Running head: Finding The Verbs

Finding The Verbs: Distributional Cues to Categories Available to Young Learners

Toben H. Mintz

University of Southern California

Address correspondence to:

Toben H. Mintz
Department of Psychology
SGM 501, MC-1061
University Park Campus
University of Southern California
Los Angeles, CA 90089-1061

Email: tmintz@usc.edu
Telephone: 213-740-2253
Fax: 213-746-9082

Acknowledgements:
This research was supported in part by a grant from the National Institutes of Health (HD40368).

Introduction

Before language learners can start to learn the meanings of verbs, they must first determine what words in their language are verbs. In this chapter I will discuss a kind of information inherent in the structure of children's linguistic input that they could use to categorize verbs together. The information is a type of *distributional information* involving the patterning of words in sentences. The general hypothesis is that words of the same grammatical form-class category (e.g., noun, verb, adjective, etc.) occur in similar distributional patterns across utterances, and that this information could be a basis for learners to identify verbs, as well as other categories. The specific distributional patterns considered are those arising from *frequent frames* (Mintz, 2003). For the purposes of this chapter, a *frame* is defined as any two words that occur in a corpus with exactly one word intervening. For example, in the sentence, “Who wants some ice-cream?”, *who__some* is a frame containing the word *wants.* The hypothesis explored here is that the words that are contained by a given *frequent* frame—a frame that occurs above some frequency threshold in a learner’s input—belong to the same grammatical category, and hence that frequent frames could provide a *bootstrap*, or initial basis, for categorizing and identifying verbs.

The focus of this chapter is a two-part examination of the degree to which the category of words can accurately be derived from distributional information. The first part motivates and discusses in detail a recent approach I have carried out involving frequent frames as a distributional context (Mintz, 2003), and along the way provides a comparison to other recent distributional approaches. The second part presents preliminary behavioral evidence that infants, indeed, categorize novel words based on distributional information, and perhaps based on frequent frames.

Deriving Categories From Distributional Information

In a classic study, Roger Brown (1957) showed that given a scene and a novel word, three- to five-year-old children's interpretation of how the word relates to the scene changed depending on the word's morphosyntactic environment. When shown an image of a pair of hands kneading a confetti-like material in a bowl, children who heard the scene described using the word *sib* as a verb, as in *to sib*, or *sibbing,* thought that *sib* referred to the kneading action; whereas if they heard *a sib*, they thought *sib* referred to the bowl. In short, the assumptions children made about word meaning depended on the grammatical category of the word.

Brown's study showed, among other things, that children pay attention to the syntactic privileges of words in determining their meanings. From a linguistic point of view, they appeared to categorize the novel word based on the morphosyntactic environment, as either verb or a noun, and inferred a meaning for the word that depended on this categorization. One way the children could have determined the category of the unknown word was to note what words surrounded it, and to note its morphological marking. For instance, children could identify as verbs, words that occurred after *to* or affixed with *–ing*.

The process by which a word's environment—the words that surround it, its morphological marking, its relative position in a sentence—is used to determine its category is called *distributional analysis*. Maratsos & Chalkley (1980) advanced a theory that children perform distributional analyses on their input to identify grammatical form-classes. On their proposal, children tracked the range of environments in which words occurred—including co-occurrence with other words and affixes—and grouped words together that occurred in overlapping environments. Semantic information played a role as well, in that an affix such as *–s* would be treated as a different element when it designates plurality than when it designates present tense. However, for Maratsos & Chalkley, a primary source of information was distributional information.

The possibility that children use distributional information to classify words is appealing. After all, linguists rely on distributional analyses to discover what words belong together as a class in newly-studied languages (Harris, 1951). If distributional information is useful for linguists in constructing a grammar of a language, couldn't it also be informative for children learning their first language? Despite the appeal of this possibility, it is not without problems. Some problems have been thought to be so serious that the approach was not widely considered viable (see especially Pinker, 1984, 1987).

One potential problem concerns how a learner is to identify the appropriate distributional contexts on which to base an analysis. To take an extreme example, the absolute position of a word in a sentence (e.g., 3rd word, 1st word, etc.), while a perfectly coherent distributional attribute, is clearly not an adequate basis for categorizing words (Pinker, 1987). Clearly, then, distributional environments appropriate for categorization must be defined relative to other words (and morphemes) in an utterance. The problem then becomes, to which environment(s) should one attend? On the surface, this is a difficult question, since the "right" context might vary from utterance to utterance. Consider the target word *monkey* in (1):

(1) a. The monkey is climbing up a tree.

b. The furry black and very funny monkey is climbing up a tree.

