Novel Words in Novel Contexts: The Role of Distributional Information in Form-class Category Learning

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

One major aspect of successful language acquisition is the ability to organize words into form class categories and generalize from properties of experienced items to novel items. Furthermore, learners must often determine how to use a new word, when there is very sparse information regarding its acceptable contexts. In this work we employ an artificial language learning paradigm to explore how adult learners, under circumstances of varying distributional cues to category boundaries, apply their knowledge of category properties to a new word. We find that in cases of strong category cues and strong category learning, adults readily generalize all of the distributional properties of the learned category to a word that shares just one context with the other category members. However, as the distributional cues regarding the target category become sparser and contain more systematic gaps, learners show more conservatism in generalizing the allowable distributional properties to the novel word. Taken together, these results show striking flexibility in learners' tendency to generalize, depending on the distributional properties of the input corpus, in a probabilistically rational way.

Introduction

The problem that learners face when they attempt to categorize items in the environment is deciding when they should treat instances as a category (thus generalizing from properties of experienced items to novel ones) and when they should treat instances separately (with no generalization from properties of experienced items to predicted properties of novel items). This problem cannot always be solved on the basis of perceptual similarity, as membership in some categories is independent of the surface features of the members.

The acquisition of grammatical categories is an example of this type of problem, but has some additional complicating factors. We hear individual words in a limited number of specific contexts. However, the rules that languages are built on involve patterns defined over categories of words, not the individual words themselves. Language input is serially presented, so we need to predict the proper contexts for words we have not yet heard. Furthermore, learners never see the entire input corpus, so they must figure out the proper contexts for new words, keeping in mind that sometimes there are lexically specific restrictions on words (such as *give* versus *donate*: despite similar meaning, Joe can *give David a book,* but Joe cannot **donate David a book*). In acquiring grammatical categories, the learner must ask whether contexts are absent by accident, or because they are ungrammatical. This question is particularly difficult to resolve when a new item is encountered in a single context and therefore overlaps only minimally with previously encountered words. For example, consider hearing the sentence: *I remembered to <u>nerk</u> yesterday*. Should one generalize from this context to another context where words of the category 'verb' are grammatical, such as *She will make him <u>nerk</u> tomorrow*, or *I saw the cat <u>nerk</u> earlier*?

One hypothesis about how learners handle this situation is that they have innately defined linguistic categories with featural and contextual information predefined, so that minimal exposure to language is needed to sort out which words belong to each category (e.g., McNeill, 1966). Another hypothesis is that learners use semantic categories to bootstrap the syntactic categories (e.g., Grimshaw, 1981). A third possibility is that learners exploit distributional information in the input to discover the category structure of natural languages (e.g., Braine, 1987). This third hypothesis is what we investigate in the present experiments.

A number of researchers have asked whether there is adequate distributional information in the input to form linguistic categories. This work uses hierarchical clustering and a computational learning mechanism to attempt to deduce grammatical categories from corpora of child-directed speech based solely on distributional analyses of the input (e.g., Mintz, Newport, & Bever, 2002; Redington, Chater, & Finch, 1998). These models have been able to use co-occurrence statistics among words to achieve relatively good categorization performance for frequent target words. To explore whether human learners can actually use this information during language learning, Mintz (2002) tested categorization in an artificial language learning environment, showing evidence that learners did engage in distributional analyses of the input in order to generalize their knowledge of previously encountered strings to grammatical novel strings. Hunt and Aslin (2010) showed that adults could learn categories embedded in sequences of visual symbols during a serial reaction time task when the only cue to category structure was distributional information among the symbol strings.

