Distributional Regularities of Form Class in Speech to Young Children

Toben H. Mintz
University of Rochester

Elissa L. Newport
University of Rochester

Thomas G. Bever
University of Rochester

Distributional Regularities of Form Class in Speech to Young Children

Toben H. Mintz, Elissa L. Newport, Thomas G. Bever

University of Rochester

The problem we address in this paper is how an infant could learn the form classes of his or her language. In particular, we focus on how the infant could learn the major form class categories Noun and Verb.

A number of solutions to this problem have been proposed in the literature (see Maratsos & Chalkley, 1980, Pinker, 1984, 1987, for discussion). One proposal is that form classes are part of the innate knowledge of the learner (Chomsky, 1965; McNeill, 1966). However, while learners might well know the abstract universal properties of major form classes (Chomsky, 1986; Jackendoff, 1977), they cannot innately know which particular words of their input language fall into each of these classes, and so a substantial learning problem remains to be explained. Other proposals hypothesize that form classes may be acquired from semantic (Pinker, 1984) or prosodic (Kelly, 1992) information which correlates with form class. However, the pertinent semantic or prosodic information may not always be available to the learner, and, even when available, correlates imperfectly with form class. These proposals therefore also run into difficulties (see Gleitman, 1990, and Pinker, 1987, for discussion). Thus, while all of these types of information might participate to some degree in the acquisition of form class, they cannot alone provide an adequate solution to the learning problem.

In the present paper we consider yet another solution to this problem, namely, that form class may be acquired from the distributional information present in the way words are sequenced within input sentences. As we will discuss, like other solutions, this one is also unlikely to work adequately for the learner on its own. However, unlike other solutions, distributional information is always present in the input the child receives, and, depending on the analytic capacities of the learner, it may therefore provide a uniquely consistently available source of information for inducing form class. On the other hand, because distributional information involves retention and analysis of rather complex corpora of word sequences, its use may require capabilities which young learners may lack. In order to evaluate this source of information, two sets of empirical information are therefore needed: first, one must ask whether the distributional information available in linguistic input is in principle adequate to solve, or at least to participate in solving, the form class problem; and second, one must ask whether the language learning child is capable of utilizing this type of information. The analyses presented in this paper form a part of the first of these questions: here we ask whether, in principle, relatively restricted and limited analyses of the input word sequences can in fact lead to inducing the basic form classes. In a separate and subsequent line of work (Mintz, in progress) we consider whether children at the age for early acquisition in fact evidence the memorial and analytic capacities which such analyses would entail.

The notion that form classes fall out of the way words distribute with one another was first made explicit by Harris (1951) and other structural linguists. To demonstrate with a simple example, in sentence (1), note that the words *dog* and *moon* are preceded by *the*, and furthermore that they are frequently preceded by *the* in other utterances.

(1) The dog is barking at the moon.

Similarly, a small class of words often precedes *the*, such as *at* in sentence (1), as well as *in*, *on*, *around*, etc., in other sentences. Such similarities in patterning, it was argued, are what, in fact, *define* words as belonging to the same form class.

More recently, Maratsos and Chalkley (1980) proposed that such distributional information may play a crucial role in the acquisition process. On their view, since the members of a form class appear in a highly overlapping set of contexts, these overlaps may be used to acquire the form class categories. Homonymous forms, like the plural -s and the possessive -s, could be correctly separated by keeping track of semantic as well as purely distributional information.

But, as outlined by Chomsky and others after him, there are clearly problems with such an approach. One potential problem concerns the erroneous inductive generalizations the learner could make from distributional similarities. For instance, what would stop a learner from classifying *white* and *barking* together from sentences like (2) (for other examples, see Pinker, 1987)?

(2) The dog is white.