Perhaps *adjacent-to-the-left* is an informative distributional environment. In (1a), a distributional analysis procedure that categorized words based on this environment would categorize *monkey* with all the other words immediately preceded by the determiner *the* in other utterances, which, one can easily imagine, will include a number of other nouns. But in (1b), while the same informative determiner is present, it is not immediately adjacent to the target word; in fact, there is no constraint on the number of words that can intercede. *Monkey* would not be correctly categorized in this case, and *furry* would be incorrectly grouped with other nouns. The problem is that informative contexts are not necessarily in the same relative positions to target words from utterance to utterance (or even within an utterance). How is a learner to know which contexts to pay attention to in any given situation? (For additional problems of this type, see Pinker, 1984, 1987).

Arguments against distributional analyses as a means of *initially* categorizing words were not blanket arguments against the importance of distributional analyses at later points in language acquisition. Rather, the idea was that learners would already have to have a considerable amount of linguistic knowledge in order to know how to treat the distributional information in a linguistically meaningful way. It was only once a considerable amount of structure was already fixed in the child's developing grammar (and, crucially, that categories had already been assigned to many words and affixes by non-distributional means) that distributional information could play a role in determining the category of an unknown word. Indeed, Pinker describes the concept *structure-dependent distributional learning* (Pinker, 1984, p. 40-42), to refer to distributional analyses that are guided by some knowledge of phrase structure, and knowledge of the category membership of some other words or inflections in a sentence. In other words, although a relatively advanced language user (e.g., those in Brown's study) could use distributional information to categorize a novel word, it was argued that the earliest categorization of words had to be based on other sources.

Deriving Categories From Semantic Information

Given the link between grammatical form and meaning, some researchers proposed that learners initially determine the grammatical category of a new word by observing what kind of entity it refers to. If it refers to an object, or a substance, then it is a noun; if it refers to an action, then it is a verb. Such a proposal forms part of Pinker's (1984) Semantic Bootstrapping Hypothesis (see also Grimshaw, 1981; MacNamara, 1972). According to Pinker, aspects of the meaning of an utterance—who and what are being talked about—are transparent to learners, even before they have acquired much knowledge of the vocabulary and structure particular to their language. This allows learners to identify the *semantic* category of a word (e.g., action-word) by observing the referential contingency of the words' use (e.g., that it is used to refer to an action). Innate linking rules then allow the child to classify the word syntactically (e.g., as a verb). The newly categorized word can then be fit into the developing grammar, and at early stages might be used as a source of information for determining certain language-specific aspects of the grammar (e.g., head branching direction, case marking, etc.).

One problematic aspect of semantic bootstrapping that has been discussed extensively in the literature has to do with the difficulty of identifying the meaning, and thus the semantic category, of unknown words, especially verbs (for example, see

Gillette, Gleitman, Gleitman, & Lederer, 1999; Gleitman, 1990). For verbs, the problem is not that the meanings cannot be recovered, clearly they can since children eventually learn verbs. Rather, because of ambiguities inherent in verb-to-world mappings, learners apparently rely on structural information in the carrier utterance to focus them on the relevant aspects of the world. For example, when confronted with scenes in which a causative and a non-causative action simultaneously occur (e.g., one character feeds another character, and the latter eats what is fed to him), two-year-olds interpret novel verbs in transitive constructions as referring to causative actions (feeding), and novel verbs in intransitive constructions as referring to non-causative actions (eating) (Fisher, Hall, Rakowitz, & Gleitman, 1994; Naigles, 1990; Naigles & Kako, 1993). But on a semantic bootstrapping account, that structural information (transitive/intransitive) would not yet be available in early stages of verb-learning, as that is precisely what is hypothesized to be ultimately deduced once the category of the word is identified. Thus, this crucial information would not be available to cue the learner as to the intended referent of the verb.

