When parents verb nouns: Resolving the ambicategoricality problem

Erin Conwell & James Morgan

Brown University

Draft: 9/12/2007

Running head: Resolving the Ambicategoricality Problem

Key words: language acquisition, grammatical categories, ambiguity

Address correspondence to:

Erin Conwell

Brown University

Department of Cognitive and Linguistic Sciences

Box 1978

Providence, RI 02912

Erin_Conwell@brown.edu

401-863-2497

Word Count: 7992

Abstract

In many languages, significant numbers of words are used in more than one grammatical category; English, in particular, has many words that can be used as both nouns and verbs. Such *ambicategoricality* potentially poses problems for children trying to learn the grammatical properties of words and has been used to argue against the logical possibility of learning grammatical categories from syntactic distribution alone. This article addresses how often English-learning children hear words used across categories, whether young language learners might be sensitive to phonetic cues that differentiate noun and verb uses of such words and how young speakers use ambicategorical words. The findings suggest that children hear considerably less cross-category usage than is possible and are sensitive to the phonetic cues that distinguish the two categories. Furthermore, in early language production, children's cross-category production mirrors the statistics of their linguistic environments, suggesting that they are distinguishing noun and verb uses of individual words in natural language exposure. Taken together, these results indicate that children may use cues in the speech stream to resolve the ambicategoricality problem.

Language makes infinite use of finite means by combining known words into novel sequences. Words are not restricted to linguistic contexts in which they have previously been used, nor may they be used freely in any context. Rather, the potential syntactic contexts in which words may occur are governed by their grammatical categories: noun, verb, adjective, or adverb. Membership in one of these categories defines how a word behaves syntactically. For example, nouns can appear as subjects of verbs, but also as objects of verbs, objects of prepositions, indirect objects, and so forth. Knowledge of category membership, in turn, allows speakers to use words productively, in contexts that vary from those in which particular words have been heard. In this article, we examine words that can be used in more than one grammatical category. We begin by explaining why such words might pose a problem for learning grammatical categories. Then we consider the nature of these words in the linguistic experience of young children in terms of the frequency with which they are used across category boundaries. We next ask whether infants are sensitive to subtle phonetic properties that distinguish noun and verb uses of the same words. Finally, we examine the nature of these words in children's own production and the influence of language experience on their usage.

A central task for language learners is to determine which words in the language belong to which categories. Unfortunately for learners, membership in a category not only defines how a word behaves syntactically, but is defined by the word's syntactic behavior. Distributionally, nouns are defined by how they may co-occur with verbs and adjectives; verbs are defined by how they may co-occur with adjectives and nouns, adjectives are defines by how they may co-occur with nouns and verbs, and so on. The circularity of this system poses a particular challenge to learners: What cues can be relied upon for learning category membership if one is learning about the syntax of the language at the same time? Known as the "bootstrapping problem," this

question is central to much of the literature on syntactic acquisition (e.g., Gleitman & Wanner, 1982; Pinker, 1984; Naigles, 1990).

One solution to this problem is that learners may use local co-occurrence cues to learn the categories of words, a process that is sometimes referred to as *distributional bootstrapping* (e.g., Höhle, Weissenborn, Kiefer, Schulz & Schmitz, 2004; Maratsos & Chalkley, 1980; Monaghan, Chater & Christiansen, 2005). Models of grammatical category learning based on distributional cues in corpora of child-directed speech are reasonably effective for categorizing words into noun and verb classes (Mintz, 2003; Mintz, Newport & Bever, 1998; Redington, Chater & Finch, 1998). However, such models typically restrict themselves to only those words that are unambiguous with regard to grammatical category, at least within the corpus in question, or take the most frequent grammatical category of a word as its only category (but see Cartwright & Brent, 1997, for an interesting exception).

The reason for such assumptions is that grammatical category ambiguity could wreak havoc with distributional learning: for such learning to be effective, it must be possible to keep separate those contexts in which, say, nouns and verbs occur. However, many languages have words that can appear in more than one grammatical category. In English, in particular, many words can be used as either nouns or verbs. Pinker (1987) pointed out the sort of problem that arises for distributional bootstrapping from such *ambicategorical* words: children using this learning strategy should take the evidence in (1a-c) and conclude that (1d) is grammatical in English (Pinker, 1987).

(1) a. I like fish.

b. I like rabbits.

c. John can fish.

d. *John can rabbits.

On the basis of distributional evidence alone, learners cannot rule out (1d); by causing noun and verb contexts to become conflated, ambicategorical words could cause learners to wildly overgeneralize the contexts in which any word might occur.

Ambicategoricality may, but need not, encompass usages that are semantically related. For example, as Clark and Clark (1979) pointed out, some words are typically nouns, but can be used as verbs, as in (2), and *vice versa*, as in (3). In other cases, homophones may belong to different categories, as in (4) or (5), with no systematic semantic relationship between two or more forms.

(2) a. The *water* on the beach stretched to the horizon.

b. John should *water* the flowers in the morning.

(3) a. I will *walk* to the park tonight.

b. Mary takes a *walk* every day.

(4) a. That dress doesn't *fit* you.

b. The toddler threw a *fit*.

(5) Hairy bears who bury berries barely bear bare bears.

Pinker (1987, 1989) used examples such as (1) to argue that children could not possibly rely on syntactic distribution to learn about category membership. The type of error in (1d) is almost never attested in child speech, as distributional theories of category learning should predict; Pinker argued, therefore, that children are using something beyond syntactic distribution to learn grammatical categories. Pinker's own theory of category learning, however, *semantic bootstrapping* (Pinker, 1984), does not resolve the problem either, as it relies on lexical semantics for categorization and many ambicategorical words refer to the same event or object

regardless of their grammatical category, as in (3). It is not clear how lexical semantics would aid learners in such cases.

Ambicategoricality may seem, at first blush, to be a rather limited problem. Certainly, in languages with richer morphology than English, derivational and inflectional affixes may often unambiguously indicate whether phonotactically constant stems are being used as nouns or verbs. However, languages may include homophonous affixes (as English –s serves as either a verb or a noun inflection), and not all languages are morphologically rich. Even if English were the only language to exhibit ambicategoricality – it is not – whatever capacities learners use to solve this problem must be available to all learners of all languages. The problem of ambicategoricality, therefore, remains a central quandary for most theories of category learning. Despite this, relatively little research has focused on explicating the facts of ambicategoricality, either in language input, or in children's own productions.

Macnamara (1982) described attempts to teach his own son, Kieran, the same word as both a noun and a verb. He reported that, at 17 months of age, Kieran was able to learn the same word to refer to both an object and an action, but that he began to introduce phonological distinctions between the noun and verb forms. For example, within two weeks of being taught the nonsense word "bel" to refer to both an action and an unrelated object, Kieran used "bam" to refer to the action and "ban" to refer to the object. In a longitudinal study, Macnamara (1982) examined the use of words as noun and verb in the speech in the Sarah corpus (Brown, 1973). He found that adults did not seem to avoid cross-category use when talking to Sarah, but that Sarah failed to use the same word as both a noun and a verb until the age of 30 months. In all of these cases, she used object words to refer to actions characteristically performed with those objects.

Nelson (1995) further explored the issue of cross-category usage in speech to children by examining six word types (*call, drink, help, hug, kiss,* and *walk*) in 12 corpora of mother/child interactions. These corpora consisted of 5 recordings per dyad. Each use of the target words was categorized as either noun or verb, and proportional use in each category was calculated. Nelson found that parents do use these words as both noun and verb, but, as her analysis focused on only 6 word types in relatively brief corpora, it is not possible to discern from these results how extensive ambicategoricality might be in children's linguistic experience. If parents use the preponderance of word types only in a single category when speaking to their children, then children might not encounter category ambiguity until their knowledge of language is robust enough to incorporate it.

