Language Learning and Development

ISSN: 1547-5441 (Print) 1547-3341 (Online) Journal homepage: www.tandfonline.com/journals/hlld20

The Effect of Zipfian Frequency Variations on Category Formation in Adult Artificial Language Learning

Kathryn D. Schuler, Patricia A. Reeder, Elissa L. Newport & Richard N. Aslin

To cite this article: Kathryn D. Schuler, Patricia A. Reeder, Elissa L. Newport & Richard N. Aslin (2017) The Effect of Zipfian Frequency Variations on Category Formation in Adult Artificial Language Learning, Language Learning and Development, 13:4, 357-374, DOI: 10.1080/15475441.2016.1263571

To link to this article: https://doi.org/10.1080/15475441.2016.1263571

Published online: 02 Aug 2017.

Submit your article to this journal

Article views: 457

View related articles

View Crossmark data

Citing articles: 9 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=hlld20

LANGUAGE LEARNING AND DEVELOPMENT
2017, VOL. 13, NO. 4, 357–374
https://doi.org/10.1080/15475441.2016.1263571

Routledge
Taylor & Francis Group

Check for updates

The Effect of Zipfian Frequency Variations on Category Formation in Adult Artificial Language Learning

Kathryn D. Schuler[a], Patricia A. Reeder[b], Elissa L. Newport[a], and Richard N. Aslin[c]

[a]Center for Brain Plasticity and Recovery, Department of Neurology, Georgetown University; [b]Department of Psychological Science, Gustavus Adolphus College; [c]Department of Brain & Cognitive Sciences, University of Rochester

ABSTRACT
Successful language acquisition hinges on organizing individual words into grammatical categories and learning the relationships between them, but the method by which children accomplish this task has been debated in the literature. One proposal is that learners use the shared distributional contexts in which words appear as a cue to their underlying category structure. Indeed, recent research using artificial languages has demonstrated that learners can acquire grammatical categories from this type of distributional information. However, artificial languages are typically composed of a small number of equally frequent words, while words in natural languages vary widely in frequency, complicating the distributional information needed to determine categorization. In a series of three experiments we demonstrate that distributional learning is preserved in an artificial language composed of words that vary in frequency as they do in natural language, along a Zipfian distribution. Rather than depending on the absolute frequency of words and their contexts, the conditional probabilities that words will occur in certain contexts (given their base frequency) is a better basis for assigning words to categories; and this appears to be the type of statistic that human learners utilize.

Introduction

Grammatical categories serve as the foundation of natural language structure. An essential part of natural language acquisition involves determining the number of grammatical categories, assigning words to these categories, and learning the rules for combining these categories to produce and comprehend grammatical utterances. A number of hypotheses have been proposed to explain the ease with which children accomplish this seemingly complex task. Some accounts claim that syntactic categories must be innately defined (e.g., Chomsky, 1965; McNeill, 1966), while others suggest that categories must be acquired from cues in the language input (e.g., semantic bootstrapping: Pinker, 1984, 1987; constructivist accounts: Tomasello, 2003). In either case, precisely how learners determine the mapping between individual words and the underlying grammatical categories remains unclear.

Distributional information is one cue in the language input that has been proposed as the solution to this mapping problem (e.g., Harris, 1954; Maratsos & Chalkely, 1980). On this account, learners use the fact that words of the same syntactic category tend to appear in highly overlapping distributional contexts as a cue to infer the category structure of the language. There is a large literature, using both corpus analyses and artificial grammar learning paradigms, demonstrating the availability and utility of such distributional information for syntactic categorization (e.g., Cartwright & Brent, 1997; Mintz, 2002, 2003; Mintz, Newport, & Bever, 2002; Redington, Chater, & Finch, 1998; Reeder, Newport, & Aslin, 2013).

CONTACT Kathryn D. Schuler kds71@georgetown.edu Department of Neurology, Georgetown University, Washington DC 20057, USA.

© 2017 Taylor & Francis Group, LLC

While these findings make important contributions toward our understanding of how categories may be acquired from distributional information, it is not yet known whether these results will scale up to natural language input. Natural languages differ from the artificial languages used in these studies in a number of important ways, including the way word frequencies are distributed. Most experimental demonstrations of categorization from distributional information rely on artificial languages with carefully balanced word frequencies within and across categories to eliminate the possibility that learners will rely on extremely superficial statistics to extract categories from the input (e.g., lexical bigram frequencies). In contrast, however, word frequencies in natural languages are known to follow a Zipfian distribution, in which a small number of words occur with very high frequency (e.g. *boy, car*), while many words occur at much lower frequencies (e.g. *filibuster*) (Zipf, 1965).

The implications of a Zipfian distribution would be unimportant if learners' sensitivity to frequency were coarse. However, research on child language acquisition and on adult sentence processing has demonstrated that comprehension, production, and learning are all sensitive to lexical frequency and to the frequency with which words occur in various sentential contexts (e.g., Blackwell, 2005; Goodman, Dale, & Li, 2008; Harris, Barrett, Jones, & Brookes, 1988; Holmes, Stowe, & Cupples, 1989; Kidd et al., 2006; Lapata et al., 2001; Naigles & Hoff-Ginsberg, 1998; Roy, Frank, & Roy, 2009; Schwartz & Terrell, 1983; Theakston, Lieven, Pine, & Rowland, 2004; Trueswell, Tanenhaus, & Kello, 1993). Thus whatever the mechanism for acquiring grammatical categories, it must be robust to variations in word frequency.

There is some evidence to suggest that a distributional learning mechanism is not only sensitive to these frequency variations but may also benefit from them. For example, learners are better at acquiring categories (Valian & Coulson, 1988) and learning both adjacent (Kurumada, Meylan, & Frank, 2013) and non-adjacent dependencies (Gomez, 2002) from distributional cues if the structures they are learning contain some high frequency elements. Researchers suggest that high frequency elements may facilitate learning because they provide an additional distributional cue to the learner. A number of studies have also found that correlated cues are advantageous for distributional learners (e.g. semantic cues: Braine et al., 1990; morphological cues: Brooks, Braine, Catalano, Brody, & Sudhalter, 1993; phonological cues: Frigo & McDonald, 1998; Gerken, Gomez, & Nurmsoo, 1999; Gerken, Wilson, & Lewis, 2005; Monaghan, Chater, & Christiansen, 2005; Morgan, Shi, & Allopenna, 1996; Wilson, 2002; shared features: Gomez & Lakusta, 2004).

Still, the variations in absolute word frequency (or absolute bigram frequency) that natural languages contain are not always relevant to determining what lexical category a word belongs to. Functional elements, like *a* and *is*, differ dramatically in frequency from items in lexical categories, like *boy* and *car*, and learners may use this dramatic frequency difference, along with other cues such as their prosodic and distributional differences, to differentiate these two types of categories. However, lexical items within the same category and across different lexical categories also differ in frequency. How does the learner determine which of these differences are important for categorization? For example, though one might refer to *milk* often and *typhoid* only rarely, these words enjoy similar syntactic privileges that come from belonging to the category **Noun**. Furthermore, on a distributional learning account, learners must be able to apply knowledge they have acquired about syntactic categories from frequent words, which they have heard in many contexts, to novel words that they have heard in only a few contexts. Upon hearing the sentence *I have a mawg in my pocket*, learners must infer that other noun contexts are grammatical for this newly encountered word *mawg*, such as *There are three mawgs over there* and *That yellow mawg is nice*.

Even more complicating, one may not hear words in a particular context just because those words are rare (low frequency), or alternatively those contexts might be absent because they are ungrammatical for this lexical item. For example, even though it is grammatical to say *give a book to the library* and *donate a book to the library*, only *give the library a book* is grammatical; the analogous **donate the library a book* is ungrammatical.