To overcome the kind of ambiguity present in the example above, when syntactic cues are unavailable, semantic bootstrapping accounts assume that a learner can deduce the referent of a novel word by observing many situations in which the same unknown word is used to describe different scenes. The learner then abstracts away the elements that are common across all uses to arrive at the correct interpretation. For example, the child may hear the word *feed* used in situations in which there is no eating (e.g., one can feed a dog without the dog eating), and may hear the word *eat* used in situation where there is no feeding (e.g., when the dog finally comes to eat the food). However, compelling arguments have been made that cross-situational comparison cannot solve some of the logical problems in identifying a word's referent, without recourse to syntax (Fisher et al., 1994; Gillette et al., 1999; Gleitman, 1990; Gleitman & Gleitman, 1997; Landau & Gleitman, 1985). A classic example involves the events of chasing and fleeing. Any situation in which there is a chasing event, there is also a fleeing event, and vice-versa, so that no amount of exposure to chase/flee events in which the word *glip*, say, is uttered could resolve this mapping ambiguity (Fisher et al., 1994; Gleitman, 1990; Gleitman & Gleitman, 1997; but see also Pinker, 1994). To be sure, the problem faced by a learner in categorizing a word as a verb is less complex than determining its meaning; nevertheless, studies with adults have shown that cross-situational observation alone is not even suitable for determining weather an unknown word is a noun or a verb (Snedeker, Brent, & Gleitman, submitted; Snedeker, Gleitman, & Brent, 1999). In sum, procedures for identifying verbs that rely initially on identifying a words' semantic type might not be feasible.

Another problem for semantically driven categorization is that, even if the learner could reliably recover the semantic type of a word, the links between semantic and grammatical categories are not one-to-one, but many-to-many. One aspect of the problem is that that there are words for which the semantic antecedent conditions (e.g., action implies verb) do not come into play. For example, 'know' and 'love' are not actions, so linking rules would not be relevant for identifying them as verbs. This state of affairs (i.e., many semantic types mapping to the verb category) is not fatal for semantic bootstrapping: structure-dependent distributional learning was proposed to solve exactly this kind of problem. A more serious difficulty is that, as Maratsos & Chalkley (1980) discuss in detail, semantic-to-syntactic linking rules are subject to one-to-many mappings as well; that is, one semantic type can be associated with several syntactic types. For example, the words ‘action’ and ‘noisy’ are not verbs, but, Maratsos & Chalkley argue, they have action-like semantics and thus would be mapped to the verb category given the linking rules that would categorize words like ‘throw’ and ‘kick’ as verbs. In other words, nouns and adjectives can have action semantics as well, so young learners could apply action-word-to-verb linking rules to words that aren't verbs. Thus, even if mapping the to-be-categorized word to the correct semantic type could be reliably achieved, many incorrect grammatical assignments would be made by purely semantic-to-grammatical linking rules. Of course, distributional analyses do not run into these problems: Since the initial categorization is not dependent on accessing word meanings, the problems associated with finding a word’s referent in the world do not come up, and the unreliability of semantic linking rules is not a factor.

Frequent Frame Approach to Distributional Bootstrapping

Recently, investigators have started to re-examine the usefulness of distributional methods in providing an initial classification of words in child directed speech (Cartwright & Brent, 1997; Mintz, 2003; Mintz, Newport, & Bever, 1995, 2002; Redington & Chater, 1998; Redington, Chater, & Finch, 1998). By analyzing transcripts of input children actually receive, researchers have endeavored to understand whether the arguments raised against distributional approaches *in principle* are indeed problematic *in practice*. In different ways, the learning algorithms employed in these recent approaches take into account frequency when grouping words with other words in the input, essentially combining statistical and distributional approaches to categorizing words.

The consideration of frequency is potentially relevant because if the problematic aspects of distributional information are relatively rare—for example, if sentences like (1a), above, are the norm, and (1b) are infrequent—it may be possible to filter out the problematic cases via a sensitivity to frequency.[1] Frequency plays an important role in the present research in that the distributional environments that are analyzed in the service of categorization are only the environments that occur frequently.

