Developmental Science 7:5 (2004), pp 567–580

PAPER

A first step in form-based category abstraction by 12-month-old infants

Rebecca L. Gómez[1] and Laura Lakusta[2]

1. *Department of Psychology, University of Arizona, USA*
2. *Departments of Cognitive Science and Psychological and Brain Sciences, The Johns Hopkins University, USA*

Abstract

The present experiments investigate how young language learners begin to acquire form-based categories and the relationships between them. We investigated this question by exposing 12-month-olds to auditory structure of the form aX and bY (infants had to learn that a-elements grouped with Xs and not Ys). Infants were then tested on strings from their training language versus strings from the other language using a preferential-listening procedure. Importantly, the X and Y elements were new at test, requiring infants to generalize to novel pairings. We also manipulated the probability of encountering grammatical structures of the training language by mixing strings from two artificial languages according to 83/17 and 67/33 percentage ratios in Experiment 2. Experiment 1 shows that 12-month-olds are capable of forming categories of X- and Y-elements based on a shared feature and, furthermore, form associations between particular a- and b-elements and these categories. Experiment 2 shows that learning was sustained even when 17% of instances from another language were present during training. However, infants failed to generalize when exposed to a larger percentage of strings from another language. The findings demonstrate that the first step of form-based category abstraction (the ability to generalize based on marker-feature pairings) is in place by 12 months of age.

Introduction

The generative power of human language stems from our ability to perceive category relationships among words in strings. As proof, an experienced English user can easily generalize from a novel string like 'The dax plipped along the moop' to 'Is the dax plipping?' This ability plays a central role in linguistic productivity because once a novel word is categorized, language learners can automatically apply syntactic constraints associated with other words in its category. Given the important role of category information in linguistic productivity, a critical question is how children might achieve such generalization.

A widely held view, referred to as the semantic bootstrapping hypothesis, is that young children discover lexical categories by first noting semantic or referential information.[1] By this view, learners are equipped with knowledge of innate categories, such as noun and verb, as well as knowledge of grammatical functions, such as subject and object (Grimshaw, 1981; Pinker, 1984). Children identify semantic referents in the world by means of perceptual processing and then link these to innate knowledge of syntactic categories and functions. For instance, people and objects are linked to nouns, actions are linked to verbs and agents of actions are linked to subjects. Once such links have been made, children can use distributional information, having to do with the relationships of words to one another in strings, to classify lexical elements that might be ambiguous, such as labels that occur as both nouns and verbs (e.g. 'reply') or words that are more abstract in nature (e.g. 'situation' or 'think'). Distributional learning involving form-based cues is peripheral on this view.

A very different view assumes that distributional relationships among form-based cues are central to category-based abstraction (Braine, 1987; Gerken, Landau & Remez, 1990; Gleitman & Wanner, 1982; Morgan & Demuth, 1996; Morgan & Newport, 1981; Redington, Chater & Finch, 1998). Examples of such cues are relative location

[1] See Gleitman, 1990; Landau & Gleitman, 1985 for an account based on syntactic, rather than semantic, bootstrapping.

Address for correspondence: Rebecca L. Gómez, Department of Psychology, University of Arizona, Tucson, AZ 85721-0068, USA; e-mail: rgomez@u.arizona.edu

of words in strings, phonological regularities within words of a class and co-occurrence relations between classes. With regard to phonological regularities within a class, functor categories tend to have shorter vowel durations, weaker amplitudes and simplified syllable structure compared to lexical categories such as noun and verb (Morgan, Shi & Allopena, 1996; Shi, Morgan & Allopena, 1998). Newborn infants are sensitive to such differences (Shi, Werker & Morgan, 1999) and by 7 months of age, infants can recognize and track specific functor elements in running speech (Höhle & Weissenborn, 2003). Nouns and verbs are also distinguishable by means of phonological cues. In English, for instance, nouns tend to have more syllables than verbs and also tend to receive first syllable stress (Kelly, 1992). Nouns and verbs also tend to differ in vowel quality such that high frequency nouns are more likely to have front than back vowels, and the reverse holds for high frequency verbs (Sereno & Jongman, 1990). Furthermore, English language users appear to be sensitive to these differences (Cassidy & Kelly, 1991; Sereno & Jongman, 1990).

Distributional cues play a secondary role on the semantic bootstrapping account. However, such cues may be a more important source of information for category-based abstraction than previously thought. While pre-linguistic infants might be limited in their ability to map semantic knowledge on to word forms, they are actively engaged in, and increasingly adept at, processing the sound properties of language (Juszczyk, 1997). If infants are able to identify categories in the speech stream by means of their phonological properties, they might then use this information to learn the predictive relationships between categories. In English, for example, children must learn that 'the' and 'a' precede nouns and not verbs, whereas 'will' and 'can' precede verbs but not nouns. An infant who has learned that particular functors predict particular lexical forms (i.e. one who has identified categories in speech and the relationships between them) will have a considerable advantage with respect to the later task of mapping between meaning and form, compared to the toddler who only begins this process once semantic knowledge is more fully in place (Gómez & Gerken, 2000; Naigles, 2002).

Indeed, there is mounting evidence that infants become increasingly sensitive to the syntactic constraints of functor elements in the second year of life. For instance, between 15 and 16 months of age, English-speaking infants become sensitive to the syntactic position of particular functors within sentences (Shady, 1996; Shady, Jusczyk & Gerken, 1998) and German-speaking infants are able to use co-occurrence restrictions between determiners and nouns to classify novel nouns syntactically (Höhle *et al.*, 2004). By 18 months, English-speaking infants are able to track syntactic relationships between functor elements, as evidenced by their ability to discriminate '*is* quickly runn*ing*' from '*can* quickly runn*ing*' (Santelmann & Jusczyk, 1998; see also Gómez, 2002). Thus, there is every reason to believe that distributional learning may play a more central role in syntactic category abstraction than previously thought.

If so, what learning processes might be involved in a distribution-based account? Artificial language studies with adults suggest several possibilities (Braine, 1987; Frigo & McDonald, 1998; Gerken, Nurmsoo & Gómez, 1999; Gómez & Schvaneveldt, 1999; Mintz, 2002; Smith, 1969; Wilson, 2002). In one approach to examining form-based abstraction, grammatical classes are given arbitrary labels such as 'a', 'X', 'b' and 'Y' (Braine, 1987; Frigo & McDonald, 1998; Gerken *et al.*, 1999; Gómez & Schvaneveldt, 1999; Smith, 1969; Wilson, 2002). Words from these classes are then combined to form legal phrases. For instance, aX and bY might be legal whereas aY and bX are illegal. Learners are exposed to most, but not all, aX and bY phrases, and then are tested to see if they will discriminate legal phrases they have not yet encountered from illegal ones. To give an example, imagine that a-elements correspond to 'a' and 'the', and b-elements to 'will' and 'can' (see Table 1). Learners will only be successful at discriminating a new legal phrase (e.g. 'a cat' denoted by the empty cell in Table 1) from an illegal one ('a eat') if they have learned that a-elements go with nouns (the Xs), but not with verbs (the Ys). As in natural language, the functor-like a- and b-categories have fewer members than lexical-like Xs and Ys. Category-size asymmetries of this type abound in language and may play an important role in learning by highlighting anchor points for distributional analysis (Valian & Coulson, 1988).