How, then, do learners handle these variations in word frequency—and their accompanying variations in context statistics—as they acquire grammatical categories from distributional information? Psycholinguistic evidence shows that learners are sensitive to these frequency variations; but in the acquisition of grammatical categories, frequency variations are not necessarily a relevant cue to the underlying category structure of a language. How do learners preserve their sensitivity to word frequency variation while not being misled into putting words with different frequencies into distinct categories or using frequency differences to form too many (or too few) grammatical categories?

The primary objective of this article is to contribute to the literature on distributional learning as a mechanism for the acquisition of grammatical categories by examining how learning is affected by variations in word frequency that are modeled after those of natural languages. In a series of three artificial language learning experiments, we ask whether lexical frequency variation within a grammatical category affects learners' determination of the category to which these items belong or their ability to generalize category information to novel words. We provide evidence that, despite substantial word frequency variation in the language, learners can make use of distributional contexts to acquire a category and can use these contexts to determine when it is appropriate to extend category membership to a novel word and when it is not.

To address these questions, we adapt an artificial language paradigm designed by Reeder et al. (2013) and modified here to incorporate variable word frequency. Reeder et al. showed that learners can group nonsense words in this language into categories and can also generalize syntactic properties of the category to novel words, based on the degree to which the surrounding linguistic contexts for those words overlap. After exposure to sentences in the language, category membership for individual words in Reeder et al. was based on the number of surrounding lexical contexts for each word that were shared with other category members (which we call *overlap*) and on the probability with which the learner hears (or fails to hear) each of the particular word-context combinations. In these studies, bigram frequencies for words and their contexts were carefully balanced so that these frequencies alone could not be the basis for generalization across test item types and across experimental conditions. When exposed to a large sample of the possible sentences in the grammar (a *high density* sampling), learners collapsed words into categories and fully generalized novel syntactic contexts to the familiar words based on the category structure they had inferred. That is, learners extended full category privileges to all members of the category, thus generalizing beyond their input. When we reduced the number of surrounding contexts that were shared across lexical items, learners still generalized from familiar to novel contexts but were less confident about extending full category privileges to all members of the category. These results thus illustrate both the ability of learners to generalize and the distributional details on which such generalization is based. In the present research we use this paradigm to ask whether these same outcomes can be achieved when lexical frequency is imbalanced.

As we describe below, learners are exposed to a set of sentences from an artificial language. These sentences have no meaning and do not contain any other cues to the category structure of the language beyond the distributional contexts that words from the target category share. Crucially, to mirror frequency variation in natural language, the absolute frequency of words in the target category varies along a Zipfian distribution. When lexical frequency varies widely within and across categories, information about contexts for low frequency lexical items will be much more sparse than that for high-frequency lexical items. Under these circumstances, how will learners use frequency and consistency of contexts to make decisions regarding categorization and generalization? Words that occur at low frequencies overall, or at low frequencies in specific contexts, could indicate the presence of a separate category (thereby leading a learner to restrict generalization). Alternatively, their rarity could simply reflect the Zipfian distribution of words within categories (and should have no effect on generalization). Learners might overcome these variations by several methods. They might use the distributional information from the most frequent words in order to form a category, and then apply the full set of contexts associated with this category to other words that share some of these contexts, regardless of their frequency. Alternatively, they might compute the conditional probabilities with which words occur in each of their possible contexts, taking the overall frequency of the word as a baseline against which its occurrence in specific contexts is assessed. As a third alternative, when words are less frequent, learners might be less certain about their status within the category or about the category as a whole. This would lead to decreased generalization, either specifically for low frequency words or for all lexical items in the category.

Since words and contexts in real languages do indeed vary dramatically in frequency, studying the effects of such frequency variations is important for understanding how statistical learning works in such circumstances (see also Kurumada et al. (2013), who used Zipfian frequency variations in statistical learning of word segmentation and found that such variations can be advantageous). By introducing large lexical frequency variations in the input, we will not only explore how this impacts category formation and generalization; we will also test whether previous distributional learning results scale-up to more naturalistic input, and precisely how computations in natural language acquisition are adjusted for frequency variations.

Experiment 1

In Experiment 1 learners are exposed to a large sample of the possible sentences in the language (a *high density* sampling), and the contexts surrounding the words in the target category have a high degree of overlap. These factors should lead learners to conclude that the target words all belong to a single category and therefore to generalize to novel grammatical contexts (Reeder et al., 2013). However, in the present experiment, in contrast to previous research on category formation, the words of the language vary in their frequency of occurrence: they are divided evenly into high-frequency, mid-frequency, and low-frequency words. This results in the word sequences that form the contexts for these words (word bigrams) also being high, mid, or low frequency. This design thus allows us to ask whether the absolute frequency of individual words and their combinations with other words—or, rather, their patterns and probabilities of occurrence with other words, regardless of how frequent the word is—determines the formation of syntactic categories and the generalization of words to novel contexts.

We also will examine the extreme case in which a novel word appears very infrequently and only in a single linguistic context. This provides a particularly strong test of generalization: its membership in a category is supported by only a single familiar context, but other words in the target category, all of which occur much more frequently, also occur in the same context (as well as in others). Learners might either collapse this rare word into the category and extend to it all of the unattested contexts—in effect, interpreting the absence of these combinations in the input as due to the overall low frequency with which the word occurs—or maintain this rare word as a lexical exception, in light of the missing information about the contexts in which it can occur. By examining how learners interpret this minimally overlapping word, we can better understand the use of lexical frequency and contextual probability in generalization.

Recall that, to be successful in acquiring categories from distributional information, learners must be sensitive to variation in word frequency but not be misled into putting words with different frequencies into distinct categories or using frequency differences to form too many (or too few) grammatical categories. We hypothesize that, while learners in Experiment 1 may show sensitivity to variation in word frequency in their sentence ratings, their overall pattern of generalization will remain the same across all word frequency levels. Under the conditions of Experiment 1 (high-density, high-overlap), learners should rate familiar and novel sentences containing a given x-word the same, even though x-words with high frequency may be rated higher overall than those with low frequency. When familiar and novel sentences are rated the same—that is, when learners rate grammatical sentences they have never heard before to be just as well formed as the familiar sentences they heard during exposure—this indicates that they have formed a category, extending the same category privileges (permitted syntactic contexts) to all members of the category.

Method

Participants

Twenty-one monolingual, native English-speaking undergraduates from the University of Rochester were paid to participate in this experiment. Six participants were excluded from the analyses due to equipment failure (4) or for failure to comply with experimental instructions (2). Of the remaining participants, eight were exposed to Language 1 and seven were exposed to Language 2, which differed only in the specific words assigned to the grammatical categories.

Stimulus materials

Sentences were generated from a grammar of the form (Q)AXB(R). Each letter corresponds to a category of nonsense words (see Table 1). Categories A and B contained three words each, X contained four words, and Q and R contained two words each. X is the target category of interest; A and B served as contexts for X, providing distributional information that can indicate whether the different X-words are part of the same category (that is, have the same privileges of occurrence in A_ and _B contexts) or, rather, have individually distinct contexts in which each of them can occur. Q and R words are optional, creating sentences of varying length in the language (between 3–5 words long) and preventing the A, X, and B words from appearing consistently at the beginning or end of the sentence.

