UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Categorizing Grammar: Differential Effects of Preceding and Succeeding Contextual Cues

Permalink

https://escholarship.org/uc/item/3cj2h4k7

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 27(27)

ISSN

1069-7977

Authors

Murai, Hajime
Toksumi, Akifumi

Publication Date

2005

Peer reviewed

Categorizing Grammar: Differential Effects of Preceding and Succeeding Contextual Cues

Michelle C. St. Clair (msc110@york.ac.uk)
Department of Psychology, University of York
York, YO10 5DD, UK

Padraic Monaghan (pjm21@york.ac.uk)
Department of Psychology, University of York
York, YO10 5DD, UK

Abstract

Distributional information has been shown to combine with phonological information in aiding categorisation of words into grammatical categories. There has been debate about the type of distributional information that is most useful in category learning, in particular whether bigrams are sufficient for category acquisition. This paper presents two experiments testing people's sensitivity to bigram information for categorisation. Sentences were composed of words with category markers occurring either before (Experiment 1) or after (Experiment 2) the category word. Sentences were presented auditorily. Categorical information was learned in both experiments, but the preceding and succeeding distributional cues contributed to learning in different ways. Furthermore, in both cases phonological information assisted the learning of low frequency words.

Introduction

How do children acquire grammatical categories in their language? In particular, what sources of information are potentially available to enable this process to occur? We present two experiments exploring the interaction between contextual information and phonological information in the child's environment, and the extent to which these sources can drive learning in artificial grammar learning experiments.

Mintz (2002, 2003) has suggested that children acquire grammatical categories based on frequently occurring "frames" in the child's language environment. Such frames are defined by co-occurring words immediately prior to and succeeding the target word, i.e., a phrase of the form aXb, where X is the category word, and a and b co-occur frequently in text. In child-directed speech in English, one example is the frame "The X is…" where X could be one of several nouns. Mintz (2002) conducted an artificial language study that incorporated frequent frames as markers for categories, and found evidence for categorisation based on this structure.

To follow this work, Mintz (2003) conducted corpus analyses of small samples of child directed speech taken from CHILDES (MacWhinney, 2000). Using the 45 most frequent frames in these corpora, Mintz (2003) found that the accuracy of the category groupings based on these frames was extremely high. However, completeness was very low, though was significantly above a random baseline.

Monaghan and Christiansen (2004) suggested that this low completeness indicated that frequent frames had very low coverage and that this was consequently a poor source of information for categorisation. Instead, they suggested that bigram information could provide richer cues about category. The trigram information in the aXb frames was conflated with aX (e.g., "The X") and Xb (e.g., "X is") bigrams, and learning in Mintz (2002) may have been driven by information at the bigram level only. Monaghan and Christiansen (2004) replicated Mintz's (2003) corpus analyses, finding high accuracy and low completeness for aXb frames, and slightly lower accuracy but much higher completeness for bigram information: categorizing 69.9% of the words, compared to only 14.3% using the aXb frame. In addition, Monaghan and Christiansen (2004), in neural network models trained to learn grammatical categories, found that when the aXb frame was decomposed as aX and Xb then learning increased.

Yet the finding that bigram information is potentially useful for category learning is not the same as demonstrating that it is useable. However, Valian and Coulson (1988) constructed an artificial language where bigram information was exploited for category learning. Sentences were of the form aAbB, where a and b were high frequency marker words, and A and B were sets of category words. They found that novel aAbB sentences were preferred to those that violated the bigram information: *aBbA. In an extension of this study, Monaghan, Chater, and Christiansen (2005) found direct evidence for words from the same category being grouped together. Both these studies have explored the learning of aX bigram information, but as yet there is no evidence that categories can be learned from Xb bigrams.

Furthermore, these artificial language studies were conducted with sentences presented visually. There may be a difference in categorisation performance when the modality of presentation is altered. In an artificial language learning experiment, Onnis, Christiansen, Chater, and Gómez (2003) found a similar trend in visually presented stimuli compared to auditory presentation, though the effect was reduced in the former.