Interestingly, although learners readily acquire the legal positions of words with respect to which occur first versus second (Smith, 1969), categories and their

Table 1 *A paradigm for investigating category abstraction. Learners are exposed to the pairings shown below except for those denoted by empty cells. Learners are then tested to see if they will generalize correctly to the withheld strings (denoted by empty cells)*

	X_1	X_2	X_3	X_4	X_5
a_1 = the	boy	girl	ball	dog	cat
a_2 = a	boy	girl	ball	dog	
	Y_1	Y_2	Y_3	Y_4	Y_5
b_1 = will	jump	run	play	sleep	eat
b_2 = can	jump	run	play	sleep	

relationships (i.e. that words belong to particular a, b, X, and Y classes, and that a-words go with Xs and not Ys) are virtually impossible to acquire unless some subset of the X- and Y-category members are marked with salient conceptual or perceptual cues (Braine, 1987; Frigo & McDonald, 1998; Gerken *et al.*, 1999; Gómez & Schvaneveldt, 1999; Wilson, 2002). This is not surprising given the importance of correlated cues in learning (Billman, 1989; Morgan, Meier & Newport, 1987).

Braine (1987) investigated whether correlating a conceptual referential cue (feminine or masculine) with a subset of the X and Y category members would aid learners in acquiring relations to a- and b-elements. Partial correlation of the conceptual cue did help. Participants not provided with the cues for distinguishing X and Y category members were unsuccessful at abstracting category relations. According to Braine (1987, see also Frigo & McDonald, 1998), there are two essential steps in aX bY category abstraction. Learners must first associate a- and b-elements with cues differentiating X and Y categories. They can then categorize a- and b-elements based on their co-occurrence (or distributional) relationships to the X- and Y-cues. In the second step, learners group (or categorize) a- and b-elements by merit of their joint association with particular distinguishing cues. Once a- and b-categories are formed, learners can rely on memory for a pair they have heard (e.g. 'the cat' in Table 1) and the fact that a_1 ('the') and a_2 ('a') belong in the same category to make inferences about a pair they have not heard (e.g. 'a cat'). By this view, Step 1 learning is evidenced by the ability to discriminate correct from incorrect pairings of functional and lexical test items with distinguishing cues present. Step 2 learning is evidenced by discrimination of test items in the absence of distinguishing cues. Frigo and McDonald (1998) found that there were no learners who had mastered Step 2, but not Step 1. Thus these steps appear to be serial in nature.

Frigo and McDonald reported similar findings to Braine (1987) with perceptual, rather than conceptual cues. In this study, adult learners had to categorize particular a- and b-elements based on their relationships to members of X- and Y-categories. Cues took the form of word beginnings and endings such that members of X began with *kais* and ended with *rish* as in *kaisemilrish* whereas members of Y began with *wan* and ended with *glot* as in *wanersumglot*. Importantly, only six out of ten category members were augmented with these cues. Once categories were made distinguishable in this manner, learners were able to induce grammatical combinations both when cues were present (e.g. *kaisaudorish* or *wanorbinglot*) and when they were absent (e.g. for *faranu* or *roosa*), consistent with Braine's two-step model. Wilson (2002) further investigated such learning with a set of Russian feminine and masculine nouns in two syntactic cases. As in Frigo and McDonald's studies, adult learners were unable to learn the category relations unless a subset (three out of six) of the nouns of each gender had distinct cues.

What about younger learners? Recent research with 17-month-old infants from an English-speaking environment found that they too were able to induce Russian gender categories (Gerken, Wilson & Lewis, 2003; see also Gerken, 2002). Gerken *et al.* created a set of stimuli in which six feminine lexical stems appeared with the case endings *–oj* and *–u* and six masculine stems appeared with the case endings *–ya* and *–em*. Case endings in these experiments were equivalent to a- and b-elements. Additionally, cues distinguishing Xs and Ys were present for a subset of category members. For instance, three out of six of the X-words contained the derivational suffix *–k* (e.g. *polkoj*, *polku*) whereas three of the Y-words contained the suffix *–tel* (e.g. *zhitelya*, *zhitelyem*). Seventeen-month-olds first familiarized with a subset of stimuli and were then tested to see if they would attend differentially to novel aX and bY stimuli versus ungrammatical aY and bX ones. Infants were able to generalize to new grammatical test items in which cues were entirely absent, e.g. generalizing to words such as *vannoj* and *pisarem* after hearing *vannu* and *pisarya*. This is critical for showing that infants of this age have mastered both steps in category-based abstraction. They have gone beyond the first step of associating functor-like markers (particular case endings) with cues distinguishing X- and Y-category members (particular derivational suffixes) to the second one of grouping or categorizing the markers such that having heard *vannu*, they were able to treat *vannoj* equivalently.

The ability to abstract categories from distributional information in speech is a significant milestone in cognitive and language development, not only for what it implies about early abstraction abilities but also because of its potentially important contribution to syntactic development. Thus it is necessary to ask about the developmental trajectory of such learning. Do infants younger than 17 months of age show similar abilities? Gerken *et al.* (2003) reported in a footnote that 12-month-olds were unable to generalize to new grammatical items, even when some distinguishing cues were present. However, the materials in their study were designed primarily for showing more advanced mastery of category abstraction (Step 2 learning) and thus did not directly test whether 12-month-olds are capable of achieving the first step in category-based abstraction of associating functor-like a- and b-elements (or markers) with X- and Y-categories based on their distinguishing cues. The purpose of the present study was to investigate the abstraction abilities of this younger age group.

We know that by 7 and 12 months of age, infants are able to abstract patterns from artificial grammars as evidenced by their ability to discriminate grammatical from ungrammatical strings in new vocabulary (Gómez & Gerken, 1999; Marcus, Vijayan, Bandi Rao & Vishton, 1999). However, the kind of abstraction demonstrated by these studies involved noting patterns of repeating or alternating elements (Gómez, Gerken & Schvaneveldt, 2001), a very different kind of abstraction than might be involved in learning form-based categories (Gómez & Gerken, 2000). We also know from Gerken *et al.* (2003) that 12-month-olds do not show Step 2 learning. However, these younger infants might be able to engage in a more preliminary form of category-based abstraction, therefore providing clues to the developmental trajectory of such learning. Thus we asked in Experiment 1 whether 12-month-olds would learn the relationship between functor-like a- and b-words and X- and Y-categories based on their distinguishing features. If so, they should subsequently generalize to a- and b-words paired with novel X- and Y-elements. We chose 12-month-olds because of research suggesting increasing sensitivity to functional elements by this age (Shady, 1996; Shafer, Shucard, Shucard & Gerken, 1998). In Experiment 2, we investigated the robustness of this learning by asking how discrimination would be affected by adding 'ungrammatical' instances during training. This manipulation should provide insights into infants' abilities to learn about grammatical categories when inconsistencies are present in linguistic input.