Building off of these findings on the importance of distributional information for category formation, we have proposed a systematic set of computational variables that can explain the types of distributional information that are important for categorization. Deciding whether to generalize across words or preserve lexical specificity appears to be determined by (at least) 3 distributional variables: the *number* of linguistic contexts in which each word in the input set occurs, the *density* or proportion of these contexts that are present in the input, and the degree of *overlap* of contexts across words. In previous work (Reeder et al., 2009) we showed that learners are remarkably sensitive to these cues, which interact with each other to determine how basic category and subcategory structure are acquired. To do this, we manipulated the distribution of contexts for a target category in the exposure set to examine how adults determine when to generalize (deciding whether gaps in their input are accidental or systematic). When participants were exposed to a dense sampling of the language where there was rich coverage of contexts for a target category and high overlap in contexts across words, adult learners showed complete generalization to all possible grammatical contexts, even those that were never heard before for particular words. But as the input to the learner became more sparse with less overlap, participants became more conservative in their generalizations. Furthermore, as we increased the frequency of recurring gaps in the input, participants became more certain that the gaps were not accidental but rather part of the structure of the language, and they decreased their generalizations to unseen grammatical contexts. In the present work we ask how, under these same varying circumstances of category strength and category learning, learners will extend their knowledge of the target category to a *novel* word, one for which they have only minimal context information. In particular, is there a point in category learning where hearing one context for a novel word is enough to obtain full category privileges for that word? Or does every novel word need to be heard in a number of overlapping contexts in order to be treated as a member of the category?

Experiment 1

In Experiment 1 of Reeder et al. (2009), the learner was exposed to a very dense sampling of the language space, with all the words in the target category appearing in many highly overlapping contexts. Under these conditions, learners represented the words as a true category, generalizing fully across the gaps in the exposure corpus. In Experiment 1 we ask whether, under the same circumstances, the target category's distributional properties will also generalize to a novel word that they have only heard in a single context. The logic of this paradigm is that, if learners acquire a strong category (called X), then novel sentences which observe even a bit of the category structure of the language might be perceived to be just as grammatical (or familiar) as sentences that have actually been heard during training.

An artificial grammar with the structure (Q)AXB(R) was used, similar to that used in Reeder et al. (2009), where each letter represents a set of 2, 3, or 4 words. In Experiment 1, the Q and R categories had 2 words each, the A and B categories had 3 words each, and the X category had 4 words. The words of the grammar were *spad, klidum, flairb, daffin, glim, tomber, zub, lapal, fluggit, mawg, bleggin, gentif, frag,* and *sep.* The words were not mapped on to any referential world, so there were no semantic cues to categorization. All studies were run with two languages that differed only in which words were assigned to each of the categories in the language, to ensure that obtained results were not due to coincidental preferences for specific sound combinations. As in Reeder et al. (2009), X was the target category of interest, A and B were "context" categories that formed the distributional cues to the category X, and Q and R were optional flanker categories that allowed strings to range from 3 to 5 words in length.

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

$A_1X_1B_1$ *	$A_1X_2B_1$	$A_1X_3B_1$ * ♦ ○	$A_1X_4B_1$ * ♦ ○
$A_1X_1B_2$	$A_1X_2B_2$ * ♦	$A_1X_3B_2$ * ○	$A_1X_4B_2$
$A_1X_1B_3$ * ♦ ○	$A_1X_2B_3$ *	$A_1X_3B_3$	$A_1X_4B_3$
$A_2X_1B_1$	$A_2X_2B_1$ * ♦ ○	$A_2X_3B_1$ *	$A_2X_4B_1$
$A_2X_1B_2$ * ♦ ○	$A_2X_2B_2$ *	$A_2X_3B_2$	$A_2X_4B_2$
$A_2X_1B_3$ * ○	$A_2X_2B_3$	$A_2X_3B_3$ * ♦	$A_2X_4B_3$
$A_3X_1B_1$ * ♦	$A_3X_2B_1$ * ○	$A_3X_3B_1$	$A_3X_4B_1$
$A_3X_1B_2$ *	$A_3X_2B_2$	$A_3X_3B_2$ * ♦ ○	$A_3X_4B_2$
$A_3X_1B_3$	$A_3X_2B_3$ * ♦ ○	$A_3X_3B_3$ *	$A_3X_4B_3$

Method

Participants 16 monolingual native English-speaking students at the University of Rochester participated in Experiment 1, eight in each of the two languages created by different assignments of words to categories. Subjects had not participated in any other categorization experiment and were paid for their participation.

