

Category induction from distributional cues in an artificial language

TOBEN H. MINTZ

University of Southern California, Los Angeles, California

The ability to identify the grammatical category of a word (e.g., noun, verb, adjective) is a fundamental aspect of competence in a natural language. Children show evidence of categorization by as early as 18 months, and in some cases younger. However, the mechanisms that underlie this ability are not well understood. The lexical co-occurrence patterns of words in sentences could provide information about word categories—for example, words that follow *the* in English often belong to the same category. As a step in understanding the role distributional mechanisms might play in language learning, the present study investigated the ability of adults to categorize words on the basis of distributional information. Forty participants listened for approximately 6 min to sentences in an artificial language and were told that they would later be tested on their memory for what they had heard. Participants were next tested on an additional set of sentences and asked to report which sentences they recognized from the first 6 min. The results suggested that learners performed a distributional analysis on the initial set of sentences and recognized sentences on the basis of their memory of sequences of *categories* of words. Thus, mechanisms that would be useful in natural language learning were shown to be active in adults in an artificial language learning task.

The words of natural languages are organized into *categories*. These categories are the fundamental building blocks with which the structure of sentences are built. Thus, an important part of an adult's knowledge of his/her native language is the ability to identify the category membership of words (whether the word is a noun, a verb, both, etc.). Numerous studies have shown that children differentiate these types of word classes by the time they start putting words together to form simple sentences (Bloom, 1970; Brown, 1973), and some studies have shown evidence of categorization in children at earlier ages (Shi, Werker, & Morgan, 1999; Waxman, 1994). Despite the considerable research showing children's very early appreciation of word class distinctions, the process by which children learn about lexical category membership is not fully understood. Yet, gaining insight into the mechanisms of word classification is an important part of understanding the processes that underlie syntactic development.

One hypothesis about how children initially start to group words into categories is that they perform a *distributional analysis* on sentences they hear and categorize words

together that appear in the same lexical co-occurrence patterns (Cartwright & Brent, 1997; Kiss, 1973; Maratsos & Chalkley, 1980; Mintz, Newport, & Bever, 1995, 2002; Redington, Chater, & Finch, 1998). Distributional information can include information about absolute position of words in a sentence (e.g., first word, second word, last word, etc.) as well as information about the relative position of words with respect to one another (e.g., following *the*, preceding *is*, two words before *going*, etc.); the latter is often referred to as a *lexical co-occurrence pattern*. (The term *distributional information* generally is used for both lexical and sublexical units, such as bound morphemes. Unless otherwise specified it will be used here to refer to lexical co-occurrence patterns, and the terms will be used interchangeably.) Thus, if over time a child observed that several different words occur in the "X" position in the pattern . . . *the X is going* . . . , the child would conclude that the words belong to the same grammatical category. Recent computational and statistical studies of child-directed speech corpora suggest that distributional information of this type is reliable enough, in principle, to form a basis for children's initial word classification (Cartwright & Brent, 1997; Kiss, 1973; Mintz et al., 1995, 2002; Redington et al., 1998). However, it has yet to be determined whether young children make use of this kind of information in the normal course of language acquisition.

In understanding unexplored mechanisms in children, it is sometimes instructive first to investigate similar processes in adults. Although the results can only speculatively be applied to children, much insight can be gained into the properties of the mechanism in question. These

This research was supported in part by an equipment grant from the Intel Corporation, and by a grant from the Zumberge Faculty Research and Innovation Fund at the University of Southern California. I thank Maryellen MacDonald and Laura Siegel, as well as LouAnn Gerken, Martin Redington, and an anonymous reviewer for their helpful comments and suggestions on earlier drafts of this paper. I also thank Rachel Wilson for helpful discussions of her research. Finally, I thank Kim Lemkuhl and Cynthia Semrow for their assistance in material preparation and testing. Correspondence should be addressed to T. H. Mintz, Department of Psychology, MC-1061, University of Southern California, Los Angeles, CA 90089-1061 (e-mail: tmintz@usc.edu).

insights should then facilitate similar investigations with young children. The goal of the present study is to take a step in addressing the role of distributional information in children's language acquisition by studying distributional mechanisms in adults. The results will show that adults, when hearing sentences in an artificial language, spontaneously form lexical categories that are based on the distributional regularities in the language. Although this finding does not directly address whether very young children categorize words in this way, it does demonstrate that the cognitive mechanisms needed to carry out such an analysis do occur in the species, and are naturally active during processing of artificial linguistic stimuli. Thus, this study provides a suggestive step forward in investigations of how very young children categorize words.

The research to date on distribution-based category learning in adults has yielded mixed results. Evidence for category learning has been found in some studies but not in others. For example, in an early investigation of adults' ability to learn co-occurrence relationships among distributionally defined classes, Smith (1966, Experiment 2) devised an artificial language consisting of four non-overlapping categories, M, N, P, and Q, each of which contained four unique members (spoken letter names). Sentences in the language were two-word sequences of M words followed by N words (e.g., *vee kay*), and P words followed by Q words. In a training phase, participants were exposed to multiple repetitions of 24 of the 32 possible MN and PQ sequences. Participants were subsequently tested on additional sequences and asked whether they had heard them in the learning phase. Participants heard four types of test sequences: (1) repetitions of trained sequences, (2) previously unheard combinations adhering to the trained MN/PQ structure, (3) sequences adhering to an MQ/PN structure, and (4) other unheard sequences (e.g., PM, QP, etc.). Smith hypothesized that if participants generalized from the training phase and learned the abstract, category-based co-occurrence structure, then they should incorrectly recognize unheard sequences more often for MN/PQ (Type 2, above) sequences than for MQ/PN (Type 3) sequences. However, Smith's participants recognized MN/PQ and MQ/PN types at equal rates, and did so at a higher rate than they did for other unheard strings (Type 4). The findings suggested that participants generalized positional dependency categories from the learning set (i.e., first vs. second/last), not co-occurrence dependency categories. Thus, there was no evidence that participants classified words into the distributionally defined M, N, P, Q categories.