Infants in all the experiments were exposed to one of two training languages. In Experiment 1, one language consisted of aX and bY pairings. The other training language consisted of aY and bX pairs. Xs were instantiated as disyllabic words and Ys were monosyllabic. During testing, infants were exposed to *new* phrases from their training language versus phrases from the other language. Thus some phrases conformed to the training language whereas others did not. Importantly, all X- and Y-words were novel at test.[2]

Infants sensitive to the category relations defined by their training language should exhibit different listening times to legal versus illegal category pairings. We cannot predict the direction of preference with absolute certainty (whether toward novel or familiar structure), but given previous research involving generalization (e.g. Gómez & Gerken, 1999) there is some reason to expect infants to orient longer towards familiar grammatical versus unfamiliar ungrammatical pairings. Regardless of the type of effect elicited (familiarity versus novelty), orienting reliably longer to one stimulus type or the other would suggest that infants have become sensitive to the relationships between specific functor-like (or marker) elements and the features differentiating X- and Y-categories (Step 1 learning).

Experiment 1

Method

Participants

Twenty-four infants with an average age of 12 months and 16 days (11 months 18 days to 13 months 21 days) participated in the experiment. Equal numbers of male and female infants were run in each condition. Eleven other infants were tested but were not included for the following reasons: unable to complete the procedure due to excessive fussiness (5); being treated for ear infections at the time of testing (1); weighed less than 2500 grams at birth or had gestational terms less than 36 weeks (2); had health histories putting them at risk for normal cognitive functioning, e.g. Downs syndrome, prenatal drug exposure (2); and insufficient number of test trials (1) (in the interest of obtaining stable estimates of looking time, a minimum of three trials for each language were required). Eighteen of the 24 infants were recruited in the Baltimore metropolitan area.

Training stimuli

Infants were exposed to strings from one of two training languages, L1 or L2. Each language contained two a-elements, two b-elements, six X-elements and six Y-elements (see Table 2). The a- and b-elements were of the form vowel, consonant, consonant (*alt*, *ush*, *ong* and *erd*) and were thus similar in structure. X-elements were disyllabic, Y-elements were monosyllabic. The elements were combined to form grammatical phrases. In L1, *alt* and *ush* were paired with two-syllable words (e.g. *alt coomo*, *ush loga*) and *ong* and *erd* were paired with one-syllable words (e.g. *erd deech*, *ong jic*). The opposite was true for L2 where *alt* and *ush* were paired with one-syllable words and *ong* and *erd* were paired with two-syllable words.

More specifically, L1 consisted of 12 grammatical aX and 12 grammatical bY phrases (likewise L2 consisted of

[2] Note that instead of using the paradigm shown in Table 1, where certain pairings were withheld, infants were exposed to completely new X- and Y-elements at test. This allowed us to ask whether infants would abstract the relations between markers and features distinguishing X- and Y-categories.

Table 2 *Training vocabulary*

Category			
a	b	X[a]	Y[b]
alt	*ong*	*coomo*	*deech*
ush	*erd*	*fengle*	*ghope*
		kicey	*jic*
		loga	*skige*
		paylig	*vabe*
		wazil	*tam*

[a] 'X' elements are disyllabic words.
[b] 'Y' elements are monosyllabic words.
Note: L1 phrases are formed by aX and bY pairings. L2 phrases take the form aY and bX.

Table 3 *Test vocabulary*

Category			
a	b	X[a]	Y[b]
alt	*ong*	*bevit*	*vot*
ush	*erd*	*meeper*	*pel*
		gackle	*tood*
		roosa	*rud*
		nawlup	*biff*
		binnow	*foge*

[a] 'X' elements are *novel* words, yet they remain disyllabic.
[b] 'Y' elements are *novel* words, yet they remain monosyllabic.
Note: L1 phrases are formed by aX and bY pairings. L2 phrases take the form aY and bX.

12 grammatical aY and 12 grammatical bX phrases). Thus, each language contained 24 unique phrases, which were then combined to form grammatical strings. L1 strings took the form, S → [aX] [bY] or S → [bY] [aX]. L2 strings took the form S → [aY] [bX] or S → [bX] [aY]. Altogether there were 288 possible strings in each of the training languages, 72 of which occurred once during training. A subset of 72 strings was used so that the training phase lasted only 3 minutes. Strings were selected such that each of the 24 phrases in the training language was presented six times during training and so that the order of phrases within strings (e.g. [aX] [bY] versus [bY] [aX]) was counterbalanced.

Phrases were recorded by a trained female speaker using a rising intonation from the first to the second word. Two-syllable words received stress on the first syllable. The stimuli were recorded using Sound Edit software. Individual word tokens were extracted and then combined to form phrases. Phrases were combined to form strings. The process of constructing strings in both languages from the same tokens ensures that any differences observed cannot be attributed to acoustic distinctions between the languages. Approximately 250 milliseconds of silence separated each phrase and 500 milliseconds of silence separated each string.

Test stimuli

In the test stimuli, the a- and b-words (*ong*, *erd*, *alt* and *ush*) were identical to those used in the training stimuli, however X- and Y-words were all novel: infants did *not* hear these words during training (see Table 3). These elements were combined to form 48 phrases: 12 aX, 12 bY, 12 aY and 12 bX, such that there were 24 phrases that were grammatical for L1 and 24 phrases that were grammatical for L2.

The phrases that were grammatical for L1 were combined to form 24 grammatical strings. Half (12) of the strings took the form S → [aX] [bY] and half (12) took the form S → [bY] [aX]. The 24 grammatical L1 strings were then divided into two stimulus sets (L1-test1 and L1-test2). Each test stimulus contained six S → [aX] [bY] strings and six S → [bY] [aX] strings for a total of 12 strings. The L2 test stimuli were constructed in the same manner (see Table 4).

The four stimulus sets were each presented three times during the test for a total of 12 test trials. The test trials were divided into three blocks so that each stimulus was presented once in each block. Test strings were recorded in the same manner as the training strings. Each stimulus lasted approximately 34 seconds.

Procedure

Each infant was tested individually while seated on the care-giver's lap in an enclosed booth using the head-turn preference procedure (see Kemler Nelson *et al.*, 1995). An observer outside the test booth monitored the infant's looking behavior using a button box connected to an Apple Powermac. The experimental control program initiated trials and scored head-turn responses. To eliminate bias, both care-giver and observer listened to masking stimuli over headphones. During training, stimuli were presented simultaneously from two loudspeakers located on either side of the infant. The infant's gaze was directed first toward a blinking middle light then toward one of two blinking sidelights (There was a light below each loud speaker). When the infant looked away from the sidelight for two seconds, his or her gaze was again directed toward the middle. There was no relationship between lights and sound during training.

