The Role of Distributional Information in Linguistic Category Formation

Patricia A. Reeder (preeder@bcs.rochester.edu)
Elissa L. Newport (newport@bcs.rochester.edu)
Richard N. Aslin (aslin@cvs.rochester.edu)
Department of Brain & Cognitive Sciences, University of Rochester
Meliora Hall, Box 270268
Rochester, NY 14627 USA

Abstract

A crucial component of language acquisition involves organizing words into grammatical categories and discovering relations between them. Many studies have argued that phonological or semantic cues or multiple correlated cues are required for learning. Here we examine how distributional variables will shift learners from forming a category of lexical items to maintaining lexical specificity. In a series of artificial language learning experiments, we vary a number of distributional variables to category structure and test how adult learners use this information to inform their hypotheses about categorization. Our results show that learners are sensitive to the contexts in which each word occurs, the overlap in contexts across words, the non-overlap of contexts (or systematic gaps), and the size of the data set. These variables taken together determine whether learners fully generalize or preserve lexical specificity.

Introduction

Language acquisition crucially involves finding the grammatical categories of words in the input. The organization of elements into categories, and the generalization of patterns from some seen element combinations to novel ones, account for important aspects of the expansion of linguistic knowledge in early stages of language acquisition. One hypothesis of how learners approach the problem of categorization is that the categories (but not their contents) are innately specified prior to experiencing any linguistic input, with the assignment of tokens to categories accomplished with minimal exposure. A second possibility is that the categories are formed around a semantic definition. A third hypothesis, explored in the present research, is that the distributional information in the environment is sufficient (along with a set of learning biases) to extract the categorical structure of natural language. While it is likely that each of these sources of evidence makes important contributions to language acquisition, this third hypothesis regarding distributional learning has often been thought to be an unlikely contributor, given the information processing limitations of young children and the complexity of the computational processes that would be entailed.

Furthermore, it has been difficult to test the importance of such a distributional learning mechanism because the cues to category structure in natural languages are highly correlated. In fact, it has been argued in many artificial language studies that the formation of linguistic categories (e.g., noun, verb) depends crucially on some perceptual property linking items within the category (Braine, 1987). This perceptual similarity relation might arise from identity or repetition of elements in grammatical sequences, or a phonological or semantic cue identifying words across different sentences as similar to one another (for example, words ending in *–a* are feminine, or words referring to concrete objects are nouns). Learners of artificial languages have been unable to acquire grammatical categories and to extend their linguistic contexts to new items correctly without such cues (Braine et al., 1990; Frigo & McDonald, 1998; Gomez & Gerken, 2000). However, this has been somewhat of a puzzle: Maratsos & Chalkley (1980) argued that in natural languages, grammatical categories do not have reliable phonological or semantic cues; rather, learners must utilize distributional cues about the linguistic contexts in which words occur to acquire such categories. Mintz, Newport & Bever (2002), as well as several other researchers, have shown that computational procedures utilizing distributional contexts can form elementary linguistic categories on corpora of mothers' speech to young children from the CHILDES database, and Mintz (2002) and Gerken et al. (2005) have shown that both adults and infants can learn a simple version of this paradigm in the laboratory, at least when there are multiple correlated distributional cues. In the present series of experiments we also begin by demonstrating that there are distributional properties that lead to successful learning of linguistic categories in artificial language paradigms. Importantly, however, in order to understand how this mechanism works in human learners and why many previous experiments have not found such learning, we present a series of experiments that manipulate various aspects of these distributional variables, in order to understand the computational requirements for successful category learning.

Experiment 1

An artificial grammar was created with the structure (Q)AXB(R), where each letter represents a set of 2 or 3 words: the Q and R categories had 2 words each, and the A, X, and B categories had 3 words each. The words of the grammar were *spad*, *klidum*, *flairb*, *daffin*, *glim*, *tomber*, *zub*, *lapal*, *fluggit*, *mawg*, *bleggin*, *gentif*, and *frag,* and there was no referential world to which the words were mapped. All studies were run with two languages that assigned different words to each of the categories. X was the target category under study, while A and B were the context elements that formed the distributional cues to the category. Q and R served as optional categories that made sentences of the language vary in length from 3 to 5 words and made words of the language observe patterning in terms of relative order but not fixed position. Focusing on just the AXB portion of the grammar, there were 3x3x3=27 possible word strings in the language. In order to study whether learners can acquire X as a category of words, rather than simply learn the specific word strings to which they have been exposed, we present some of these AXB's but withhold others; and then we ask during post-exposure testing whether learners recognize the withheld AXB's as grammatical.