In this experiment, training sentences were selected such that the words in the target X category had highly overlapping contexts. During exposure, X_1, X_2, and X_3 all occurred with every A word and every B word (though not with every A_B context). This means that, in aggregate, the contexts in which X_1, X_2 and X_3 occurred were *completely* overlapping in terms of the preceding A or the subsequent B word. In contrast, X_4 occurred in only one context: A_1X4B1. Because of this, X_4 was *minimally* overlapping with the other words in the X category (see Table 2). Focusing on the target X-category and its immediate A and B context cues, there were 3x4x3 = 36 possible AXB strings in the language. Of these, learners were exposed to 19 AXB combinations: 6 with X_1, 6 with X_2, 6 with X_3, and 1 with X_4. The rest were withheld for testing generalization. Reeder et al. called this a *dense* sampling of the target category, since learners were exposed to more than half of the possible $AX_{1\text{-}4}B$ combinations.

In order to test learners' sensitivity to variations in lexical frequency, we systematically varied the exposure to each X-word along a Zipfian distribution to create high-, medium-, and low-word frequency groups. AXB strings containing X_1 were presented 3 times each (*low frequency*) for a total of 18 strings, strings containing X_2 were presented 11 times each (*medium frequency*) for a total of 66 strings, and strings containing X_3 were presented 22 times each (*high frequency*) for a total of 132 strings. As in a Zipfian distribution, our second most frequent X-word (medium) occurred half as often as our most frequent X-word (high). In a Zipfian distribution, the word frequencies continue to follow this pattern, with the next most frequent X-word occurring approximately half as often as the word one frequency rank above it. We chose to present our lowest frequency X-word, X_1, 18 times, which corresponds approximately to the fourth most frequent word in a Zipfian distribution for our corpus size. We selected this value because the single-context, minimally-overlapping X_4 string was presented 18 times. Crucially, then, X_4 was heard just as often as the low-frequency X_1, but the contexts surrounding X_1 strings were broader than the single context surrounding the X_4 strings. The possible X_1 strings were densely sampled (two-thirds of the possible strings were in the input) and included all of the possible A and B contexts. X_4, however, was sparsely sampled, seen with only one of the 9 possible A_ and _B contexts. Of special interest, then, is how well learners are able to generalize to X_4 as compared with X_1. In previous work, if

Table 1. Assignment of words to categories for Languages 1 and 2.

Language 1				
Q	A	X	B	R
spad	flairb	tomber	fluggit	Gentif
klidum	daffin	zub	mawg	Frag
	glim	lapal	bleggin	
		norg		
Language 2				
Q	A	X	B	R
frag	gentif	spad	zub	Lapal
daffin	mawg	fluggit	tomber	Flairb
	klidum	bleggin	glim	
		sep		

Table 2. All possible strings generated from the (Q)AXB(R) grammar. Strings that were presented in the input for Experiment 1 are denoted with *; strings presented in the input for Experiment 2 are denoted with #.

Strings with X1	Strings with X2	Strings with X3	Strings with X4
A1 X1 B1 *	A1 X2 B1	A1 X3 B1 * #	A1 X4 B1 * #
A1 X1 B2	A1 X2 B2 *	A1 X3 B2 * #	A1 X4 B2
A1 X1 B3 * #	A1 X2 B3 *	A1 X3 B3	A1 X4 B3
A2 X1 B1	A2 X2 B1 * #	A2 X3 B1 *	A2 X4 B1
A2 X1 B2 * #	A2 X2 B2 *	A2 X3 B2	A2 X4 B2
A2 X1 B3 * #	A2 X2 B3	A2 X3 B3 *	A2 X4 B3
A3 X1 B1 *	A3 X2 B1 * #	A3 X3 B1	A3 X4 B1
A3 X1 B2 *	A3 X2 B2	A3 X3 B2 * #	A3 X4 B2
A3 X1 B3	A3 X2 B3 * #	A3 X3 B3 *	A3 X4 B3

participants acquired X_1–X_3 as a strong category, they also readily included X_4 in the category; but when X_4 in the present experiment is both rare and narrowly distributed, it is not clear whether participants should so readily generalize all of the target category's distributional properties to it.

Because of the optional Q and R flanker words, each AXB string could be presented in multiple contexts: AXB, QAXB, AXBR, or QAXBR. The frequency of these different Q/R contexts was divided equally for each frequency group, though not every AXB string was seen with each of the Q's and R's. Altogether the exposure set consisted of 234 strings.

Test strings consisted of *familiar* grammatical AXB strings that were presented during training, *novel* but grammatical AXB strings that were withheld from exposure, and *ungrammatical* strings that were of the form AXA or BXB (with no word repetitions in any string). Test strings were presented to participants in a pseudorandom order: the first half of the test contained 10 familiar, 13 grammatical novel, and 12 ungrammatical strings; the second half repeated the 10 familiar and 13 novel strings, but presented 12 new ungrammatical strings and presented all of these test strings in a different randomized order. Of the 10 familiar test strings, there were three containing each of X_1, X_2, and X_3 and one containing X_4 (recall that there is only one familiar X_4 string possible); of the 13 novel test strings, there were 3 containing each of X_1, X_2, and X_3, and four containing X_4; and of the 12 ungrammatical strings in each test half, there were three containing each of X_1, X_2, X_3, and X_4 strings. The difference in ratings of familiar and ungrammatical strings tells us whether participants have learned the basic properties of the language. We use the *difference between ratings of familiar and novel grammatical strings* to indicate whether learners have collapsed X-words into a single category. If this difference is large, learners are not generalizing to the novel, unheard contexts for each X-word. If this difference is small, learners are generalizing beyond their input, which suggests that they have formed an X-category that allows every X-word to appear in the same contexts as every other X-word.

During the test, participants were asked to rate only a subset of the possible novel AXB strings. To ensure that there was nothing special about the particular subset of novel strings that were tested, we divided subjects into two testing groups. Each testing group received a different subset of novel items to rate.

To create the training and test strings, nonsense words were recorded separately, each with terminal and with non-terminal intonation, by a female native English speaker. These recordings were adjusted with Praat (Boersma, 2001) to achieve relatively consistent pitch, volume, and duration among words. Words were then concatenated into sentences in Sound Studio with 50 ms of silence inserted between them. Sentence-initial and medial words had non-terminal intonation, whereas sentence-final words had terminal intonation.

Procedure

Prior to exposure, participants were instructed to listen carefully while they heard sentences from a made-up language, because they would be tested on their knowledge of the language later. During exposure, participants listened passively via headphones as a custom software package presented the training strings with 1500 ms of silence between sentences. After training, participants were presented with individual test sentences and asked to rate each sentence based on whether the sentence came from the language they heard during training: 5 meant the sentence definitely did come from the language, and 1 meant the sentence definitely did not come from the language.

Results

We found no significant differences in ratings of the two sets of novel grammatical strings ($F < 1$), suggesting that we did not inadvertently select a biased set of novel grammatical strings to test. We therefore collapsed ratings across the two testing groups for all subsequent analyses. Additionally, there was no main effect of how words were assigned to categories in Language 1 vs. Language 2 (see Table 1) ($F < 1$), so we collapsed participants' ratings across the two languages.

As in Reeder et al. (2013), we analyzed ratings of strings containing X_1, X_2, and X_3 separately from strings containing X_4.[1] Although the raw number of exposures to X_1 and X_4 were the same during training, the nature of the exposure and test for these strings was quite different: X_1 was heard 18 times across 6 different contexts, whereas X_4 was heard 18 times in just one context; thus there was only one familiar X_4 string to test, whereas there were 6 familiar X_1 strings to test. Given this difference, we first focus on the patterns of generalization across $X_{1\text{-}3}$ test strings, and then consider generalization to novel AX_4B test strings separately. Because individual learners may have used our rating scale in different ways, subject ratings were examined as raw scores and also were transformed to z-scores for each individual. There were no differences in results across analyses using the raw vs. transformed ratings; we therefore report only the raw ratings here.