Stimulus Materials Of the possible 36 AXB sentence types in the language, 19 were presented to participants, and the remainder were withheld for testing generalization (see Table 1). The presence of the 2 Q and 2 R words was varied evenly such that the exposure set was expanded to 76 possible (Q)AXB(R) sentences. The exposure set contained only four X_4 strings: $A_1X_4B_1$, $Q_1A_1X_4B_1$, $A_1X_4B_1R_1$, and $Q_2A_1X_4B_1R_2$, which presented the X_4 word in only one context ($A_1X_4B_1$); the remaining 72 sentences included equal numbers of sentences containing X_1, X_2, and X_3. Training consisted of 4 times through this exposure set, forming 22 minutes of exposure. Importantly, every X_1, X_2, and X_3 was seen with every A and every B word, but X_4 was only seen in one context. Thus, the training set for Experiment 1 was dense for X_1-X_3 such that participants were exposed to a high proportion of the possible strings for those three X words, but very sparse for X_4. Additionally, there was complete overlap of contexts among X_1, X_2, and X_3, but X_4 shared only one context with X_1-X_3.

A female native English speaker recorded the words in isolation with both non-terminal and terminal intonation. Words were then adjusted in Praat such that pitch, volume, and duration were roughly consistent. Sentences were constructed by splicing words sequences in Sound Studio such that all words except the last had non-terminal intonation, with 50ms silence between each word. The final word in each sentence had terminal intonation contour. The order of sentences in the exposure set was randomized for each subject and presented via a custom software package on a Dell PC. Each sentence was separated by 1.5s of silence. Participants wore headphones and passively listened to the exposure sentences during training.

Immediately after exposure, participants heard a series of test strings and were asked to rate each on a scale from 1 to 5, where 1 meant it definitely did *not* come from the language they were exposed to, and 5 meant it definitely *did* come from the exposure language. All test strings were 3-word sentences of one of the following forms: a grammatical familiar string (10 AXB strings presented during training), a grammatical novel string (13 AXB strings withheld during training), or an ungrammatical string (of the form AXA or BXB). Of the grammatical novel test strings, 4 of the 13 were strings testing generalization of X_4: $A_2X_4B_2$, $A_2X_4B_3$, $A_3X_4B_2$, and $A_3X_4B_3$. With these strings we can ask whether learners have generalized X_4 to the full range of grammatical contexts for X-words, judging the familiar and novel grammatical sentences for X_4 to be equivalent, even though they have only seen X_4 in one of these contexts. These strings can then be compared to the 6 ungrammatical strings that contain X_4 (3 AX_4A, 3 BX_4B).

Results

A repeated measures ANOVA with condition (familiar, novel, ungrammatical) as the within subjects factor and language as the between subjects factor showed no significant effects of language ($F<1$). For test items without X_4, the mean rating of grammatical novel strings was 3.87 (SE=0.14), the mean rating of grammatical familiar strings was 3.85 (SE=0.13), and the mean rating of ungrammatical strings was 2.89 (SE=0.15). We found no significant difference between ratings of grammatical novel items and grammatical familiar items ($F(1,14)=0.24$, $p=0.63$). These items were rated significantly higher than ungrammatical test strings ($F(1,14)=26.40$, $p<0.005$). For the test items that contained X_4, the mean rating of grammatical novel strings was 3.28 (SE=0.18), the mean rating of grammatical familiar strings was 3.59 (SE=0.24), and the mean rating of ungrammatical strings was 2.61 (SE=0.21). These items showed the same pattern as the without-X_4 items: there was no significant difference between ratings of grammatical novel X_4 items and familiar X_4 items ($F(1,14)=1.71$, $p=0.21$), however there was a significant difference between these items and ungrammatical X_4 strings ($F(1,14)=13.10$, $p<0.01$). [1]