Distributional information is not the only cue determining grammatical categorisation; phonological information has also been implicated in the learning of grammatical and gender categories (Braine et al., 1990; Brooks, Braine, Catalano, Brody, & Sudhalter, 1993). Indeed Braine (1987) has claimed that support from phonological information is vital for categories to be learnable. Kelly (1992) has shown

correspondences between grammatical categories and a range of phonological cues, and Cassidy and Kelly (2001) indicated that such phonological cues could be utilised for grammatical category judgements. Monaghan et al. (2005) investigated the combination of phonological and distributional information in grammatical categorization based on the frequency of words. Using corpus analyses, they found that distributional information was a highly reliable cue for high frequency words, but that reliability reduced for lower frequency words. In addition, the reliability of phonological information was highest for the low frequency words. In an artificial language experiment, Monaghan et al. (2005) found that phonological information provided most assistance to categorising words that occurred with low frequency in the language.

The experiments we present below extend previous studies of learning categories from bigrams. In Experiment 1 we tested whether the effects observed in Monaghan et al. (2005) pertain when sentences are presented auditorily. In Experiment 2, we tested whether category learning can proceed on the basis of Xb bigram information alone.

Experiment One

Method

Participants Twenty-nine University of York undergraduate and postgraduate students participated. Five participants were judged to have misunderstood the task and were excluded. All participants were native English speakers and were either paid £2 or given course credit.

Stimuli and Materials The artificial language used was an adapted auditory version of Valian and Coulson's (1988) written language. There were 12 words in the language divided into two categories, A and B. Each word was always preceded by a marker word for its category. One marker word always preceded category A words whereas the marker word preceded category B words. The two marker words, *alt* and *ong,* were counterbalanced across participants, in terms of whether *alt* marked A category words or B category words. Half of the category words were of high frequency and occurred twice as often (8 times in each training session) as the low frequency category words (4 times per training session).

Table 1: The category words used in the Experiments.

Frequency	Category A	Category B
High	Tweand	Foth
	Dreng	Vawse
	Klimp	Suwch
Low	Gwemb	Zodge
	Prienk	Thorsh
	Blint	Shufe

Table 2. Examples of the test sentences

Type I	Type II	Type III
aAbB	*aAaB*	*bAaB*
bBaA	*bBbA*	*aBbA*

There were two conditions to test the role of phonological cues in categorisation. In the phonologically coherent condition, words within the same category shared phonological properties. Category A words had consonant clusters at the onset and offset, rounded low vowels, and contained nasals and stops. Category B words had no consonant clusters, contained unrounded high vowels, and fricatives. For the phonologically incoherent condition the low frequency words from the Category B were exchanged with the high frequency words from Category A, creating categories with no common phonological cues. Table 1 lists the words.

During training, 18 sentences were presented in the form *aAbB*, where *a* and *b* were the marker words and *A* and *B* were category words. Category words appeared equally often in first and second position.

The test session comprised of 24 sentences. 12 sentences had not occurred during training, but conformed to the artificial language's regularities (Type I sentences). 6 were composed of high frequency category words, and 6 of low frequency. Of the 12 sentences that violated the artificial language's regularities, 6 had the same marker word preceding both category words (e.g., *aAaB*, Type II sentences), and 6 had both marker words preceding the wrong category word (e.g., *aBbA*, Type III sentences). The violation occurred in half the cases for high frequency and half for low frequency category words for Type II sentences. Table 2 summarises the three sentence types.

There were also 12 cards with the category words printed on them which were used in a card sorting task. The language was synthesized with Festival Speech Synthesizer using male British English diphones (Black, Taylor, & Caley, 1990). Headphones were used to deliver the stimuli presented on a Hewlett Packard Pavilion laptop using E-Prime.

Procedure Participants were instructed to pay attention to the patterns within a made-up language. They first heard all the words in the language. Then the first training session began, where they heard two sets of the 18 training sentences in a random order. For each trial the sentence was heard twice with approximately 1s interval between the two repetitions. Participants were instructed to repeat the sentence aloud after the repetition. During the test phase, participants were instructed that half the sentences were similar to the training language and half dissimilar. They then had to judge whether each of the 24 test sentences was similar or dissimilar by pressing a keyboard key. Participants were then given two more sets of the 18 training sentences, and then testing was repeated using a different set of 24 test sentences. After the

second test session, the participants were given instructions to sort the 12 category cards into two groups according to which words they thought went together.