$X_{1\text{-}3}$ analyses

Figure 1 shows the mean ratings for the grammatical familiar, grammatical novel, and ungrammatical test strings containing X_1, X_2, and X_3. The mean rating of familiar strings was 3.54 ($SE = 0.08$), the mean rating of novel grammatical strings was 3.52 ($SE = 0.12$), and the mean rating of ungrammatical strings was 2.71 ($SE = 0.15$). As in Reeder et al. (2013), this pattern of results provides compelling evidence that learners learn the basic structure of the grammar and generalize fully from familiar to novel grammatical test strings.

To examine this generalization effect in more detail, we ran a repeated-measures ANOVA with test-item type (familiar, novel, ungrammatical) and X-word (X_1, X_2, X_3) as within-subjects factors. This allowed us to determine how ratings differed as a function of word frequency (see Figure 2). Mauchly's test indicated a violation of the sphericity assumption for both test type ($\chi^2(2) = 9.35$, $p < 0.01$) and X-word (χ^2 $(2) = 7.53$, $p < 0.05$), so degrees of freedom were corrected using Greenhouse-Geisser estimates ($\varepsilon_{\text{TestType}} = 0.65$, $\varepsilon_{\text{X-word}} = 0.68$). The results revealed significant main effects of test-item type ($F(1.30,16.87) = 14.26$, $p < 0.001$) and $X_{1\text{-}3}$-word ($F(1.36,17.74) = 4.94$, $p < 0.05$). There was also a significant interaction between test-item type and X-word ($F(4,52) = 2.99$, $p < 0.05$). However, this interaction was not due to a changing effect of word frequency on generalization. Planned comparisons showed that ratings of familiar and novel grammatical strings did not significantly differ ($F(1,13) = 0.03$, $p = 0.86$) for any of the X-word types. However, ungrammatical strings were rated significantly lower than either familiar or novel strings ($F(1, 13) = .15.96$, $p < 0.01$), and this difference increased over the X-word types (i.e., as X-word frequency increased).

These results suggest that learners are just as willing to generalize from the familiar to the novel grammatical combinations for each $X_{1\text{-}3}$-word, regardless of lexical frequency, which varied by a factor of 7. Learners also correctly reject ungrammatical strings for each frequency level. However, strings containing the low-frequency X_1 word are rated lower overall (a planned comparison reveals that strings containing X_1 are rated significantly lower than strings containing X_3, $p = 0.014$), demonstrating sensitivity to lexical frequency but no disruption to the pattern of ratings to novel grammatical test-

[1]Analyzing all X-word ratings together does not qualitatively change the results.

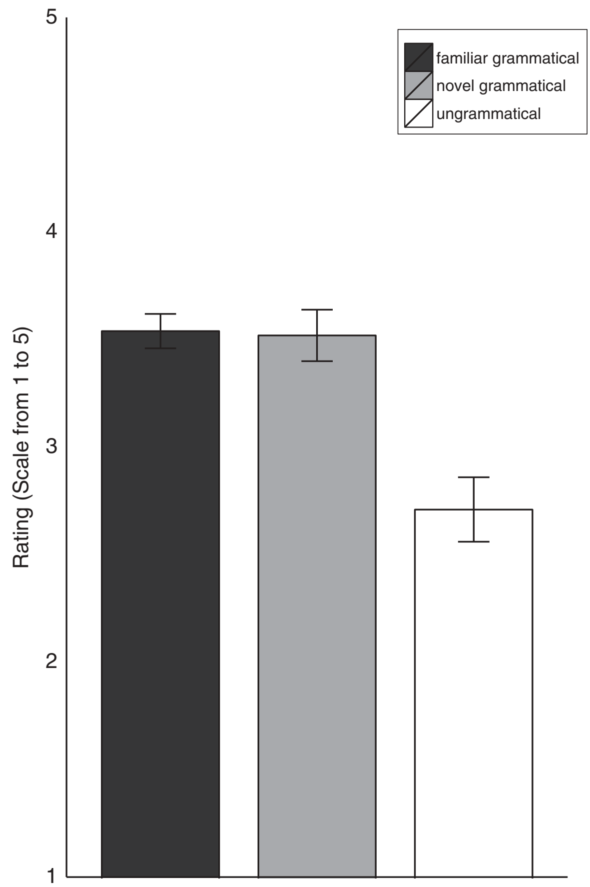

Figure 1. Mean ratings from Experiment 1, comparing familiar, novel, and ungrammatical test strings for X_1, X_2, and X_3 words combined. Participants rate on a scale from 1–5 sentences presented during exposure (familiar grammatical), sentences that are of the form AXB but were not presented during exposure (novel grammatical), and sentences that were ungrammatical. Error bars are standard error.

items across lexical frequency. That is, there was similar use of distributional cues to category membership across lexical items that differ dramatically in frequency.

X_4 analyses

As shown in Figure 3, for test items containing X_4, the mean rating of familiar strings was 3.77 (SE = 0.16), the mean rating of novel grammatical strings was 3.18 (SE = 0.10), and the mean rating of ungrammatical strings was 2.38 (SE = 0.15). A repeated measures ANOVA on these test items, with test type (familiar, novel, ungrammatical) as the within subjects factor, revealed a significant main effect of test type ($F(2,26) = 29.89$, $p < 0.0001$). Planned comparisons show significant differences between all three test types (for familiar vs. novel grammatical, $F(1,13) = 19.05$, $p < 0.001$; for novel grammatical vs. ungrammatical, $F(1,13) = 17.07$, $p < 0.001$). Generalization for the single-context X_4 category to novel grammatical strings was thus less robust than generalization for the X_1-X_3 categories. It is important to note that X_4 appears in the exposure corpus exactly the same number of times as X_1. Despite this, there was more generalization to novel grammatical test items for X_1, apparently due to its occurrence with a broader set of A_ and _B contexts. It is therefore not the frequency of occurrence of a word, but rather its occurrence across distributional contexts, that is more important for category formation and generalization.

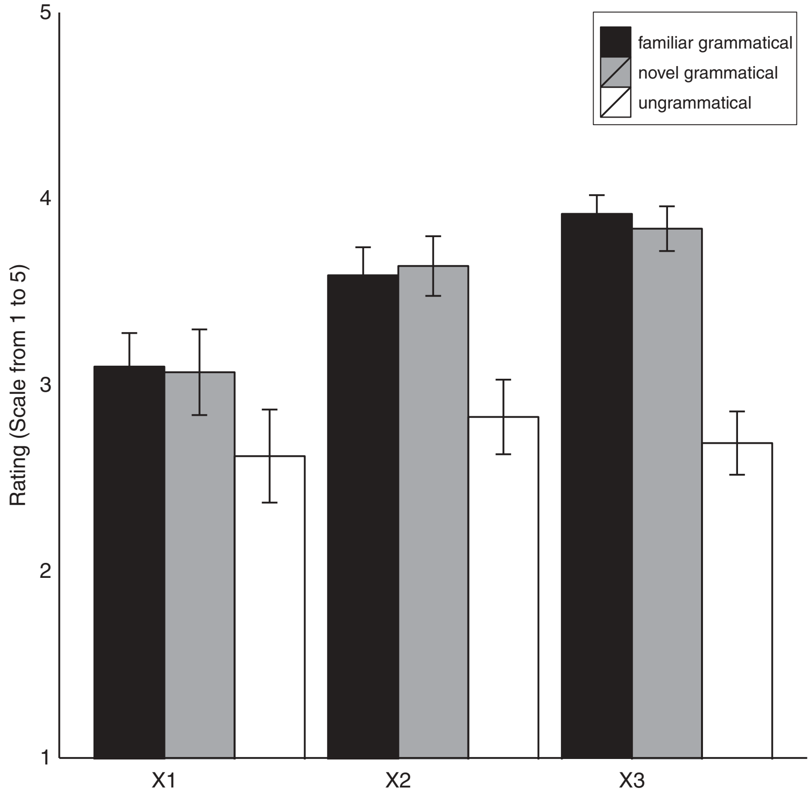

Figure 2. Mean ratings of familiar, novel, and ungrammatical test strings for Experiment 1, separated by X-word. Participants rate on a scale from 1–5 sentences presented during exposure (familiar grammatical), sentences that are of the form AXB but were not presented during exposure (novel grammatical), and sentences that were ungrammatical. Error bars are standard error.