While considering frequency might increase the informativeness of analyzing a given distributional context (e.g., *adjacent-to-the-right*) by filtering out deviant or misleading cases, the antecedent question of which distributional patterns should be analyzed remains. Distributional analyses can cover many (indeed, infinitely many) types of patterns and relationships among a variety of linguistic units (e.g., phonemes, syllables, morphemes, words). A challenge in developing an account of grammatical learning that incorporates distributional analyses at an early stage is discovering which distributional patterns might be particularly informative, and determining whether learners are sensitive to these patterns. One way to approach this challenge is to look to the behavioral literature for reasonable conjectures about the types of distributional

[1] One might question the practice of ignoring or discounting evidence because it is rare. For example, in some areas of syntax acquisition, rare constructions can provide crucial information (for example, Gibson & Wexler, 1994; Wexler & Culicover, 1980). But there is no reason that the same considerations must hold throughout all domains and all stages of acquisition. It is reasonable to posit a system that builds an initial categorization of words in which relative frequency of occurrence matters, and later stages in which importance is placed on different aspects of the input.

environments very young learners are likely to attend to, and then determine whether those environments embody grammatically relevant properties. For instance, could they be used to group verbs with other verbs, nouns with other nouns, etc? The logical problem of which distributional context(s) to attend to can be circumvented if one can demonstrate that the distributional contexts that learners *do* attend to can support categorization. Part of the motivation for studying the distributional properties of frequent frames (Mintz, 2003) arose from such a consideration. Another motivation is a logical analysis of how frequent frames could provide an informative categorizing context. These points will be addressed in turn.

Sensitivity to Frames in Processing

There is a growing body of evidence that infants are sensitive to patterns of regularity in their linguistic input. For example, infants notice patterns in syllable sequences and make conjectures about the words in their language based on sequences that are highly predictable (Saffran, 2001). Patterns of this type involve computing relations between adjacent elements, however languages also give rise to patterns of regularity that involve non-adjacent elements. For example, present progressive sentences in English involve some form of the copula (e.g., *is*) followed by the affix -*ing*, with a variable amount of intervening material, as in the sentence *She is washing the car*, or *She is slowly washing the car*. Eighteen-month-olds are apparently sensitive to these non-adjacent dependencies, as they notice when an ungrammatical auxiliary verb (e.g., *can*) occurs in place of the copula in the examples above (Santelmann & Jusczyk, 1998). In other words, infants store information about elements that co-occur at a distance and surround other material, which is just the type of pattern that constitutes frames, as defined here.

Further evidence that infants track frame-like co-occurrence patterns comes from artificial language learning studies with 18-month-olds. In one study, Gómez (2002) tested 18-month-olds' sensitivity to the contingency between two non-adjacent words separated by exactly one intervening word. She exposed learners to multiple sequences of nonsense words that each followed the pattern *aXb*, where *a* and *b* stand for particular words, and the word in the *X* position varies across sequences. She found that infants noticed the nonadjacent dependency between words *a* and *b* when the *X* word alternated between 24 different words (but not when it alternated between 3 or 12 different words). That is, a high degree of variability in *X* caused infants to noticed the relationship between *a* and *b.* Taken together, the Gómez, and Santelmann & Jusczyk (1998) studies demonstrate that 18-month-olds attend to just the kind of non-adjacent dependencies that define frames.

Additional evidence comes from work by Childers & Tomasello (2001) in which 2 ½-year-olds were instructed to produce novel action-verbs in transitive frames. Experimenters modeled the task using English verbs in transitive frames. Children performed better when the task was modeled with verbs that were embedded in the frequently occurring pronoun frames—e.g., *He's pulling it*, which creates the frame *he__it*—versus the less frequent frames created as a consequence of common noun arguments—e.g., *The cow's pulling the chair*. In other words, the verb category was more strongly activated when the verbs fell within frequent frames.

Of course, frames more broadly construed have been shown to play an important role in verb learning (Fisher, Gleitman, & Gleitman, 1991; Fisher et al., 1994; Gillette et al., 1999; Gleitman, 1990; Landau & Gleitman, 1985; Naigles & Hoff-Ginsberg, 1998). Finally, research from my own lab (Mintz, 2002) has shown that adults categorize nonsense words based on the words' distribution within frames.

In sum, a number of different studies from different lines of research support the notion that infants attend to frame-like non-adjacent co-occurrences of words (and morphemes), and that for children and adults, the frame influences how an intervening word is processed.[2] Together, these independent lines of research converge in support of the notion that frames are a salient distributional environment for young language learners. I now consider why frequent frames might logically be thought to provide informative contexts for categorizing words, and in particular verbs.