Barner (2001) examined all of the denominal verbs and deverbal nouns in nine corpora of mother/child speech. His analysis focused mainly on the ways in which lexical semantics (e.g., action-denoting vs. object-denoting) affected use of words as both noun and verb. He found that both adults and children use some words as both noun and verb, but to a lesser extent than they could. He also found that object-denoting words were more likely to be used flexibly as both noun and verb than were abstract words. However, his analysis examined only data from Brown's (1973) Stage 1. It is possible that the rate of cross-category word use by both caregiver's and children might increase with grammatical ability. Furthermore, restricting the analysis to denominal verbs and deverbal nouns neglects the potential contributions of words that are accidental homonyms (e.g., *fit, leaves,* etc.) to the ambicategoricality problem.

In this article, we begin by examining the incidence of ambicategoricality in early English language input. We tabulate usages of several hundred word types from six longitudinal corpora of caregiver speech to children, providing a more complete picture of the nature of

ambicategoricality in early linguistic experience. While our results show that the problem is not as great as it might be, children are exposed to a non-trivial amount of cross-category word use. Therefore, the question of how they might deal with category ambiguity persists.

We next ask whether early language learners are sensitive to acoustic cues to grammatical category that may be present even in ambicategorical words. Previous work (Sorenson, Cooper & Paccia, 1978) indicates that noun tokens of words are reliably longer than verb tokens of the same words in adult-directed speech. Given the exaggerated prosody of infant-directed speech (Ferguson, 1964; Fernald, Taeschner, Dunn, Papousek, Boysson-Bardies & Fukui, 1989; Fisher & Tokura, 1996), these cues should also be available in speech to infants (Kelly, 1992). Our habituation study indicates that 13-month-olds are able to distinguish noun and verb uses of the same words based on differences in pronunciation alone. We propose that this sensitivity may allow children to separate noun and verb uses of the same word for the purposes of category learning.

Finally, we ask whether children use words across categories in their earliest combinatorial speech. Previous research on early word learning suggests that they should not. Children are known to prefer to use a single word form for only one linguistic purpose (Slobin, 1973; Nelson, 1995). Children are also highly adept at regularizing variable input, even imposing structure where there is none (Goldin-Meadow & Mylander, 1984; Hudson Kam & Newport, 2005). Such work predicts that children should avoid using words across categories and use words only in their prevalent category. However, if children use information available in the speech stream to distinguish noun and verb uses of the same word, they may learn two distinct, homophonous forms, rather than a single word that is ambicategorical. If this is the case, they should use words across categories. Because the statistics of the language children hear is often reflected in their

own productions (Demuth, Machobane & Maloi, 2003; Lieven, Pine & Baldwin, 1997), children's cross-category word use should mirror that of their caregivers. Our results show that young speakers not only use words as both nouns and verbs, but their cross-category usage of particular words is strongly predicted by their caregiver's usage, a pattern that we would not expect unless children were able to discriminate noun and verb uses of the same word.

Study 1

To determine the scope of the ambicategoricality problem for language learners, we examined six longitudinal corpora of child-directed speech. If caregivers regularly use words only in a single category when speaking to young children, the problem of category ambiguity in early acquisition would be rendered moot. If, however, caregivers use some word types as both noun and verb, the problem remains, and we must find a means by which language learners might resolve it. Previous work indicates that mothers do use some words as both noun and verb when talking to their children (Nelson, 1995). However, that work focused on only six words (*call, drink, help, hug, kiss* and *walk*) and reviewed only a limited number of recordings (five per mother/child dyad). This limits the extent to which Nelson's findings can be generalized to the larger problem of category ambiguity in language learning. Examination of more natural, longitudinal corpora will result in a more complete understanding of cross-category usage in speech to children and allow us to determine the extent to which ambicategoricality is a problem for learners.

Method

Corpora. Six longitudinal corpora of maternal speech were examined. Five of these corpora came from the Demuth Providence Corpus (Demuth, Culbertson & Alter, 2006). The sixth was

the Nina corpus (Suppes, 1974) from the CHILDES database (MacWhinney, 2000), which was included to provide evidence that our results generalize beyond the dialect of English spoken in Providence, Rhode Island. The ages and number of recordings for each corpus are presented in Table 1. Children in the Providence corpus were recorded every other week beginning as soon as they uttered their first words. The Lily corpus is an exception, as a sudden, rapid increase in her language production created a need for weekly recordings approximately a year after recording commenced. For completeness, all of the Lily files are included in this analysis. Nina was recorded approximately weekly. In all of these corpora, the child's mother is the primary caregiver and interlocutor. This age range (approximately 1-3 years) is of particular interest because it provides a comprehensive view of the child's productive language development from the very first utterances to complete, well-formed sentences. It also captures any changes in parental speech that may accompany the child's shift from language receiver to active conversationalist.

Insert Table 1 about here

Procedure. For each corpus, the number of maternal uses of each word type was counted, with morphologically complex words treated as individual types (e.g., *run*, *runs* and *running* were each counted separately). Because each corpus contained over 3,000 word types, it was impractical to examine every single one for cross-category use. Therefore, three frequency ranges were chosen as "core samples" for analysis. High frequency words were those used more than 150 times by the mother, middle frequency words were those used 40-60 times and low frequency words were those used 3-10 times. Within each frequency range, every word type was placed in one of two categories: "noun or verb" and "neither noun nor verb". Then, all those words that were nouns or verbs were further categorized as potentially ambicategorical or not.

These categorizations were based on the intuitions of native speakers of English[1]. For every word type that was potentially ambicategorical, each utterance including one or more tokens of that type was extracted from the corpus, and each token was classified as a noun, a verb or "other." Single word utterances, proper nouns and metalinguistic uses were classified as "other." A token was considered a noun if it was modified by an adjective, appeared as the head of a noun phrase, was an argument of a verb or could be replaced with a pronoun. A token was counted as a verb if it was modified by an adverb, took noun phrase or prepositional phrase arguments or could be replaced with a pro-verb. The breakdown of number of types analyzed in each corpus is shown in Table 2. Classification was done by trained coders. For consistency, 5% of all word types were reclassified by a second coder. Reliability between coders was very high (Cohen's K=.93).

Insert Table 2 about here

The total proportion of potentially ambicategorical words that were actually used across category was calculated for each mother as the number of words used at least once as both noun and verb divided by the total number of potentially ambiguous words analyzed. To obtain a better idea of how ambicategoricality relates to frequency of use, for each frequency range for each mother, the same kind of calculation was done on only those word types within a given frequency range. These numbers provide an estimate of how many of the word types each child heard were used across category boundaries at least once.

[1] Ideally, potential for cross-category use would be defined by a standardized database or corpus tagged for grammatical category. However, such databases and corpora are small and often based on written texts and do not, therefore, capture common, informal cross-category uses (e.g., noun use of *quack*, as in *The new doctor was a* quack.).

Results

No mother used more than a quarter of the potentially ambicategorical words across category boundaries. The proportions of ambicategorical use for all types used by a given mother ranged from .15-.21. This overall rate of cross-category usage is similar to that found by Barner (2001). Figure 1 shows the results broken over frequency ranges. Because the particular word types within each frequency range are different for each mother, these data cannot be directly compared. However, all mothers showed a similar relationship between the frequency of a potentially ambicategorical word and the likelihood that it would be used as both noun and verb. Specifically, words in the high and middle frequency ranges were more likely to be used across category than were words in the low frequency range. For all but Alex's mother, words in the middle frequency range were the most likely to be used across category.

Insert Figure 1 about here

The consistency of the relationship between frequency and cross-category use across mothers is not surprising when the composition of each frequency range is examined. The high frequency range for all mothers contained words that *can* be used as both noun and verb, but very rarely are, even in adult-directed speech (e.g., *go*). The middle frequency range contained more words that are intuitively ambicategorical, mostly concrete nouns and action verbs (e.g., *whistle, run*). For low frequency words, it is likely that there were simply not enough tokens to show cross-category use.