Method

Participants 17 monolingual native English-speaking students at the University of Rochester participated in Experiment 1 and were paid for their time. Eight subjects were exposed to language 1, and nine subjects were exposed to language 2.

Stimulus Materials Out of the 27 basic AXB sentence types in the language, 18 were presented and 9 were withheld (see Table 1). By varying whether the 2 Q words and the 2 R words were present or absent, the 18 AXB types used for exposure were enlarged to a total of 72 different (Q)AXB(R) sentences. This exposure set of 72 sentences was presented 4 times, forming 20 minutes of exposure to the language. The 18 sentence types used in exposure included each X word in the presence of each A word and each B word. Thus the exposure set for this language is *dense* (covering a high proportion of the overall language space), and has complete *overlap* of contexts among the various X words within the target category.

Words were read in isolation by a native English-speaking female and were spliced together to form the sentences of the language. Each word was recorded with both non-terminal and terminal intonation, and the words were adjusted in Praat so the pitch, volume, and duration of words were fairly consistent. Sentences were constructed by assembling words in sequences in Sound Studio, with 50ms silence between each word, and using the word token with a terminal intonation contour as the final word in the sentence. Exposure strings were recorded to mini-disc, with approximately 1.5s of silence between sentences.

After exposure, participants were asked to rate test strings on a scale of 1 to 5, where 1 meant that the string sounded like it definitely did not come from the exposure language and 5 meant that the string definitely came from the exposure language. Test strings were all 3-word sentences of three types: grammatical familiar (9 AXB strings presented during training), grammatical novel (9 AXB strings withheld during training), and ungrammatical (strings of the form AXA or BXB).

Table 1: Possible AXB strings in Exp. 1-4. Items withheld in Exp. 1 are denoted *; items withheld in Exp. 2 are denoted ♦; items withheld in Exp. 3 & 4 are denoted ○.

A1 X1 B1 ♦ ○	A1 X2 B1 * ♦ ○	A1 X3 B1
A1 X1 B2 * ♦ ○	A1 X2 B2 ○	A1 X3 B2 ♦
A1 X1 B3	A1 X2 B3 ♦ ○	A1 X3 B3 * ♦ ○
A2 X1 B1 * ♦ ○	A2 X2 B1	A2 X3 B1 ♦ ○
A2 X1 B2	A2 X2 B2 ♦ ○	A2 X3 B2 * ♦ ○
A2 X1 B3 ♦	A2 X2 B3 * ♦ ○	A2 X3 B3 ○
A3 X1 B1 ○	A3 X2 B1 ♦	A3 X3 B1 * ♦ ○
A3 X1 B2 ♦ ○	A3 X2 B2 * ♦ ○	A3 X3 B2
A3 X1 B3 * ♦ ○	A3 X2 B3	A3 X3 B3 ♦ ○

Results

A repeated measures ANOVA was conducted with condition (familiar, novel, and ungrammatical) as the within subjects factor and language as the between subjects factor. There were no significant effects of language (F<1), so the two languages have been collapsed. The mean rating of grammatical familiar strings was 3.78 (SE=0.13), the mean rating of grammatical novel strings was 3.69 (SE=0.13), and the mean rating of ungrammatical strings was 2.58 (SE=0.19). There was no significant difference between ratings of grammatical novel items and grammatical familiar items ($F(1,15)=1.845$, $p=0.1$). However, these items were rated significantly higher than ungrammatical items ($F(1,15)=45.651$, $p<0.001$).

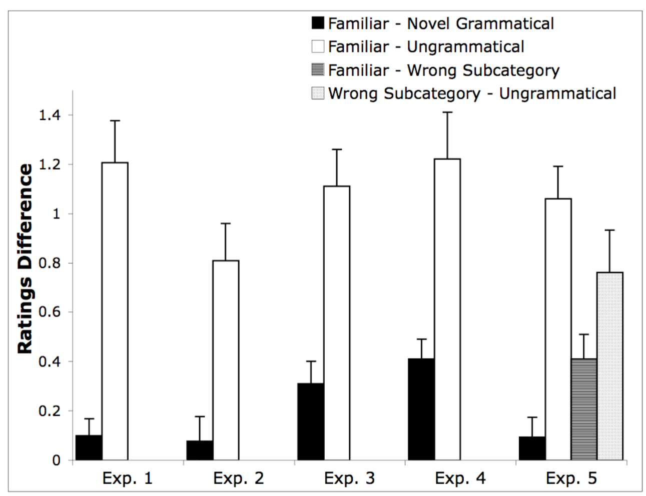

Figure 1: Difference scores of grammatical familiar items and grammatical novel items, and grammatical familiar items and ungrammatical items from Experiments 1-5.