A number of subsequent studies have used variants of Smith's (1966) grammar (MN/PQ type distinctions) to examine the effects of adding converging category cues on participants' ability to categorize using distributional information. The results from these studies can be summarized as follows¹: When (and only when) converging cues specify the categories of some portion of the words, learners can use distributional information to categorize all

words. For instance, Braine (1987) provided a consistent semantic cue (natural sex, i.e., male or female) to half the words in the N and Q categories. Frigo and McDonald (1998; see also Gomez & Gerken, 1999) marked a portion of words in N and Q categories with salient affixes—prefixes and suffixes—that differentiated the two classes. Examining sublexical distributional categorization, Wilson (2000) and Wilson, Gerken, and Nicol (2000) marked a portion of words in a category with consistent morphological endings (here M/P types were stems and N/Q types were suffixes). In all of these studies, a word for which alternative category information was provided occurred in all possible distributional environments for that word's category. For example, in Braine's study, half of the N words that participants were trained on referred to males (gender category) and co-occurred during training with all possible M words (all possible distributional environments for N words); likewise, half of the Q words referred to females and co-occurred during training with all possible P words. (The remaining words in each category referred to inanimate objects.) Finally, in all of these studies, there was no evidence of distributional categorization when converging cues were absent. Only when the converging cues were present for some of the words were learners able to use distributional information to categorize the other words. Plausibly, then, learners notice the distributional environments of the words that were independently categorized and associate the environments with categories; then they categorize the remaining words according to the learned distributional environments they fall into. On this account, the converging cues act as a bootstrap into the distributional patterns that are relevant for categories.²

These studies clearly demonstrate the usefulness of converging cues, but it is premature to conclude from them that distributional information alone is not a viable basis for early categorization in language acquisition. The necessity of a prior bootstrap might have been due to properties of the distributional patterns themselves, and parameters of the mechanisms that perform distributional analyses. The distributional cues provided in those studies were somewhat limited due to the fact that MN/PQ-type grammars produce sentences that are only two words long (or words in isolation, in the case of Wilson, 2000, and Wilson et al., 2000). Natural languages generally provide a richer variety of distributional environments. Although formally the distributional patterns in MN/PQ grammars are systematically linked to categories, they might need to be made more robust to engage distributional learning mechanisms. Given that converging cues are not always present in natural language or are sometimes misleading (see Gleitman, 1990; Maratsos & Chalkley, 1980), and given the potential usefulness of distributional information in categorization, it is worthwhile to investigate learning situations that might be better suited for observing the operation of distributional mechanisms alone. In doing so, the goal is not to pursue a theory of acquisition that relies solely on distributional evidence. Rather, it is to isolate and

understand the limits of purely distributional mechanisms so that their role in language learning can be clarified.

The present experiment reexamined the question of whether lexical co-occurrence patterns alone could be a sufficient source for categorization when the distributional cues are made more robust. Participants were familiarized to three-word sentences in an artificial language and told that they would later be tested on their memory of the sentences. The language was designed to provide redundant distributional cues in support of a specific partitioning of words into categories, providing no other bases for categorization. In particular, the first and final words constituted a static frame (i.e., the first word and final word were mutually predictive of each other) in which one class of words could occur. The basis for category generalization (classifying medial words based on the words that surround them) was purely distributional, but the distributional information was somewhat richer than what was available in the studies reviewed above. In a sense, like prior studies, this experiment provides converging cues in that a middle word's category was determined by the preceding word *and* by the following word. Crucially, however, unlike in other studies, the source of all cues was distributional lexical co-occurrence information.

Participants were then presented a set of additional sentences and asked whether they had heard each one in the training set and how confident they were of their responses. Even though the task did not require categorization, participants' responses indicated that they had categorized words on the basis of distributional patterns in the training set and that their subsequent responses were driven by their memory of the resulting patterns of words and categories. Thus, by using rich distributional patterns and no other cues to categories, this experiment goes beyond previous studies in providing evidence of categorization based purely on the distributional patterns of words in sentences.

METHOD

Participants

Forty-nine undergraduates from the psychology department participant pool at the University of Southern California served in the study. Of these, 9 participants were excluded from the data analysis due to equipment failure or experimenter error, leaving 40 participants in the data analysis. Of these, half were assigned to counterbalance Group A and the other half to counterbalance Group B. Participants received course credit for their participation in the study.

Materials and Design

A total of six experimental sets were created, each consisting of a training list and a test list. Experimental sets were designed in counterbalanced pairs (A-set and B-set). Thus, the six experimental sets were composed of three unique A-sets and three associated B-sets. The design of the A-sets will be described first, followed by the counterbalancing procedure.

