

Resolving Grammatical Category Ambiguity in Acquisition

Erin Conwell and James Morgan
Brown University

English, like many languages, has words that can be used as both nouns and verbs. In some cases, as in (1), the word is taken to be inherently a member of one category and the other use is derived (Clark & Clark 1979). In other cases, the two words are merely homophones, at least synchronically, as in (2).

- (1) I will *walk* to the store.
I will take a *walk* tonight.
- (2) I will *fly* to Chicago tomorrow.
I swatted the *fly*.

This kind of category ambiguity has been cited as evidence against the logical possibility of learning grammatical categories from distribution (Pinker 1987). Indeed, if we are to take seriously the many proposals that children may learn which words should be categorized together based on such cues as local co-occurrence (e.g., Redington, Chater & Finch 1998; Mintz 2003), we must address the potential pitfalls posed by such words. The difficulty with these words lies in the fact that their ambicategorical status may cause learners to conflate co-occurrence information across categories. That is, if a single word occurs with a distribution that is consistent with both noun and verb categories, learners could conclude that this distribution characterizes a single word class and thereby form a single category that contains all nouns and verbs. As Pinker (1987) pointed out, a child who hears the sentences in (3-5) would be expected to make errors such as (6).

- (3) I like fish.
- (4) I like rabbits.
- (5) I can fish.
- (6) * I can rabbits.

* This work was supported by NIH grant HD-32005 to JM. The authors wish to thank Katherine Demuth for access to the Demuth Providence Corpus, Lori Rolfe, Megan Blossom and Glenda Molina for their assistance running experimental subjects, Melanie Soderstrom and Ben Balas for helpful discussions and all infant participants and their families.

While the logic of this critique is sound, there is little empirical work on the nature of ambicategorical word use in the linguistic environments of children (see Nelson 1995 for an exception). Previous work on the learning of grammatical categories based on distribution has analyzed only words that are unambiguously nouns or verbs, thereby avoiding this problem. However, a child learning language must have some other means of sidestepping this problem of ambicategoricity, as the syntactic cues used by researchers are unavailable to very young learners. This paper, therefore, asks three questions. First, what is the experience of young learners with regard to these words? If parents are wholly unicategorical in their uses of nouns and verbs, children may not even encounter this problem until they have a robust understanding of syntax. If, however, children are exposed to some cross-category usage of nouns and verbs, alternative means of solving this problem must be found. Our second question, therefore, is whether infants are sensitive to low-level acoustic differences in the pronunciation of noun and verb uses of the same words. If they are, they may be able to avoid this problem by learning two distinct word types, one with the distribution of a noun and one which behaves syntactically like a verb, rather than a single type that can be used across category boundaries. Our final question is whether children actually produce words ambicategorically. One potential learning strategy would be to ignore the noise in the system and use a word only in its more prevalent category. However, if children use words across categories, this suggests that they have dissociated noun and verb uses of these words and are using that distinction in their own productions. The issue of whether children's cross-category use is related to that of their parents will also be examined.

1. The Nature of the Input

Previous work on parents' use of ambicategorical words addressed how parents would use words that are frequently used across category boundaries in adult speech (Nelson 1995). In this study, parents were recorded talking to their children in a laboratory setting where particular toys were provided to elicit certain ambicategorical words. In particular, the words of interest were *call*, *drink*, *help*, *hug*, *kiss* and *walk*. Nelson found that parents did use these words as both nouns and verbs in the context of this study. However, the generalizability of these results is limited by focusing on only these 6 verbs in a single recording session.

To address the issue of ambicategoricity in parental speech at a larger scale, we looked at maternal speech to two children in the Demuth Providence Corpus (Demuth, Culbertson & Alter 2006). This corpus consists of hour-long, bimonthly recordings of naturalistic interactions between mothers and their children beginning at the time of the child's first words (~12 months of age) and continuing for 2 years. The extensive nature of this corpus allows for the examination of a greater number of word types while the longitudinal sample controls for possible effects of context.

1.1 Method

Maternal word use frequency was calculated for each mother using the CLAN program (MacWhinney 2000). To make the analysis more manageable, only words in 3 frequency ranges were coded. These ranges were >150 uses (high frequency), 40-60 uses (medium frequency) and 3-10 uses (low frequency), based on the number of tokens in the speech of the mother. Morphologically complex words were treated as unanalyzed wholes (e.g., *run* and *runs* were treated separately). Within each frequency range, each use of all potentially ambicategorical words was hand-coded for syntactic category (noun, verb or other). Single word utterances and those cases where category was not readily determined by the syntax of the sentence were classified as “other.”