Discussion

As in Reeder et al. (2013), we found that when there was dense sampling and complete overlap among contexts of the words in the X-category, learners rated familiar and novel $X_{1\text{-}3}$ test strings the same, indicating that they collapsed $X_{1\text{-}3}$ into a single category and generalized the allowable contexts across these words. This suggests that lexical frequency differences as large as 7:1 do not significantly impact how learners form categories based on distributional cues like context overlap and sampling density. Given sufficient exposure to fully overlapping context cues, learners will collapse words into a category and generalize across gaps in their input. This is not because learners entirely ignore lexical frequency information. Learners were sensitive to the lexical frequency differences: strings with low frequency X_1 words were rated significantly lower than strings with high frequency X_3 words, but the same pattern of generalization to novel grammatical test items was seen across all three word frequencies. Importantly, the results from the X_4 word emphasize a similar point: participants show a somewhat diminished tendency to generalize X_4 to all of the X-word contexts; but this is due to the reduced range of contexts in which it appeared in the exposure corpus, not to its reduced frequency, which was identical to that of the low-frequency X_1 words.

Experiment 2

The previous experiment demonstrated that large lexical frequency imbalances do not prevent learners from using distributional cues to discover categories in their input, provided that the target

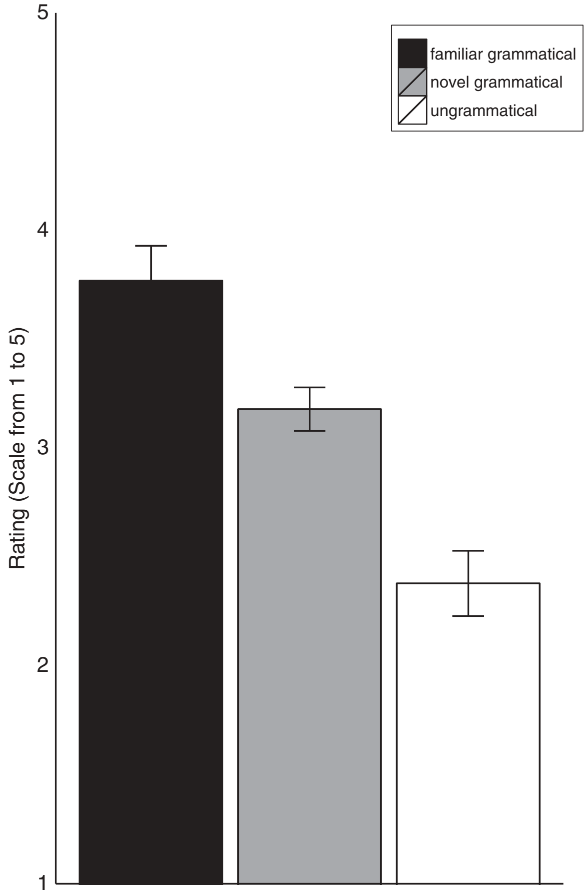

Figure 3. Mean ratings from Experiment 1, comparing familiar, novel, and ungrammatical test strings for X_4 word. Participants rate on a scale from 1–5 sentences presented during exposure (familiar grammatical), sentences that are of the form AXB but were not presented during exposure (novel grammatical), and sentences that were ungrammatical. Error bars are standard error.

category members are surrounded by a dense and highly overlapping set of context words. However, during natural language acquisition, learners do not always have access to dense, highly overlapping samples of input for every word and category they must learn. Rather, because learners hear only a sample of the possible sentences in their language, they often need to infer what category a word belongs to after hearing only a few of the syntactic contexts in which that word can occur. When there are gaps in the input (missing syntactic contexts), learners must decide whether those contexts are absent by chance or because that particular construction is ungrammatical. This task is further complicated when the words in a category (and, as a result, the permitted syntactic contexts for that word) occur with unequal frequencies. When word frequency varies along a Zipfian distribution, with some being highly frequent and others highly infrequent, gaps in the input may be due to a third possibility: low frequency.

In Experiment 2 we tested whether lexical frequency would have a larger impact on generalization when the exposure corpus contained systematic gaps, created by reducing the overlap of contexts in which different category members appeared. In this corpus, each X-word appeared with only 2 of the 3 possible A-words and 2 of the 3 possible B-words; the X-words differed in which specific A- and B-words they combined with. Learners could reasonably interpret these patterns as suggesting that the X-words were a single category, or that each X-word had its own subcategorization restrictions. In Reeder et al. (2013) this incomplete overlap among words resulted in somewhat decreased generalization to novel strings containing $X_{1\text{-}3}$ words and also decreased generalization to a minimally overlapping X_4 word. Learners did continue to generalize from familiar to novel grammatical contexts, but their ratings of novel contexts were lower than those of familiar contexts (although substantially higher than their ratings of ungrammatical sequences). Here we explore whether this occurs when the same reduction in context overlap appears in X-words of varying lexical frequencies, or whether highly variable word frequencies buffer the learner from restricting generalization when contextual gaps occur or, on the other hand, reduce certainty and generalization overall.

We hypothesize that learners in Experiment 2 will show sensitivity to variation in word frequency in their sentence ratings, but the pattern of generalization shown in Reeder et al. (2013) for incomplete overlap will remain the same across all word frequency levels. That is, learners will rate novel sentences somewhat lower than familiar sentences across all word frequency levels, although X-words with high frequency may be rated higher overall than X-words with low frequency. As compared to Experiment 1, this lower rating for novel sentences would suggest that learners with gaps in the exposure corpus are more uncertain about generalizing to novel contexts, but we expected that this pattern would not be altered by variable word frequency.

Method

Participants

Thirty-one adults were paid to participate in Experiment 2. Eleven subjects were excluded from all analyses because they did not follow instructions (5), had participated in a similar experiment (1), were bilingual (1), were outside of our target age range (3), or because of equipment failure (1). All of the remaining participants were monolingual, English-speaking undergraduates who had not participated in Experiment 1. Eleven participants were exposed to Language 1, and nine were exposed to Language 2, which differed only in which specific words were assigned to each grammatical category.

Stimulus materials and procedure

Stimulus materials were constructed in the same manner as Experiment 1. However, exposure strings were now selected to create incomplete overlap in contexts across $X_{1\text{-}3}$-words (see Table 2). This design creates systematic gaps in the contexts that support forming a single X-category. X_1 was only heard with A_1, A_2, B_2, and B_3 context words; X_2 was only heard with A_2, A_3, B_1, and B_3; X_3 was only heard with A_1, A_3, B_1, and B_2; and X_4 was only heard in one context (A_1X4B1). As in Experiment 1, X-word input frequencies followed the ratio 18:66:132:18 for X_1:X_2:X_3:X_4, which correspond to the first, second, and fourth ranked words in a Zipfian distribution of a corpus this size. With the addition of optional Q and R flanker words, the total exposure consisted of 234 strings (as in Experiment 1).