Frequent Fames As an Informative Context

To understand the specific benefits frequent frames might provide, it will be instructive to overview an alternative type of distributional information that has recently been studied—information from *bigrams* (Mintz et al., 1995, 2002; Redington et al., 1998). A bigram is any two word sequence in an utterance, and, typically, one word provides the distributional context for the other. Two of the many differences between the frequent frames approach are worth mentioning because they demonstrate potential advantages to using frequent frames: 1) the use of more restricted contexts, 2) the use of

[2] For further research on the properties of human and non-human primates' sensitivity to non-adjacent linguistic dependencies, see Newport & Aslin (2004), and Newport, Hauser, Spaepen, & Aslin (2004).

the frequency of the distributional context as opposed to the frequency of the target word to select which target words to categorize.

Comparing Frame and Bigram Contexts

The difference between bigrams and frames is illustrated by comparing how the context of the target word is represented under each type of system. Recall that a frame is defined here as the two end elements of a three-word sequence. Thus, the target word, *W*, in the sequence ‘ ... X W Y ...’, occurs in the frame *X__Y.* In other words, it “jointly follows *X* and precedes *Y*”. In contrast, bigram contexts would record two independent co-occurrence patterns, namely, “follows *X*”, and “precedes *Y*”. To understand the potential consequences of such a distinction, consider the three word sequence “ … to put it …” that might occur in a child-directed utterance. Under the current approach, the word *put* would be recorded as falling within the frame *to__it*, and it would be classified with other words that fall within that frame throughout the corpus. When this frame was analyzed in six corpora of child-directed English (see below), the resulting category contained exclusively verbs. In contrast, in a bigram analysis, the representation of the word *put*’s distribution would overlap somewhat with the representation of all words that are immediately preceded by *to*, and all those that are immediately followed by *it*. For example, sequences like “… to the store …”, and “… know about it …”, would give rise to representations of *the* and *about* that overlap somewhat with the representation of *put*. *Put* would not necessarily be categorized with *the* and *about*—its category would depend on details of the specific computational model and corpus—but the possibility is greater than in a case in which distributional information comes from frames.

The previous example demonstrates that frame contexts are structurally more restrictive than bigrams because they involve a relationship between the context elements themselves (the framing words), in addition to the relationship between context and target word. Moreover, the frames used for categorization are required to be *frequent,* which imposes an additional constraint on the context (the frame). It is reasonable to assume, *a priori*, that if a given frame occurs frequently in a corpus of natural language, the co-occurrence of the frame words (i.e., the existence of the frame) is likely to be caused by some systematic aspect of the language, rather than by accident. Therefore, words that, throughout a corpus of speech, occur inside instances of a frequent frame are likely to have some linguistically pertinent relationship, such as grammatical category membership. This is not a necessary outcome, but arguably a likely consequence of using frequent frames as contexts. Thus, the added restrictive nature of frame contexts versus bigrams, while not specifically linguistic, plausibly provides a more linguistically relevant context.

Verbs, in particular, might benefit from frame-like contexts, because in most simple transitive sentences, the verb's arguments form a frame around it. Given the relatively high frequencies of pronouns compared to full nouns, frequently occurring framing elements in subject and object position are likely to be the very small and frequent set of pronouns (and perhaps determiners in the object noun phrase). Hence, frames consisting of pronouns are likely to be very frequent and informative verb environments. With this in mind, recall that Childers & Tomasello (2001) reported that learning novel verbs is facilitated by pronoun-frame contexts.

Determining Which Words to Categorize

Another important difference between the present approach and prior distributional approaches concerns the criteria used to select which words are categorized. Although the selection criteria are logically independent of the categorization contexts (i.e., frame, bigram, etc.), in the bigram studies by Mintz et al. (2002) and Redington et al. (1998), the most frequent words in the corpus were categorized, regardless the frequency of the contexts in which they occurred. A potential drawback to such "target-centered" selection criteria is that the contexts of the most frequent words may not be among the most informative. For example, the word *and* is very frequent, but it occurs in a variety of contexts because it is relatively free in the kinds of words and phrases it can conjoin (nouns, noun phrases, relative clauses, etc.). As a result, its immediate contexts are extremely varied and overlap with the distributional patterns of words from different grammatical categories. In contrast, in the present approach, only words occurring in frequent contexts (the frequent frames) are categorized. This imposes a frequency criterion on the context itself, rather than on the target words. Under the assumption that frames that occur frequently are governed by linguistically relevant phenomena, this selection criterion is potentially advantageous because it selects target words that have linguistically informative contexts. Again, the issue of target word selection is logically independent of the distributional contexts used for categorization; it just so happens that the recent bigram approaches used target-centered selection criterion, and the frequent frames approach uses a context-based criterion (but see Christiansen & Monaghan, this volume, for a different approach).