For each potentially ambicategorical word, the number of uses of that item in each category was calculated. The proportion of potentially ambicategorical words that were actually used in both categories was calculated as the number of such words with at least one noun use and one verb use over the total number of potentially ambicategorical words.

1.2 Results

In the Lily corpus, *potentially* ambiguous words accounted for 42% of all nouns and verbs in the high frequency range, 52% in the middle frequency range and 34% in the low frequency range. In the Ethan corpus, these numbers were 54%, 52% and 42%, respectively. Both mothers used some of these potentially ambicategorical words as both noun and verb. In the Lily corpus, 28% of potentially ambiguous words in the high frequency range were actually used in both categories. For the middle frequency range, this percentage was 38% and for the low frequency range it was 18%. In the Ethan corpus, 20% of potentially ambicategorical words in the high frequency range were actually used across category. For the middle and low frequency ranges, the percentages were 29% and 14%, respectively. In both corpora and in all frequency ranges, the majority of the potentially ambiguous words were only used in a single category. The results of this analysis are presented in Figure 1.

For both mothers, words in the middle frequency range were more likely to be used in both categories than words in the other two ranges. This is unsurprising when the composition of the frequency ranges is considered. Very high frequency words that can be used across category tend to strongly favor one category (e.g., *Let's go to the store.* vs. *Let's have another go at that.*). Lower frequency words may simply not have had enough exemplars to provide evidence of cross-category usage. However, words in the middle frequency range included several of the words previously examined by Nelson (1995) and words that are more intuitively used across category in adult directed speech (e.g., *whistle*, *paint*, etc.) and also included enough tokens of each word that cross-category usage was likely to be observed.

Given these results, it would appear that cross-category usage of nouns and verbs in child-directed speech is not so great as it might be, but that children nevertheless do hear words used across category boundaries. Therefore, it is reasonable to consider whether there might be cues beyond the distribution that would help language learners learn these words without conflating noun and verb categories on a larger scale. In particular, if young children are sensitive to acoustic distinctions between categories, this information maybe used to categorize words without relying on syntax.

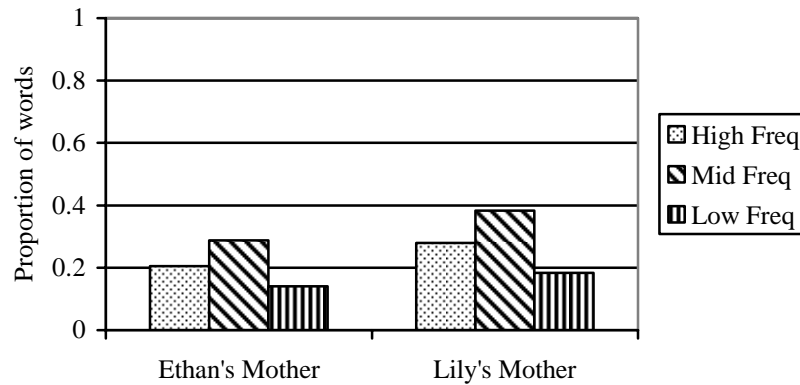


Figure 1: For each mother, the proportion of potentially ambicategorical words that are actually used across categories is reported above for each of 3 frequency ranges.

2. Infant perception of ambicategorical words

Because language learners must contend with at least some cross-category uses of nouns and verbs, we now ask whether other cues to categorization might be available to infants. In multisyllabic forms, different patterns of stress are often correlated with grammatical category; for example, *REcord* (noun) versus *reCORD* (verb). Sorenson, Cooper and Paccia (1978) found that acoustic cues to grammatical category may exist even for monosyllabic ambicategorical words, despite their cross-category homonymy. These cues are due largely to the position of the word in an utterance and may allow learners to distinguish grammatical category without understanding the syntax of the sentence. To resolve the potential pitfalls of cross-category usage in the early language environment, infants may actually form two distinct representations of the same phonemic string. That is, rather than a single word *kiss* that can behave both like a noun and a verb, infants may learn two acoustically distinct items: one that behaves like a noun and one that behaves like a verb. We now turn to the question of whether presyntactic infants are sensitive to acoustic distinctions as a means of lexical categorization.