After all, a barking dog barks, but a white dog doesn't white. Maratsos & Chalkley's distributional-semantic procedure might surmount this particular difficulty by the discovery that *barking* contains an internal morpheme *-ing*, or by noting other crucial contexts (e.g., question forms) which differ for these two words. Nevertheless, more fundamental problems remain. Sentences are not simply linear strings of words; sentence structure is not completely represented by, nor can it be built up from, the patterning of words with their immediate neighbors. Moreover, once one admits more complex distributional contingencies than immediate co-occurrences, the problem of fixing on just the *right ones* becomes enormous.

These problems have generally been viewed as so serious, that, until recently, little empirical work has been done to investigate the potential form class information provided by the distributional properties of speech to children. However, while a distributional analysis may not totally solve the problem, it is nevertheless possible that it could provide very useful initial information as to what the grammatical categories of the language are. As mentioned above, this kind of information is crucial even if the learner is innately endowed with fairly rich knowledge about the grammatical structure of the target language; knowing how at least some sound sequences map onto grammatical categories is necessary before any innate syntactic constraints (which are formulated in terms of these grammatical categories) could come into play to determine the rest. Thus, it could be that although distributional analyses may not provide the learner with complete form class structure, they could provide a necessary bootstrap onto the process. Other sources of information (such as prosody, semantics, and innate syntactic constraints) could build on and modify early distributional classifications.

We therefore conducted a series of analyses to determine just how useful distributional information might be, in principle, for constructing the initial form class categories from the input. Our aim in these analyses was not to model the actual procedures a child might use, but rather to examine the information available in children's input. For this reason we focused exclusively on distribution, and did not consider semantic or prosodic information at all. In the interest of making our estimate of the distributional information realistic, however, we used actual input corpora directed to young children of the appropriate age. Other researchers (Brill, 1991; Finch & Chater, 1992, 1994; Schütze, 1993) have recently conducted very similar analyses, but instead used written English text as their corpora. To our knowledge, only one other study (Redington, Chater, & Finch, 1994), conducted simultaneous with our own analyses, has also examined children's input in this fashion.

As we describe in more detail below, our analyses (and the others cited above) use the similarity of the immediate lexical contexts of words to classify them into groups: target words that are preceded and followed by the same set of words are judged to belong to the same class. Recall that these are just the kind of simple co-occurrences that should fail to represent the full structure of the grammar. Nonetheless, as we show below, we have found that nouns and verbs can be identified to a surprising degree by this type of information. This finding accords with those of Brill (1991), Finch & Chater (1992, 1994), Redington et al. (1994), and Schütze (1993), using different corpora, different windows of information, and somewhat different methods of analysis.

We also then proceeded to manipulate our input corpora in various ways, to better approximate the limited information that very young children (rather than tape recorders or computers) might perceive and represent from their input. We now turn to a fuller description of our procedures and the results of all these analyses; for complete details, see Bever, Newport, Aslin, Mintz, Juliano, & LaMendola (1994) and Mintz, Newport, & Bever (in preparation).

Our input corpora were taken from the CHILDES database (MacWhinney & Snow, 1985, 1990), which consists of transcriptions of speech by and to young children. Our analyses focus entirely on the speech directed to the children; in other words, we examine the linguistic input which language learners actually receive. We were interested in looking at speech directed at children less than 2 years old. By 2 years the child already produces 2- and 3-word utterances that display some rudimentary syntax and knowledge of grammatical categories. Therefore, any initial category learning must take place before this point. We looked at samples of speech directed at 3 different children: Peter, from the Bloom (1970) corpus, Eve, from the Brown (1973) corpus, and June, from the Higginson (1985) corpus. Typically, each sample contains speech from 4 or 5 adults (primarily the mother or the experimenter) addressed to one child. The number of utterances for each sample averages approximately 6,000. Because of the relatively small size of the samples, and because, as we will see presently, our analyses involve examining the set of varying lexical contexts of individual words, we base our analyses on the 200 most frequent words.