Each A-set consisted of 22 three-word training sentences and 10 three-word test sentences. Training and test sentences were created using 63 monosyllabic nonsense words, divided into three groups of 21 words each (one group for each of the three A-sets and their corresponding B-sets). The structure of the training sentences will

be described in detail, followed by a detailed description of the test sentences.

Training Sentences. Training sentences were divided into four subtypes: full-paradigm (12 sentences), partial-paradigm sentences (3), alternate-paradigm sentences (3), and no-paradigm sentences (4). Broadly, the full-paradigm sentences provided a basis for participants to form category-like generalizations, the partial-paradigm sentences provided a means to test whether participants formed abstract categories, and the alternate-paradigm and no-paradigm sentences provided a means to control for superficial surface factors in testing for generalization and abstraction.

The full-paradigm lists were composed of 12 sentences that demonstrated a consistent pattern of medial word substitutions in three different three-word environments; for example, a subset of four full-paradigm sentences from one experimental set was *bool nex jiv*, *bool kwob jiv*, *bool zich jiv*, and *bool pren jiv*. Full-paradigm sentences consisted of three such sublists of four sentences each. The initial and final words (henceforth *initial/final frame*) for each sublist were unique, and all sublists had the same four medial words (from the previous example, *nex*, *kwob*, *zich*, and *pren*). In this way, the set of mutually substitutable medial words can be taken as a category. An example of the full-paradigm structure is given in Table 1.

The partial-paradigm sentences provided a basis for testing whether learners formed a category on the basis of distributional patterns in the full-paradigm sentences. The three partial-paradigm sentences consisted of a new initial/final frame and three of the medial words that occurred in the full-paradigm lists. For example, partial-paradigm sentence corresponding to the full-paradigm example might be *sook nex runk*, *sook kwob runk*, *sook zich runk*. Thus, the partial-paradigm sentences followed the same substitution pattern as the full-paradigm sentences, except that one sentence in the paradigm—*sook pren runk*—was “missing.” The missing sentence was used in the test phase to provide a test for categorization. An example of the partial-paradigm structure is given in Table 1.

The three alternate-paradigm sentences resembled the partial-paradigm sentences in that they were composed of one initial/final frame occurring with three different medial words (e.g., *choon pux wug*, *choon yult wug*, *choon plif wug*, repeated in Table 1). Structurally, the alternate- and partial-paradigm sentences were the same on the surface: Both sentence types consisted of a single initial/final frame type occurring with each of three different medial words. However, unlike partial-paradigm sentences, the alternate-paradigm sentences did not have the medial words from the full-paradigm sentences. Thus, there was no relationship between words and sentences in the alternate-paradigm sentences and the words and sentences in the full- or partial-paradigm sentences. Crucially, the medial words in the alternate-paradigm sentences (*pux*, *plif*, and *yult*) could not be taken as members of the medial word category in the full- and partial-paradigm sentences.

The no-paradigm sentences increased the overall variability and complexity of the entire training set and provided additional material for the testing phase. No-paradigm sentences were made from two of the three medial words from the alternate-paradigm sentences, as well as four additional words that did not appear elsewhere in the training set. The sentences were designed to have no consistent relationship to each other, although the medial word in two of the sentences occurred as medial words in the alternate-paradigm list. An example of the no-paradigm structure is given in Table 1.

Test Sentences. There were four types of test sentences: repeated (4), novel (4), category conforming (1), and control (1). The critical experimental contrast was between category-conforming and control sentences. Both category-conforming and control sentences were, in fact, novel sentences, but their structure was systematically different from that of the sentences called “novel” (described below). The category-conforming sentence was the sentence “missing” from the partial-paradigm sentences in the training list (*sook*

Table 1
Sentences for Experimental Sets A1 and B1

Set A1			
Training			
<u>Full Paradigm</u>	<u>Partial Paradigm</u>	<u>Alternate Paradigm</u>	<u>No Paradigm</u>
bool nex jiv bool kwob jiv bool zich jiv bool pren jiv zim nex noof zim kwob noof zim zich noof zim pren noof poz nex fen poz kwob fen poz zich fen poz pren fen	sook nex runk sook kwob runk sook zich runk	choon pux wug choon yult wug choon plif wug	fimp pux vot plif daik fimp pux ferd daik vot plif ferd
Test			
<u>Category Conforming</u>	<u>Control</u>	<u>Repeated</u>	<u>Novel</u>
sook pren runk	choon pren wug	bool pren jiv zim nex noof choon plif wug fimp pux vot	daik vot plif ferd fimp pux noof fen poz jiv bool choon
Set B1			
Training			
<u>Full Paradigm</u>	<u>Partial Paradigm</u>	<u>Alternate Paradigm</u>	<u>No Paradigm</u>
Same as A1	choon nex wug choon kwob wug choon zich wug	sook pux runk sook yult runk sook plif runk	Same as A1
Test			
<u>Category Conforming</u>	<u>Control</u>	<u>Repeated</u>	<u>Novel</u>
choon pren wug	sook pren runk	bool pren jiv zim nex noof sook plif runk fimp pux vot	Same as A1

pren runk, repeated in Table 1); it consisted of the same initial/final frame as the sentences in the partial-paradigm structure (*sook-runk*), but the medial word (*pren*) did not occur in this frame during training. Nevertheless, patterns in the full-paradigm sentences could be taken to show that *pren* belonged to the category of words that occurred with *soon-runk*. Thus, if participants categorized these words, they might find category-conforming sentences familiar.