All subjects were given a ratings test. As in Experiment 1, the first half of the test contained 10 familiar, 13 novel, and 12 ungrammatical strings, presented in pseudo-random order; the second half repeated the 10 familiar and 13 novel strings with 12 new ungrammatical strings, all in a different random order. As in Experiment 1, of the 10 familiar test strings, there were three containing each of X_1, X_2, and X_3 and one containing X_4 (recall that there is only one familiar X_4 string possible); of the 13 novel test strings, there were three containing each of X_1, X_2, and X_3, and 4 containing X_4; and of the 12 ungrammatical strings in each test half, there were 3 containing each of X_1, X_2, X_3, and X_4 strings.

Results

There was no main effect of language ($F < 1$), so we collapsed the results across the two languages for all subsequent analyses.

$X_{1\text{-}3}$ analyses

Figure 4 shows the mean ratings for the grammatical familiar, grammatical novel, and ungrammatical test strings containing X_1, X_2, and X_3. The mean rating of $X_{1\text{-}3}$ familiar strings was 3.96 ($SE = 0.09$), the mean rating of novel grammatical strings was 3.63 ($SE = 0.09$), and the mean rating of ungrammatical strings was 2.61 ($SE = 0.12$). As in Reeder et al. (2013), then, when X-words did not overlap completely in the contexts in which they appeared, learners did not fully generalize from familiar to novel grammatical strings; however, they rated both much higher than ungrammatical strings.

To examine how ratings differed as a function of word frequency, we ran a repeated-measures ANOVA with test-item type (familiar, novel, ungrammatical) and X-word (X_1, X_2, X_3) as within-subjects factors. The analysis revealed significant main effects for both factors (test item type: $F(2,36) = 51.81$, $p < 0.0001$; X-word: $F(2,36) = 9.71$, $p < 0.0001$). Planned comparisons showed that ratings of familiar and novel grammatical strings differed significantly across the X-word types ($F(1,18) = 9.38$, $p < 0.01$). Ungrammatical strings were rated significantly lower than either familiar or novel grammatical strings ($F(1, 18) = 50.20$, $p < 0.0001$). There was also a significant difference between the ratings of the low-frequency X_1 strings and the mid frequency X_2 strings ($F(1,18) = 14.90$, $p < 0.001$) (see Figure 5). Importantly, however, there were no significant interactions ($F < 1$). For all frequency levels, familiar strings were rated somewhat higher than novel strings, and both were rated quite substantially higher than ungrammatical strings. There was no difference in this pattern across different word frequency levels.

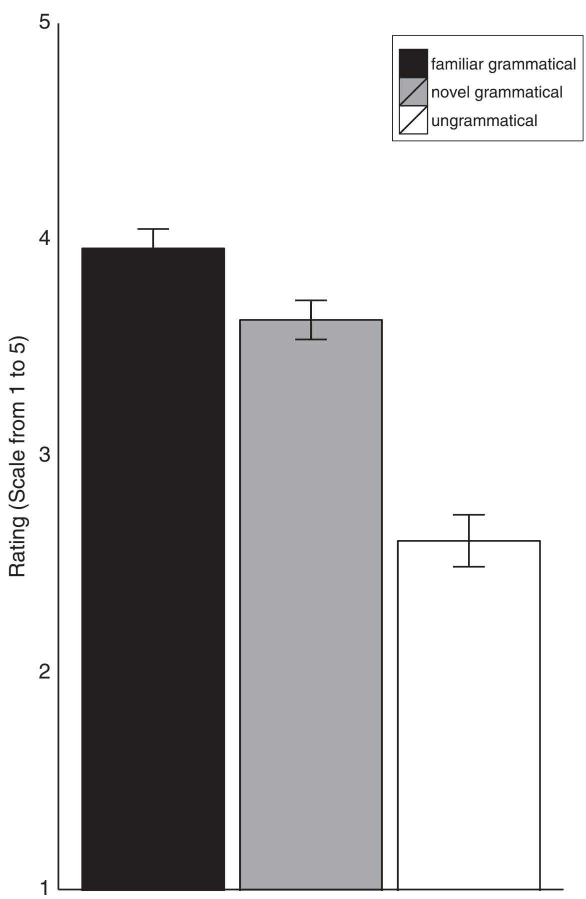

Figure 4. Mean ratings from Experiment 2, comparing familiar, novel and ungrammatical test strings for X_1, X_2, and X_3 words combined. Participants rate on a scale from 1–5 sentences presented during exposure (familiar grammatical), sentences that are of the form AXB but were not presented during exposure (novel grammatical), and sentences that were ungrammatical. Error bars are standard error.

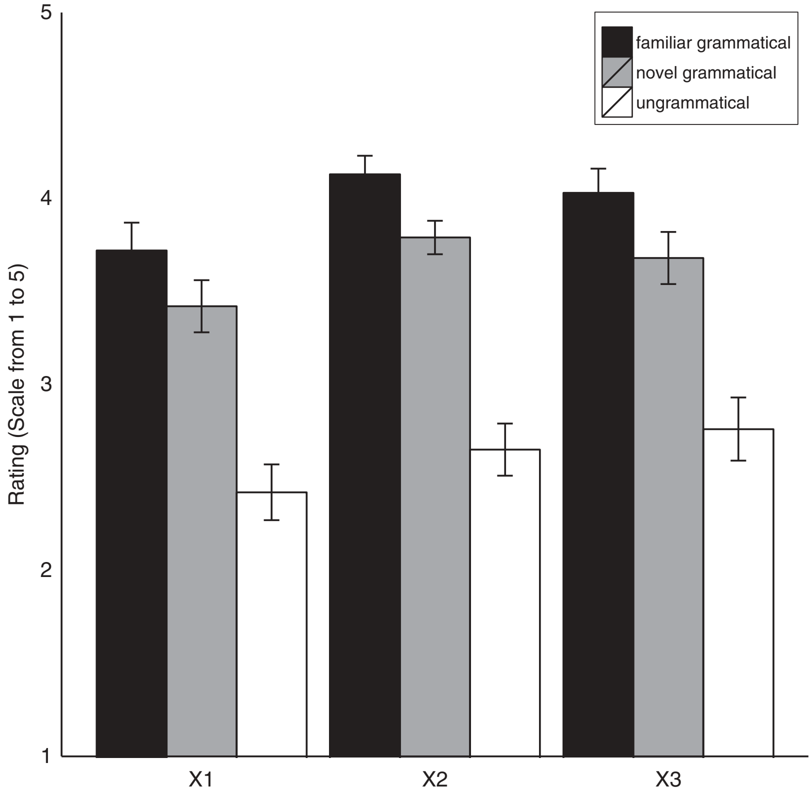

Figure 5. Mean ratings of familiar, novel, and ungrammatical test strings for Experiment 2, separated by X-word. Participants rate on a scale from 1–5 sentences presented during exposure (familiar grammatical), sentences that are of the form AXB but were not presented during exposure (novel grammatical), and sentences that were ungrammatical. Error bars are standard error.

X_4 analyses

The mean rating of X_4 familiar strings was 3.45 (*SE* = 0.23), the mean rating of novel grammatical strings was 2.95 (*SE* = 0.18), and the mean rating of ungrammatical strings was 2.36 (*SE* = 0.19) (see Figure 6). The ratings for these test items were submitted to a repeated-measures ANOVA with test type as the within-subjects factor. Mauchly's test revealed that the sphericity assumption was violated ($\chi^2(2) = 8.56$, $p < 0.05$), so degrees of freedom were corrected using the Greenhouse-Geisser estimate ($\varepsilon = 0.72$). There was a significant effect for test-item type ($F(1.43, 25.80) = 10.25$, $p < 0.001$). Planned comparisons revealed that familiar X_4 strings were rated marginally higher than novel grammatical X_4 strings ($F(1,18) = 3.53$, $p = 0.08$), which were rated significantly higher than ungrammatical X_4 strings ($F(1,18) = 14.88$, $p < 0.001$).