Those words that were used as both noun and verb were only rarely used equally as both. That is, many words were only used once or twice in their minority category. Figure 2 shows the percent noun use of high and middle frequency words. Words used as nouns 100% of the time are unambiguously nouns, while those used as nouns 0% of the time are unambiguously verbs.

As previously discussed, such words constitute the majority of potentially ambiguous words in child-directed speech. Words that were used at least once as both noun and verb were sorted into three bins: words that were predominantly used as nouns, with some verb uses (100-66% noun use), words that were predominantly verbs with some noun uses (0-33% noun use) and words that were used roughly evenly in both categories (33-66% noun use). The patterns of use across mothers are somewhat consistent. All mothers use very few words equally in both categories. This shows that young children do not hear many words that are perfectly ambiguous between noun and verb. Rather, words appear in a single category the majority of the time with a few uses in the alternate category. For all mothers, verbs are more likely to be occasionally used across category than nouns are. This may be due to the high frequency of "light verb" constructions in speech to children (Barner, 2001; Theakston, Lieven, Pine & Rowland, 2004).

Insert Figure 2 about here

We also considered the distribution of cross-category usages. It is possible that usages in the non-predominant category might occur in a handful of clusters, as might happen if, for example, the game of Go were introduced during a particular recording session. To address this possibility, we examined the high and middle frequency words in the speech of Nina's mother. For words that had more than one usage in the non-predominant category, two transitional probabilities were calculated: the likelihood that the previous token of that word was also from the non-predominant category and the likelihood that the next token was in the non-predominant category. Because considerable time elapsed between recordings, these calculations were all made within a recording and averaged over the corpus. First and last tokens in a recording had only one value (category of the following token and category of the preceding token, respectively), while all other tokens had two values. Of the twenty words that Nina's mother

used as both noun and verb, 19 were used more than once in the non-predominant category. For seven of these words, the tokens in the non-predominant category generally appeared in clusters, set apart from tokens of the predominant category. That is, the tokens following and/or preceding a minority use were more likely than chance ($p>.66$) to also be from the non-predominant category. Three of these words showed no such clustering at all; tokens of the non-predominant category never appeared adjacent to other tokens of that category. The minority tokens of the remaining 8 word types were equally likely to follow or preceded tokens from either category ($.33<p<.66$). These mixed results suggest that tokens from the non-predominant category do not reliably appear in clusters, nor are they evenly interspersed with tokens from the predominant category.

These findings indicate that cross-category word use is not as prevalent in speech to young children as it might be, given that roughly one third of the nouns and verbs they hear *can* be used across category. Neither, however, is it so rare as to be irrelevant to the problem of language learning. Because mothers do use some words as both noun and verb when speaking to their children, learners must have some means of coping with this source of noise in the data if they are to use distributional cues for learning grammatical categories. While meaning has been proposed as a way to solve this problem, it is not clear how meaning would be useful for distinguishing between noun and verb uses of words such as *kiss* or *smile*, which have the same referent regardless of the grammatical category that they are used in. Alternatively, young language learners might be able to use acoustic cues to differentiate noun and verb uses of the same word. If they do so, they might learn two homophonous forms, one that is a noun and one that is a verb, rather than one word that is used across category. To assess this possibility, we

now ask whether infants are able to differentiate noun and verb uses of the same word based only on their acoustic properties.

Study 2

We have demonstrated that young language learners hear some word types used as both noun and verb. Now we turn to the question of how children avoid conflating noun and verb categories given this fact about their linguistic experience. If they can differentiate noun and verb tokens of the same word based on their acoustic properties, they might learn two distinct, homophonous forms instead of one.

This idea is supported by work indicating that nouns, on the whole, are longer and have greater pitch change than verbs, even in cases of homophony (Sorenson, et al., 1978). While this difference is largely a function of the fact that nouns appear clause and phrase finally more often than verbs do, these prosodic cues are available in the speech stream and learners would not need to understand the syntax of the language to make use of them. Since the prosodic properties of child-directed speech are somewhat exaggerated relative to adult-directed speech (Ferguson, 1964; Fernald, et al., 1989; Fisher & Tokura, 1996), the cues found by Sorenson and colleagues should be present, if not even more apparent, in the language that infants hear (Kelly, 1992). If infants are sensitive to these differences, they might be able to use this information to differentiate noun and verb uses of the same word, thereby avoiding the problem of ambicategoricality, or at least postponing it until later in acquisition.

Method

Participants. A total of 36 13-month-old infants from the Providence, Rhode Island, area participated (12 male and 24 female). The mean age was 391 days (range 358-432 days). Previous work has shown that infants at this age are able to categorize words based on

distribution (Gómez & Gerken, 1999; Mintz, 2006). If they distinguish between noun and verb uses of the same word at this age, they may be able to use such information in real-world lexical categorization. An additional 15 infants participated in the study but were excluded due to excessive fussiness or squirminess (9), failure to habituate (1) or a looking time on either test trial that was more than two standard deviations from the group mean (5).

Procedure. Infants' ability to distinguish noun and verb uses of the same words was tested via an infant-controlled habituation paradigm. Each infant was seated in a testing room on a caregiver's lap, while the caregiver listened to masking music over headphones. The infant's gaze was coded by an experimenter observing via video camera from a separate control room where the audio stimuli could not be heard. At the beginning of each trial, a computer monitor mounted on the wall in front of the infant displayed a flashing yellow ball to attract the infant's attention. Once the infant oriented toward the monitor, the yellow ball was replaced with a static black and white checkerboard pattern and the audio stimulus began to play. The audio stimulus was contingent on the infant's looking and played only when the infant looked at the monitor. Each trial lasted for a minimum of 2 seconds and a maximum of 15 seconds or until the infant looked away for at least 2 continuous seconds, whichever came first. The average looking time on the first three habituation trials was the baseline looking time for the infant. The habituation criterion was reached when the average looking time on three contiguous trials (not including the baseline trials) declined to less than 65% of the baseline looking time. Two test trials followed the same format as the habituation trials. The dependent measure was the length of time the infant listened to each of the two test trials.

Design. For each of 7 monosyllabic word types, 4 noun tokens and 4 verb tokens were extracted from the audio recordings of the mother in the Lily corpus. The word types were

dance, *drink*, *help*, *kiss*, *rest*, *slide* and *swing*. Tokens were selected on the basis of having little extraneous noise and low co-articulation with surrounding words. The audio track for each token was extracted from the video using SoundConverter and edited using PRAAT. Intensity was normalized across tokens. Two sets of noun stimuli were created by randomly assigning two noun tokens of each word type to each set. Two sets of verb stimuli were created the same way. All stimulus sets contained unique tokens of the same word types. Each infant was habituated to one stimulus set. One-quarter of the participants was habituated to each of the four stimulus sets. In each trial, a new ordering of tokens was presented, created by randomly sampling from the token set without replacement. An interstimulus interval of 500 ms was used.

When the infant reached the habituation criterion, two test trials were presented. On the "same" test trial, the infant heard the tokens from the other stimulus set that were of the same grammatical category. On the "switch" test trial, the infant heard the tokens from one of the stimulus sets containing items from the non-habituated category. Importantly, all tokens in both test trials were novel, but in the "same" trial they were tokens of the habituated category and in the "switch" trial, they were tokens of the other category. Order of test trials was counterbalanced across subjects.