During the test, each trial began with the light blinking at center. Once the infant fixated on the center light the experimenter pressed a button extinguishing it. This action initiated blinking of one of the sidelights (the one

Table 4 *Test materials*

Test sets			
L1-test1	L1-test2	L2-test1	L2-test2
[*alt nawlup*] [*ong pel*]	[*alt roosa*] [*ong tood*]	[*alt biff*] [*ong meeper*]	[*alt rud*] [*ong gackle*]
[*alt gackle*] [*ong foge*]	[*alt meeper*] [*ong biff*]	[*alt vot*] [*erd gackle*]	[*alt pel*] [*ong nawlup*]
[*alt bevit*] [*erd tood*]	[*alt binow*] [*erd rud*]	[*alt tood*] [*ong binnow*]	[*alt foge*] [*erd roosa*]
[*ush binnow*] [*erd vot*]	[*ush gackle*] [*erd foge*]	[*ush rud*] [*ong roosa*]	[*ush vot*] [*erd meeper*]
[*ush roosa*] [*ong rud*]	[*ush nawlup*] [*ong vot*]	[*ush pel*] [*erd nawlup*]	[*ush tood*] [*erd binnow*]
[*ush meeper*] [*erd biff*]	[*ush bevit*] [*erd pel*]	[*ush foge*] [*erd bevit*]	[*ush biff*] [*ong bevit*]
[*ong vot*] [*alt roosa*]	[*ong rud*] [*ush roosa*]	[*ong gackle*] [*ush vot*]	[*ong roosa*] [*ush rud*]
[*ong tood*] [*ush bevit*]	[*ong pel*] [*alt bevit*]	[*ong bevit*] [*alt rud*]	[*ong meeper*] [*alt vot*]
[*ong biff*] [*ush gackle*]	[*ong foge*] [*ush meeper*]	[*ong nawlup*] [*ush tood*]	[*ong binnow*] [*ush pel*]
[*erd foge*] [*alt meeper*]	[*erd biff*] [*ush binnow*]	[*erd binnow*] [*alt pel*]	[*erd nawlup*] [*ush foge*]
[*erd pel*] [*alt binow*]	[*erd tood*] [*alt gackle*]	[*erd meeper*] [*alt foge*]	[*erd gackle*] [*alt tood*]
[*erd rud*] [*ush nawlup*]	[*erd vot*] [*alt nawlup*]	[*erd roosa*] [*ush biff*]	[*erd bevit*] [*alt biff*]

Note: Test strings were presented in sets (e.g. L1-test1). Notice, X- and Y-categories consist of new one- and two-syllable words. The strings in the test sets were presented in a random order.

associated with the source of sound for that trial). When the infant turned his/her head in the direction of the sidelight by 30 degrees, the test-set for that trial played until the infant looked away for 2 seconds (or until the test stimulus ended). The observer recorded the direction of the infant's head turns. The computer program tracked looking times, the amount of time looking away from the source of sound (terminating trials after 2 seconds) and controlled the randomization and presentation of stimuli. A significant difference in listening time to strings from the training language versus the other language would indicate that infants have acquired some sensitivity to the marker-feature pairings defined by their training language.

Results and discussion

Phrases lasted approximately two seconds. Under the assumption that infants must register a minimum of one phrase in a test stimulus to determine grammaticality, listening times of two seconds or greater only were included in the analyses. By this criterion 13% of the trials (38 out of 288) were discarded. The dependent measures were median listening times to strings from the training language versus those from the other language. We chose to use medians rather than means as the most stable measure of central tendency given the variability inherent in infant listening-time data. Preliminary analyses showed that there were no effects of gender or training language, thus medians were collapsed over these variables. We used non-parametric tests for these and subsequent analyses as these were deemed most appropriate for median data.

The median difference in listening time to strings from the training language versus the other language was 1.618. We computed exact nonparametric confidence intervals around the difference between medians using the Hodges-Lehman method (Lehman, 1975). The 95.1% confidence interval was 0.665–2.605. A Wilcoxon Signed Ranks Test showed that infants listened significantly longer to strings from their training language than to strings from the other language ($T = 251$, $p = .004$). Eighteen out of 24 infants showed this pattern, suggesting that they had acquired some sensitivity to the category-based structure of their training grammar.

The ability to discriminate legal from illegal marker-feature pairings, despite the fact that X- and Y-elements were novel at test, reflects sensitivity to the co-occurrence relations between markers and X- and Y-categories based on their distinguishing features. Such learning is striking given its complexity – infants had to track four similar sounding markers, associate them with particular distinguishing features and generalize to pairings containing novel words. The fact that infants were able to generalize to novel X- and Y-elements suggests that learning was to some degree abstract (involving grouping of the X- and Y-elements according to syllable number). Does such grouping count as *categorization*? We would argue 'yes' to the extent that categorization involves distinguishing elements according to some feature or set of features. When individuals treat elements similarly, they are responding to these elements *as if* they were members of the same category.

Our next step in investigating such learning was to test its limits. Linguistic input is noisy. All children are exposed to inconsistencies of one type or another in adults' informal speech, in children's own ungrammatical utterances and in the ungrammatical utterances of other learners (such as playmates and siblings). Inconsistencies also occur naturally in language, e.g. in English the degree to which verbs take the regular *–ed* ending for the past tense, or in Spanish the extent to which feminine nouns end in *–a*. Other instances of noise in linguistic input are less widespread, such as when deaf children are exposed to non-native ASL signers (Newport, 1999; Ross & Newport, 1996; Singleton & Newport, 1994) or when normal hearing children are exposed to pidgin languages. In all of these instances, children must distinguish grammatical from ungrammatical language data, and they must generalize beyond the data to which they are exposed.

Thus, in Experiment 2 we investigated whether young learners are able to separate more probable from less probable structure by exposing them to artificial languages with varying degrees of probabilistic structure. In Condition 83/17, approximately 83% of the training strings were from the infants' 'predominant' training language, whereas 17% of the strings were from the other language. In Condition 67/33, the split between the predominant and non-predominant training languages was 67% and 33%, respectively. Infants who are able to distinguish more probable from less probable structure should show the same pattern of discrimination observed in Experiment 1 (from now on the 100/0 condition). However, for such learning to play a useful discriminatory role, it should begin to break down at some point. Because of the greater prevalence of consistent relations in Condition 83/17 compared to Condition 67/33, we anticipated a higher rate of learning in the former than in the latter condition.

Experiment 2

Method

Participants

The participants were 48 infants (half were run in Condition 83/17, the other half in Condition 67/33). There were 12 male and 12 female infants in each condition. Infants were an average age of 12 months, 3 days in Condition 83/17 (11 months 13 days to 12 months 28 days) and 12 months, 2 days (11 months 3 days to 12 months 28 days) in Condition 67/33. Twenty-one additional infants were tested (13 in Condition 83/17), but were not included for the following reasons: were unable to complete the procedure due to excessive fussiness (6 in Condition 83/17 and 8 in Condition 67/33); were being treated for ear infections at the time of testing (1 in Condition 83/17); weighed less than 2500 grams at birth (1 in Condition 83/17); had older siblings diagnosed with language delay (2 in Condition 83/17); had health histories putting them at risk for normal cognitive functioning (e.g. one child in Condition 83/17 had suffered a stroke); and participated in an insufficient number of test trials (2 in Condition 83/17). All infants were recruited in the Tucson metropolitan area and were from monolingual English-speaking families.