Discussion

As in Reeder et al. (2009), learners strongly preferred familiar and novel grammatical sentences to ungrammatical sentences. Learners also showed generalization to the novel grammatical X_4 strings, but not to the ungrammatical X_4 strings. Thus they generalized X_4 to the full range of grammatical contexts for X words, even though they heard X_4 in only one of these contexts. These results show that, when learners are exposed to a dense sampling of the language space for words in the target category (X_1-X_3) and presented with many overlapping contexts, they generalize their knowledge within the category X_1-X_3 and also extend it to X_4. Importantly, the generalized contexts are novel contexts for X_4, but are strongly represented by the learner's exposure to the permissible contexts for X_1-X_3. Learners did not require semantic or perceptual cues to indicate that the X words form a category.

Experiment 1 provided the learner with a dense sampling of the language space for most of the words in the target category. In the remaining experiments we systematically manipulated the density, overlap, and number of contexts for X_1-X_3 in the exposure set while restricting exposure to contexts for X_4, in order to explore the impact of these distributional variables on the generalization of category knowledge.

Experiment 2: Sparseness

In Experiment 2, we decrease the density of the contexts for X_1-X_3 words, but we keep the number and overlap among X_1-X_3 contexts the same. We still present only one context for X_4 and explore what the increase in sparseness for X_1-X_3 does to learners' generalizations to the novel X_4 item.

Method

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

Stimulus Materials The strings of the language were constructed in the same manner as Experiment 1, with two languages that had different assignments of words to categories. Here, however, the exposure set contained only 10 (versus 19 in Exp. 1) of the 36 possible AXB combinations (see Table 1). As in Experiment 1, every X_1-X_3 word was heard in combination with every A and every B. With the addition of AXB strings with optional Q and R

[1] We did not compare ratings of the X_1-X_3 test items with the X_4 items because of the lower statistical power of the X_4 means. For all experiments, we take the pattern of learning for familiar and novel grammatical items to be more informative than the size of the differences between X_1-X_3 and X_4.

flanker words, there were 40 sentences in the exposure set. The exposure set was repeated 4 times through so that each sentence type was presented with the same frequency as in Experiment 1, for an exposure of about 12 minutes. The test phase was the same as described for Experiment 1.

Procedure The procedure was the same as in Experiment 1.

Results

A repeated measures ANOVA with condition as the within subjects factor and language as the between subjects factor revealed no difference between the two languages (F<1). For test items without X_4, the mean rating of grammatical novel strings was 3.55 (SE=0.09), the mean rating of grammatical familiar strings was 3.54 (SE=0.10), and the mean rating of ungrammatical strings was 2.63 (SE=0.14). Just as in Experiment 1, as well as Experiments 1 and 2 from Reeder et al. (2009), we found no significant difference between ratings of grammatical novel items and grammatical familiar items without X_4 (F(1,14)=0.008, p=0.93), but grammatical sentences were rated significantly higher than ungrammatical test strings (F(1,14)=25.37, p<0.001). For the test items that contained X_4, the mean rating of grammatical novel strings was 3.27 (SE=0.15), the mean rating of grammatical familiar strings was 3.53 (SE=0.22), and the mean rating of ungrammatical strings was 2.55 (SE=0.16). This is the same trend as demonstrated by the without-X_4 items and the analyses in Experiment 1. While there was a significant difference between grammatical X_4 strings and ungrammatical X_4 strings (F(1,14)=9.87, p<0.01), there was no significant difference between ratings of grammatical novel X_4 items and familiar X_4 items (F(1,14)=1.59, p=0.23).

Discussion

These results mirror those in Experiment 1, demonstrating that reduced density does not greatly affect learners' performance when there is full overlap of contexts among X_1-X_3 words. The generalization to X_4 is maintained despite greatly reduced exposure due to a sparser sampling of the language space. We next explore how learners behave when there is reduced overlap of X_1-X_3 word contexts.