In summary, frames provide a more restricted distributional context than bigrams, and *frequent* frames might provide a way of constraining distributional analyses to contexts that are likely to be determined by structure in the grammar, and hence to be linguistically informative. This kind of restriction could be beneficial for filtering out accidental and potentially misleading co-occurrence patterns (Pinker, 1984, 1987), by providing a means of capturing relevant contexts *before* learning the grammatical details of a language. Finally, frequent frames might be especially well suited for identifying verbs in English, as frames might occur around transitive verbs more reliably than around other categories.

Frequent Frame Analysis of Six Corpora

Analyses reported in Mintz (2003) evaluated the usefulness of frequent frames in categorizing words by analyzing transcriptions of speech to six different children, taken from the CHILDES database (MacWhinney, 2000): Eve (Brown, 1973), Peter (Bloom, Hood, & Lightbown, 1974; Bloom, Lightbrown, & Hood, 1975), Naomi (Sachs, 1983), Nina (Suppes, 1974), Anne (Theakston, Lieven, Pine, & Rowland, 2001), and Aran (Theakston et al., 2001). The frequent frames in each corpus were used to classify the words contained therein, and the resulting categories were compared to the target grammatical categories by visual inspection and by quantitative measures. Only sessions of each corpus in which the target child was 2;6 or younger were analyzed, and only utterances of the adults were analyzed. For a more complete description of the input corpora, see Mintz (2003).

Distributional Analysis Procedure

The following procedure was carried out separately on each corpus. First, an exhaustive tally was made of all the frames—where a frame is the first and third word of a three-word sequence—and the number of times each frame occurred in the corpus. Sequences that spanned two utterances could not contribute to a frame. Although utterance boundaries were explicitly marked in the corpora, there is evidence that infants perceive utterance boundaries from prosodic cues (Hirsh-Pasek et al., 1987), so restricting frames within utterances is not unreasonable. Next, a subset of these frames was selected as the set of frequent frames. The principles guiding inclusion in the set of frequent frames were that frames should occur frequently enough to be noticeable, and that they should also occur enough to include a variety of intervening words. Recall that variability within a frame was a crucial aspect of what made frames salient in Gómez's (2002) study. Several versions of the analysis were performed using different frequency criteria—one using an absolute frequency threshold, and another using a threshold that was relative to the total number of frames in the corpus. The outcomes varied little, so the results using only the relative threshold are discussed here. In that version, the set of frequent frames was selected to include all frames whose frequency in proportion to the total number of frames in the corpus surpassed a pre-determined threshold. (Complete details can be found in Mintz, 2003). This resulted in fewer than 50 frequent frames for each corpus (out of on average 2×10^4 unique frames). From the set of frequent frames that satisfied the frequency criterion, frames that had only one or two intervening word types were removed,[3] to ensure that there was some variability within the frames (Gómez, 2002). On average, only four frequent frames per corpus failed to satisfy this variability criterion.

Next, each instance of a given frequent frame was located in the corpus, and the intervening word was recorded and grouped together with the other intervening words for that frame, creating a frame-based category. The number of times each word occurred in a frame was also recorded. One can therefore distinguish between the number of word types that occur in a frequent frame, and the number of word tokens.

Quantitative Evaluation Measures

To obtain a standard measure of categorization success, comparable across corpora and to a control condition (described below), a quantitative measure of categorization called *accuracy* was calculated for each corpus. Accuracy was calculated by taking all the possible pairs of the words that were categorized together, and computing the proportion of pairs in which the items belonged to the same category. For example, if a frame-based category contained 30 words, there were 435 possible word pairs (30*29/2). If all words were verbs but one, all but 29 possible word pairings would result in pairs of the same category (since the one non-verb could pair with 29 verbs), so the accuracy would be 406/435, or .93.

Linguistic category membership was determined by hand, using one of two different labeling protocols. In Standard Labeling, each categorized token was labeled as

[3] A w*ord type* is a particular word form: e.g., *dog* and *cat* are different word types. A *word token* refers to a specific instance of the type: e.g., each instance the word *dog* in a corpus is an individual token of the type *dog*.

noun (nouns and pronouns), verb (verbs, auxiliaries, and copula forms), adjective, preposition, adverb, determiner, wh-word, “not,” conjunction, or interjection. In Expanded Labeling, nouns and pronouns were labeled as distinct categories, as were verbs, auxiliaries, and the copula. In situations where the grammatical category of the word was ambiguous (for example, if it was unclear whether “walk” was used as a noun or a verb) the corpus was consulted to disambiguate and appropriately label the word.