Training stimuli

We inserted irregularities into the two training languages from Experiment 1 by swapping X- and Y-elements. In Condition 83/17, for instance, every occurrence of one X-word (*coomo*) and one Y-word (*deech*) was swapped so that in L1 the pairs *alt-coomo* and *erd-coomo* became *alt-deech* and *erd-deech*, whereas *ong-deech* and *ush-deech* became *ong-coomo* and *ush-coomo*. Therefore in L1, a-words occurred with two-syllable words 83% of the time and with one-syllable words 17% of the time. Similarly, b-words occurred with one-syllable words 83% of the time and with two-syllable words the remaining 17% of the time.[3] In Condition 67/33, two of the X- and Y-elements were swapped so that in L1, a-words occurred with two-syllable words 67% of the time and with one-syllable words 33% of the time, whereas b-words occurred with one-syllable words 67% of the time and with two-syllable words the remaining 33%. Half of the infants in each experiment were trained on strings that were predominantly from L1. The other half were trained on strings that were predominantly from L2.

Test stimuli

The test stimuli were identical to those used in Experiment 1. Thus, the X- and Y-words were all novel.

Procedure

The procedure was identical to that used in Experiment 1.

Results and discussion

As in Experiment 1, we discarded trials less than 2 seconds in duration. By this criterion, 8% (or 24 out of 288) trials were discarded in Condition 83/17 and 17% (or

[3] Another method of inserting irregularities would have been to swap a percentage of all of the pairs (instead of just one pair) so that all pairs would adhere to the predominant language 83% of the time and to the non-predominant language the remaining 17% of the time. This method might be more like inserting random noise whereas the method used here resulted in systematic irregularities. We chose to insert systematic noise because we judged it to be more salient, thus presenting a greater challenge to learners.

Table 5 *Median differences and 95.1% confidence intervals for listening times (in seconds) to strings from the predominant versus the non-predominant training language. Infants showed significant discrimination even when 17% of the strings encountered during training were from the non-predominant training language*

Probability ratio	Median difference	95.1% confidence interval
Exp. 1: 100/0	1.618*	(0.665–2.605)
Exp. 2: 83/17	1.248*	(0.130–2.850)
Exp. 2: 67/33	−0.125	(−0.925–0.835)

Note: * Listening time differences in these conditions were statistically significant, $p \leq .05$.

49 out of 288) trials were discarded in Condition 67/33. Preliminary analyses showed that there were no effects of gender or training language, thus the data were collapsed over these variables. Median listening time differences to strings from the predominant training language versus those from the other language are shown in Table 5, along with exact non-parametric confidence intervals. A Wilcoxon Signed Ranks Test in Condition 83/17 showed that infants listened significantly longer to strings from their training language than to strings from the other language ($T = 223$, $p = .037$). Seventeen out of 24 infants showed this pattern. In contrast, infants in Condition 67/33 failed to discriminate strings from the two languages, ($T = 145$, $p = .8996$). Only 10 out of 24 infants showed a preference for strings from the predominant training language.

We next compared median differences in listening times to strings from the predominant versus the non-predominant language across experimental groups. A Mann-Whitney U test (appropriate for use with independent groups) showed a significant decrease in listening time differences for infants in Condition 100/0 versus Condition 67/33, ($U = 173.5$, $p < .018$, 95.2% $CI = .210 – 3.060$). However, there was no significant decrease in listening time differences for infants in Conditions 100/0 versus 83/17, ($U = 265$, $p < .635$, 95.2% $CI = -1.230 – 1.860$). The decrease in listening time differences for Conditions 83/17 versus Condition 67/33 was marginally significant, ($U = 201$, $p < .073$, 95.2% $CI = -.150 – 2.840$), suggesting that learning decreases gradually with increases in noise.

Experiment 1 showed that 12-month-olds are able to use marker-feature pairings to generalize to new category members. Experiment 2 further explored the extent to which young learners are able to separate more probable from less probable structure during form-based category abstraction. Learning appears to be fairly robust given that infants were able to tolerate some degree of irregularity in the input, showing learning even when 17% of the phrases encountered during training were from an alternate language. Generalization diminished, however, with the increased irregularity in Condition 67/33, demonstrating constraints on learning. We will explore the form such constraints might take in the general discussion.

General discussion

The purpose of the present experiments was to investigate form-based category abstraction in an artificial language-learning paradigm. Such generalization is important because once a novel word is categorized, learners can automatically extend it by means of the same syntactic constraints that apply to other words in its category. Although semantic information is likely to play an important role in such learning (Grimshaw, 1981; Pinker, 1984), there is every reason to believe that sensitivity to form-based cues is also instrumental. Language is rich in perceptual cues to syntactic structure (Cassidy & Kelly, 1991; Kelly, 1992; Sereno & Jongman, 1990) and infants are adept at tracking such cues in language (Jusczyk, 1997).

For learning to succeed, however, lexical categories must be differentiated by means of distinguishing cues, as when learners distinguish nouns and verbs by means of their characteristic stress patterns (Kelly, 1992), number of syllables (Cassidy & Kelly, 1991), or vowel quality (Sereno & Jongman, 1990). According to Braine (1987; see also Frigo & McDonald, 1998), there are two important steps in form-based category abstraction. Learners first note co-occurrence relationships between functional elements and cues differentiating lexical categories. Learners may then group (or categorize) functional elements based on their co-occurrence (or distributional) relationships to lexical cues. Once functional elements are grouped, they alone are sufficient for cuing category membership, as in the case of the word ‘the’ in English, which signals an upcoming noun even when distinguishing cues such as stress, syllable number or vowel quality are absent. In an experimental paradigm, Step 1 learning is evidenced by the ability to discriminate correct from incorrect pairings of functional and lexical test items with distinguishing cues present. Step 2 learning is evidenced by discrimination of test items in the absence of distinguishing cues.

Consistent with the mastery of Step 2 conditions, Gerken *et al.* (2003) found that 17-month-olds from an English-speaking environment were able to induce Russian gender categories when distinguishing cues were present during training, but absent in the test stimuli.

However, 12-month-olds were not able to generalize under these conditions. Yet there is some evidence for abstraction at younger ages (e.g. Gómez & Gerken, 1999; Marcus *et al.*, 1999), suggesting important developmental changes in generalization abilities between 12 and 17 months of age. The purpose of the present study was to investigate the abstraction abilities of these younger infants, who, by 12 months old, are sensitive to the kind of information that can be used in Stage 1 learning, such as frequently occurring function morphemes (Höhle *et al.*, 2004; Shady, 1996; Shafer *et al.*, 1998) and phonological regularities in speech (Juszcyk, 1997). Investigating the abstraction abilities of these younger infants should provide important clues to the developmental trajectory of such learning.