Experiment 3: Overlap

Similar to Experiment 2, we present the learner with only 10 of the 36 possible AXB combinations. However, in order to test how overlap in contexts influences generalization of category knowledge to new X-words, we now reduce the overlap of contexts among members of X_1-X_3. Individual X-words do not fully share all of their contexts with other X-words, though the set of X-words as a whole occurs in all A and B contexts. By reducing the overlap in contexts across X words, we can assess the degree to which learners restrict generalization within X_1-X_3, and also how they extend the category knowledge to X_4.

Method

Participants 16 monolingual native English-speaking students at the University of Rochester participated in Experiment 3, eight in each of the two possible languages. Participants had not been in any other categorization experiment and were paid for their participation.

Stimulus Materials Strings were assembled in the same way as Experiment 1, with two languages that had different assignments of words to categories. Exposure consisted of only 10 of the 36 possible AXB combinations, as in Experiment 2; however now X_1, X_2, and X_3 were heard with 2 of the 3 A-words and 2 of the 3 B-words each. X_1 occurred with A_1, A_2, B_1, and B_2, but not A_3 or B_3; X_2 was heard with A_2, A_3, B_2, and B_3, but not A_1 or B_1; X_3 was heard with A_1, A_3, B_1, and B_3, but not A_2 or B_2. Thus, the overlap among contexts is maintained over the X_1-X_3 category as a whole, but individual X-words 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 every A and every B. X_4 was still only seen with one context (see Table 1).

Procedure The procedure was the same as in Experiment 1.

Results

A repeated measures ANOVA with condition as the within subjects factor and language as the between subjects factor showed no significant difference between the two languages (F<1). For test items without X_4, the mean rating of grammatical novel strings was 3.71 (SE=0.12), the mean rating of grammatical familiar strings was 3.91 (SE=0.09), and the mean rating of ungrammatical strings was 2.55 (SE=0.15). Unlike Experiments 1 and 2, but in line with results from Reeder et al. (2009), we found significant differences between ratings of grammatical novel items and grammatical familiar items (F(1,14)=9.12, p<0.01). Additionally, both of these items were rated significantly different from ungrammatical test strings (F(1,14)=26.82, p<0.001). For the test items that contained X_4, the mean rating of grammatical novel strings was 3.25 (SE=0.16), the mean rating of grammatical familiar strings was 3.66 (SE=0.24), and the mean rating of ungrammatical strings was 2.21 (SE=0.16). Unlike the without- X_4 items, we do not see any significant difference between novel grammatical X_4 strings and familiar X_4 strings (F(1,14)=2.98, p=0.11), perhaps due to the lower statistical power for these test items; there is still a significant difference between ratings of grammatical and ungrammatical X_4 items (F(1,14)=26.21, p<0.001).

Discussion

In Experiment 3, we reduced the overlap among contexts in the exposure set by a third, but we kept the number of contexts in the input the same as in Experiment 2. The results indicate that despite full coverage of contexts across lexical items, the incomplete overlap between X_1-X_3-words led to decreased generalization. However, learners still showed a much higher rating for grammatical novel items than ungrammatical items, indicating that they were still willing to generalize, though more conservatively than in Experiments 1 and 2. Additionally, learners were much less likely to generalize their knowledge of grammatical X_1- X_3 contexts to X_4 given the systematic gaps in the Experiment 3 exposure set. Thus, as we move along the dimensions of sparseness and overlap explored in Experiments 2 and 3, we can see how learners weigh the likelihood that X_4 shares the same contexts as X_1- X_3 and use this as a diagnostic for how strongly the X category has been formed.