Results

The results of this study are presented in Figure 3. Infants listened to "switch" test trials for a mean of 5.4 s (SD=2.6) and to "same" test trials for a mean of 4.4 s (SD=1.9). This difference is significant ($t(35)=1.89$, $p=.033$, one-tailed[2], $d=.44$). A 2 (test trial) by 2 (habituated category) by 2 (sex) ANOVA revealed no three-way interaction ($F(1, 1, 34)=1.42$, $p=.242$), no interaction of trial by habituated category ($F(1, 34)=.688$, $p=.413$) and no interaction of trial by sex ($F(1,$

[2] Because an underlying assumption of the habituation paradigm is that infants will prefer novel stimuli to habituated stimuli, a one-tailed t-test is statistically appropriate for these data.

34)=.087, p=.77). These results show that infants can discriminate between noun and verb tokens of the same words using only the acoustic cues available in the words themselves.

Insert Figure 3 about here

When the stimuli themselves are examined, it appears that the infants may be making this discrimination based on both duration and pitch cues. Using PRAAT, mean pitch, minimum pitch and maximum pitch were measured in Hertz and token duration as well as vowel length were measured in milliseconds. Pitch change was calculated in semitones, using the minimum and maximum pitch measured for each token. Noun tokens were reliably longer (mean duration=468 ms; SD=193) than verb tokens (mean duration=366 ms; SD=192; t(27)=1.99, p=.05). Mean pitch (242.8 Hz; SD=77.75) and minimum pitch (163.97 Hz; SD=75.55) of noun tokens was lower than mean pitch (280.5 Hz; SD=84.56) and minimum pitch (202.4 Hz; SD=75.94) of verb tokens. These differences are marginal (t(27)=1.74, p=.088; t(27)=1.90, p=.063, respectively). Noun tokens had greater pitch change (13.6 ST; SD=9.92) than did verb tokens (9.53 ST; SD=7.48), although this difference was also marginal (t(27)=1.75, p=.086). These cues may arise from the characteristic syntactic positions in which nouns and verb occur. Nouns are more likely than verbs to be phrase and clause final and therefore are more likely to be subject to phrase final lengthening and falling pitch (Sorenson, et al., 1978). However, it is also possible that speakers may intentionally lengthen noun tokens to distinguish them from verb tokens when speaking to children. A more extensive analysis of these cues and their relationship to sentence position in a larger corpus study is necessary to completely address this possibility. Regardless, these cues are available in the speech that children hear and infants' sensitivity to such cues may help them segregate noun and verb uses of the same word.

Because infants are able to make this distinction based only on prosodic cues, it is possible that they could use this information to avoid the problem of ambicategoricality in acquisition. Rather than learning a single word that can be used as both a noun and a verb, they might learn two distinct, homophonous forms, one that appears in noun environments and one that appears in verb environments. If children are learning distinct forms, ambicategorical words should pose no problems for language learning. If they do not, however, such words should be more difficult to learn or may be used in only one category. This raises the issue of whether children will use word types in both noun and verb contexts. If they do, this suggests that they are able to distinguish between uses in each category. To address this, we examined the child speech from the six corpora analyzed in Study 1 for cross-category word use.

Study 3

Thus far, we have demonstrated that caregivers use some words as both nouns and verbs when talking to their children. Furthermore, infants are able to discriminate noun from verb uses of the same words based only on their acoustic properties. If this information is not useful to them in their normal language learning experience, they should use a single word only in one category, due to the well-documented tendency of children to restrict a single form to a single function (Slobin, 1973) and/or to regularize irregular language input (Goldin-Meadow & Mylander, 1984; Hudson Kam & Newport, 2005). If, however, children use acoustic information to learn not one word that is ambicategorical but two homophonous forms, one that is a noun and one that is a verb, they should be able to use words across category boundaries in their early productions. While there are many anecdotal reports of preschool-aged children *productively* using nouns as verbs (e.g., Clark, 1982), previous analyses of children's use of words that are ambicategorical to adults have focused primarily on the ways in which lexical

semantics interact with children's cross-category use (Barner, 2001) or have examined only a very small number of words (Nelson, 1995) or a restricted age range (Barner, 2001).

To address this issue, we performed two different corpus analyses. The first assessed whether children use words across category boundaries at all and the second compared children's use of individual ambicategorical words to their mothers' use of those words. If children use words across category boundaries, then they must be able to incorporate grammatical category ambiguity into their earliest combinatorial speech. Furthermore, unless children are distinguishing noun and verb uses of the same word, there should be no relationship between their use of a word across category and their caregivers' use of that word. If their use of ambicategorical words is well-predicted by that of their caregivers, this would suggest that children are able to dissociate noun and verb tokens and that their use of these words as nouns and verbs is a product of the statistics of their linguistic environments.

Method

The child speech from the six longitudinal corpora used in Study 1 was examined. For each corpus, the numbers of child uses of each word type were counted, as in Study 1, with morphologically complex words treated as individual types. Again, three frequency ranges were chosen for analysis. High frequency words are those used more than 150 times by the child, middle frequency words are those used 40-60 times and low frequency words are those used 3-10 times. Within each frequency range, words were classified as in Study 1. The breakdown of number of types analyzed in each corpus is shown in Table 3. Classification was done by trained coders. To ensure reliability, 5% of all word types were reclassified by a second coder. Reliability between coders was high (Cohen's K=0.81).

Insert Table 3 about here

The total proportion of potentially ambicategorical words that were actually used across category was calculated for each child as the number of words used at least once as both noun and verb divided by the total number of potentially ambiguous words analyzed. To obtain a better idea of how ambicategoricality relates to frequency of use, the same kind of calculation was done on only those word types within each frequency range for each child. These numbers provide an estimate of what proportion of the word types used by each child were used across category boundaries at least once.

However, because a given frequency range for a particular child does not necessarily contain the same word types as that frequency range for his or her mother, it is not possible to directly compare these data with the data from Study 1. To determine whether a child's use of a particular word across category boundaries is well-predicted by his or her mother's use of that word, a second corpus analysis was performed. Within a mother-child dyad, all tokens of each word type in the high and middle frequency ranges for the mother were extracted from the child's speech and coded as noun, verb or other, as described above. Likewise, all tokens of each word type in the high and middle frequency ranges for the child were extracted from the mother's speech and coded as described above. This allowed us to calculate the proportion of noun uses of each of these word types for each speaker. Proportion of noun use for a given word type was calculated as the number of noun uses of that type divided by the total number of noun and verb uses of that word type. If a word was only used as a noun, it would, therefore, have a proportional noun use of 1. Words used only as verbs would have a proportional noun use of 0. For each word type, this calculation was performed on maternal tokens to obtain maternal proportion of noun use and on child tokens to obtain child proportion of noun use. A correlation

analysis on these values within a dyad will reveal the extent to which maternal use of a word in a given category predicts child use in a category.

Results

When all of the word types analyzed were considered, the proportion of potentially ambiguous words used across category boundaries by the children ranged from .09-.13. This rate is slightly higher than that described by Barner (2001). The discrepancy may be the result of increased cross-category use with improved grammatical ability; our corpora included speech from children at older ages than those considered by Barner. Figure 4 shows how frequency of use relates to the likelihood that a child will use a word as both noun and verb. Recall that the proportions of ambicategorical use by mothers ranged from .15-.21. The children used a smaller proportion of words across categories than their mothers did, although this may be a function of a smaller number of word types in the children's speech than in their mothers. Like their mothers, these children do not use as many words as they could across category boundaries, but do show some cross-category word use. This indicates that children are not strictly adhering to the principle of one-form/one-function, but rather flexibly using some words as both noun and verb. This does not, however, demonstrate that children use the same words across category as their mothers do.