2.1 Stimuli

Using the speech of one of the two mothers previously discussed, we extracted 4 noun tokens and 4 verb tokens of each of 7 monosyllabic ambiguous word types. These words were *dance*, *drink*, *help*, *kiss*, *rest*, *slide* and *swing*. Because they would be presented in isolation, rather than in a sentence context, tokens were selected on the basis of having little or no extraneous noise and no coarticulation with surrounding words. Four sets of words were created, each comprising two unique tokens of each word type. Within a set, all words were from the same category, with two sets of noun uses and two sets of verb uses. In a particular trial, participants heard one of the stimulus sets presented in semi-random order with a 1 second interstimulus interval (ISI).

2.2 Method

Infants were habituated to one set of words, containing either all noun tokens or all verb tokens. Baseline looking time was provided by mean looking time on the first three trials with a minimum of 6 seconds looking time. Habituation was reached when mean looking time on three consecutive trials reached 65% of baseline. Once an infant reached the habituation criterion, two test trials were presented. All test tokens were novel. On the “same” test trial, infants heard a set of 2 novel tokens of each word type from the habituated category (noun or verb). On the “switch” test trial, infants heard a set of 2 novel tokens of each word type from the category to which they were *not* habituated. Order of presentation of test trials was counterbalanced across subjects.

Participants were 24 13-month-old infants (mean age of 389 days; range 358-432 days), 13 female and 11 male. Infants were recruited from birth records and referrals. All lived in the Providence, Rhode Island, area. An additional 10 infants were tested and excluded, 6 due to excessive fussiness and 4 due to looking times that were greater than 2 standard deviations from the group mean on test trials.

2.3 Results

The results of this study are presented in Figure 2. Overall, infants looked longer to “switch” over “same” test trials (5.4s and 4.2s, respectively). This difference was significant ($t(23)=2.37$, $p<.05$, two-tailed). Furthermore, 15 of 24 participants preferred switch trials to same. This preference is seen regardless of the category to which the infants were habituated. Infants habituated to verb tokens looked an average of 5.7s to “switch” trials and 4.5s to “same” trials. For infants habituated to nouns, these numbers were 5.1s and 3.9s, respectively. These results suggest that infants are sensitive to acoustic-level cues that distinguish noun from verb uses of the same word. This

sensitivity could be used to resolve or at least avoid the problem of ambicategoricity in early language development.

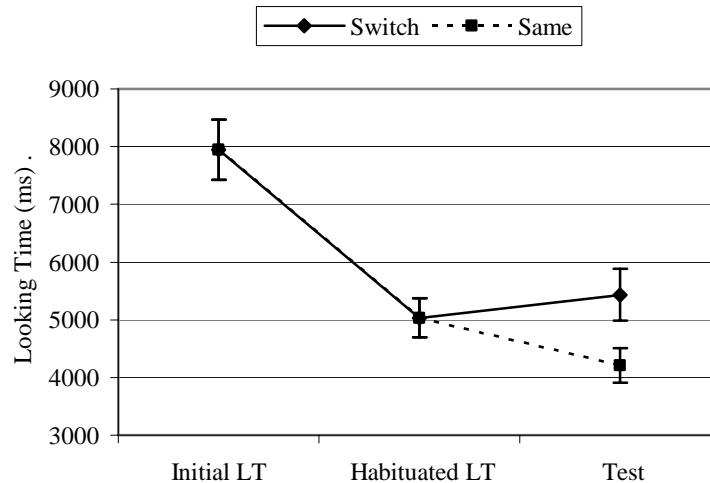


Figure 2: Results of our habituation study indicate that infants prefer word usages from a new category over those from the habituated one, even when the word forms themselves are the same.

3. Children's productions: One category or two?

To this point, we have provided evidence that mothers do use some words as both noun and verb when speaking to their children and also that infants are sensitive to the acoustic distinctions between noun and verb uses of the same word form. However, it is possible that children do not make use of these distinctions and instead assign words to a single grammatical category in their early uses. While there is anecdotal evidence that preschool-aged children will productively use nouns as verbs and vice versa (Clark 1982), there has been no systematic, longitudinal examination of the ways in which children use words that are ambicategorical for adults.