Experiment 4: Overlap with extended exposure

The decision to generalize over a gap in the input or maintain lexical distinctness may also be influenced by the frequency of contexts (and gaps) in the input. If a context is consistently absent as in Experiment 3, learners start to show conservatism in their generalizations. If this gap is made even more prominent by creating an exposure set that has repeated instances of sparse contextual information, learners might develop even more certainty that gaps in the input are systematic and not accidental (e.g., Wonnacott, Newport & Tanenhaus, 2008; Xu & Tenenbaum, 2007). This will be particularly important with regard to X_4, where we can explore how an increase in the exposure to the one context for X_4 (and potentially a perceived increase also in the gaps at the non-occurring contexts for X_4) affects how learners generalize their knowledge of the category X_1- X_3. If the category X_1- X_3 is strongly defined (as in Experiment 1), we would expect that a very large increase in frequency of the one context of X_4 (and perceived increase in exposure to gaps for X_4) might be required before there is a decrease in generalization and a lessening of X_4 membership in the X-word category. However, if the X-category is weakly defined as in Experiment 3, the small increase in the number of repetitions in Experiment 4 might be enough to make learners conservative in their generalizations.

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. Participants had not been in any other categorization experiment and were paid for their participation.

Stimulus Materials The language was the same as in Experiment 3, except that exposure to the language was tripled by presenting the corpus 12 times rather than 4. Training lasted for approximately 22 minutes (as in Experiment 1), but contained only 10 contexts (as in Experiments 2 & 3). Test strings were the same as in Experiment 3.

Procedure The procedure was the same as in Experiment 1.

Results

A repeated measures ANOVA with condition as the within subjects factor and language as the between subjects factor showed no significant difference between the two languages (F<1). For test items without X_4, the mean rating of grammatical novel strings was 3.86 (SE=0.12), the mean rating of grammatical familiar strings was 4.05 (SE=0.10), and the mean rating of ungrammatical strings was 2.61 (SE=0.21). These results show a significant difference between ratings of grammatical novel items and grammatical familiar items ($F(1,14)=8.60$, $p=0.01$). Additionally, these items were rated significantly higher than ungrammatical test strings ($F(1,14)=35.83$, $p<0.001$). For the test items that contained X_4, the mean rating of grammatical novel strings was 3.44 (SE=0.19), the mean rating of grammatical familiar strings was 4.06 (SE=0.21), and the mean rating of ungrammatical strings was 2.37 (SE=0.21). Similar to the without-X_4 items, we now find a significant difference between novel grammatical X_4 strings and familiar X_4 strings ($F(1,14)=8.33$, $p=0.011$), along with a significant difference between these and ungrammatical X_4 items ($F(1,14)=31.04$, $p<0.001$).

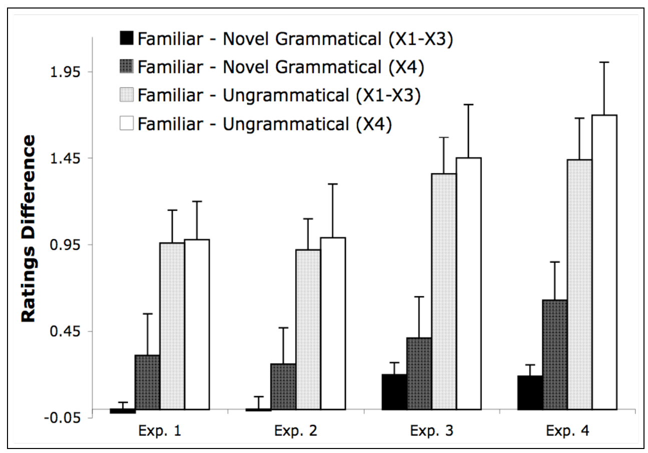

Figure 1: Experiment 1-4 difference scores of ratings of grammatical familiar items and grammatical novel items (for X_1-X_3 words and X_4), and grammatical familiar items and ungrammatical items (for X_1-X_3 words and X_4).

Discussion

These results indicate that, when we increase exposure to the same sparse data (with recurring gaps that may also become more prominent), learners act rationally and are even less likely to generalize over such gaps. Furthermore, learners apparently view the category formed by X_1- X_3 as weakly defined due to the sparse sampling of the language and incomplete overlap among words, which also seems to increase learners' uncertainty about the status of the withheld grammatical X_4 items. While we still see that novel grammatical test strings are judged more grammatical than the ungrammatical strings, we hypothesize that increasing exposure to the sparse input set even longer might push learners to judge all novel items as ungrammatical. In contrast, if we increased the number of unsystematic gaps in the input, we expect that learners would show more generalization, especially for the X_4 word.