Discussion

As was found in Reeder et al. (2013), the systematic gaps created by incomplete overlap of contexts leads learners to be more conservative in generalization. As shown in Figure 4, these gaps in Experiment 2 led participants to judge familiar and novel grammatical test strings as more different from each other than did participants in Experiment 1 where there was complete overlap. These results suggest that learners did not fully collapse $X_{1\text{-}3}$ into a single category when context overlap was reduced. Since the exposure in Experiments 1 and 2 contained the same number of strings and the same ratio of frequency imbalances, variable word frequency cannot explain the change in behavior across the two experiments. Instead, only the shift in context overlap could be responsible

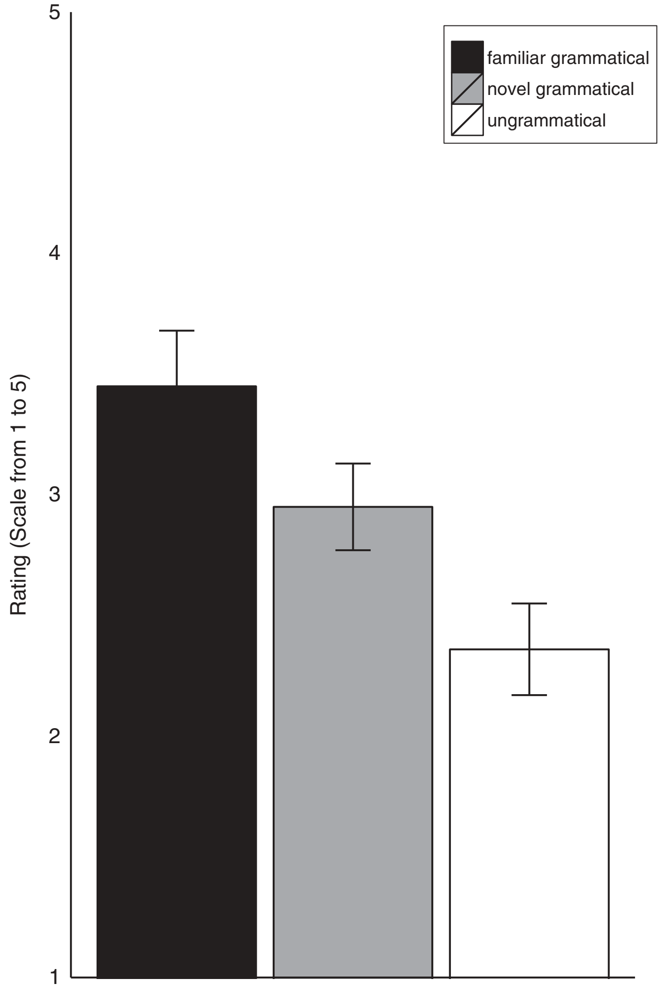

Figure 6. Mean ratings from Experiment 2, comparing familiar, novel, and ungrammatical test strings for X_4 word. Participants rate on a scale from 1–5 sentences presented during exposure (familiar grammatical), sentences that are of the form AXB but were not presented during exposure (novel grammatical), and sentences that were ungrammatical. Error bars are standard error.

for the observed change in generalization behavior. Most important, as we saw in Experiment 1, while there is sensitivity to lexical frequency as shown by the significant differences in overall mean ratings of individual X-words, the large lexical frequency variations (7:1 ratio) in the X-category do not alter the patterns of ratings across different types of test items. That is, lexical frequency variations do not alter how learners interpret distributional cues to categorization.

While X_4 and $X_{1\text{-}3}$ strings showed the same pattern (that is, mean rating of familiar items was greater than that for novel grammatical items for both), the difference between novel and familiar X_4 test items in this experiment was only marginally significant. This is most likely due to the fact that there were many fewer X_4 test items compared with $X_{1\text{-}3}$ items and thus lower statistical power in this comparison. Reeder et al. (2013) also found a significant difference between novel and familiar $X_{1\text{-}3}$ items but not X_4 items when learners were exposed to reduced overlap. However, both differences became significant when learners received additional exposure to this reduced overlap. Perhaps when overlap is reduced and frequency is highly variable, as is the case in our Experiment 2, learners require more evidence of systematic gaps before they restrict generalization to a low-frequency novel word.

Overall, in this experiment learners restrict generalization based on the reduced overlaps among the contexts in which items occur. This pattern does not change with large differences in item frequency.

General discussion

Our results provide further evidence that a powerful statistical learning mechanism is sufficient to enable adult learners to acquire the latent category structure of an artificial language, even without correlated phonetic or semantic cues. More importantly, our results also show that, while learners are sensitive to lexical frequency, substantial lexical frequency variations do not alter how learners interpret distributional cues to categorization. That is, when learners are exposed to words that have completely overlapping contexts, as in Experiment 1, they generalize (participants rated novel sentences no differently than familiar sentences). When learners are exposed to words that have partially overlapping contexts (i.e., lexical gaps), as in Experiment 2, they restrict generalization somewhat (participants rated novel sentences slightly but significantly lower than familiar sentences) but still distinguish novel grammatical sentences from ungrammatical strings. These results suggest that in a more naturalistic learning environment, where lexical and bigram frequencies are not uniform and instead mirror the Zipfian lexical frequency variations present in natural languages, distributional learning is still a viable mechanism for category acquisition. Importantly, they also suggest that learners do not form their categories based only on high-frequency lexical items, leaving lower-frequency items aside or judging them with greater uncertainty. Rather, they apparently conduct distributional analyses similarly on lexical items of varying frequency levels, conditioning their expectations about context occurrence based on the frequency of the lexical items. In other words, the type of statistic that learners use to determine categorization is not the absolute frequency with which words occur in each context, but rather the conditional probabilities that words occur in these contexts, given their overall frequency. This is consistent with learners' use of conditional rather than absolute frequency statistics in other experiments as well (see, for example, Aslin, Saffran, & Newport, 1998; Kurumada et al., 2013).

As we have discussed, when there are gaps in the input (missing syntactic contexts), learners must decide whether those contexts are absent by chance or because that particular construction is ungrammatical. Previous work has suggested that a distributional learner could accomplish this via Bayesian inference (Reeder et al., 2013; Tenenbaum & Griffiths, 2001). That is, the learner could compute the likelihood that a particular context is missing by chance, given the data he or she has experienced. In this framework, the larger the corpus—or in the present case, the more frequently a lexical item occurs in the corpus without appearing in a particular context—the less likely it is that the context is absent by chance, and the more likely the learner should be to rate these withheld contexts as ungrammatical. Indeed, when gaps are persistent, both human learners (Reeder et al., 2013) and Bayesian-inspired models (Qian, Reeder, Aslin, Tenenbaum, & Newport, 2012) are more likely to identify withheld contexts as ungrammatical. Importantly, the results of the present experiments suggest that learners do not merely compute the likelihood that a context is absent by chance based on its absolute frequency in the corpus. Were this the case, our learners should have concluded that low frequency words (e.g. X_1), with their low frequency of appearance in all contexts, were much less likely to be part of the X category and its contexts compared to high frequency words (e.g. X_3); but we observed no such interaction in our results. Instead, learners apparently condition their expectations about context occurrence based on the frequency of the lexical items and determine whether to generalize to novel contexts based on these frequency-adjusted probabilities.