The control sentence was identical to the category-conforming sentence except that the initial/final frame was taken from the alternate-paradigm sentences (*choon-wug*). Thus, the similarity between category-conforming sentences and control sentences was equated on surface dimensions that could play a role in familiarity judgments. Specifically, the initial and final words in the test sentences occur in an equal number of training sentences and occupy the same absolute position in the training sentences. Since the medial word is the same for both test sentences, they are equated on that dimension as well. Finally, the word-to-word transitions in both category-conforming and control sentences are unattested in the training sentences (i.e., the transitional probabilities are zero). Thus, on positional, frequency, and sequential dimensions relating to words only, participants should perceive category-conforming and control sentences to be equally familiar. However, crucially, if participants' judgments involve categories, and if they categorize

the medial words in the full- and partial-paradigm structures, then category-conforming sentences should be perceived as more familiar than control sentences.

The four repeated sentences were repetitions of four training sentences: two full-paradigm sentences, one alternate-paradigm sentence, and one no-paradigm sentence. Unlike category-conforming and control sentences (which were both novel), novel sentences consisted of novel three-word combinations of training words. Each word in a novel sentence occurred in an absolute position (i.e., initial, medial, or final) in which it never occurred during training. (Sometimes this resulted in the occurrence of bigrams that occurred in the training sentences; this characteristic was not systematically controlled.) Repeated and novel sentences provided a means of ensuring that participants followed directions and could perform the required judgment tasks. Specifically, it was expected that participants would be more likely to recognize repeated sentences than novel sentences.

A total of three A-sets were created in this way, each with an entirely different set of nonsense words. The words for all sets are listed in the Appendix.

Counterbalancing. For each A-set, a corresponding counterbalanced B-set was created. B-sets were created from A-sets by switching the initial/final frames in the partial-paradigm sentences with those of the alternate-paradigm sentences. All full-paradigm

and no-paradigm sentences remained the same. Hence, the category-conforming test sentence in an A-set was the control sentence in the associated B-set, and vice versa. The novel test sentences remained the same, as did the majority of repeated sentences. Set A1's counterbalanced pair, Set B1, is shown in Table 1.

Stimulus recording. Training and test sentences were recorded onto audiocassette by a female native English speaker who was not aware of the purpose of the study. Training and test sentences were printed individually on strips of paper and the strips were mixed together and placed in a closed container. Sentences were withdrawn and recorded, one by one, by the speaker. The speaker was instructed to read all of the sentences with a prosodic contour similar to that of the English sentence *I see you*. This resulted in the middle word being slightly more stressed than the other words. Sentences were mixed up and read three times, resulting in three randomized lists of the sentences. Each sentence was then digitized and stored on the computer that controlled the experiment. When a given sentence was presented during the experiment, the computer randomly selected one of the three versions to play. This recording and sentence presentation procedure ensured that the position of the sentence in the recording list would not be confounded with sentence type, thus avoiding the possibility of list effects. Unique recordings were made for repeated sentences for the training and test items.

Procedure

Before the experiment began, participants were informed that they would be listening to sentences from a made-up language. They were told that after listening to the language for about 6 min, they would be tested on their memory for what they had heard. During the training phase, participants were given the opportunity to use a computer drawing program to draw pictures while they listened to the training stimuli. This option was provided to keep participants from becoming bored during the task, and has been used successfully (for slightly different purposes) by Saffran, Newport, and Aslin (1996).

The experiment was carried out on a personal computer equipped with a digital-to-analog sound card. The output from the sound card was amplified and played to the participant through headphones. After the experimenter read the instructions to the participant, the participant was presented with digitized recordings of each word in the training set, spoken in isolation. The aim was to improve the participants' parsing and representation of the training stimuli by first familiarizing them with the individual words in the training set. After hearing the words, the participant clicked a button on the screen to begin the training phase.

Twenty-two unique training sentences were played in blocks of 10. For each block, the computer randomized the order of the training sentences and played them to the participant with 300 msec of silence between sentences. Each sentence was approximately 1,400 msec long; thus, the total training time was between 6 and 7 min.

After the training phase was over, the computer displayed a dialog box instructing the participant to prepare to answer questions about a series of sentences. The computer then played the first test sentence and asked the participant to judge whether or not the sentence was one of those played during the training phase; the participant was then asked to rate how confident he/she was in the judgment, on a 7-point scale, 7 being the most confident. The participant entered the responses directly into the computer; the computer then played another test sentence and asked for the participant's *yes/no* judgment and confidence rating. The process continued until all the test sentences were played. The order in which the test sentences were played was randomized by the computer. The training and testing procedure was carried out three times, once for each experimental set. All participants received the sets in the same order.

Data Analysis and Predictions

If participants recognized repeated sentences but not completely novel ones, they should give "yes" responses at a higher rate to repeated sentences than to novel sentences. Likewise, if participants

categorized medial words on the basis of the distributional evidence in the full- and partial-paradigm sentences, they should be more likely to recognize (incorrectly) category-conforming sentences than control sentences and respond yes at a higher rate to the former.