General Discussion

The present experiments add grammatical category learning to a large literature showing that learners are highly sensitive to many types of distributional information in their input. We have replicated Experiments 1-4 of Reeder et al. (2009), demonstrating that learners are able to extract the category structure of an artificial language based on distributional information alone, and we show that learners are quite rational, statistically speaking, in how much and when they generalize across gaps in the input. Importantly, the current experiments also show that learners can skillfully transfer their knowledge of category structure and category cues to a novel item that is only weakly represented in the input. When given a dense sampling of the language space with almost complete overlap of contexts for many words in a target category X, learners generalize a novel word (X_4) to the full range of grammatical contexts of the other X-words, even when they have only seen X_4 in one of those contexts. This willingness to add X_4 to the strongly established X_1- X_3 category is strongest when the X_1- X_3 contexts are dense and overlapping; when contexts are more sparse and less overlapping across different X words, we also see more conservative generalization to a new X_4 word. The most extreme case is when we increase the number of times the learner hears the sparse exposure set, thus increasing also the frequency of recurring gaps in the input for X_1- X_3: learners in this situation rate the withheld X_4 contexts as more unfamiliar, while rating as highly familiar only the one context in which X_4 was actually heard. These findings are in line with results from Wonnacott, Newport and Tanenhaus (2008) in the area of verb-argument learning, where if the language is generally lexically specific, participants do not show generalization of the minimal exposure item (i.e., X_4) to other contexts. In contrast, if the language has the same contexts permitted for all verbs, then participants show strong generalization for the minimal exposure item.

We are in the process of modeling these results to determine the type of information learners might encode in order to accomplish these outcomes; storing any simple statistics – such as word, bigram, or trigram frequencies – would not be adequate to account for generalization to the novel X_4 strings. Instead, learners must be forming a more abstract representation of the data in order to generalize their knowledge to novel strings.

In contrast to our experiments, as learners face the problem of inferring category membership from sparse and incomplete data in natural languages, there are a number of correlated cues that they could use to help them extract category information, such as phonological, prosodic, or semantic cues as well as distributional cues. Indeed, many studies have shown that category learning is enhanced when category membership is correlated with such surface cues (e.g., Monaghan, Chater, & Christiansen, 2005). But an important question in this literature has been whether category learning can utilize distributional information, either alone or when very poorly correlated with other cues. While natural languages do sometimes contain multiple cues to grammatical categories, our work indicates that learners are able to skillfully employ a statistical learning mechanism as a primary tool with which to extract category information from the input, even in cases where other correlated cues are incomplete or absent.

Acknowledgments

We would like to thank Josh Tenenbaum for valuable discussions of this work. 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.* (pp. 65-87). Hillsdale, NJ: Lawrence Erlbaum Associates.

Grimshaw, J. (1981). Form, function, and the language acquisition device. In C.L. Baker and J.J. McCarthy (Eds.), *The logical problem of language acquisition*. (pp. 183-210). Cambridge, MA: MIT Press.

Hunt, R.H., & Aslin, R.N. (2010). Category induction via distributional analysis: Evidence from a serial reaction time task. *Journal of Memory and Language*, *62,* 98-112.

McNeill, D. (1966). Developmental Psycholinguistics. In F. Smith & G. Miller (Eds.), *The Genesis of Language: A Psycholinguistics Approach* (pp. 69-73). Cambridge, MA: MIT Press.

Mintz, T.H. (2002). Category induction from distributional cues in an artificial language. *Memory and 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*, 393-425.

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

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22,* 435-469.

Reeder, P.A., Newport, E.L., & Aslin, R.N. (2009). The role of distributional information in linguistic category formation. In N.A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2564-2569). Austin, TX: Cognitive Science Society.

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.