Accuracy was the primary outcome measure. A second measure, completeness, assessed the degree to which the analysis grouped in the same frame-based category words that belong to the same grammatical category. High completeness scores result when words that belong to the same linguistic category are concentrated in one distributional category. For example, completeness would be higher if all the verbs in an analysis occurred in just one frame rather than two or three, regardless of how accurate the groupings were. A complete description of the evaluation metrics are given in Mintz (2003).

Computing Chance Categorization

Chance-categories were created for each corpus as a baseline control against which to compare the accuracy of the frame-based categories. For a given corpus, a chance-category was generated for each frame-based category such that it contained the same number of tokens as the corresponding frame-based category. The content of the chance-categories was determined by selecting the word tokens from all the frame-based categories at random, and randomly assigning them to chance-categories. Token and type accuracy and completeness were computed (for both Standard and Expanded labeling) on the chance-categories to yield baseline measures.[4] The baseline essentially indicates the accuracy and completeness that could be achieved by randomly assigning the words in a manner that superficially matches the actual analysis (in the number and size of resulting categories), but ignores the distributional structure of the corpus.

Results of Frequent Frame Analysis[5]

The frequent frames contained, on average, 450 different word types per corpus, comprising approximately 4,000 word tokens per corpus. The number of tokens categorized in a given corpus was only about 5% of the total number of tokens in that corpus, but the types constituted approximately half of all the tokens in each corpus. In other words, for each instance a word occurred in a frequent frame, it occurred about 9 times in a position that was not within a frequent frame. This means that a large portion of word types in a corpus pass through the most frequent frames, even if only a relatively small portion of tokens of each type do, allowing robust categorization with minimal analysis. But how well did frequent frames categorize words? And, in particular, how well did they categorize verbs?

Tables 1 & 2 provide representative examples of the several of the frame-based categories computed from two of the corpora. Frequent frames contained words from a range of categories, including nouns, verbs, adjectives, pronouns, adverbs, and auxiliaries. As the tables show, the words contained in each frame-based category were almost exclusively from one grammatical category, and many of these categories were

---

[4] See footnote 3.

[5] The following presents the results in summary form; for details, and for results of related analyses, refer to Mintz (2003).

verb categories. In fact, for all corpora, the plurality of frames contained verbs (and nearly only verbs).

Quantitative accuracy measures for frequent frame analyses were also very good, averaging .98 for tokens and .94 for types (significantly higher than chance baseline scores of .49 and .50, respectively), reflecting the fact that words within a given frame-based category were almost exclusively members of the same grammatical category. Completeness scores (.08 for token and type), although significantly higher than in the baseline analyses (.04 for token and type) were relatively low, reflecting the fact that, although the derived categories were accurate, it was not the case that there was a single noun category, a single verb category, etc. Rather, words from a given grammatical category often occurred in several frames. This was especially true for verbs since, as previously mentioned, the plurality of frequent frames contained verbs. This issue will be taken up in a later section.

Not only did frequent frames produce very accurate categories, but there was a considerable amount of consistency across the six corpora in the frequent frames that occurred. On average, 45% of the frequent frames that occurred in any given corpus occurred in at least three of the five additional corpora. Depending on the specific threshold criterion for frequent frames, approximately 20-30% of the frequent frames for a given corpus occurred in all six corpora; for verbs, these included frames like *you__it*, *you__the,* and *to__it*. The consistency is appealing because it indicates that informative contexts are not idiosyncratic in the input to any particular child; this, in turn, means that a learner would not have to adjust his or her frame-based processing from one speaker to another. Whatever factors determine the structure of frames in a sample of speech apparently have broad similarity across corpora, as would be the case for reflexes of the grammar, as opposed to influences from idiosyncrasies of the speaker and situation.

Further Consolidation of Frame-Based Categories

These findings suggest that frequent frames could provide an extremely informative context for categorizing words. However, several issues need to be addressed to further understand how frames could provide a basis for finding the verbs. One issue has to do with how comprehensive the frame-based categories are: As noted above, while it is clearly the case that the groupings that the frame-based analyses produces are very accurate, there are multiple frame-based categories that correspond to a given linguistic category. The learner would need a way of consolidating a number of frame-based categories to come up with a uniform class that corresponds to a linguistic category. Fortunately, there is at least one plausible way that further consolidation can be achieved using information inherent in the frame categories themselves.