We arbitrarily defined four categories (a, b, X and Y). The a- and b-elements were smaller in number and were reliably associated with Xs and Ys, thus playing the role of markers or functional elements (see Valian & Coulson, 1988). Xs and Ys were differentiated via syllable number. We asked whether 12-month-olds would learn the marker-feature relationships and generalize these to *new* X and Y vocabulary at test (thus showing evidence for Step 1 learning). The answer was affirmative. Infants were able to discriminate legal from illegal marker-feature pairings (exhibited by longer looking times to the former than the latter). These findings are noteworthy given the brief training period (three minutes) and the complexity of the artificial language. Infants had to track four similar sounding markers and associate them with particular distinguishing features. The fact that infants were able to generalize to novel X- and Y-elements suggests grouping of these elements according to syllable number and hence categorization by means of this feature. It is important to stress that unlike previous studies, the X- and Y-elements were all novel at test, whereas the features used to group X- and Y-elements in previous studies were physically identical. For instance, for the Step 1 learning demonstrated in Frigo and McDonald (1998), a subset of X-words began with *kais* and ended with *rish*, whereas in Gerken *et al.* (2003) X-words contained the derivational suffix *–k*. Importantly, the 12-month-old infants in this study were not simply learning associations between a- and b-elements and physically identical features. Rather, they were generalizing based on the abstract feature of syllable number, demonstrating that they are capable of categorizing at a level at least one step removed from physical identity. Such generalization is an important precursor to that shown by the older infants in the Gerken *et al.* study, who, by 17 months old, can form a- and b-categories comprised of elements with no common features other than their co-occurrence patterns with X- and Y-categories.

Infants in Experiment 2 of our study also showed learning when some percentage of the training strings originated from another language (17% in the 83/17 signal-to-noise Condition). Infants in this condition had to discriminate new strings obeying the rules of the predominant training language from strings generated by the other language. The results are important for determining whether infant learners are equipped to tolerate some degree of inconsistency in their linguistic input. Infants were indeed able to focus on the predominant pattern in their training language and generalized to new strings on this basis. There appear to be limits on such learning, however. Infants exposed to greater noise in Condition 67/33 failed to show learning.

These findings pose intriguing questions with regard to possible constraints on learning. Were infants in Condition 83/17 learning two forms of structure simultaneously or only the more predominant abstract structure? According to the first possibility, infants were learning specific marker-word phrases from the non-predominant language (there were only two of these) and the more abstract pairings of markers and features (i.e. syllable number) from the predominant language. According to the second possibility, infants were ignoring phrases from the non-predominant language entirely. Because infants were tested on their ability to generalize to new marker-word phrases (rather than old marker-word phrases), we are unable to distinguish these explanations in the present studies.

What about learning in Condition 67/33? Infants in this condition were clearly not generalizing the marker-feature pairing. Nor were they engaged in learning two forms of structure simultaneously (abstract marker-feature pairings and specific marker-word phrases), otherwise they would have shown discrimination on the test as in Condition 83/17. An alternative possibility is that the greater presence of phrases from the non-predominant language disrupted learning entirely. Yet another possibility is that infants learned *only* specific marker-word phrases from the non-predominant language. However, we are unable to distinguish these possibilities with the present data because we did not test infants on marker-word phrases from training. Nevertheless, the latter possibility is consistent with studies on adult categorization in which learners are more likely to generalize a prototype when category instances (or exemplars) are many, but remember specific exemplars when category instances are few (see Homa, 1973). By this logic, markers in Condition 83/17 were paired with a large enough number of different X- and Y-words to permit generalization to marker-feature pairings, but in Condition 67/33, markers were paired with a smaller number of Xs and Ys. Thus, the tendency to generalize in this latter condition may not have been great enough to override learning of particular marker-word phrases, leading to poor discrimination on the generalization test.

Although the present findings do not distinguish these competing explanations, they clearly demonstrate that infants in the 83/17 Condition were able to abstract the predominant structure of their training language even when there was some inconsistency present. This finding is noteworthy because it shows that at a very young age, infants are able to separate more probable from less probable structure and are able to generalize on this basis. What is the basis of such learning?

Mounting evidence points to a tendency on the part of learners to seek out structure that remains constant across different instances and contexts (Gibson, 1991). For instance, 8-month-olds can use the higher conditional probabilities of syllables within words versus the lower ones spanning words to identify word-like units in continuously running speech (Aslin, Saffran & Newport, 1998, Saffran, Aslin & Newport, 1996), and by 12 months of age, can track the conditional probabilities of words in strings (Gómez & Gerken, 1999; Saffran & Wilson, 2003). Infants also show some selectivity in terms of their tendency to focus on different types of structure. Given two sources of statistical information, infants will favor the source of greater statistical regularity. Gómez (2002) found that when conditional probabilities between adjacent words are high, learning will reflect a focus on this source of structure. However, when conditional probabilities are low, learners will focus on some other, more reliable, source of statistical information, such as non-adjacent dependencies (see also Gómez, Welch & Lany, 2004). As such, it is reasonable to hypothesize that learners will only focus on a particular source of information to the extent that it yields some degree of statistical regularity. Beyond this point, learners will seek out alternative sources of information (Gómez, 2002).

Children's tendency to focus on the most reliable (or regular) sources of structure is apparent not only in artificial language studies, but also in the learning patterns observed in natural language, such as in English when children add regular endings to verbs with irregular past tense (for instance producing utterances like *runned* and *wented*). This tendency is also observed in the case of Simon (Newport, 1999; Ross & Newport, 1996; Singleton & Newport, 1994), a congenitally deaf child of two deaf parents. Simon's parents were not exposed to ASL until their teen years, thus their mastery is not proficient and the forms they use are inconsistent. Simon was exposed to both infrequent, inconsistent patterns and frequent, consistent ones, yet his tendency was to converge on the latter.

The ability to focus attention on one source of information over another as a function of the statistical characteristics of the input could be very useful for guiding learning and might be important for explaining how learners converge on certain aspects of linguistic structure.[4] This is an important issue from a learning perspective because it begins to address long-time concerns that learning might either be powerless for dealing with inconsistencies found in human language or too unconstrained (resulting in acquisition of irrelevant structure). Instead, we see that learning is both powerful and constrained. Learners can tolerate some degree of inconsistency in their linguistic input, but learning diminishes as inconsistency increases.

Remaining issues

It is important to note that there are different types of inconsistencies in natural language. Many errors in natural speech, for instance, are sporadic. Other inconsistencies are more systematic, as in the case of the English irregular past tense. Sometimes the input contains both types of inconsistencies. An example is the process of creolization in which children are exposed to pidgin languages, where certain forms may be more frequent and consistent than others, or cases in which children are exposed to ASL acquired by non-native speakers (Singleton & Newport, 1994; Ross & Newport, 1996). The more frequent forms are thought to be the ones on which learners converge (Newport, 1999). Importantly, the type of inconsistencies learners encounter may lead to very different patterns of learning. For instance, Hudson Kam and Newport (2004) manipulated the type of inconsistency they introduced to children learning a miniature artificial language such that, under some conditions, noise was distributed broadly across many instances versus other conditions in which noise was distributed systematically. Six- and 7-year-old children were more likely to regularize when noise was spread out because this acted to highlight the more regular structure. The questions asked by Hudson Kam and Newport differ somewhat from the questions asked here. They were investigating whether learners would regularize their production of determiners,

[4] We do not mean to imply that statistical cues are the only source of information available to learners. As examples, infants are particularly attuned to the rhythmic patterns of their native language from very early on (Mehler *et al.*, 1988; Nazzi, Bertoncini & Mehler, 1998), and under certain circumstances attend to stress patterns and coarticulatory cues at the expense of statistically ordered elements in sequence (Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003). However, sensitivity to multiple converging cues is likely to be critical for negotiating the variety of structure found in natural language (Jusczyk, 1997). The ability to track probabilistic structure will play an important role in such learning.

whereas we were investigating generalization. However, their findings raise important issues for our research, suggesting for example, that infants might show a higher tolerance for probabilistic structure if it were distributed more randomly compared to the method used in the present studies.