Results

We performed a repeated measures ANOVA on accuracy with time (1st and 2nd test session), and frequency (high and low) as repeated measures, and coherency (phonologically coherent or incoherent) as a between subjects factor. Figure 1a shows number correct for high and low frequency sentences for coherent and incoherent groups. Score in each column was out of a maximum of 24, with chance level 12 correct. However, there were no significant effects or interactions (time by frequency: $F(1, 22) = 1.616$, $p = .217$; all others: $F<1$). Though performance was better than chance, this could be due to the participants learning to respond to the syntax of the sentences alone, effectively rejecting all Type II sentences but accepting all sentences which had two different marker words (Type I and III). This would result in performance around 18 out of 24 correct, though not necessarily as a consequence of category learning.
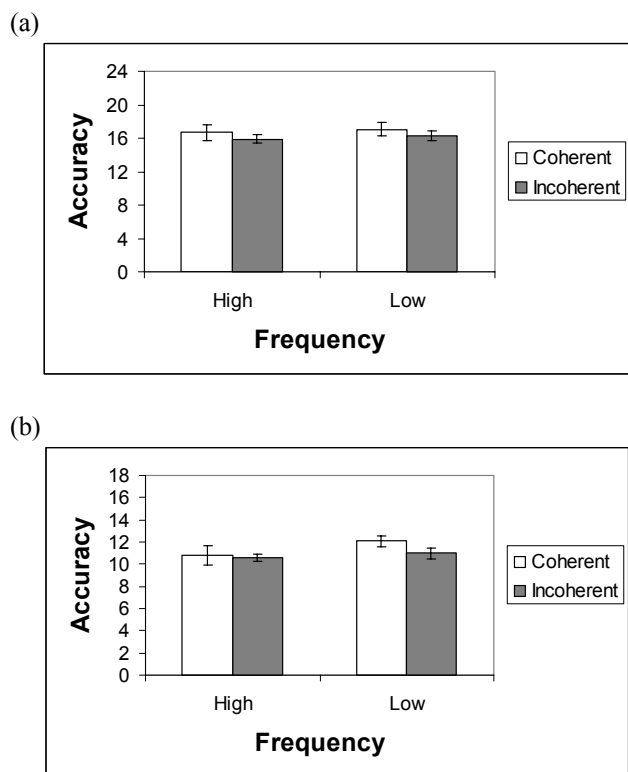
(a)

(b)

Figure 1. Performance for low and high frequency sentences in the coherent and incoherent condition for (a) all test sentences, (b) Type I and Type III sentences only.

To test the extent to which categories had been learned, we looked at performance on correctly accepting Type I sentences and correctly rejecting Type III sentences. The results are shown in Figure 1b. Again, we performed an ANOVA on accuracy with time and frequency as within subjects factors and coherency as a between subjects factor. There was a marginally significant effect of frequency, $F(1, 22) = 4.207$, $p = .052$, with higher accuracy for low frequency sentences than high frequency sentences. No other effects were significant (frequency by coherency: $F(1, 22) = 1.052$, $p = .316$; all others: $F < 1$).

For the card-sorting task, we scored the number of category A words grouped together. The coherent and incoherent groups did not significantly differ from each other, $t(22) = 1.969$, $p = .062$. The coherent group did not differ from chance level of 3.9, $t(11) = .064$, $p = .950$; mean 3.917 of 6). The incoherent group scored significantly below chance level, $t(11) = -3.987$, $p < .01$; mean 3.333 of 6).

Discussion

There was some evidence that category learning was taking place in this study. The card-sorting task showed that, in making judgements about category, phonological properties interfered with correct classification when there was a mismatch between phonological cues and the categories of the words. However, we did not find a significant difference between the coherent and incoherent groups.

In terms of performance on the similarity judgements for sentences, the absence of coherency effects was in contrast to the visual presentation of the experiment in Monaghan et al. (2005). They found a significant main effect of coherency, with coherent sentences responded to more accurately than incoherent sentences. They also found a significant main effect of frequency which was in the reverse direction to that found here. The greater accuracy for low frequency sentences in the current experiment was perhaps due to an effect of greater familiarity of high frequency category words, which may have influenced incorrect acceptance of Type III sentences as similar to the training items.