Again, the general idea behind our categorization procedure is that it will group words based on the similarity of their immediate lexical contexts. First we construct a list of all the different words that appear in the corpus. Then for each word, we record what words come immediately before and after it throughout the corpus, and how many times. For example, in Figure 1, *w3* represents a target word; the list to the left labeled *W-1* represents how many times all other words come before *w3* throughout the corpus, and the list to the right labeled *W+1* represents how many times all other words come after *w3*.

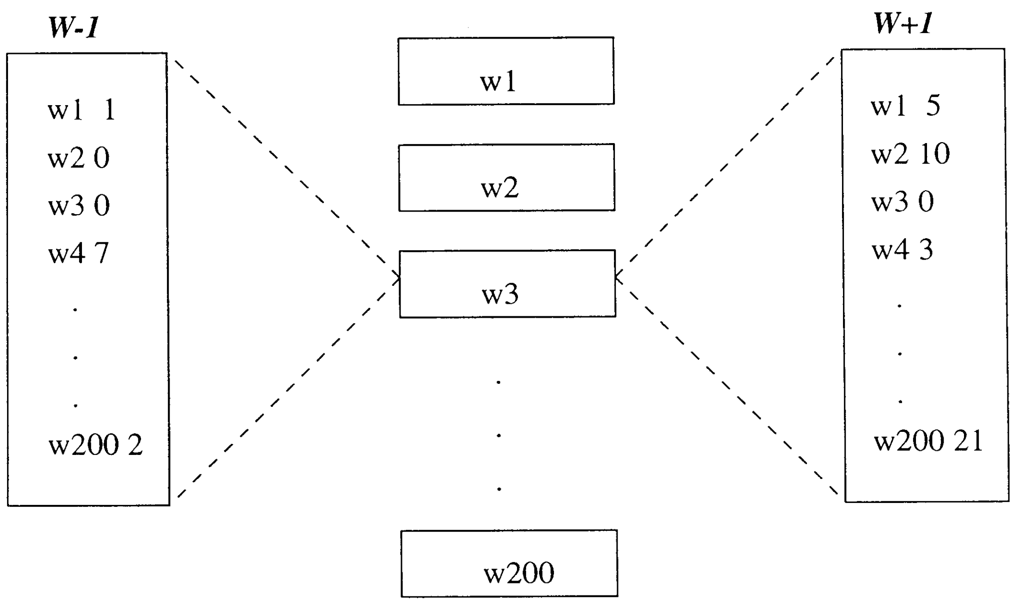

Figure 1

As mentioned above, we only look at the contexts of the 200 most frequent words; in addition, only the 200 most frequent words are considered in tallying contexts. For example, if an infrequent word appears before *w3*, it will not be entered in the *W-1* list; for that instance of *w3* in the corpus, nothing will be entered for the preceding context.

This procedure then gives us a list of numbers for each word, representing the frequencies with which all other words appear before and after it. Table 1 shows the context representation of *w3* rewritten as one list, compared with the context representation of another word *wn*. We can see how similar the lexical contexts of any two words are by numerically comparing these lists: the more similar the lists, the more similar the lexical contexts. (In particular, we treat each numeric list as a vector and compute the angle between two vectors as a measure of *dis*similarity; the vector comparison formula we use is given in Equation 1 in the Appendix.) What we calculate, then, is a similarity measure for each pair of words, based on context representations like those in Table 1. These word pairs with their similarity measures are then submitted to a hierarchical cluster analysis, which forms a hierarchy of word clusters, based on their degree of context similarity. Figure 2 shows an example of a simplified hierarchical cluster representation.