It is also important to note that categories in natural language vary in the regularity with which cues are present. We know from studies of acquisition patterns of word order in human children (Slobin & Bever, 1982) and neural networks (Christiansen & Devlin, 1997; Lupyan & Christiansen, 2002) that the consistency of a pattern or cue will affect learning. We also know from work in artificial language learning that greater correlations of cues can enhance learning (Billman, 1989; Morgan *et al.*, 1987), thus there is every reason to expect that when certain categories or structures are more regular than others, the former will be more easily acquired than the latter. Additionally, the cues used in the present study were fairly simple, whereas cues in natural language tend to be correlated in more complex ways. Thus, it is quite possible that the consistency of cues combined with the degree to which they correlate will interact in potentially powerful ways during learning.

Another issue raised by this research is the question of whether learning degrades gradually or catastrophically with increases in noise. The present findings suggest that learning degrades gradually in that there were no significant decreases in learning from the 100/0 to the 83/17 conditions, and then a marginal decrease between Conditions 83/17 and 67/33. In future studies it will be important to consider finer-grained intervals with regard to introducing probabilistic structure, as well as the type of probabilistic structure, whether it is in any way systematic (as in the present studies) or whether it is more random in nature (as in Hudson Kam & Newport, 2004). The full answer will depend in part on whether infants are simultaneously learning the non-predominant language and whether abstract marker-feature pairings are harder to learn than specific marker-word pairings. There is some reason to think that certain forms of structure are easier to acquire than others (see Gómez, 2002; Gómez, Welch & Lany, 2004), hence the degree to which we might see gradual degradation in learning will likely be a function of the extent to which learners are attracted to one type of structure or another (specific marker-word combinations or more abstract marker-feature pairs).

Finally, what accounts for the developmental changes observed between 12 and 17 months of age? Successful learning with 12-month-olds in the present study (Step 1 learning) reflects two milestones: (1) grouping based on a shared, abstract feature, and (2) the ability to associate a particular element with an abstract feature. Successful learning with 17-month-olds (Step 2 learning) reflects a third milestone: the ability to abstract a- and b-categories with no features present except a pattern of associations between these elements and another category. Such changes reflect an ability to capitalize on increasingly abstract structure. At 12 months of age, infants abstract X- and Y-categories over a feature at least one step removed from the actual phonemes instantiating category members. By 17 months of age, abstraction is at least two steps removed in that the feature need not be directly attached to a- and b-category members. Rather, it can be inferred by means of association to X- and Y-categories. Another way of construing this is that younger infants associate specific a- and b-elements with abstract X- and Y-categories, whereas older infants can associate two abstract categories (a-words with X-words and b-words with Y-words). It is unlikely that these changes are confined to the domain of language, given that they encompass very general functions such as association formation and grouping based on similarity (in the case of X- and Y-elements grouping is based on a shared feature, in the case of a- and b-elements it is based on a shared association). It is also unlikely that development reflects a change in association formation, given ample evidence that this ability is present in infants much younger than those tested here (e.g. Kirkham, Slemmer & Johnson, 2002; Saffran *et al.*, 1996). Rather, development appears to reflect a progression in the ability to form abstractions from features that must be present at the time of generalization (e.g. syllable number) to those that may be inferred based on memory of a previously encountered instance. This likely reflects increases in basic memory and abstraction abilities in the second year of life. For instance, we know that infants become progressively less dependent on specific cues for triggering memory (e.g. Hayne, MacDonald & Barr, 1997; Herbert & Hayne, 2000a). Infants also make increasingly complex inferences in domains such as word learning and categorization (Waxman, 2003), conceptual learning (Mandler & McDonough, 2000), and memory (e.g. Hayne *et al.*, 1997; Herbert & Hayne, 2000a, 2000b; Rovee-Collier, 1999). Thus, the changes observed in form-based category abstraction may reflect a host of cognitive changes taking place in the second year of life.

Conclusion

The present experiments investigate the important question of how learners begin to acquire form-based categories and the relationships between them. For more advanced learners, the acquisition of such categories involves processes such as differentiation, grouping and inference. The first step in such learning is the acquisition of co-occurrence relationships between functor-like elements and categories distinguished by abstract features. The present experiments show that 12-month-olds are capable of learning such relationships. Learning is sustained even when some ungrammatical instances are present and thus this aspect of categorization appears to be quite robust. Although the eventual task of categorizing a novel word based purely on the presence of a functor-like element is far more complex, the present study provides some initial evidence that, by 12 months of age, preliminary aspects of form-based categorization are well underway.

Acknowledgements

This research was supported by NSF SES-9910203 and NIH R01 HD42170-01 to Rebecca L. Gómez. We would like to thank LouAnn Gerken, Roger Schvaneveldt and three anonymous reviewers for helpful comments and suggestions. We would also like to thank the parents and infants who generously gave of their time to participate in our studies.

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.

Billman, D. (1989). Systems of correlations in rule and category learning: use of structured input in learning syntactic categories. *Language and Cognitive Processes*, **4**, 127–155.

Braine, M.D.S. (1987). What is learned in acquiring words classes: a step toward acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 65–87). Hillsdale, NJ: Erlbaum.

Cassidy, K.W., & Kelly, M.H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language*, **14**, 333–352.

Christiansen, M.H., & Devlin, J.T. (1997). Recursive inconsistencies are hard to learn: a connectionist perspective on universal word order correlations. In *Proceedings of the 19th Annual Cognitive Science Society Conference* (pp. 113–118). Mahwah, NJ: Erlbaum.

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

Gerken, L.A. (2002). Early sensitivity to linguistic structure. *Annual Review of Language Acquisition*, **1**, 1–40.

Gerken, L.A., Landau, B., & Remez, R.E. (1990). Function morphemes in children's speech production and perception. *Developmental Psychology*, **26** (2), 204–216.

Gerken, L.A., Nurmsoo, E., & Gómez, R.L. (1999). The role of meaning and form in the formation of syntactic categories. Paper presented at the Society for Research in Child Development, Albuquerque, New Mexico, USA, April.

Gerken. L.A., Wilson, R., & Lewis, W. (2003). Seventeen-month-olds can use distributional cues to form syntactic categories. Manuscript under review.

Gleitman, L., & Wanner, E. (1982). The state of the state of the art. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art.* Cambridge: Cambridge University Press.

Gleitman, L. (1990). Structural sources of verb learning. *Language Acquisition*, **1**, 1–63.