However, it is possible that the recognition responses are too coarse to reveal a categorization pattern. Consider that in both category-conforming and control sentences, the initial and final words are in the trained positions and co-occur 100% of the time during the training phase. In addition, the medial word is in an attested position, although it has never occurred in either initial/final frame. Thus, participants might tend to respond yes to control and category-conforming sentences because of superficial similarities to sentences heard during training. However, in such cases confidence ratings might reveal perceived differences between category-conforming and control sentences that are not apparent in recognition responses. Although participants may respond yes to both sentence types, if they categorized medial words they may, nonetheless, be less confident in their responses to control sentences due to the questionable status of the medial word.³

To test for this possibility, a second measure was created that incorporated confidence ratings. Yes/no and confidence ratings were transformed into a single scale by multiplying confidence ratings for "no" responses by -1 . Thus, a rating of -7 indicated the participant was certain that the sentence had not occurred in the training phase and a rating of 7 indicated that the participant was certain that the sentence had occurred. A graded measure such as this allows for more subtle comparisons than a strict binary-choice measure. Henceforth, the term *familiarity* will be used to refer to this translated response scale.⁴ Familiarity should be higher for repeated than for novel sentences, and higher for category-conforming than for control sentences.

RESULTS

"Yes/No" Responses

The total number of "yes" responses was tallied individually for each of the four test sentence types and then averaged across participants. A maximum of three "yes" responses is possible for both category-conforming sentence types and control sentence types. A maximum of 12 "yes" responses is possible for each of repeated and novel sentence types (three sets of four sentences for each type). To ensure that there was no difference in response pattern between the two counterbalanced groups and to test for an effect of experimental set on "yes"/"no" responses, two $2 \times 3 \times 2$ analyses of variance (ANOVAs) were performed with group (counterbalance A-set or B-set) as a between-subjects factor, and Set (1-3) and sentence type (category-conforming and control for one ANOVA, repeated and novel for the other) as within-subjects factors. (Two separate analyses are required since the range of possible responses differs across sentence types; only comparable sentence types are included in each analysis.) For the analysis with repeated and novel sentences, there was a main effect of sentence type [$F(1,38) = 73.6, p < .001$], no main effect of group or set, and no interactions. Pooling across group and summing across set, participants gave "yes" responses more often to repeated sentences ($M = 8.95$) than to novel sentences ($M = 4.75$), as expected. The mean number of "yes" responses was also greater than chance (6) for repeated sentences [$t(39) = 6.3, p < .001$], and less than chance for novel sentences [$t(39) = -2.8, p < .01$]. For the analysis with category-conforming and control sentences, there was a main effect of sentence type

$[F(1,38) = 4.34, p < .05]$, no effect of group or set, and no interactions. Participants gave “yes” responses more often to category-conforming sentences ($M = 2.3$) than to control sentences ($M = 1.8$), pooling across group and summing across set. The mean number of “yes” responses was also greater than chance (1.5) for category-conforming sentences [$t(39) = 5.10, p < .001$], but did not differ from chance for control sentences [$t(39) = 1.70, p = .11$].

In summary, participants by and large reported that they had heard repeated and category-conforming sentences during training, and they reported that they had not heard novel sentences during training. Their responses to control sentences were not distinguishable from chance, but participants were less likely to respond yes to control sentences than to category-conforming sentences.

Familiarity

For each participant, the derived familiarity scores for the four repeated sentences in an experimental set were averaged together, as were scores for the four novel sentences, yielding a total of four scores for each experimental set (one score each for category-conforming, control, repeated, and novel). In order to ensure that there was no difference between the counterbalanced groups, as well as to test for any difference in scores relating to experimental set, a $2 \times 3 \times 4$ ANOVA was performed with group (counterbalanced A-set or B-set) as a between-subjects factor and set (1–3) and sentence type (category-conforming, control, repeated, and novel) as within-subjects factors. The only main effect to reach significance was sentence type [$F(3,114) = 23.35, p < .001$]. There were no significant interactions, so the data for the comparisons below were pooled across group and averaged across set. Figure 1 shows mean responses to the four sentence types.

Planned comparisons were made between responses to repeated and novel sentences, and between category-conforming and control sentences. Subjects’ familiarity scores for repeated sentences were reliably higher than those for novel sentences. Mean scores for repeated sen-

tences and novel sentences were 3.20 and $-.50$, respectively [$t(39) = 8.01, p < .001$]. The next planned comparison tested the hypothesis that category-conforming sentences would have higher familiarity scores than control sentences, although participants had heard neither sentence type during training. Mean scores for category-conforming and control sentences were 3.3 and 1.3, respectively [$t(39) = 3.11, p < .005$]. Thus, although both sentences were novel, participants were more confident that they had previously heard category-conforming sentences.

Post hoc pairwise comparisons between sentence type conditions revealed several other significant differences (Bonferroni corrected, $p < .01$ for all): category-conforming and control sentences were rated more familiar than novel sentences [$t(39) = 7.23, t(39) = 3.68$, respectively], and repeated sentences were rated more familiar than control sentences [$t(39) = 3.38$]. The only comparison that did not reach significance was between repeated and category-conforming sentences [$t(39) = 0.16$].

In summary, participants did not respond in the same way to all previously unheard sentences, and they responded equally to some unheard (category-conforming) sentences as they did to repeated ones.