w3:	1	0	0	7	.	.	.	2	5	10	0	3	.	.	.	21
wn:	1	5	7	9	.	.	.	4	2	0	0	9	.	.	.	9

Table 1
Context vector representations for two words

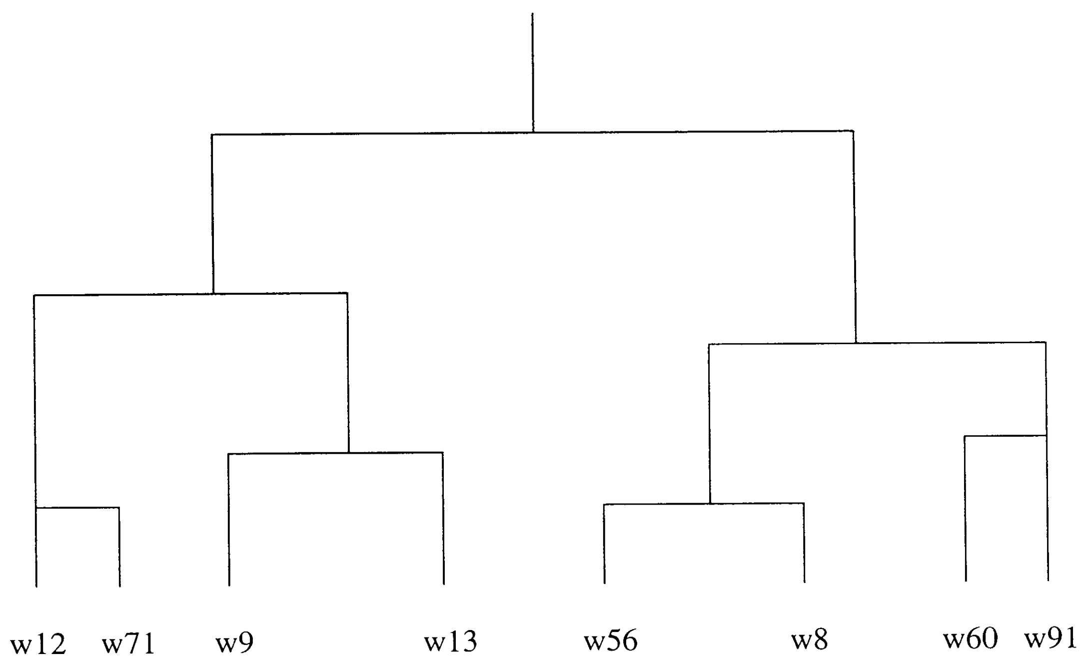

Figure 2
Example of a simplified hierarchical cluster representation

To evaluate the results of the analysis, we want to see how well the clusters that the model constructs match the actual linguistic categories. Since our analysis yields a hierarchy of clusters, we must pick a level in the hierarchy at which we want to consider the clusters to be our obtained categories. For present purposes we examine a middle level of clustering.[1] We have developed a set of quantitative methods for assessing how well the obtained distributional categorization coincides with the true linguistic categorization, which space does not permit us to go into here. See Bever,

[1] As is customary for any hierarchical clustering analysis, the lowest level of clustering gives many very narrow categories; the highest level of clustering gives few very broad categories. A middle level is, of course, in between on these characteristics. It is important to note, however, that for most of the results presented below, within fairly broad limits the choice of precisely where in the hierarchy to place our categorization turns out not to matter. The results across our manipulations are approximately the same at any of several levels of clustering.

Newport, Aslin, Mintz, Juliano, & LaMendola (1994) and Mintz, Newport, & Bever (in preparation) for further details. In the present paper we will instead show the actual clusters we obtain at a middle clustering level, so that the success of this procedure can be examined qualitatively.

Table 2 shows how the majority of the nouns (79%) cluster in the speech to one child, June. Note that there is only a small number of clusters containing nouns, and that these clusters contain almost exclusively nouns. (The complete depiction of the selected mid-level clusters for this sample is given in Table A1 in the Appendix.) Table 3 shows the clustering of *all* of the nouns in speech to another child, Peter. Notice that these clusters are even more uniformly nouns.