3.1 Method

Using the same two corpora that were analyzed in section 1, we examined the children's use of potentially ambicategorical words. As with the maternal speech, only words in 3 frequency ranges were coded. These ranges were the same as for the maternal uses. Each use of a potentially ambicategorical word in one of these frequency ranges was coded for use as either noun or verb. Words that were used at least once as a noun and once as a verb were counted as being

used across categories in the child's speech. The proportion of potentially ambicategorical words that were actually used in both categories was calculated by taking the number of words used at least once in each category and dividing it by the total number of potentially ambiguous words in that frequency range. Note that the words types in each frequency range for the children will not necessarily be the same word types in that frequency range for their mothers.

3.2 Results

For the Lily corpus, *potentially* ambiguous words accounted for 33% of all high frequency nouns and verbs, 43% of the middle frequency range and 28% of the low frequency ranges. For the Ethan corpus, these numbers are 50%, 55% and 30%, respectively. Lily used 25% of these potentially ambicategorical words in the high frequency range as both noun and verb. For the middle and low frequency ranges, this number is 12% and 10%, respectively. Ethan used 33% of the potentially ambicategorical words in the high frequency range across category. In the middle frequency range, this number is 24% and in the low frequency range, it is 10%. The results of this analysis are presented in Figure 3.

Overall, the children, like their parents, used most potentially ambiguous words in only one category. However, children not only used some words in both categories, they tended to do so with a greater proportion of high frequency words than did their mothers. Since the types in these frequency ranges do not match between a mother and her child, a more direct analysis is necessary to determine the role of the input language on children's cross-category word use.

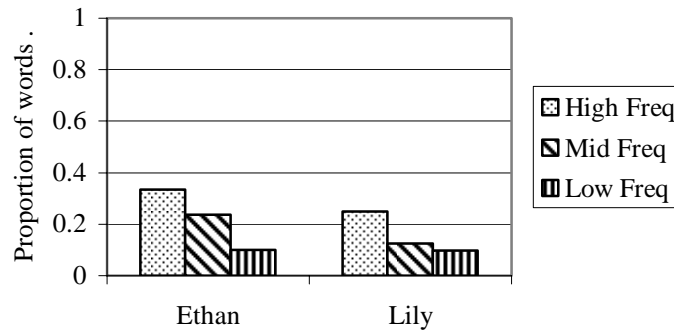


Figure 3: For each child, the proportion of potentially ambicategorical words that are actually used across categories is reported above for each of 3 frequency ranges.

To determine the relationship between the way in which a child heard a word used and the way the child used the word, individual word types within these frequency ranges were directly compared. For a given word, maternal noun

proportion was calculated as the number of noun uses of that word divided by its total uses as either noun or verb by the mother. The same calculation was made using the child data to determine child noun proportion. Words that were used only as nouns would have a noun proportion of 1 and words that were used only as verbs would have a noun proportion of 0. For each mother/child dyad, the correlation of child noun proportion with maternal noun proportion was evaluated. Ethan's proportional noun use correlated significantly with his mother's ($r=.96$, $p<.01$). Likewise, for Lily and her mother, the correlation was significant ($r=.98$, $p<.01$). These data are presented in Figures 4 and 5.

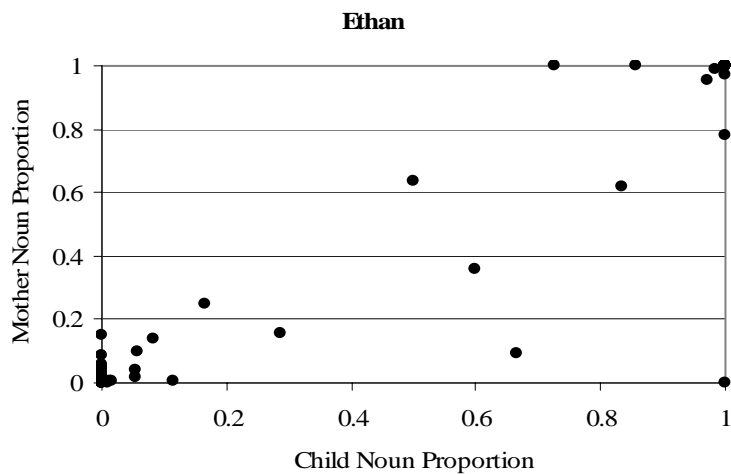


Figure 4: For each ambicategorical word analyzed in the Ethan corpus, proportional noun use by the child is plotted against proportional noun use by his mother. The correlation is significant, $r=.96$, $p<.01$.