DISCUSSION

The overall pattern of results shows that participants were sensitive to the position of words in the training sentences. Repeated, control, and category-conforming sentences matched training sentences in word position in at least two of the three words (initial and final for category-conforming and control; all three positions for repeated sentences). Participants judged these sentences to be more familiar than novel sentences that did not match the positional regularity of the training set. However, the results further show that participants’ representations of the training sentences involved more than encoding word position information. On that dimension, category-conforming and control sentences were equally similar to the set of

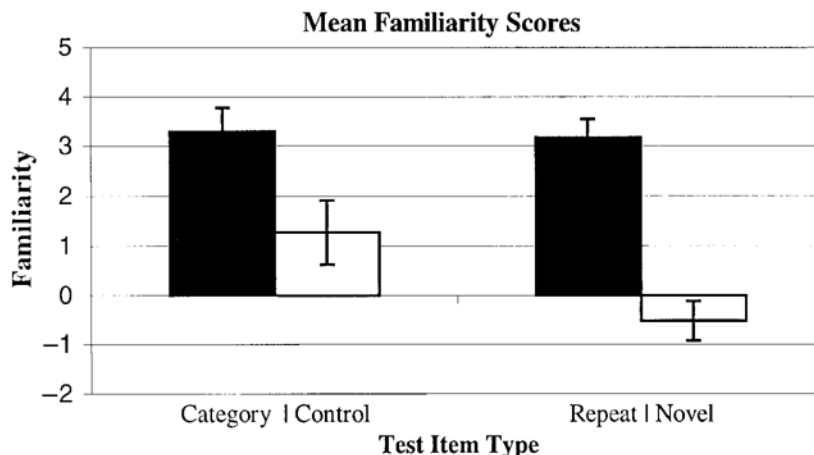


Figure 1. Mean familiarity by sentence type.

training sentences, yet participants were more likely to report having heard category-conforming sentences than they were control sentences, and they were more confident that they had heard category-conforming sentences during training. These results are best explained by an account in which participants remembered the training sentences as sequences of words and categories. Specifically, the lexical co-occurrence properties of the medial words in the full- and partial-paradigm frames led participants to treat *nex*, *kwob*, *zich*, and *pren* (Table 1) as members of the same category. Participants gave relatively high ratings to category-conforming sentences because the medial word—*pren*—was an exemplar of the category whose position in the category-conforming frame (*sook X runk*) conformed to the partial-paradigm sentences. In other words, the word and category pattern *sook CAT runk* was familiar, and this category-based pattern was the basis for their response. Hence, these results suggest that participants carried out a distributional analysis on the repeated list of training sentences and that their later judgments were influenced by the lexical category they induced.

Participants' scores for repeated versus novel sentences were in the predicted direction on both measures, demonstrating that participants followed instructions and that they had paid attention in the training phase. Furthermore, the similarity between familiarity scores for the repeated and category-conforming sentences underscores the degree to which the distributional information influenced participants' perception of the novel category-conforming sentences. Learners scored category-conforming sentences to be just as familiar as the repeated sentences that they actually heard. That is, when the unheard sentences conformed to the category patterns in the training set, sentences they had never heard before were as familiar as those they had.

Interestingly, even for sentences that participants found least familiar overall (novel sentences), the average familiarity score for this type was near the middle of the derived scale and yes/no responses were at chance, reflecting uncertainty rather than out-and-out rejection of the strings. This is perhaps because all sentences, regardless of type, sounded somewhat familiar since they were composed of familiar words. Thus, participants might have been less likely to give a definitive *no* response due to the familiarity of the words themselves.

A comment is in order about the nature of the category that learners formed. Although learners must have formed some kind of abstract category, it is not clear from the present study exactly what kind of category it was. One possibility is that participants formed a category of the medial words (e.g., *nex*, *kwob*, *zich*, and *pren* in Table 1), much like a grammatical category. On this view, category-conforming sentences were familiar because the medial word belonged to the same category as the other medial words that occurred within the same frame during training (in the partial-paradigm sentences). However, another possibility is that participants formed a category of the relevant initial/final frames, on the basis of the patterning of the frames with the same set of medial words. According to that

hypothesis, the frames in the full- and partial-paradigm lists were grouped together because they occurred with the same medial words. On this view, category-conforming sentences were accepted as having been heard before because the initial/final frame (*sook-runk*, in Table 1) belonged to the same category as the other frames that occurred with the common medial words during training. Although it is impossible to determine from this study alone which of these two possibilities is correct, or if both are, both possibilities provide evidence of distributional learning mechanisms that have a significant effect on learners' processing of linguistic input.

In summary, the results of the study suggest that, while listening to sentences in the training phase, participants performed a distributional analysis. The analysis brought together sentences in the full- and partial-paradigm frames by virtue of the lexical co-occurrence properties of the medial words in those lists. This process led to the formation of a category and a representation of sentences of the language that involved the category. The relatively short amount of time that participants were given to learn the sequences—just over 6 min—means that the distributional mechanisms that were involved must have begun to function early in the training phase and suggests that these mechanisms are activated fairly readily. Thus, going beyond previous studies, this experiment shows evidence of categorization mechanisms that function from distributional cues alone.