Insert Figure 4 about here

The results of the analysis comparing children's cross-category use of particular lexical items to that of their mothers are presented in Figure 5. For all children, use of a particular word was well-predicted by their mother's use of that word (all $R>.94$, all $p<.01$, two-tailed). However, it is possible that because most potentially ambicategorical words were not used across category by either the mother or the child, these words are driving the correlation. That children's very early

utterances do not include spontaneous (as opposed to attested) cross-category use is unsurprising. Overgeneralizations and creative word use often do not appear until the third or fourth year of life (Clark, 1982; Tomasello, 2000). Therefore, an important test of the extent to which maternal use of a word across category boundaries predicts the child's use of the word is whether these correlations remain strong when only those words that are used as *both* noun and verb are included in the analysis. To this end, we removed those words that are used in only one category by both the mother and the child and ran the correlations again. For five of the six children, this had only a small effect on the correlations (all $R>.91$, all $p<.01$, two-tailed). However, in the William corpus, the correlation decreased notably, although it remained highly significant ($R=.73$, $p<.01$, two-tailed). While it is difficult to tell exactly why William's cross-category word use was less correlated with his mother's than that of the other children, it is important to note that, of all the children analyzed, William was the only one with siblings much older than himself. It is possible that the presence of more interlocutors resulted in a more variable linguistic environment for William, which would explain the lower correlation of his word use with that of his mother.

Insert Figure 5 about here

Taken together, these results suggest that using a single word as both a noun and a verb is not impossible for young children. Furthermore, there is a clear and strong relationship between a child's use of a word and his or her mother's use of that same word. Specifically, children use words as nouns or verbs based on the way their mothers use those words. They are sensitive to the statistics of their language environment in a way that is reflected in their own productions, indicating that they are likely distinguishing between noun and verb uses of the same words during their natural language experience.

General Discussion

This paper set out to address the potential problem posed to language learners by cross-category word use. If children are to learn about grammatical categories based on their distribution relative to other words, ambicategorical words should cause learners to conflate category distributions and create categories that contain both nouns and verbs. To address this potential problem, this paper first asked whether children hear words used as both noun and verb. We find that mothers do use words across category boundaries when talking to their children, although not to the fullest possible extent. Secondly, we asked whether infants might be able to use cues other than distribution to separate noun from verb uses of the same word, such as acoustic differences between the two categories. Our habituation study shows that they are able to do so. Finally, to assess whether language learners might be able to use this information in natural language learning, we examined children's use of words across categories, both whether children did so at all and how their use of a given word type was determined by their mothers' use of that word. We find that children not only use across category boundaries in their natural productions, but that they also do so in a way that reflects their mothers' patterns of use. That is, the proportion of the time that a word is used as a noun by a child is closely correlated with the proportion of the time that the child's mother uses the word as a noun. Taken together, all of these results suggest that cross-category usage need not pose a major problem to language learners because there are cues available to them in the speech signal that allow them to distinguish noun and verb uses of the same word. They can, therefore, effectively learn two words, one that is a noun and one that is a verb, rather than having to contend with the problem of words that can be used across category boundaries.

The results of our habituation study indicate that children are able to use the information available in the speech stream to differentiate noun from verb tokens of the same word. This, in turn, allows them to avoid conflating noun and verb distributions while they are learning the syntax of their native language. That they themselves are able to use the same word as both a noun and a verb is further evidence that they are not confused by cross-category usage. If cross-category usage were a hindrance to word learning, children's acquisition of such words would likely be delayed. If children were to restrict the use of ambicategorical words to a single category, this would suggest that they are unable to incorporate cross-category behavior into their developing language system. Neither of these predictions is borne out by the results of our corpus study. However, these results run counter to predictions that would be made given several well-attested phenomena in child language development.

First of all, there is evidence that children prefer to restrict their use of a particular form to a single grammatical function (sometimes called the principle of one-form/one-function; Slobin, 1973). Language learners also tend to resist homonymy and prefer to assign a novel word to a referent for which they do not already have a label (e.g., Markman & Wachtel, 1988; Clark, 1988; Golinkoff, Mervis & Hirsh-Pasek, 1994). Given these two tendencies of language learners, it seems surprising that we would observe any cross-category word use at all by young children. Furthermore, children should have great difficulty learning the meanings of words that are applied to more than one referent and their acquisition should be delayed. However, this is not the case.

Children are also reported to regularize irregular input (Goldin-Meadow & Mylander, 1984; Hudson Kam & Newport, 2005). In some learning situations, such as when primary exposure to a language comes from a non-native speaker, children receive linguistic input that is irregular or

where the grammar is inconsistently used. These children tend to produce language with more consistent grammar, indicating that they are regularizing their language rather than reproducing the statistics of their experience (Goldin-Meadow & Mylander, 1984). In laboratory studies of artificial language learning, children who receive irregular or inconsistent information appear to derive consistent rules or schema for use in production, rather than learning the frequency patterns of the language (Hudson Kam & Newport, 2005). These kinds of results appear to be in direct conflict with our findings that young children not only use words across category boundaries, but that they do so in a way that reflects the statistics of their language environments. That is, children do not regularize irregular linguistic experience by restricting a word to a single category. Rather, they reproduce the statistics of the language that they hear.

However, our results can be interpreted in such a way that they complement these well-known findings, rather than contradict them. Because children are able to differentiate noun from verb tokens of the same word, based only on the acoustic cues available in that word, perhaps they are not learning a single word that can be both a noun and a verb. Instead, they may be learning two words, differentiated by these prosodic cues, one that appears in noun contexts and one that appears in verb contexts. This strategy would allow them to avoid the problem of ambicategoricality until their knowledge of syntax is robust enough to accommodate cross-category usage of the same word.

The findings that children can discriminate noun and verb tokens of the same word based on their acoustic properties alone bolsters this claim, but does not provide unequivocal support for it. The more important finding in support of this hypothesis is that young children will use particular ambicategorical words across category boundaries about as often as their mothers do. This suggests that in real world learning situations children are able to differentiate their

mothers' noun and verb uses of ambicategorical words and that they treat these uses as separate in their interpretation and production of language. They may, therefore, be learning two words rather than one that is used across categories. How (indeed, whether) children eventually collapse these two words into a single word remains an empirical question.

This study focused on the acoustic cues to grammatical category that children might use to learn about ambicategoricality. These cues might arise because talkers intend (at some level) to distinguish between noun and verb uses of phonotactically identical forms or because talkers characteristically produce noun and verb tokens in distinct syntactic contexts and positions. Regardless, particular prosodic cues, most notably duration, are associated with noun and verb categories; Study 2 shows that infants can use these cues to distinguish tokens presented in isolation. In context, learners may use these cues for other purposes as well, such as locating phrase, clause, or utterance boundaries, but there is no principled reason why this should preclude the simultaneous use of such information for inferring grammatical categories of words.

Other sources of information may also be relevant. Meaning may also offer cues that would allow children to distinguish noun and verb tokens of the same word, especially in those cases where the two words are homophonous, rather than derived. How meaning could cue grammatical category in situations where the noun and verb forms of a word refer to the same event or action remains unclear. In languages other than English, derivational morphology may provide information that words are being used across category boundaries. However, the absence of overt morphology for many noun/verb pairs in English may limit the extent to which such cues are useful. Furthermore, some English morphology is itself ambiguous with regard to grammatical category (e.g., *-s* may mark either plural nouns or third person singular present tense verbs). Nevertheless, the potential contributions of these two sources of information to the

problem of resolving ambicategoricality in acquisition are potential avenues of further research. Also, this paper focused only on words that are ambiguous between noun and verb uses. There are also words that are ambiguous between noun and adjective or adjective and verb uses. The issues of how mothers and children use such words and what cues might be available to children for learning about them remain unaddressed, but could also provide fodder for further work in this area.

Conclusion

Although mothers do use some words as both nouns and verbs when speaking to their children, the problem of cross-category usage of nouns and verbs in language learning is more apparent than real. Instead, even very young children appear able to use subtle acoustic cues available in the language they hear to differentiate noun and verb uses of the same words without relying on syntactic information. This allows children to learn and use ambicategorical words in a manner commensurate with the statistics of their linguistic environments.