It is possible that we observed robust categorization in this paradigm because our frequency manipulations were smaller (a maximum ratio of 7:1) than frequency ratios in natural language input. Despite this difference in scale, however, our results provide important insight into how a mature statistical learner interprets frequency information, especially given that exposure was limited to only 234 sentences. Research on priming and adaptation in language have demonstrated that even short exposures to new linguistic environments can bias how language users interpret information in natural languages (e.g., Fine, Jaeger, Farmer, & Qian, 2013; Thothathiri & Snedeker, 2008; Traxler, 2008) and in artificial languages (Fedzechkina, Jaeger, & Newport, 2012). These results are striking because they demonstrate that in certain situations, recent frequency information can rapidly outweigh a lifetime of language experience. Learners in our experiments are naïve to both the structure of the artificial grammar and the assignment of words to categories. Therefore, we might expect frequency to have a *larger* effect than it would with natural language input, as our learners have no prior biases in this language to overcome. Given this, it is unlikely that our results are solely due to scaled-down frequency variations.

Of course, in future work the same questions must be studied in child learners. But at least for adult learners, the present results support the relevance of distributional learning for grammatical categorization by confirming that the same patterns of learning occur across lexical frequency variations. Because natural languages do exhibit extreme variations in lexical frequency, these results take an important step in suggesting that findings from artificial grammar learning experiments of categorization may well apply to natural language learning. Our results also suggest the type of statistics that learners utilize as they acquire grammatical categories. As has been shown in studies of word segmentation (Aslin et al., 1998; Kurumada et al., 2013), statistical learning does not appear to depend primarily on simple frequency statistics (such as lexical frequency or bigram frequency), but rather utilizes more complex calculations (such as conditional probabilities or Bayesian statistics) that involve the expected frequency with which element combinations should occur, given their individual element frequencies. While it might seem unlikely that infants and young children could be capable of these complex calculations, our studies to date with young learners (Aslin et al., 1998; Schuler et al., 2017) support the notion that statistical learning involves such computations at many levels of analysis.

Acknowledgments

We would like to thank Neil Bardhan, Cory Bonn, Alex Fine, Davis Glasser, and Ting Qian for helpful comments on the analysis and interpretation of this work.

Funding

This research was supported by NIH Grant HD037082 to R.N.A. and E.L.N, by NIH Grant DC00167 and DC014558 and funds from Georgetown University to E.L.N., and by an ONR Grant to the University of Rochester.

ORCID

Richard N. Aslin http://orcid.org/0000-0003-4092-1712

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*, 321–324. doi:10.1111/1467-9280.00063

Blackwell, A. A. (2005). Acquiring the English adjective lexicon: Relationships with input properties and adjectival semantic typology. *Journal of Child Language*, *32*, 535–562. doi:10.1017/S0305000905006938

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Braine, M. D. S., Brody, R. E., Brooks, P., Sudhalter, V., Ross, J. A., & Catalano, L. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language*, *29*, 591–610. doi:10.1016/0749-596X(90)90054-4

Brooks, P. B., 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*, 79–95. doi:10.1006/jmla.1993.1005

Cartwright, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition*, *63*, 121–170. doi:10.1016/S0010-0277(96)00793-7

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, *109*, 17897–17902.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *Plos ONE*, *8*(10), e77661. doi:10.1371/journal.pone.0077661

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. doi:10.1006/jmla.1998.2569

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, *32*, 249–268. doi:10.1017/S0305000904006786

Gerken, L. A., Gomez, R., & Nurmsoo, E. (1999, April). The role of meaning and form in the formation of syntactic cateogries. *In*: *Paper presented at the Society for Research in Child Development*, Albuquerque, NM.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436. doi:10.1111/1467-9280.00476

Gomez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science*, *7*(5), 567–580. doi:10.1111/desc.2004.7.issue-5

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531. doi:10.1017/S0305000907008641

Harris, M., Barrett, M., Jones, D., & Brookes, S. (1988). Linguistic input and early word meaning. *Journal of Child Language*, *15*, 77–94. doi:10.1017/S030500090001206X

Harris, Z. S. (1954). Distributional structure. *Word*, *10*, 146–162. doi:10.1080/00437956.1954.11659520

Holmes, V. M., Stowe, L., & Cupples, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language*, *28*, 668–689. doi:10.1016/0749-596X(89)90003-X

Kidd, E., Lieven, E., & Tomasello, M. (2006). Examining the role of lexical frequency in the acquisition and processing of sentential complements. *Cognitive Development*, *21*, 93–107. doi:10.1016/j.cogdev.2006.01.006

Kurumada, C., Meylan, S. C., & Frank, M. C. (2013). Zipfian frequency distributions facilitate word segmentation in context. *Cognition*, *127*, 439–453. doi:10.1016/j.cognition.2013.02.002

Lapata, M., Keller, F., & Schulte Im Walde, S. (2001). Verb frame frequency as a predictor of verb bias. *Journal of Psycholinguistic Research*, *30*, 419–435. doi:10.1023/A:1010473708413

Maratsos, M. P., & Chalkely, 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* (Vol. 2, pp. 127–214). New York, NY: Gardner Press.

McNeill, D. (1966). Developmental psycholinguistics. In F. Smith, & G. Miller (Eds.), *The genesis of language* (pp. 15–84). Cambridge, MA: The MIT Press.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory and Cognition*, *30*, 678–686. doi:10.3758/BF03196424

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, *90*, 91–117. doi:10.1016/S0010-0277(03)00140-9

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. doi:10.1207/s15516709cog2604_1

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, *96*, 143–182. doi:10.1016/j.cognition.2004.09.001

Morgan, J. L., Shi, R., & Allopenna, P. (1996). Perceptual bases of grammatical categories. In J. L. Morgan, & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 263–283). Mahwah, NJ: Lawrence Erlbaum Associates.

Naigles, L. R., & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. *Journal of Child Language*, *25*(1), 95–120. doi:10.1017/S0305000997003358

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1987). The bootstrapping problems in language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. New York, NY: Springer-Verlag.

Qian, T., Reeder, P. A., Aslin, R. N., Tenenbaum, J. B., & Newport, E. L. (2012). Exploring the role of representation in models of grammatical category acquisition. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 881–886). Austin, TX: Cognitive Science Society.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, *22*, 425–469. doi:10.1207/s15516709cog2204_2

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology*, *66*, 30–54. doi:10.1016/j.cogpsych.2012.09.001

Roy, B. C., Frank, M. C., & Roy, D. (2009). Exploring word learning in a high-density longitudinal corpus. In N. A. Taatgen, & H. Van Rijn (Eds.), *Proceedings of the 31st Annual Meeting of the Cognitive Science Society* (pp. 2106–2111). Austin, TX: Cognitive Science Society.

Schuler, K. D., Reeder, P. A., Lukens, K., Newport, E. L., & Aslin, R. N. (2017). *Learning grammatical categories in an artificial language by 5- to 7-year-olds using distributional information. Manuscript in preparation*.

Schwartz, R. G., & Terrell, B. Y. (1983). The role of input frequency in lexical acquisition. *Journal of Child Language*, *10*, 57–64. doi:10.1017/S0305000900005134

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*, 629–641. doi:10.1017/S0140525X01000061

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2004). Semantic generality, input frequency and the acquisition of syntax. *Journal of Child Language*, *31*, 61–99. doi:10.1017/S0305000903005956

Thothathiri, M., & Snedeker, J. (2008). Give and take: Syntactic priming during spoken language comprehension. *Cognition*, *108*(1), 51–68. doi:10.1016/j.cognition.2007.12.012

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Traxler, M. J. (2008). Lexically independent priming in online sentence comprehension. *Psychonomic Bulletin & Review*, *15*(1), 149–155. doi:10.3758/PBR.15.1.149

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*, 528–553.

Valian, V., & Coulson, S. (1988). Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language*, *27*, 71–86. doi:10.1016/0749-596X(88)90049-6

Wilson, R. (2002). *Syntactic category learning in a second language*. Unpublished doctoral dissertation. The University of Arizona.

Zipf, G. (1965). *Human behavior and the principle f least effort: An introduction to human ecology*. New York, NY: Hafner.