BABY BALL CANDY CHOO-CHOO DOGGIE DUCK FALL FOOD HAT HEAD KEYS KITTY LONG MOUTH NOSE PAGE PEOPLE PUPPET SHOE SOCK SOCKS STORY STRINGS THEM TICKET WOLF
CHILDREN CRADLE LITTLE MAN OTHER PUPPETS TREE
BUG CINDERELLA OUTSIDE POTATOHEAD SHE THAT THEY THIS
GIRL PIG PIGGIE

Table 2
Clusters containing Nouns for speech to June

BAG SLIDE TRAIN MICROPHONE SOLDIER FLOOR CAR BALL BRIDGE MOUSE DOGGIE WHEELS DONKEY AIRPLANE BLOCKS PEN BOYS BOX SEE WHEEL CHAIR PAPER TOYS MAN TOP OTHER MILK MEN HORSE TWO TAPE BOY MORE MAXINE WAY
BACON POCKET FOOT HEAD BOOK RING HOUSE PUZZLE SUITCASE ONE GONE
RECORDER

Table 3
Clusters containing Nouns for speech to Peter

Table 4 shows the majority of verbs (53%) (including Modals) for speech to June. As with nouns, clustering follows the actual linguistic categories quite well, although perhaps not as well as for nouns.

FIND GOING HAVE KNOW LIKE SING STARTED TELL WANNA WANT WENT
COME DANCE DOING GO GOT ROCK SAY SEE SLEEP THINK
CAN COULD DO DON'T FOR JIMINY SHOW THANK WOULD
AROUND BACK DOWN GOES IN OFF ON OVER ROUND WAS

Table 4
Clusters containing Verbs for speech to June

We see that nouns are clustered very well, and verbs not as well, but they are still quite good. Of course, this method also sometimes groups words together that form no single linguistic category, as seen in Table 5. Yet, while one might have expected this kind of poor result across the board, overall nouns and verbs appear much better than this.

We have just seen that immediate lexical context can indeed provide at least some basic information about form class to a language learner. However, it is possible

AND ANYTHING BOY BUT HE HERE'S HMM HOW I I'LL IT JUST LOOK MY NO OF OH OOPS POOR ROCK-A-BYE SAND SHALL THAT'S THERE'S TRY TWO UH UHHUH UP WELL WHAT WHAT'S WHEN WHERE WHERE'S WHO WHO'S WHY YEH YES
A ANOTHER ISN'T PUT REALLY TAKE TURN VERY WITH

Table 5
Mixed clusters for speech to June

that an infant learner is not be able to represent every word in the utterance as our procedure does. For instance, as Roger Brown (1973) first noted, speech produced by 2-year olds lacks functional categories; perhaps, then, they likewise do not have a representation of functional categories in perception. We were therefore interested in seeing what would happen to the obtained categories if function words were not used by our procedure in computing context similarity. To examine this, we ran a manipulation in which function words were treated in the same way as infrequent words, that is, their frequencies were not entered in the list of numbers representing the contexts of words. (This is conceptually equivalent to eliminating function words from the corpus.) There were two versions of this manipulation, one in which we eliminated all the function words, and one in which we eliminated only those which are the most prosodically reduced, that is, those which are unstressed monosyllables. Our reason for this latter variation of the manipulation was that this class of words is those which are least likely to be perceived and represented by infants.

However, researchers such as Shipley, Smith and Gleitman (1969), Peters (1977), Gerken, Landau and Remez (1990), and Echols (1993) have suggested that, while young language learners may not have a fully articulated representation of function words, they might nonetheless have some reduced representation of them (for example, as an unstressed syllable '@' or 'd@'). Therefore we ran one final manipulation where, instead of leaving the function words out of the analysis altogether, we reduced the different function words to a common symbol. We also varied this manipulation by employing the two definitions of function words (grammatical and prosodic) described above.