Discussion

In this first experiment, learners did not discriminate between the presented and the withheld AXB's, both of which were rated as highly grammatical and strongly preferred to ungrammatical sentences AXA or BXB. These findings show, therefore, that when the input densely samples the language space and words within a category appear in highly overlapping contexts, learners will fully generalize within the category to novel contexts and novel strings, even without any perceptual or semantic cues to indicate that the words form a single category. In our subsequent experiments, we investigate the degree to which category generalization is affected by manipulating these distributional variables, in learning a single category and in learning subcategories.

Experiment 2: Sparseness

In Experiment 2, we explored what happens if we keep the number and overlap among X-word contexts in the language the same, but during learning we present learners with substantially fewer of the contexts that are possible in the language. We refer to this as *reducing the density* (or *increasing the sparseness*) of the contexts for X words that are presented during learning.

Method

Participants 16 monolingual native English-speaking students at the University of Rochester participated in Experiment 2, eight in each of the two possible languages. Subjects had not participated in any other categorization experiment and were paid for participation.

Stimulus Materials Strings were created in the same manner as Experiment 1. Here, however, out of the 27 possible AXB combinations, only 9 were presented during exposure (see Table 1). Crucially, every X-word was still heard in combination with every A and every B. As in Exp. 1, each sentence type was presented with optional category elements Q and R present or absent, producing 36 sentences in the exposure set. The exposure set was presented 4 times, for a total exposure of about 10 minutes. (Each input sentence type was thus presented with the same frequency in this experiment as in Exp.1; the overall exposure was reduced in time and number of strings by reducing the size of the exposure set.) The test was the same as in Exp. 1, except that the grammatical novel test items were counterbalanced such that half of the participants in each language were tested on one subset of nine of the withheld (grammatical novel) items, and the other participants were tested on the other nine grammatical novel items.

Procedure The procedure was the same as in Experiment 1.

Results

As in Experiment 1, there was no difference between the two counterbalanced languages ($F<1$), so all further analyses combine the languages. The mean rating of grammatical familiar strings was 3.54 (SE=0.12), the mean rating of grammatical novel strings was 3.47 (SE=0.12), and the mean rating of ungrammatical strings was 2.73 (SE=0.14). A repeated measures ANOVA showed that grammatical novel strings were rated just as highly as grammatical familiar strings and there was no significant difference between the two types of items ($F(1,14)=.558$, $p>0.5$). The analysis further revealed that the ungrammatical items were rated significantly lower than the grammatical items ($F(1,14)=28.767$, $p<0.001$).

Discussion

These results show that learners' performance is unchanged from Experiment 1 when density/sparseness is reduced but other properties of the distributional information are maintained, despite the fact that the exposure is half as rich and half as long. This permits us to ask what happens, in contrast, when the amount of *overlap* in the contexts of the X-words is reduced.

Experiment 3: Overlap

In Experiment 3, as in Experiment 2, we presented only 9 of the 27 possible AXB combinations. Here, however, we presented particular AXB combinations that reduced the degree of overlap among members of X in the contexts in which they were heard, in order to assess the importance of the overlap in distributional information for category formation and generalization. In the present experiment, the set of X-words, taken together, occurred in all of the A and B contexts, and the different X-words overlapped in part with all the other X-words. However, individual X-words did not fully share all their contexts with one another. The question we address, then, is the degree to which learners will restrict their generalization across the category as a function of this reduction in overlap.

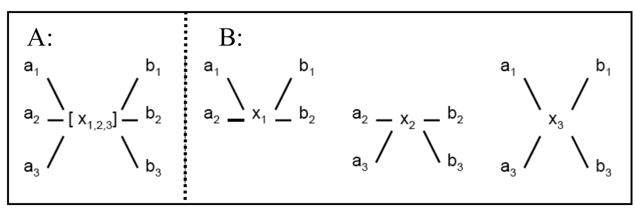

Figure 2: Full overlap in the grammar space for the X-words in Experiment 2 (Fig 2A), compared to the partial overlap in Experiment 3 (Fig 2B).