Although one cannot say exactly what aspects of these stimuli made them amenable to a distributional analysis, in light of the previous studies that failed to find evidence of purely distributional categorization, it is likely that the degree of distributional overlap between sentences (the robust redundant distributional information provided by the frame) has an important effect on how the mechanisms function. One reason that this might be so has to do with the nature of distributional information and distributional analyses. Specifically, a given word can be a word-to-be-categorized (target word) while also being a word that functions as a categorizing environment. To perform an effective analysis, learners must track a target word with respect to all of its environments across sentences; likewise, learners must register a word as an environment for all the relevant target words across sentences. While logically it would be possible for an ideal learner to track words simultaneously as targets and environments, without some way to ground a subset of words (e.g., in a category), it might be difficult for human learners to treat a word in a consistent way across sentences, and this might lead to difficulty in tracking the appropriate distributional contingencies. This may be particularly problematic with two word MN/PQ type sentences where there is no basis for making this distinction in distributional role. Absolute position (e.g., first vs. last) might be more salient as a category cue than lexical co-occurrence patterns in these cases (see Smith, 1966). In contrast, because of redundant distributional cues, the majority of stimuli that were used here provided a natural distinction: those words that

made up a frame and those that occurred in the middle of a frame. This distinction might function like a figure-ground distinction to naturally lead learners to track the patterns of middle words in reference to frames (or frames in reference to middle words), thereby providing a grounding for the distributional analysis. Perhaps what was crucial about the converging cues in prior studies was that they selected a group of words as a target/environment reference point to start distributional learning, not necessarily that they directly (nondistributionally) categorized a set of words. If, here, the initial/final frames played a grounding role, it nevertheless is an open question whether natural language input incorporates functionally equivalent framing features to a significant degree. It is worth noting, however, that researchers have proposed elsewhere that frames play an integral role in language acquisition (Gleitman, 1990; Gleitman, Landau, & Wanner, 1988; Lieven, Pine, & Baldwin, 1997; Olguin & Tomasello, 1993; Tomasello, 1992, 2000).

Extending the present research to the case of natural language learning is subject to the same caveats that apply to most artificial language learning studies. For instance, the patterns created by the sample sentences are much simpler and much more regular than they are in natural language. This is in part because the grammar here is so much simpler, and also because the sentences are not subject to factors that affect word choice (such as semantic content), and other factors that can alter the distributional patterns that occur. Although in principle the same kinds of mechanisms in evidence here could induce category structure from more complex natural input (Cartwright & Brent, 1997; Kiss, 1973; Mintz et al., 1995, 2002; Redington et al., 1998), the parameters needed to make them effective might be outside the range of normal human performance. How would these mechanisms perform when the distributional patterns are not so perfectly predictive of category membership, as is the case in natural languages? What is the relationship between the predictive power of the environments and the number of exposures a learner needs to the relevant sentences? How does learning change if there is evidence for multiple categories? Through further experimentation with more complex grammars, one can begin to address these kinds of questions. It may turn out that in more complex cases distributional information alone is insufficient for categorization, and that categorization mechanisms must rely on converging cues from other sources, similar to the ones shown to be influential in previous studies. In any case, investigating the parameters of these mechanisms will be an important part of understanding what role they might play in learning natural languages.

Carrying out these studies with adults can yield important insights into fully developed adult learning mechanisms. However, to fully assess the role of distributional analyses in normal language acquisition, one needs to investigate how these mechanisms function in very young children. Since categorization is a ubiquitous human process, it is likely that similar mechanisms are in operation for 1-year-olds and younger, but the parameters under

which these mechanisms operate—the issues that the questions at the end of the last paragraph raised—are likely to be different from those of adults.⁵ These differences can have an important effect on how distributional analyses are incorporated into theories of language acquisition, so extending these findings to infants and young children is critical (for related discussions see Braine et al., 1990; Elman, 1993; Newport, 1990).

To conclude, this study adds to the literature in artificial language learning by demonstrating that learners will naturally form categories solely on the basis of distributional patterns. In previous studies evidence for distributional analyses was found only when learners were provided with other sources of categorization information, whereas here learners formed abstract categories solely from distributional patterns within 6 min of exposure. These findings provide evidence of rapidly engaged distributional mechanisms that could play a role in early stages of language acquisition. Further research with adults and very young children is needed to determine the limits of these mechanisms and how they interact with information from other sources in natural language acquisition.

REFERENCES

- BLOOM, L. (1970). *Language development: Form and function in emerging grammars*. Cambridge, MA: MIT Press.
- 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: Erlbaum.
- BRAINE, M. D. S., BRODY, R. E., BROOKS, P. J., SUDHALTER, V., ROSS, J. A., CATALANO, L., & FISCH, S. M. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory & Language*, **29**, 591-610.
- BROWN, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- CARTWRIGHT, T. A., & BRENT, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition*, **63**, 121-170.
- CHAN, C. (1991). *Implicit cognitive processes: Theoretical issues and applications in computer systems design*. Unpublished doctoral dissertation, University of Oxford.
- DIENES, Z., ALTMANN, G. T. M., KWAN, L., & GOODE, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology*, **21**, 1322-1338.
- ELMAN, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, **48**, 71-99.
- FRIGO, L., & McDONALD, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory & Language*, **39**, 218-245.
- GERKEN, L. A., GOMEZ, R. L., & NURMSOO, E. (1999, April). *The role of meaning and form in the formation of syntactic categories*. Paper presented at the Society for Research in Child Development, Albuquerque.
- GLEITMAN, L. R. (1990). The structural sources of verb meaning. *Language Acquisition*, **1**, 3-55.
- GLEITMAN, L. R., GLEITMAN, H., LANDAU, B., & WANNER, E. (1988). Where learning begins: Initial representations for language learning. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey, Vol. 3. Language: Psychological and biological aspects* (pp. 150-193). New York: Cambridge University Press.
- GOMEZ, R. L., & GERKEN, L. A. (1999, November). *Generalization from limited input*. Poster presented at the 40th annual meeting of the Psychonomic Society, Los Angeles.
- KISS, G. R. (1973). Grammatical word classes: A learning process and