Because these data include a large number of words that were used only in a single category by both mother and child, it is possible that these data points are driving the correlation. This would be unsurprising, as it indicates that very young language learners do not spontaneously use words across category boundaries. To ensure that these data points are not driving the correlation, they were removed from the data set and the correlations were recalculated. For both children, this slightly reduced the strength of the correlation, although both correlations remained significant. For Ethan, his production of actually ambicategorical words was significantly correlated with that of his mother ($r=.91$, $p<.01$). The same was true for Lily ($r=.94$, $p<.01$).

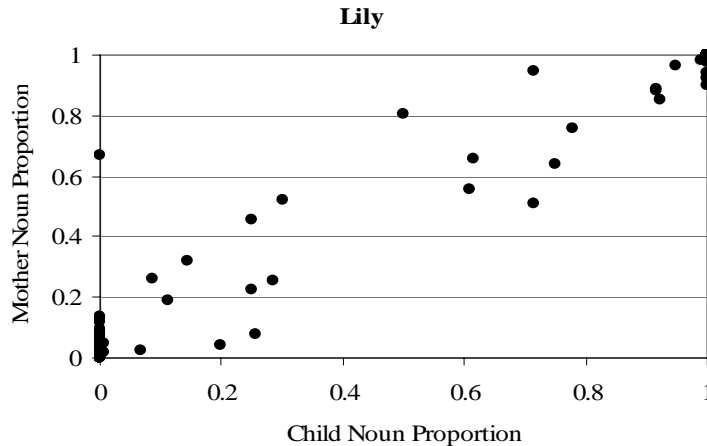


Figure 5: For each ambicategorical word analyzed in the Lily corpus, the child’s proportional noun use is plotted here against her mother’s proportional noun use. The correlation is significant, $r=.98$, $p<.01$.

That these exceptionally strong correlations persist suggests that children are not only able to learn to use single words in more than one category, but further indicates that this behavior is driven by the statistics of their language environment. These data are somewhat surprising given results of other studies that indicate that children will regularize irregular language input (e.g., Hudson Kam and Newport 2005; Goldin-Meadow and Mylander 1984). Taken with these other results, our results suggest that children are treating noun and verb tokens of ambicategorical words as two distinct forms, rather than a single form with two uses.

4. General Discussion

The present study addressed the problem of ambicategorical words in language acquisition. If young children hear words used across category boundaries (e.g., as both noun and verb), how might they avoid conflating distributional information about the categories of noun and verb? In examining a corpus of child-directed speech, we found that parents do show some cross-category usage of nouns and verbs, although most potentially ambicategorical words are used only in a single category. Furthermore, when a word is used as both a noun and a verb, there are subtle acoustic correlates of grammatical category to which infants are sensitive. In their own productions, children show similar patterns of overall cross-category use to that of their parents: most potentially ambicategorical words are used in only a single category. Children

are most likely to use a word as both noun and verb if that word is highly frequent in their productions. Parents show greater ambicategoricity in their use of medium frequency words. Nevertheless, for individual word types, parental cross-category usage is a strong predictor of child usage: proportional use in a given category by the mother correlates strongly with that of her child. This is a longitudinal pattern and cannot, therefore, be explained by context effects.

These ambicategorical words, then, do not pose the insurmountable problem that they have been theorized to (e.g., Pinker 1987). Not only are they less frequent in the language that children hear than they might be, very young language learners are sensitive to low-level acoustic cues to category present in these words. Children use words as both noun and verb in their early combinatorial speech, in spite of previous suggestions that they would prefer to restrict a single form to a single function (Slobin 1973, but see Nelson 1995). This is in keeping with work by Casenhiser (2005) indicating that preschool-aged children more willingly accept homonyms when they are used across category boundaries than when they are used in a single category.

While the best predictor of grammatical category is syntactic distribution, children who are only beginning to learn the syntax of English may not be able to use that information to categorize words. When words are used as both noun and verb in speech to children, acoustic cues exist that indicate whether the word is being used as a noun or as a verb. Because we have found that pre-syntactic infants are sensitive to these cues, it is possible that they might use this information to avoid conflating distributional frames. That is, rather than learn a single word form that is used in multiple categories, young language learners might acquire two acoustically distinct forms: one that has the distribution of a noun and one that has the distribution of a verb. In this way, a learner could avoid the problem of ambicategoricity until his or her syntactic competence is robust to cross-category use.