Method

Participants 24 monolingual native English-speaking students at the University of Rochester participated in Experiment 3 (12 in each language). Subjects had not participated in any other categorization experiment and were paid for their participation.

Stimulus Materials Strings were composed in the same way as in Experiment 1, and only 9 of the 27 possible AXB combinations were heard. X1 was heard in the context of A1, A2, B1, and B2, but not in the context of A3 or B3. X2 was heard in the context of A2, A3, B2, and B3, but not A1 or B1. X3 was heard in the context of A1, A3, B1, and B3, but not A2 or B2. Thus, the overlap among contexts is maintained over the X category as a whole, but individual words in X do not have the degree and type of overlap in distributional contexts that they do in Experiments 1 and 2, where each X word occurs with each A and each B.

Procedure The procedure was the same as Experiment 1.

Results

A repeated measures ANOVA was conducted and revealed no differences between languages one and two ($F(2,44)=1.581$, $p>0.1$); therefore, all of the following analyses collapse the two languages. The mean rating of grammatical familiar items was 3.79 ($SE=0.1$), the mean rating of grammatical novel items was 3.48 ($SE=0.16$), and the mean rating of ungrammatical items was 2.85 ($SE=0.15$). The ANOVA revealed significant differences between grammatical familiar and grammatical novel items ($F(1,22)=19.191$, $p<0.001$) and between grammatical and ungrammatical items ($F(1,22)=70.271$, $p<0.001$).

Discussion

Whereas in Experiment 2 we tested how subjects would respond to fewer contexts but full overlap of the context environment, Experiment 3 greatly reduced the overlap in the exposure while keeping number the same (see Figure 2A as compared to Figure 2B). It is important to note that, at some point along the sparseness and non-overlap dimensions, learners must stop concluding that X is a category and must acquire lexical restrictions or shift to word-by-word learning. The results of Experiment 3 give insight into the computational details of how this occurs by showing that, despite full coverage over lexical items, the incomplete *overlap* between words led to a slight decrease in generalization. At the same time, however, learners did continue by and large to generalize, showing a much higher rating for grammatical novel items than for ungrammatical items. These results suggest that learners take into account both the overlap and the non-overlap among items, modestly reducing their willingness to generalize when the data supporting generalization are less strong.

Experiment 4: Overlap with extended exposure

One more variable that may impact generalization versus lexical distinctness is the frequency or consistency with which each type of context is presented (and therefore the frequency with which contextual gaps recur). If learners operate in an optimal way when using the statistics of their input corpus, the prediction is that very high frequencies of sparse distributional information, with systematic and recurring gaps, should lead learners to increased certainty that the gaps are meaningful and should restrict generalization. Indeed, this is the result obtained in work by Wonnacott, Newport and Tanenhaus (2008) in a miniature verb-argument structure learning paradigm, as well as in work on concept acquisition by Xu and Tenenbaum (2007). In Experiment 4, we explored how an increase in the amount of exposure to the very same corpus used in Experiment 3 would affect categorization.

Method

Participants 16 monolingual native English-speaking students at the University of Rochester participated in Experiment 4 (8 in each language). Subjects had not participated in any other categorization experiment and were paid for their participation.

Stimulus Materials The corpus was the same as in Experiment 3; however exposure was doubled, by presenting the exposure corpus 8 times rather than 4. (The exposure thus lasted for approximately 20 minutes, as in Experiment 1, but contained only 9 contexts, as in Experiments 2 and 3). The same test as in Experiment 3 was given after exposure.

Procedure The procedure was the same as Experiment 1.

Results

A repeated measures ANOVA revealed no significant differences between languages one and two ($F<1$), so they have been combined for all following analyses. The mean rating of grammatical familiar items was 4.05 ($SE=0.14$), the mean rating of grammatical novel items was 3.64 ($SE=0.16$), and the mean rating of ungrammatical items was 2.83 ($SE=0.24$). There were highly significant differences between all conditions. Novel items were rated significantly different from familiar items ($F(1,14)=26.865$, $p<0.001$), and ungrammatical items were rated significantly lower than novel items ($F(1,14)=39.756$, $p<0.001$).