- its simulation. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 7, pp. 1-41). New York: Academic Press.
- LIEVEN, E. V. M., PINE, J. M., & BALDWIN, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, **24**, 187-220.
- MARATSOS, M., & CHALKLEY, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. Nelson (Ed.), *Children's language* (Vol. 2, pp. 127-214). New York: Gardner.
- MINTZ, T. H., NEWPORT, E. L., & BEVER, T. G. (1995). Distributional regularities of grammatical categories in speech to infants. In J. Beckman (Ed.), *Proceedings of the North East Linguistics Society 25* (Vol. 2, pp. 43-54). Amherst: GLSA.
- 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.
- MORGAN, J. L., MEIER, R. P., & NEWPORT, E. L. (1987). Structural packaging the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, **19**, 498-550.
- MORGAN, J. L., MEIER, R. P., & NEWPORT, E. L. (1989). Facilitating the acquisition of syntax with cross-sentential cues to phrase structure. *Journal of Memory & Language*, **28**, 360-374.
- MORGAN, J. L., & NEWPORT, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning & Verbal Behavior*, **20**, 67-85.
- NEWPORT, E. L. (1990). Maturation constraints on language learning. *Cognitive Science*, **14**, 11-28.
- OLGUIN, R., & TOMASELLO, M. (1993). Twenty-five-month-old children do not have a grammatical category of verb. *Cognitive Development*, **8**, 245-272.
- REDINGTON, M., CHATER, N., & FINCH, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, **22**, 435-469.
- SAFFRAN, J. R., NEWPORT, E. L., & ASLIN, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory & Language*, **35**, 606-621.
- SHI, R., WERKER, J. F., & MORGAN, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, **72**, B11-B21.
- SMITH, K. H. (1966). Grammatical intrusions in the recall of structured letter pairs: Mediated transfer or position learning? *Journal of Experimental Psychology*, **72**, 580-588.
- TOMASELLO, M. (1992). *First verbs: A case study in early grammatical development*. New York: Cambridge University Press.
- TOMASELLO, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences*, **4**, 156-163.
- WAXMAN, S. R. (1994). The development of an appreciation of specific linkages between linguistic and conceptual organization. *Lingua*, **92**, 229-257.
- WILSON, R. (2000). *Category learning in second language acquisition: What artificial grammars can tell us*. Unpublished master's thesis, University of Arizona.
- WILSON, R., GERKEN, L. A., & NICOL, J. (2000, November). *Artificial grammar research extends to natural language: Implicit learning of categories*. Paper presented at the annual meeting of the Psychonomic Society, New Orleans.

NOTES

1. The summaries focus on the essential, formal connections between Smith's (1966) study and the cited works. For expositional simplicity, differences are glossed over—for example, differences in the number of items in the categories, the specifics about the length of the training phase, the way categorization was assessed, and so forth. The reader is encouraged to refer to the cited works for those details.

2. There are additional studies in artificial grammar learning in which other sources of information were *completely* redundant with distributional cues and in which adults were shown to categorize. For example, Gerken, Gomez, and Nurmsoo (1999) provided learners with phonological cues to lexical categories, whereas Morgan and Newport (1981) and Morgan, Meier, and Newport (1987, 1989) provided completely predictive semantic/referential cues. In these studies it is unclear to what degree, if any, learners were aided by the correlated distributional patterns.

3. It is assumed here that confidence ratings are meaningfully related to participants' knowledge. Circumstances have been reported elsewhere in the artificial grammar-learning literature in which confidence is *not* correlated with learners' accuracy in judging the grammaticality of novel strings. For example, Dienes, Altmann, Kwan, and Goode (1995) reported that for longer strings (e.g., six items), confidence and accuracy were unrelated. However, for shorter, three-item strings, participants' accuracy and confidence were correlated. Dienes et al. cited a similar pattern of results in Chan (1991). Since the sentences in this study are all three-word sequences, there is reason to expect that participants' confidence ratings will reflect knowledge they have induced during training.

4. *Familiarity* is used as a convenient term to indicate how a participant responds to a sentence along this derived scale. Directly asking for a familiarity judgment comes up against the problem that different individuals might focus on different dimensions of the stimulus—for example, whether the words are familiar, whether the order of the first two words is familiar, the last two, and so on—with the result that the judgments measure different things. Determining whether a sentence was played before presumably invokes all dimensions relevant to identity, and thus has a better chance of yielding a uniform response criterion within and between subjects. Therefore, in the reporting of results, "X is judged more familiar than Y" simply means that the derived scores for X were higher than those for Y.

5. Wilson et al. (2000) have found evidence that 17-month-olds can classify words on the basis of morphological co-occurrence cues when provided with convergent cues.

APPENDIX

Complete List of Words Used in Each Experimental Set

Set 1

bool choon daik fen ferd fimp jiv kwob nex noof plif poz
pren pux runk sook vot wug yult zich zim

Set 2

chim cluz dex drak flom foost glak gloff guf neech nep pim
roov rud sec shep sut thok treg vim zid

Set 3

chaib cleb deef feg filt frud glat glek gloost kuch mant nafe
nam nint noz plu sij spik tibe treek voke