The absence of other effects, in contrast with the previous visual studies may have been due to reduced learning in the auditory modality. In Monaghan et al. (2005), overall performance on the task was 74% correct. For the current study, performance was at 69% overall. If chance performance is 67%, after correctly learning the syntax of the sentences (so rejecting Type II sentences), then performance may well be at floor levels for the auditory version of the experiment. We discuss confounding factors that may have reduced the effect with speech synthesised stimuli in the General Discussion.

First, though, the next experiment tested whether category learning could occur on the basis of Xb bigram information.

Experiment Two

Method

Participants Twenty-five University of York undergraduate and postgraduate students and staff members participated in the study. One participant was excluded as he/she was judged to have misunderstood the task. None had participated in the

previous experiment. All participants were native English speakers and were paid £2.

Stimuli & Materials The stimuli and materials were exactly the same as Experiment 1, except all sentences now had the marker word in the second and fourth position in the sentence as opposed to the first and third in Experiment 1. For example, the sentence *aAbB* in Experiment 1 became *AaBb* in Experiment 2. All training and test sentences were changed to reflect the new sentence structure.

Procedure The procedure was the same as in Experiment 1.

Results
Performance in this study was slightly more accurate than for Experiment 1, with 70% overall correct performance. The means for overall score for low and high frequency sentences in the coherent and incoherent condition are shown in Figure 2a. A repeated measures ANOVA was performed on accuracy, with time and frequency as repeated measures and coherency as a between subjects variable.
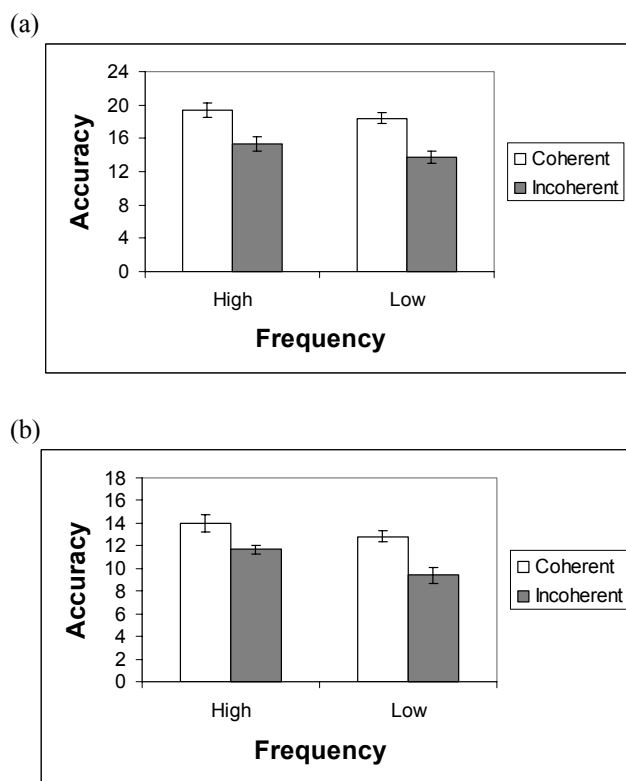
(a)

(b)

Figure 2. Mean accuracy by high and low frequency sentences for the coherent and incoherent groups. (a) overall accuracy; (b) accuracy for Type I and Type III sentences only.

There was a significant main effect of frequency, $F(1, 22) = 6.079$, $p < .05$, with high frequency sentences responded to with more accuracy than low frequency sentences. There was also a main effect of coherency, $F(1, 22) = 19.265$, $p < .001$,

with higher accuracy in the coherent group. No other main effects or interactions were significant (time: $F(1, 22) = 1.671$, $p = .210$; time by frequency: $F(1, 22) = 1.287$, $p = .269$; time by frequency by coherency: $F(1, 22) = 1.287$, $p = .269$; all others: $F<1$).