Discussion

The results from Experiment 4 reveal that increased exposure to a corpus containing incomplete overlap reduces the likelihood that learners will generalize based on this input. Instead, they are more likely to assume that gaps in the input are intentional. Nevertheless, the novel grammatical test strings are judged to be more grammatical than the ungrammatical strings. Presumably, even more exposure to highly consistent gaps would confirm the ungrammaticality of the novel grammatical strings. In contrast, more unsystematic gaps with extended exposure should lead learners to generalize more.

Experiment 5: Subcategorization

Experiments 1-4 tested whether learners can acquire a single category, generalizing from hearing some instances of the distributional contexts of individual words (with some withheld) to the full range of contexts for all the individual words in the set. As previously noted, a large body of work has concluded that linguistic categories in artificial language experiments cannot be formed on the basis of distributional contexts alone, and that additional information (such as phonological or semantic cues) are required for successful learning. Experiments 1-4 showed that additional cues are not necessary for adults to induce a category from distributional contexts alone. However, in some cases the category learning problems observed by other experimenters have been when the language contained *subcategories* – subsets of words with distinct privileges of occurrence (such as nouns of different genders). Experiment 5 explores whether subcategories are also learnable from distributional information, if the learner is given adequate overlap inside each subcategory and adequate non-overlap between subcategories.

Method

Participants 24 monolingual native English-speaking students at the University of Rochester participated in Experiment 5 (12 in each language). Subjects had not participated in any other categorization experiment and were paid for their participation.

Stimulus Materials Experiment 5 utilized the same grammar as in Experiments 1-4, but more words were added to the language in order to allow for a subcategory structure (*mib*, *bliffin*, *zemper*, *roy*, *nerk*, *prog*, and *dilba*). Categories Q and R still had 2 words each, but categories A and B had 6 words each, and category X had 4 words. A subcategory structure was devised such that $A_{1,2,3}$ and $B_{1,2,3}$ were only seen with $X_{1,2}$. $A_{4,5,6}$ and $B_{4,5,6}$ were only seen with $X_{3,4}$ (see Figure 3).

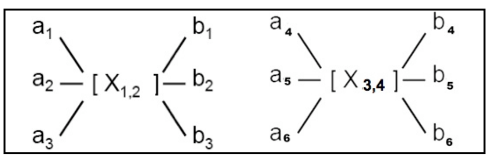

Figure 3: Subcategorization structure for Experiment 5.

In this language, there are 6x4x6=144 possible combinations of A, X, and B, but only 36 of those strings are legal according to the subcategory structure. Of those legal strings, 24 AXB combinations were presented during exposure and 12 AXB combinations were withheld. Optional Q and R elements were applied as in previous experiments, to create a training set of 96 strings. The sparseness and overlap within each subcategory were proportional to the sparseness and overlap of Experiment 1. Pilot testing revealed that keeping exposure to 20 minutes (similar to Experiment 1) did not lead to systematic learning of the language (this is unsurprising given that the language is much larger). Therefore, exposure was increased to about 45 minutes (5 times through the training set).

The test stimuli were comprised of 12 grammatical familiar items, 12 grammatical novel items, 12 ungrammatical AXA or BXB items, and 12 ungrammatical subcategory violation items. The subcategory violation items had either the A word or the B word from the opposite subcategory as the X item. Crucially, the subcategory violation items would be grammatical if learners ignored the subcategory structure of the language and generalized to form a single X category. A difference in ratings between grammatical items and subcategory violation items therefore indicates that participants have learned the subcategories in the language and are not generalizing across the gaps created by the subcategory structure.

Procedure The procedure was the same as Experiment 1.

Results

A repeated measures ANOVA revealed no differences between language one and two (F<1), so the two languages were combined. The mean rating of grammatical familiar items was 3.61 (SE=0.1), the mean rating of grammatical novel items was 3.7 (SE=0.11), the mean rating of subcategory violation items was 3.31 (SE=0.12), and the mean rating of ungrammatical items was 2.55 (SE=0.12). Grammatical familiar and grammatical novel items were not significantly different from each other ($F(1,22)=1.559$, $p>0.1$). However, subcategory violation items were rated significantly lower than grammatical items ($F(1,22)=11.698$, $p<0.01$). Ungrammatical items were rated the lowest, significantly lower than subcategory violation items ($F(1,22)=19.648$, $p<0.001$).