Gibson, E.J. (1991). *An odyssey in learning and perception.* Cambridge, MA: MIT Press.

Gómez, R.L. (2002). Variability and detection of invariant structure. *Psychological Science*, **13**, 431–436.

Gómez, R.L., & Gerken, L.A. (1999). Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. *Cognition*, **70**, 109–135.

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

Gómez, R.L., & Schvaneveldt, R.W. (1999). Abstraction of distributionally defined categories and grammatical rules. Poster presented at the 40th Annual Meeting of the Psychonomic Society, Los Angeles, November.

Gómez, R.L., Gerken, L., & Schvaneveldt, R.W. (2000). The basis of transfer in artificial grammar learning. *Memory and Cognition*, **28**, 253–263.

Gómez, R.L., Welch, K., & Lany, J.A. (2004). Statistically guided learning. Manuscript under review.

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

Hayne, H., MacDonald, S., & Barr, R. (1997). Developmental changes in the specificity of memory over the second year of life. *Infant Behavior and Development*, **20**, 233–245.

Herbert, J., & Hayne, H. (2000a). Memory retrieval by 18–30-month-olds: age-related changes in representational flexibility. *Developmental Psychology*, **36**, 473–484.

Herbert, J., & Hayne, H. (2000b). The ontogeny of long-term retention during the second year of life. *Developmental Science*, **3**, 50–56.

Höhle, B., & Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science*, **6** (2), 154–159.

Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional elements in infants' speech processing: the role of determiners in the syntactic categorization of lexical elements. *Infancy*, **5** (3), 341–353.

Homa, D. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology*, **101**, 116–122.

Hudson Kam, C.L., & Newport, E.L. (2004). Regularizing inconsistencies in language: constraints on acquisition from inconsistent input. Manuscript under review.

Johnson, E.K., & Jusczyk, P.W. (2001). Word segmentation by 8-month-olds: when speech cues count more than statistics. *Journal of Memory and Language*, **44**, 548–567.

Jusczyk, P.W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P.W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development*, **64**, 675–687.

Jusczyk, P.W., Luce, P.A., & Luce, C.J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, **33**, 630–645.

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

Kemler Nelson, D.G., Jusczyk, P.W., Mandel, D.R., Myers, J., Turk, A., & Gerken, L.A. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*, **18**, 111–116.

Kirkham, N.Z., Slemmer, J.A., & Johnson, S.P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, **83**, B35–B42.

Landau, B., & Gleitman, L. (1985). *Language and experience: Evidence from the blind child*. Cambridge, MA: Harvard University Press.

Lehman, E.L. (1975). *Non-parametrics: Statistical methods based on ranks.* San Francisco: Holden-Day.

Lupyan, G., & Christiansen, M.H. (2002). Case, word order, and language learnability: insights from connectionist modeling. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 596–601). Mahwah, NJ: Erlbaum.

Marcus, G.F., Vijayan, S., Bandi Rao, S., & Vishton, P.M. (1999). Rule learning by seven-month-old infants. *Science*, **283**, 77–80.

Mandler, J.M., & McDonough, J.M. (2000). Advancing downward to the basic level. *Journal of Cognition and Development*, **1**, 379–403.

Mehler, J., Jusczyk, P.W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, **29**, 144–178.

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

Morgan, J.L., & Demuth, K. (Eds.) (1996). *Signal to syntax*. Mahwah, NJ: Erlbaum.

Morgan, J.L., & Newport, E.L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior*, **20**, 67–85.

Morgan, J.L., Meier, R.P., & Newport, E.L. (1987). Structural packaging in 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., Shi, R., & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories. In J. Morgan & K. Demuth (Eds.), *Signal to syntax*, (pp. 263–283). Hillsdale, NJ: Erlbaum.

Naigles, L.R. (2002). Form is easy, meaning is hard: resolving a paradox in early child language. *Cognition*, **86**, 157–199.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, **24**, 756–766.

Newport, E.L. (1999). Reduced input in the acquisition of signed languages: contributions to the study of creolization. In M. DeGraff (Ed.), *Language creation and language change* (pp. 161–178). Cambridge, MA: MIT Press.

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

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

Ross, D.A., & Newport, E.L. (1996). The development of language from non-native linguistic input. In A. Stringfellow, D. Cahana-Amitay, E. Hughs, & A. Zukowski (Eds.), *Proceedings of the 20th Annual Boston University Conference on Language Development*, Vol. 2, 634–635. Boston: Cascadilla Press.

Rovee-Collier, C. (1999). The development of infant memory. *Current Directions in Psychological Science*, **8**, 80–85.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by eight-month-old infants. *Science*, **274**, 1926–1928.

Saffran, J.R., & Wilson, D.P. (2003). From syllables to syntax: multi-level statistical learning by 12-month-old infants. *Infancy*, **4**, 273–284.

Santelmann, L.M., & Jusczyk, P.W. (1998). Sensitivity to discontinuous dependencies in language learners: evidence for limitations in processing space. *Cognition*, **69**, 105–134.

Sereno, J.A., & Jongman, A. (1990). Phonological and form class relations in the lexicon. *Journal of Psycholinguistic Research*, **19**, 387–404.

Shady, M.E. (1996). Infants' sensitivity to function morphemes. Unpublished doctoral dissertation, State University of New York at Buffalo, USA.

Shady, M.E., Jusczyk, P.W., & Gerken, L.A. (1998). Infants' sensitivity to function morphemes. Boston University Conference on Language Development, Boston, November.

Shafer, V.L., Shucard, D.W., Shucard, J.L., & Gerken, L.A. (1998). An electrophysiological study of infants' sensitivity to the sound patterns of English speech. *Journal of Speech, Language, and Hearing Research*, **41**, 874–886.

Shi, R., Morgan, J.L., & Allopenna, P. (1998). Phonological and acoustic bases for early grammatical category assignment: a cross-linguistic perspective. *Journal of Child Language*, **25**, 169–201.

Shi, R., Werker, J.F., & Morgan, J.L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, **72**, B11–B21.

Singleton, J.L., & Newport, E.L. (1994). *When learners surpass their models: the acquisition of American Sign Language from impoverished input.* Unpublished manuscript, University of Illinois, USA.

Slobin, D.I., & Bever, T.G. (1982). Children use canonical sentence schemas: a cross-linguistic study of word-order and inflections. *Cognition*, **12**, 229–265.

Smith, K.H. (1969). Learning co-occurrence restrictions: rule learning or rote learning? *Journal of Verbal Behavior*, **8**, 319–321.

Thiessen, E.D., & Saffran, J.R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, **39**, 706–716.
Valian, V., & Coulson, S. (1988). Anchor points in language learning: the role of marker frequency. *Journal of Memory and Language*, **27**, 71–86.
Waxman, S.R. (2003). Links between object categorization and naming: origins and emergence in human infants. In D.H. Rakison & L.M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 213–241). London: Oxford University Press.
Wilson, R. (2002). Category induction in second language learning: what artificial grammars can tell us. Unpublished doctoral dissertation, University of Arizona, USA.

Received: 25 June 2003
Accepted: 15 January 2004