To test whether these effects were driven by participants learning to correctly reject Type II sentences, we conducted an ANOVA on correct responses to Type I and III sentences, with time and frequency as repeated measures factors, and coherency as a between subjects factor. We again found a significant main effect of frequency, $F(1, 22) = 11.861$, $p < .005$, and of coherency, $F(1, 22) = 15.554$, $p < .001$. No other effects or interactions were significant (frequency by coherency: $F(1, 22) = 1.192$, $p = .287$; time by frequency: $F(1, 22) = 1.148$, $p = .296$; all others: $F < 1$). The mean scores are shown in Figure 2b.

For the card-sorting task, the phonologically coherent group performed better than the incoherent group, $t(22) = 2.152$, $p < .05$. The coherent group did not differ from chance level (mean 4.167 from 6), $t(11) = .985$, $p = .346$, but the incoherent group was significantly below chance level, mean 3.5 out of 6, $t(11) = -2.653$, $p < .05$.

Discussion
Similar to the results from Experiment 1, the card-sorting task indicated that coherency influenced performance in learning to group words into categories according to the distributional information present in the language. As in Experiment 1, the main result from the card-sorting task was the interference resulting from a mismatch between phonological coherence and category.

The effect of coherency for sentence judgements was in accordance with previous studies in the visual modality. Consistent phonological information aided performance on the task. The main effect of frequency was also in line with that of previous studies of category learning, with better performance on high frequency sentences than low frequency sentences. However, there was no significant interaction between frequency and coherency, as found in Monaghan et al. (2005). So, coherency contributed an even advantage for both low and high frequency sentences.

Though performance was slightly more accurate than Experiment 1, overall, the results appeared to be noisier than in studies where sentences were presented in the visual modality. To increase power, we combined the results of Experiments 1 and 2 to examine category learning across the two studies.

Combined analyses
We performed an ANOVA on correct responses with time ($1^{st}/2^{nd}$ test) and frequency as within subject measures, and Experiment (1/2) and coherency as between subjects factors. There was a significant main effect of coherency, $F(1, 44) = 15.089$, $p < .001$, with better overall performance for the coherent condition. No other main effects were significant (frequency: $F(1, 44) = 1.440$, $p = .237$; time: $F<1$). There was

a significant interaction between Experiment and frequency, $F(1, 44) = 5.489$, $p < .05$. This was due to the advantage for low frequency sentences in Experiment 1 and better performance for high frequency sentences in Experiment 2. There was also a significant interaction between Experiment and coherency, $F(1, 44) = 7.549$, $p < .01$. Phonological coherency had a larger effect for Experiment 2 than for Experiment 1. The interaction between time, frequency, and Experiment was marginally significant, $F(1, 44) = 2.840$, $p = .099$, but no other interactions were significant (time by frequency by coherency: $F(1, 44) = 1.219$, $p = .276$; all others: $F < 1$).

An ANOVA on performance on Type I and Type III sentences revealed the same main effects and interactions.

For the card-sorting task, an ANOVA was conducted with Experiment and coherency as between subjects factors. There was a significant main effect of coherency, $F(1, 44) = 8.505$, $p < .01$). The coherent group correctly sorted more category cards correctly than the incoherent group. There was no main effect of Experiment and no significant interaction between the two factors ($F < 1$). The coherent group did not score above chance level of 3.9 cards, $t(23) = .764$, $p = .452$, mean is 4.042 of 6. However, the incoherent group sorted significantly fewer cards than chance, $t(23) = -4.702$, $p < .001$, mean is 3.417 of 6.

General Discussion

The auditory presentation of the stimuli in Experiment 1 resulted in several differences from Monaghan et al.'s (2005) study which used similar stimuli presented in the visual modality. The effects of frequency and phonological coherency were not found in the auditory presentation, and furthermore no significant interaction between frequency and coherency was present. We attempted to equate the timing of presentation in the two studies. In the written version, the training sentences were presented for 10 seconds. In the current experiments the inter-stimulus interval was also 10 seconds, but the participant only heard the sentence during the first 5 seconds, leaving the rest of the time for repeating the sentence aloud, so reducing the opportunities for rehearsal. A further difference between the auditory and visual presentation experiments was that in the former, participants had to form a connection between the auditory stimuli and the words written on the cards. Some participants reported visualizing how the words ought to be spelled which were different to those written on the cards. The speech-synthesised stimuli may have also contributed to reduced learning in the auditory study. We removed prosodic information from the stimuli, as we did not want uncontrolled cues for categorisation to be present. However, this absence of prosody also contributed to the speech stimuli sounding unnatural, which may have impacted on learning.