The pattern of results for all four manipulations was similar, and can be summarized as follows: The categorization of nouns sometimes gets worse when function words are eliminated or reduced. This is not surprising, seeing that, included in the function words are articles such as *a* and *the*, which are highly predictive of nouns. At the same time, while nouns do not fall out as favorably as in the original analyses, they still cluster reasonably well, as seen in Table 6. The categorization of verbs, on the other hand, never gets worse when function words are eliminated or reduced; and for samples where there is any significant change, the result is *better* than in our original analyses with full corpora. Table 7 shows an example of Verb categorization when the full corpus is analyzed, versus when function words are reduced. Notice first that, for this sample, the full corpus analysis does not yield results as good as what we have seen before for verbs. (Indeed, the outcomes do vary over the samples we have analyzed.) But what is interesting here is that, where the full corpus analysis is not ideal, the same corpus yields much better results after function words are reduced. This is most likely due to the fact that verbs in English occur before a variety of different prepositions, and thus in this sense have fairly dissimilar contexts. However, when function words are reduced to a common symbol, the dissimilar verb contexts become more similar.

IT TRAIN SLIDE BLOCKS BALL CAR TOGETHER BAG BOOK CHAIR FLOOR RING BOX HOUSE TOP
FOOT HEAD BACON POCKET
DONKEY MOUSE PETER SOLDIER SEE-SAW WHEEL PEN MICROPHONE MAXINE
GONE TURNING UPSIDE PUZZLE SUITCASE
HM THERE HERE PEEKABOO THAT TOYS NOW DONKEY

Singletons
MILK WHEELS PAPER BRIDGE MAN RECORDER WAY HORSE'S DOGGIE AIRPLANE BOYS BOY MEN

Table 6
Clusters containing Nouns for speech to Peter,
after function word elimination

Full Analysis	Reduced Function Words
BY CRACK CRACKING FIND HUMM IN ON UNDER	CRACK CRACKING FIND GET GOING GOT HAVE JUST LIKE LOOK PUT SITTING TAKE WANT WRITE
SITTING	
GET PUT TAKE	
GOING	
GOT	
CAN'T LIKE MAY SAY WANT	
LOOK	
SIT WRITE	

Table 7
Clusters containing Verbs for speech to Eve,
before and after function word reduction.

Taken together, these analyses yield two important results: First, speech to young children does contain significant distributional information as to the major form classes. These results are in accord with the findings of Brill (1991), Finch & Chater (1992, 1994), Redington, et al. (1994), and Schütze (1993). In fact, using much larger corpora and different analytic procedures, the results of these other investigators appear to be even cleaner than our own. However, the differences between these studies and ours are worth noting. These studies each use a window of two words to either side of the target word to compute context, whereas our study uses a window of only one word on either side. We discuss the significance of this factor below. In addition, excepting Redington et al. (1994), each of these studies investigates corpora of written (adult) text. In contrast, our corpora consist of spoken utterances to individual children of just the age at which initial form class learning occurs. Redington et al. (1994) also use CHILDES corpora, but they pool samples of adult speech directed to many children (and adults) spanning many age groups. Distributional information from large pooled samples of more complex and varied sentences, and from larger windows of information, may in fact provide cleaner induction of form classes than that derived from more restricted samples of the simple speech addressed to very young children. However, we believe that our method more closely approximates the actual linguistic environment of human learners, and therefore provides a stronger and more conservative test of distributional information for acquisition.

Our second result is that this kind of distributional information is apparently quite robust. This is demonstrated to some degree by the surprisingly good categorization which results under the limitations on input discussed in the previous paragraph. We did not initially expect that the immediate contexts of words, including only superficial adjacencies, could possibly provide information closely relevant to the abstract categories required in language learning. Nevertheless, it appears to be the case that analyses over a sizeable corpus of utterances, which utilize recurring sets of contexts, can provide enough predictive information about form class that a roughly accurate rendering of linguistic categories can be retrieved. In addition, the results of our analyses involving function word manipulations show that limitations on representation, of the type we might expect for young children, does not interfere with obtaining an approximately correct linguistic categorization, and in some cases can actually improve upon it. This could be an example of Newport's (1988, 1990) Less is More phenomenon, where constraints on young learners' information processing abilities may lead to more successful learning.