Acknowledgements

This work was supported by NIH grant HD-32005 to JM. We wish to thank Katherine Demuth for access to the Demuth Providence Corpus, Rachel Ostrand, David Richardson, David Mittelman and Sally Grapin for assistance with data coding, Lori Rolfe, Megan Blossom and Glenda Molina for their help recruiting and running subjects, and Melanie Soderstrom for comments on an earlier draft. An earlier version of this work was presented at the 31st Annual Boston University Conference on Language Development.

References

Barner, D. (2001). Light verbs and the flexible use of words as noun and verb in early language learning. Unpublished Master's thesis, McGill University.

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.

Clark, E. V. (1982). The young word maker: A case study of innovation in the child's lexicon. In Wanner, E. & Gleitman, L. R. (eds.), *Language acquisition: The state of the art*. New York: Cambridge University Press.

Clark, E. V. (1988). On the logic of contrast. *Journal of Child Language 15*, 317-335.

Clark, E. V. & Clark, H. H. (1979). When nouns surface as verbs. *Language, 55,* 767-811.

Demuth, K., Culbertson, J. & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the acquisition of English. *Language and Speech, 49,* 137-174.

Demuth, K., Machobane, M., & Moloi, F. 2003. Rules and construction effects in learning the argument structure of verbs. *Journal of Child Language*, 30, 1-25.

Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist, 66,* 103-114.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16,* 477-501.

Fisher, C., & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development, 67,* 3192-3218.

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

Goldin-Meadow, S. & Mylander, C. (1984). Gestural communication in deaf children: The effects and noneffects of parental input on early language development. *Monographs of the Society for Research in Child Development, 49(3/4),* 1-151.

Golinkoff, R. M., Mervis, C. V., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language, 21,* 125-155.

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

Harris, Z. (1951). *Methods in structural linguistics*. Chicago: University of Chicago Press.

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

Hudson Kam, C. & Newport, E. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1,* 151-195.

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

Lieven, E. V. M., Pine, J. M. & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language, 24,* 187-219.

Macnamara, J. (1982). *Names for things: A study of human learning*. Cambridge, MA: MIT Press.

MacWhinney, B. J. (2000). The CHILDES project: Tools for analyzing talk. 3rd edition. Mahwah, NJ: Erlbaum.

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

Markman, E. & Wachtel, G. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20,* 121-157.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90,* 91-117.

Mintz, T. H. (2006). Finding the verbs: Distributional cues to categories available to young learners. In Hirsh-Pasek, K. & Golinkoff, R. M. (eds.), *Action meets word: how children learn verbs.* Oxford: Oxford University Press.

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

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

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language, 17,* 357-374.

Nelson, K. (1995). The dual category problem in the acquisition of action words. In Tomasello, M. & Merriman, W. E. (eds.), *Beyond names for things: Young children's acquisition of verbs*. Hillsdale, NJ: Erlbaum.

Pinker, S. (1987). The bootstrapping problem in language acquisition. In MacWhinney, B. (ed.), *Mechanisms of language acquisition.* Hillsdale, NJ: Erlbaum.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

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

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In Ferguson, C. A. & Slobin, D. I. (eds.), *Studies in child language development*. New York: Holt, Rinehart, Winston.

Sorenson, J. M., Cooper, W. E. & Paccia, J. M. (1978). Speech timing of grammatical categories. *Cognition, 6,* 135-153.

Suppes, P. (1974). The semantics of children's language. *American Psychologist, 29,* 103– 114.

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

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*, 209-253.

Table Captions

Table 1. Descriptions of corpora used for this study.

Table 2. Number of types analyzed for each frequency range in each maternal corpus. High frequency words are those with >150 tokens, middle frequency words are those with 40-60 tokens and low frequency words are those with 3-10 tokens in the given corpus. The total is the sum of these three ranges.

Table 3. Number of types analyzed for each frequency range in each corpus of child speech. High frequency words are those with >150 tokens, middle frequency words are those with 40-60 tokens and low frequency words are those with 3-10 tokens in the given corpus. The total is the sum of these three ranges.

Table 1.

Child	Sex	Age Range (years;months)	# of Files
Alex	M	1;5-3;5	52
Ethan	M	0;11-2;11	50
Lily	F	1;1-4;0	80
Nina	F	1;11-3;3	52
Violet	F	1;2-3;11	52
William	M	1;4-3;4	44

Table 2.

	# Noun or Verb Types				# Potentially Ambicategorical				# Used Across Categories			
	High	Middle	Low	Total	High	Middle	Low	Total	High	Middle	Low	Total
Alex	72	81	908	1061	27	38	344	409	8	11	61	80
Ethan	82	113	1111	1306	44	59	469	572	9	17	66	92
Lily	120	156	1780	2056	50	81	612	743	14	31	112	157
Nina	90	104	751	945	24	49	291	364	6	14	36	56
Violet	45	72	1222	1339	19	32	429	480	5	13	85	103
William	62	82	878	1022	32	45	384	461	8	13	67	88

Table 3.

	# Noun or Verb Types				# Potentially Ambicategorical				# Used Across Categories			
	High	Middle	Low	Total	High	Middle	Low	Total	High	Middle	Low	Total
Alex	19	29	433	481	12	16	152	180	4	5	13	22
Ethan	12	38	694	744	6	21	211	238	2	5	21	28
Lily	36	56	821	913	12	24	229	265	3	3	22	28
Nina	46	76	627	749	15	43	269	327	3	7	23	33
Violet	10	12	484	506	7	4	190	201	0	0	19	19
William	14	19	407	440	7	7	142	156	0	6	14	20

Figure Captions

Figure 1. For each mother, the proportion of potentially ambicategorical words that are actually used across categories is reported for each of 3 frequency ranges.

Figure 2. For each mother, the number of high and middle frequency words used as nouns a given percentage of the time are shown. The left and rightmost bars for each mother represent those words that were never used across categories.

Figure 3. Results of the habituation study indicate that infants prefer word usages from a new category over those from the habituated one ($t(31)=1.87$, $p=.035$, one-tailed).

Figure 4. For each child, the proportion of potentially ambicategorical words that are actually used across categories is reported for each of 3 frequency ranges.

Figure 5. For each ambicategorical word analyzed in each corpus, proportional noun use by the child is plotted against proportional noun use by the mother. All correlations are significant ($R>.73$, $p<.01$, two-tailed).

Figure 1.