Our corpus data support this idea in two key ways. First, children's earliest combinatorial speech includes cross-category use of nouns and verbs. This suggests that children are capable of understanding category ambiguity from very early on. Taken with anecdotal and case-study evidence of children's novel cross-category usages, the corpus data indicate that ambicategoricity poses less difficulty for children than previously thought. Children do not find these words any more difficult to learn than those that are unicategorical. Furthermore, work on children's regularization of irregular input would suggest that children exposed to cross-category usage should use words in only one category (Hudson Kam and Newport 2005). Research has also indicated that children restrict use of a given form to a single function (Slobin 1973). However, our corpus data show that children's cross-category word use is well-predicted by their mothers' use. That is, children do not use words in a single category, but instead mirror the statistics of the input language. Rather than contradicting these predictions, the patterns in our corpus data converge with the findings of our experimental study to indicate that children do not, in fact, learn single word forms that are

used across categories, but that they learn distinct word forms with distinct distributions. How and when these distinct representations might eventually merge into single, ambicategorical words, remains a question for future research.

Beyond acoustic distinctions, children certainly receive other information regarding category use. Morphology and semantic context probably also play some role in helping young learners distinguish noun and verb uses of the same word. In particular, unrelated homonyms (e.g., Does that dress *fit*? vs. The child threw a *fit*.) may be learned very differently than deverbal nouns and denominal verbs. Nouns are much more likely than verbs to appear after the determiner 'the' (Redington, et al. 1998). Likewise, only verbs will appear with the past tense morpheme *-ed*. The ways in which these types of cues interact are not addressed by the present research, but certainly offer rich prospects for further work.

The present study clearly indicates that ambicategorical words need not pose a major obstacle to distribution-based theories of syntactic category acquisition. By using other cues available in the speech stream, very young language learners can distinguish noun from verb uses of such words and thereby avoid potential pitfalls of cross-category word use. Corpus data indicate that these words are not used later than other forms nor are they only used in a single category, suggesting that they pose no greater difficulty to learners than unambiguous words. However, considerable work remains in addressing how children might learn the relationship between the two forms and how other cues interact to help learners avoid the potential problems presented by ambicategoricity.

References

- Casenhiser, Devin. (2005). Children's resistance to homonymy: An experimental study of pseudohomonyms. *Journal of Child Language*, 32, 319-343.
- Clark, Eve V. (1982). The young word maker: A case study of innovation in the child's lexicon. In Wanner, Eric & Gleitman, Lila R. (eds.), *Language acquisition: The state of the art*. New York: Cambridge University Press.
- Clark, Eve V. & Clark, Herbert H. (1979). When nouns surface as verbs. *Language*, 55, 767-811.
- Demuth, Katherine, Culbertson, Jennifer & Alter, Jennifer. (2006). Word-minimality, epenthesis and coda licensing in the acquisition of English. *Language and Speech*, 49, 137-174.
- Goldin-Meadow, Susan & Mylander, Carolyn. (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.
- Hudson Kam, Carla. & Newport, Elissa. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1, 151-195.
- MacWhinney, Brian J. (2000). The CHILDES project: Tools for analyzing talk. 3rd edition. Mahwah, NJ: Erlbaum.
- Mintz, Toben H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90, 91-117.

- Nelson, Katherine. (1995). The dual category problem in the acquisition of action words. In Tomasello, Michael & Merriman, William E. (eds.), *Beyond names for things: Young children's acquisition of verbs*. Hillsdale, NJ: Erlbaum.
- Pinker, Steven. (1987). The bootstrapping problem in language acquisition. In MacWhinney, Brian (ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.
- Redington, Martin, Chater, Nick, & Finch, Steven. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, 22, 425-469.
- Sorenson, John M., Cooper, William E. & Paccia, Jeanne M. (1978). Speech timing of grammatical categories. *Cognition*, 6, 135-153.
- Slobin, Dan I. (1973). Cognitive prerequisites for the development of grammar. In Ferguson, Charles. A. & Slobin, Dan I. (eds.), *Studies in child language development*. New York: Holt, Rinehart, Winston.