We would like to re-emphasize that we are not proposing that our analytic procedures are a model of form class acquisition. Rather, we have used these analyses to ask about the information available in the input. We believe we have demonstrated that there is potentially very useful information about form class structure in children's input. Initial category assignments could be based on this information, and then refined and augmented by more linguistically articulate knowledge—a kind of syntactic bootstrap into syntax. The categories we obtain by these methods are not perfect, nor complete, and we have seen that there are some clusters that emerge that do not coincide with linguistic categories. But, as we have stated several times, our procedure operates on very restricted types of information. How might one improve the results, while simultaneously making them more realistic? First, our procedures only consider the distributional evidence of the *immediately* surrounding words; a wider window, as used by other investigators, may help to reduce noise and erroneous results. The problem raised by widening this window, however, is the one with which we began: additional information increases the number of computations, and raises the issue of directing the computations to those which allow proper generalizations. Equally if not more important, therefore, are the other correlated sources of information that could be brought to bear on this problem by a young language learner. Prosodic cues (Gleitman & Wanner, 1982; Morgan & Newport, 1981; Morgan, Meier & Newport, 1987) could modify the analyses, perhaps restricting them to stretches of the speech stream (e.g., syntactic phrases) that contain the most consistent and relevant distributional information. Semantic information, of the type Maratsos & Chalkley (1980) and also Pinker (1984, 1987) suggest, could also be brought to bear. Gleitman and colleagues (Gleitman & Gillette, 1995) have shown that the nonlinguistic environment may provide contextual information for learning some nouns, though not for verbs. Finally, deeper and more abstract syntactic information could rule out erroneous category assignments: it is conceivable that as the grammar of the language develops out of the correct assignments and as parameters are fixed, the incorrect assignments may be ruled out. This bears a general resemblance to the constraint satisfaction approach outlined in Pinker (1987).

In sum, all by themselves, distributional analyses would not be sufficient (nor, it would seem, would any of the other just mentioned sources taken alone). But in concert with these other sources, it appears that distributional information can bear its share of the burden in the acquisition of form class and grammar, and perhaps plays a necessary role. This is a surprising result, given that superficial distributional evidence has usually been considered to be misleading or non-existent with respect to significant linguistic categories. Research we are beginning now (Mintz, in progress) investigates what aspects of the distributional characteristics infants are actually sensitive to, and what use they are able to make of them in processing speech sounds.

Appendix

$$\theta = \cos^{-1}\left(\frac{\textbf{context\_vector}_i \cdot \textbf{context\_vector}_j}{|\textbf{context\_vector}_i||\textbf{context\_vector}_j|}\right) \quad (1)$$

BETTER BUNNY CRY ELSE GIPETTO GITA GOOD GOODNESS GREAT HERE HIM HOME HUH IT JUNE MUCH NICE NOW OK ONE OUT PINNOCHIO RIGHT SIDE STROMBOLI TALK THERE THESE THOSE TOO WAY WONDER-FULL
BABY BALL CANDY CHOO-CHOO DOGGIE DUCK FALL FOOD HAT HEAD KEYS KITTY LONG MOUTH NOSE PAGE PEOPLE PUPPET SHOE SOCK SOCKS STORY STRINGS THEM TICKET WOLF
CAN COULD DO DON'T FOR JIMINY SHOW THANK WOULD
AND ANYTHING BOY BUT HE HERE'S HMM HOW I I'LL IT'S JUST LOOK MY NO OF OH OOPS POOR ROCK-A-BYE SAND SHALL THAT'S THERE'S TRY TWO UH UHHUH UP WELL WHAT WHAT'S WHEN WHERE WHERE'S WHO WHO'S WHY YEH YES
A ANOTHER ISN'T PUT REALLY TAKE TURN VERY WITH
FIND GOING HAVE KNOW LIKE SING STARTED TELL WANNA WANT WENT
CHILDREN CRADLE LITTLE MAN OTHER PUPPETS TREE
ABOUT ARE DID DOES HAPPENED IS
GIRL PIG PIGGIE
AROUND BACK DOWN GOES IN OFF ON OVER ROUND WAS
COME DANCE DOING GO GOT ROCK SAY SEE SLEEP THINK
BUG CINDERELLA OUTSIDE POTATOHEAD SHE THAT THEY THIS

