: Mean word-picture mapping scores at Test1 (proportion correct), with standard error.

## Discussion

This study aimed to investigate the possibility that high frequency words in speech may assist segmentation, while also simultaneously informing grammatical categorisation (Monaghan & Christiansen, 2010; Monaghan et al., 2007).

Data from the segmentation task indicate that all participants were able to identify target words in the speech, regardless of whether that speech contained bisyllabic target words only, or target words plus monosyllabic marker words. This is especially noteworthy given the complexity of speech in the Marker1 and Marker2 conditions. These findings document a rare demonstration of adults' ability to use statistical information to segment continuous speech that contains words of varying length (see Johnson & Tyler, 2010). That participants were able to segment around the high frequency marker words supports the possibility that learners were able to use these as anchors for segmentation (Altvater-Mackensen & Mani, 2013; Mersad, & Nazzi, 2012; Monaghan & Christiansen, 2010), although in this instance they did not significantly boost segmentation compared to a stream containing just targets. Perhaps increasing pre-exposure frequency to these marker words would result in better performance (see, e.g., Bortfeld et al., 2010, for benefit of prior exposure).

Critically, only participants in the Marker1 condition showed evidence of using the marker words to help form grammatical categories, demonstrated by their giving of higher similarity ratings for targets that came from the same grammatical category, and lower similarity ratings for items that came from different grammatical categories (as denoted by the marker words that preceded these targets in speech). These findings indicate the possibility that the same high frequency words that may assist with segmentation could also inform the formation of grammatical categories – providing crucial behavioral support to the claims of prior research (Monaghan & Christiansen, 2010; Monaghan et al., 2007).

While it was not expected that the Marker0 group would differentiate between the A and B categories due to the absence of category cues in this condition, the results of Marker2 demonstrated that too many high frequency words may be an impediment to learning. One potential reason for this is that due to the increased variability in Marker2: marker words and target words appeared alongside each other half as often in Marker2 than Marker1 speech – thus, it is possible that the category distinction may emerge with more training. However, Valian and Coulson's (1988) studies suggest that a high-ratio between marker words and category words is required to result in effective categorisation of an artificial language, consistent with our results. Representing the differentiation in frequency for function words and content words that is present in natural language is not possible in a small artificial language, but the Marker0 condition is the closest approximation to this of all the three conditions.

Results from the word-picture mapping task illustrate that the prior knowledge of grammatical categories for seen Marker1 participants may have influenced performance in the initial stage of the transfer test; demonstrated by greater performance for participants receiving consistent, compared with inconsistent, mappings. However, these effects are subtle and dissipate after further exposure to the word-

picture pairings, which could be due to the rapid learning that occurs in both consistent and inconsistent conditions of the study as exposure proceeds. Thus, the categorisation effect resulting from the same information sources as inform word segmentation may only be apparent at early stages of language development.

Interestingly, the data suggest that inconsistent, rather than consistent, word-picture pairings yielded better learning for the Marker2 condition at the initial test. This raises the possibility that the increased variability in this language may have led participants to prefer to use the language flexibly immediately after exposure. However, as with the transfer effect seen for the Marker1 group, these effects also dissipate across the remaining phases of the test.

To conclude, findings provide evidence to support the suggestion that the same high frequency words might be helpful for informing learners about word boundaries, and grammatical categories in continuous speech. Previous computational models of segmentation and of word learning have (separately) shown that these same high-frequency words prove useful to each task (Monaghan & Christiansen, 2010). The question then arises as to whether these information sources are used at different stages of language development first to solve the problem of identifying words, then later to discover the grammar. The view suggested by our current results is rather that segmentation and grammatical categorisation are not temporally distinct, but rather operate simultaneously (Frost & Monaghan, 2016).

### Acknowledgments

This work was supported by the International Centre for Language and Communicative Development (LuCiD). The support of the Economic and Social Research Council [ES/L008955/1] is gratefully acknowledged.

### References

- Altwater-Mackensen, N. & Mani, N. (2013). Word-form familiarity bootstraps infant speech segmentation. *Developmental Science, 16* (6), 980-990.
- 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 & 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. Edinburgh, UK: Centre for Speech Technology Research (CSTR), University of Edinburgh. <http://www.cstr.ed.ac.uk/projects/festival.html>
- Bortfeld, H., Morgan, J.L., Golinkoff, R.M., & Rathbun, K. (2005). Mommy and me: familiar names help launch babies into speech-stream segmentation. *Psychological Science, 16*, 298-304.
- Dunn, L., & Dunn, L. (1997). *Peabody picture vocabulary test III*. Circle Pines, MN: American Guidance Service.
- Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition, 147*, 70-74.
- 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.
- Hay, J.F., Pelucchi, B., Graf Estes, K., Saffran, J.R. (2011). Linking sounds to meanings: infant statistical learning in a natural language. *Cognitive Psychology, 63*, 93-106.
- 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.
- Mersad, K. & Nazzi, T. (2012). When mommy comes to the rescue of statistics: Infants combine top-down and bottom-up cues to segment speech. *Language Learning and Development, 8*(3), 303-315.
- 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.
- 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.
- 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.
- 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.
- 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).
- Valian, V., & Coulson, S. (1988). Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language, 27*, 71-86.