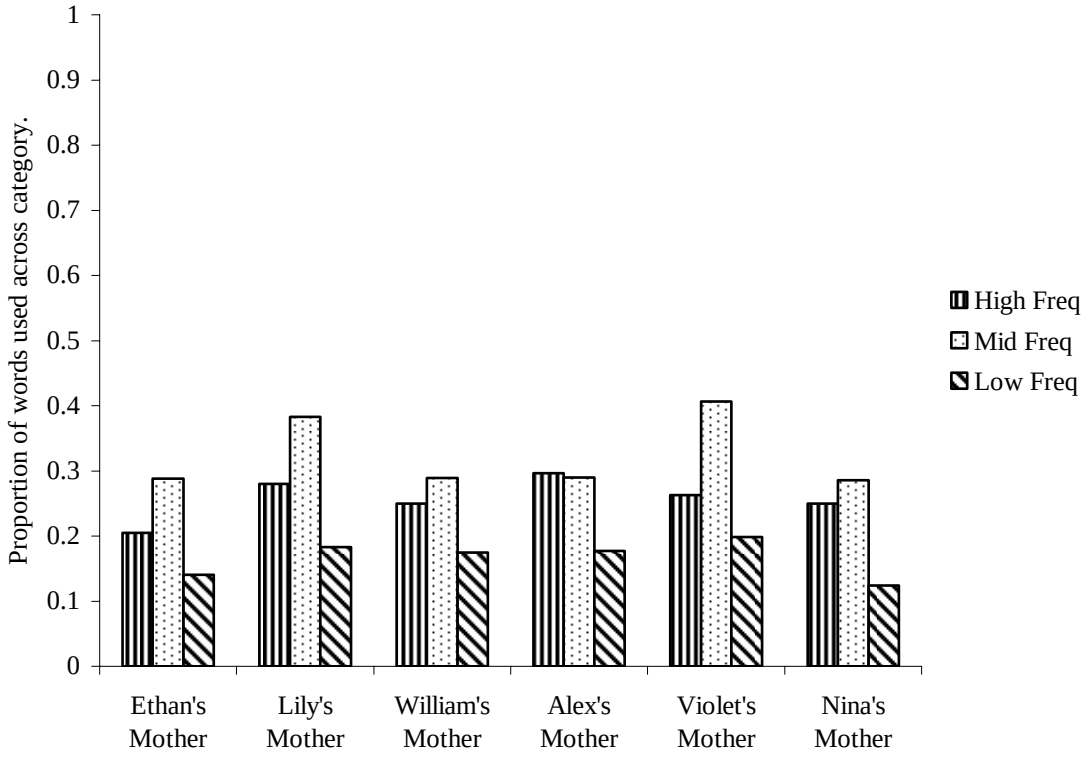

Proportion of words used across category.
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Ethan's Mother
Lily's Mother
William's Mother
Alex's Mother
Violet's Mother
Nina's Mother
High Freq
Mid Freq
Low Freq

Figure 2.

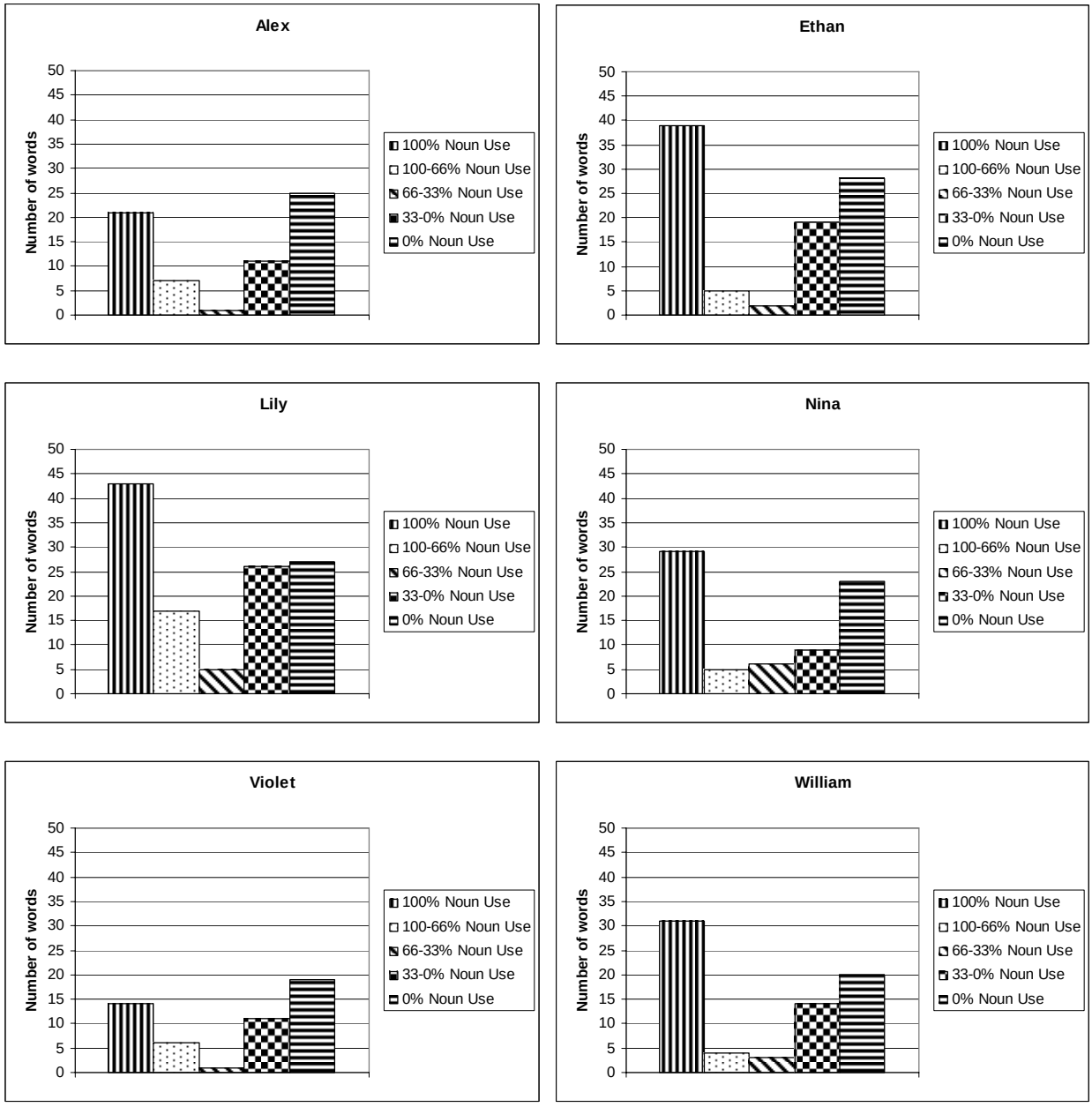

Alex
Ethan
Lily
Nina
Violet
William
Number of words
0
5
10
15
20
25
30
35
40
45
50
100% Noun Use
100-66% Noun Use
66-33% Noun Use
33-0% Noun Use
0% Noun Use

Figure 3.

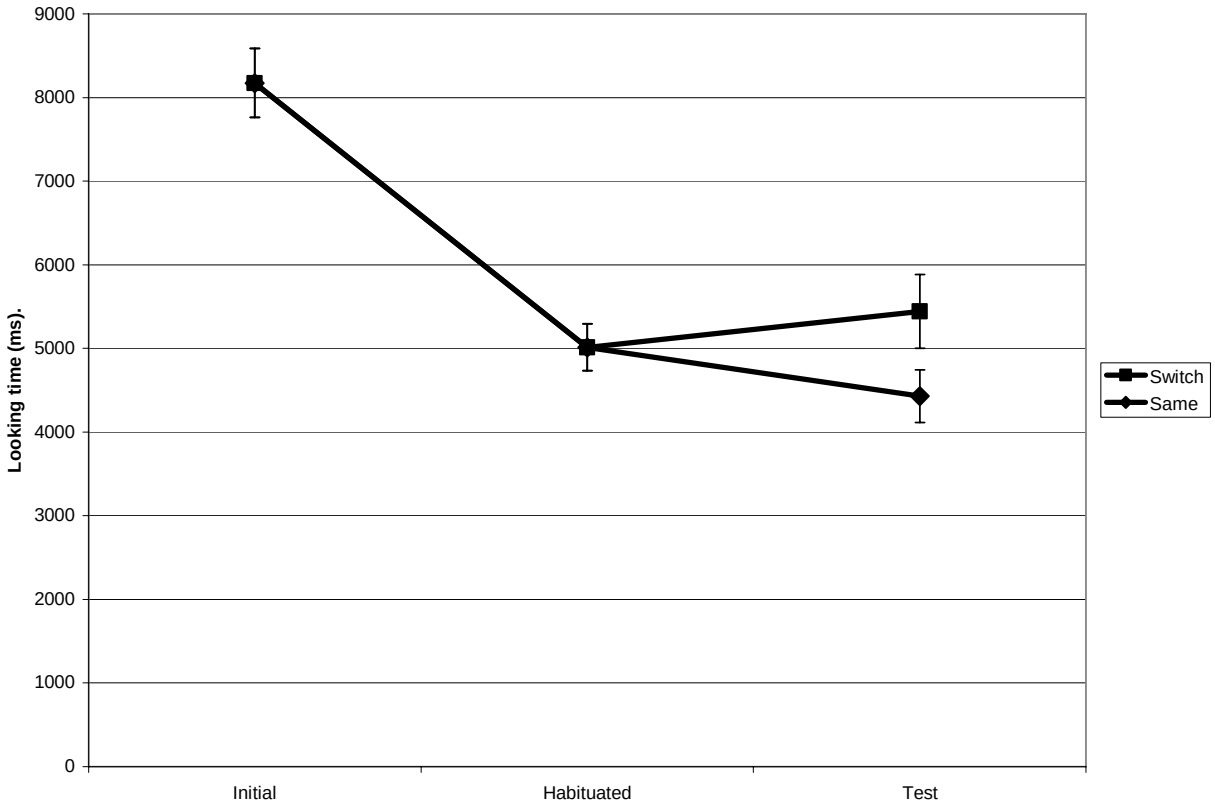

9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Looking time (ms).
Initial
Habituated
Test
Switch
Same

Figure 4.