Singletons
GET DIDN'T GONNA WE SOMETHING WILL MOVE LOOKS HAD TOP HIS YOUR ME SAID YOU

Table A1
Full analysis for speech to June

References

Bever, T. G., Newport, E. L., Aslin, R. N., Mintz, T. H., Juliano, C. & LaMendola, N. (1994). Computational studies of motherese. Submitted.

Bloom, L. (1970). *Language development: Form and function in emerging grammars.* Cambridge, Mass.: MIT Press.

Brown, R. (1973). *A first language: The early stages.* Cambridge, Mass.: Harvard.

Brill, E. (1991). Discovering the lexical features of a language. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics.* Berkeley, Ca.

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

Chomsky, N. (1986). *Knowledge of language : its nature, origin, and use.* New York: Praeger.

Echols, C. H. (1993). A perceptionally-based model of childrens' earliest productions. *Cognition,* **46,** 245-298.

Finch, S. P. & Chater, N. (1992). Bootstrapping syntactic categories. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society of America.* Hillsdale, New Jersey: LEA.

Finch, S. P. & Chater, N. (1994). Distributional bootstrapping: From word class to proto-sentences. In *Proceedings of the 16th Annual Meeting of the Cognitive Science Society of America.* Hillsdale, New Jersey: LEA.

Gerken, L., Landau, B., & Remez, R. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology,* **27,** 204-216.

Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition,* **1,** 3-55.

Gleitman, L. R., & Gillette, J. (1995). The role of syntax in verb learning. In P. Fletcher and B. MacWhinney (Eds.), *The Handbook of Child Language.* Cambridge, Mass.: Blackwell.

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

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

Higginson, R. P. (1985). Fixing-assimilation in language acquisition. Unpublished doctoral dissertation, Washington State University.

Jackendoff, R. S. (1977). *X-Syntax: A Study of Phrase Structure.* Cambridge, Mass.: MIT Press.

Kelly, M. (1992). Using Sound to Solve Syntactic Problems: The Role of Phonology in Grammatical Category Assignment. *Psychological Review,* **99,** 349-364.

MacWhinney, B. & Snow C. E. (1985). The child language data exchange system. *Journal of Child Language,* **12,** 271-296.

MacWhinney, B. & Snow C. E. (1990). The child language data exchange system: an update. *Journal of Child Language,* **17,** 457-472.

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

McNeill, D. (1966). Developmental psycholinguistics. in F. Smith and G. Miller (Eds.), *The genesis of language: a psycholinguistic approach.* Cambridge, Mass.: MIT Press.

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., & 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.

Newport, E. L. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Language Sciences,* **10,** 147-172.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science,* **14,** 11-28.

Peters, A. M. (1977). Language learning strategies: does the whole equal the sum of the parts?. *Language,* **53,** 560-573.

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

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. WacWhinney (Ed.) *Mechanisms of language acquisition.* Hillsdale, New Jersey: Erlbaum.

Redington, M., Chater, N., & Finch, S. (1994). The potential contribution of distributional information to early syntactic category acquisition. Unpublished ms.

Shipley, E. F., Smith, C. S., & Gleitman, L. R. (1969). A study in the acquisition of language: Free responses to commands. *Language,* **45,** 322-42.

Schütze, H. (1993). Part-of-speech induction from scratch. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics.* Columbus, Ohio.

Department of Psychology - Meliora Hall
University of Rochester
Rochester, NY 14627

mintz@psych.rochester.edu