Yet, the results of Experiment 2 demonstrated that learning can occur in the auditory modality, and furthermore that learning from Xb information is possible as well as learning from aX information. Indeed, learning appeared to be more effective when the marker word followed the category word, though there was no significant main effect of Experiment. Yet, the properties of the stimuli exerted a different effect in the two Experiments. The significant Experiment by frequency interaction reflected the influence of familiarity in the aX learning, resulting in high-frequency Type III sentences being incorrectly accepted as similar to the training sentences. Yet, in the Xb experiment, high frequency occurrence of the category word resulted in better classification of sentences containing these words, which supported previous visually presented versions of aX stimuli.

The influence of phonological cues was another difference between the two Experiments. The significant interaction between coherency and Experiment indicates the large effect of phonological cues in the Xb study, but little impact in the aX learning. Yet, phonological cues were found to have an influence in the card-sorting task, which directly measured the grouping of words of the same category, as defined by the distributional information. In both studies, a mismatch between phonological cues and word categories resulted in performance lower than chance. Where there was a match, categorisation was above chance, but not significantly so.

We predicted that the position of the contextual cue would not influence categorization. However, we found a greater influence of frequency and phonological coherence in Experiment 2, where the distributional cues succeeded the category words. One possibility for the greater sensitivity of learners to phonological information in Experiment 2 was that the category word occurring first in the sentence drew attention to the phonological characteristics to a greater extent than when it was presented after the marker word, as in Experiment 1. A further possibility is that category learning was better facilitated from a distributional cue occurring after the category word. Marker words occurring in this position may have operated more like an inflection than as a distributional cue. In English, inflections appear to be more reliable markers of syntactic category than preceding words such as articles for nouns, or pronouns for verbs (Maratsos & Chalkley, 1980). An adjective may intervene between an article and a noun, and adverbs can separate subjects from verbs, reducing the reliability of preceding word cues for grammatical categories. In contrast, material intervening between a word stem and an inflection is rare, and so the inflection cue is more reliable.

We have found that learning can take place on the basis of both aX and Xb bigram information. Monaghan and Christiansen (2004) showed that bigram information is more useful than frames in learning categories, and the current studies have shown that such information is not only useful but useable by adults determining word categories in an artificial language. However, the results suggest that distributional information preceding the target word plays a subtly different role in categorisation than information following the word.

References

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.

Braine, M.D.S. (1987). What is learned in acquiring word classes: A step toward an acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition* (pp.65-87). Hillsdale, NJ: Lawrence Erlbaum Associates.

Braine, M.D.S., Brody, R.E., Brooks, P.J., Sudhalter, V., Ross, J.A., Catalano, L., & Fisch, S.M. (1990). Exploring language acquisition in children with miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Cognition, 29,* 591-610.

Brooks, P. J., Braine, M. D. S., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language, 32*, 76-95.

Cassidy, K. W., & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language, 30*, 348-369.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99,* 349-364.

MacWinney, B. (2000). *The CHILDES project: Tools for analyzing talk.* Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Maratsos, M.P. & Chalkley, M.A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K.E. Nelson (Ed.), *Children's Language* Volume 2, pp.127-214. New York: Gardner Press.

Mintz, T.H. (2002). Category induction from distributional cues in an artificial language. *Memory and Cognition, 30*, 678-686.

Mintz, T.H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*, 91-117.

Monaghan, P., Chater, N., & Christiansen, M.H. (2005). The differential contribution of phonological and distributional cues in grammatical categorization. *Cognition*, in press.

Monaghan, P. & Christiansen, M.H. (2004). What distributional information is useful and usable in language acquisition? *Proceedings of the 26th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.

Onnis, L., Christiansen, M.H., Chater, N., & Gómez, R. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial grammar learning. *Proceedings of the 25th Cognitive Science Society Conference* (pp.887-891). Mahwah, NJ: Lawrence Erlbaum.

Valian, V. & Coulson, S. (1988). Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language, 27*, 71-86.