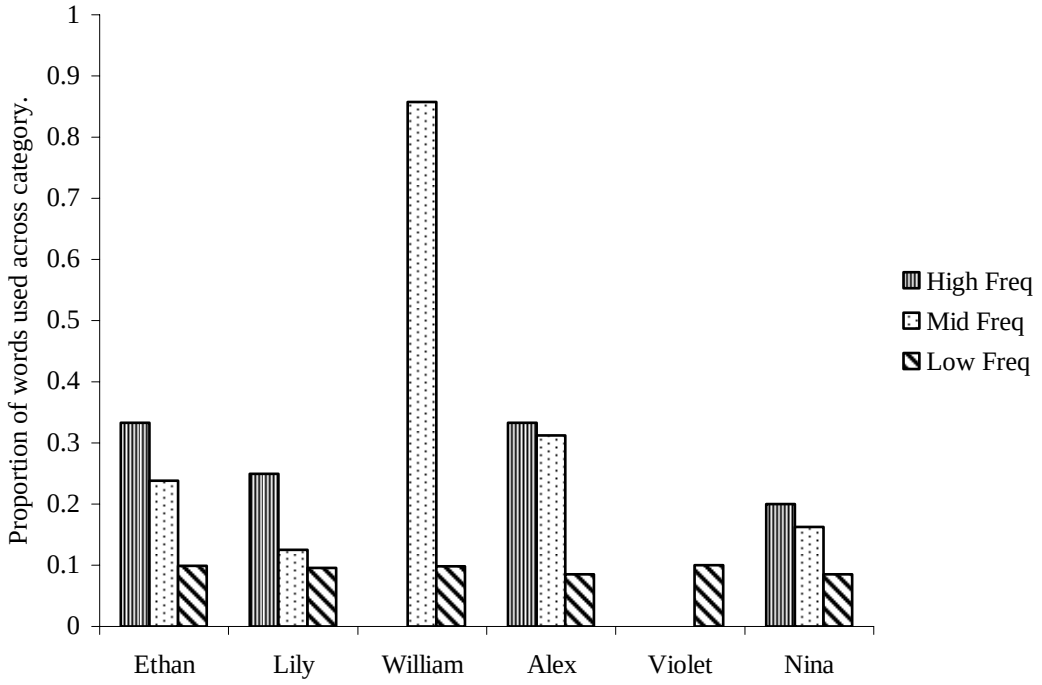
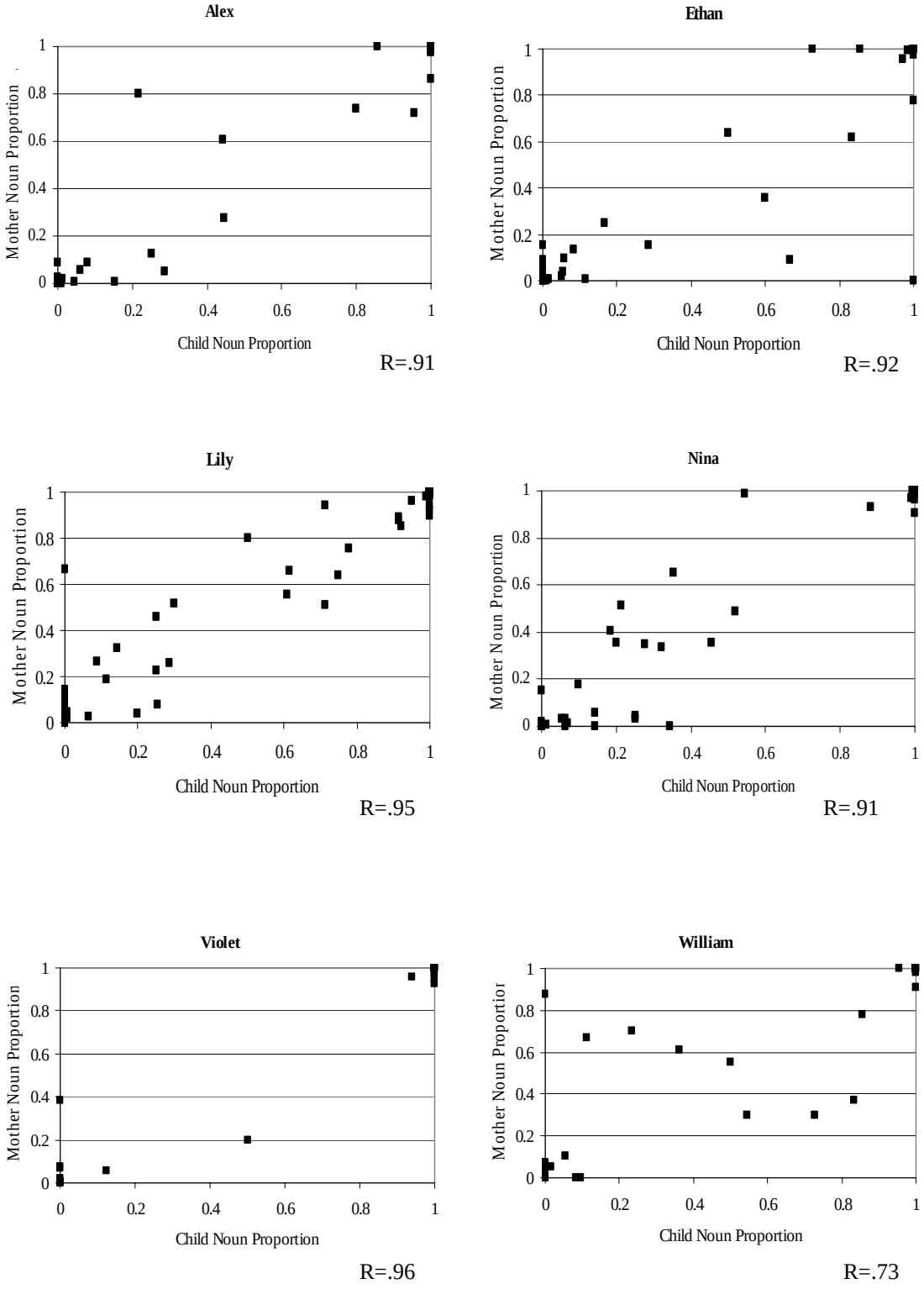

Proportion of words used across category.
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Ethan
Lily
William
Alex
Violet
Nina
High Freq
Mid Freq
Low Freq

Figure 5.


Alex
Mother Noun Proportion
Child Noun Proportion
R=.91
Ethan
Mother Noun Proportion
Child Noun Proportion
R=.92
Lily
Mother Noun Proportion
Child Noun Proportion
R=.95
Nina
Mother Noun Proportion
Child Noun Proportion
R=.91
Violet
Mother Noun Proportion
Child Noun Proportion
R=.96
William
Mother Noun Proportion
Child Noun Proportion
R=.73