Discussion

Once again, learning effects were observed based solely on distributional cues to subcategory structure. While the subcategorization results are weaker than the categorization results (as shown by the significant difference between subcategory violation items and ungrammatical items), it is important to keep in mind that this task involves a conflict of cues. The subcategory problem has an important distributional property that differentiates it from a single category problem: in the subcategory case, some of the distributional cues (e.g., word order) signal that there is only one category, while other distributional cues (A and B context words) signal that there is subcategorization within this larger category. Not only must the learner figure out that there are categories, as in Experiments 1-4, but now the learner must also decide which gaps are systematic (the gaps that create the subcategory structure) and which are accidental (the gaps that are legal but withheld items).

General Discussion

Across five experiments, we observed robust evidence that learners can extract the category and subcategory structure of an artificial language based solely on the distributional patterning of the words and their surrounding contexts. We saw no great difference between Experiments 1 and 2 when only the number of contexts differed, but not the overlap in contexts across words. However, learners began to reduce their likelihood of generalizing (that is, increased the difference in their ratings for familiar versus unfamiliar grammatical sentences) when the overlap in contexts was reduced. Furthermore, they restricted generalization quite sharply in Experiment 4, when the same exposure corpus (and its gaps) was repeated. These results show that adult learners can skillfully use the data in the input to determine whether to ignore gaps in the input or whether to generalize over them. Participants in these experiments were able to take account of a rich set of variables to aid them in this task – degree of overlap among category members, amount of input, consistency of gaps and overlaps, and conflicts or consistency among cues.

These results also highlight some types of information that learners might be encoding or computing during learning and other types that they do not appear to be relying on. If learners were encoding the full set of exposure sentences, or the trigrams or quadrigrams (e.g., AXB, AXBR) and their frequencies of occurrence during exposure, they could discriminate between the familiar and novel grammatical sentences in all three experiments. In contrast, if they were only keeping track of simple word frequencies, they would fail in all experiments, since these are carefully controlled. The results suggest that learners are keeping track of word co-occurrences at a mid-sized grain, such as bigram frequencies or probabilities (e.g., AX, XB). Alternatively, they could be keeping track of the network of occurring contexts for individual words (as in Figures 2 and 3) and collapsing the individual words into a category when these networks bear enough quantitative as well as qualitative similarities to one another.

This process can be idealized in terms of a Bayesian model estimating whether sample data are drawn from one hypothesis space or another. But there are potentially a number of models, in addition to a Bayesian model, that could simulate such results, and we are in the process of testing which types of models perform as well as actual human learners.

Another question raised by these results is whether infants and young children coordinate multiple variables as adults do. We are in the process of testing child learners to determine how they weigh the large number of variables involved in forming categories in these tasks. One possibility is that young children are as skillful as adults at weighing variables to decide how to generalize. Another possibility is that they are more likely to follow one or a few of these variables only (as found in related studies), or that they are more likely overall to generalize than adults are, regardless of the input.

These experimental results suggest that the number of categories and their functional roles in a grammar are determined, at least in part, by a form of constrained statistical learning. The patterning of tokens in a substantial corpus of linguistic input appears to be sufficient, with a small set of learning biases, to extract the underlying structural categories in a natural language. At the same time, we expect, along with other researchers (cf. Monaghan, Chater & Christiansen, 2005), that distributional variables combine with other types of information in natural language acquisition, and that the integration of multiple imperfect and uncertain cues – including the distributional ones we have studied here – can serve to help learners determine when to generalize and when to restrict generalization in a complex problem space.

Acknowledgments

This research was supported by NIH Grants HD037082 to RNA and DC00167 to ELN, and by an ONR Grant to the University of Rochester.

References

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.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Frigo, L., & McDonald, J.L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory & Language, 39,* 218-245.

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language, 32,* 249-268.

Gomez, R., & Gerken, L.A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4*, 178-186.

Maratsos, M., & Chalkley, M.A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. Nelson (Ed.) *Children's language, Vol 2*. New York: Gardner Press.

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

Mintz, T.H., Newport, E.L., & Bever, T.G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science, 26,* 393-424.

Monaghan, P., Chater, N., & Christiansen, M.H. (2005). The differential role of phonological and distributional cues in grammatical categorization. *Cognition, 96*, 143-182.

Wonnacott, E., Newport, E.L. & Tanenhaus, M.K. (2008). Acquiring and processing verb argument structure: distributional learning in a miniature language. *Cognitive Psychology, 51*, 165-209.

Xu, F. & Tenenbaum, J.B. (2007) Word learning as Bayesian inference. *Psychological Review, 114*, 245-272.