Not surprisingly, there is considerable overlap in the words contained in many of the frame-based categories. For example, the verb categories defined by frames *you__to*, *she__to*, *you__the*, etc., will generally have a number of member words in common because many of the same verbs can appear in each environment. Hence, multiple frame-based categories could be unified if they surpass a threshold of lexical overlap, and the resulting *consolidated* category might be a much more comprehensive collection of, say, all the categorized verbs in a corpus. This procedure was tested on the results from the six analyzed corpora, using a criterion of 20% token overlap. Specifically, each frame-based category was exhaustively compared to all the others; for a comparison between any frame-based categories A and B, if 20% of the tokens in category A were also in category B, and 20% of the tokens in category B were also in category A, then A and B were tagged to be joined into one consolidated category. All such comparisons between categories were performed only once, before any consolidation was carried out, and consolidation was transitive: if A and B were tagged to be joined, and B and C were tagged to be joined, then the consolidation process would place tokens in A, B, and C, in the same consolidated category.

The results produced consolidated categories that were as accurate as the categories defined by the individual frequent frames. In addition, the procedure came considerably closer to comprehensively including all the analyzed words of a given grammatical category. Interestingly, the largest consolidated category for each corpus (in tokens, and in types in all but one case) contained predominantly verbs. For example, in the Peter corpus, 24 frame-based categories containing primarily verbs were consolidated into one category containing 2904 tokens (254 types) that were main verbs, auxiliaries, or copulas, out of a total of 3191 tokens (283 types). These high proportions were actually the lowest out of all of the corpora. Table 3 gives descriptive information for the largest consolidated category for each corpus (always verbs) resulting from the initial frame-based categories The table includes the number of frame categories that were joined, and the number and proportion of verbs (including modals, auxiliaries, and copula forms) in tokens and in types. As the table shows, on average, across all corpora, 97% of the tokens in the largest consolidated category were verbs. Tables 4 & 5 list the largest category for two of the corpora: The Peter corpus, which yields the poorest verb category, and the Eve corpus, which yields one of the best and is representative of the other corpora.

In sum, these findings suggest that frequent frames could provide a robust cue for categorizing words, and could be particularly informative for categorizing verbs. But so far the discussion has been focused on how the distribution of words in frequent frames could be used as grouping cue, one that would group nouns with other nouns, verbs with other verbs, etc. But grouping is only a part of true linguistic categorization: The learner must also identify *which* of the frame-based categories are the verbs, which are the nouns, etc. It was not enough that the children in Brown's (1957) study knew that *sib* was a different kind of word in the environments *to sib* and *sibbing* compared to *a sib* or *any sib*. They had to know what the specific linguistic consequences were; essentially, they had to know that *sib* was a verb in one case, and a noun in the other. How could children identify the verbs in categories derived from frequent frames?

<u>Linking Frame-Based Categories to Syntactic Categories</u>

One solution assumes that part of children's innate linguistic knowledge is that there are categories such as noun, verb, and adjective. The problem then becomes one of labeling or associating distributionally derived categories with the innately specified system. Although it is questionable whether verb referents can be identified by learners without access to sentential structural information (Gleitman, 1990), the referents of concrete nouns have been argued to be recoverable from observations of the circumstances in which they are used (see also discussions in Fisher et al., 1994; Gillette et al., 1999). If this is so, then the distributional category that contains nouns could be readily identified based on the concrete nouns that are its members. Note that using a semantic-to-syntactic generalization to label an independently derived category avoids the one-to-many mapping problem encountered when attempting to derive syntactic categories from semantic ones, since the semantic information is simply used to determine a general tendency of a group of words that is independently categorized. (Indeed, this combined use of distributional and semantic information approaches more closely Maratsos & Chalkley's (1980) proposal.)

Once the noun category (or categories) is labeled, identifying the distributional class which contains verbs becomes much more straightforward, and is perhaps achievable without recourse to additional semantic information. A cross-linguistically viable procedure would be to label as verb the frame-based category whose members satisfy one of a predetermined set of possible relationships with already identified nouns, specifically, the category whose members take the nouns as arguments. A coarse representation of the argument structure of a set of utterances—the position of the nouns and a limited set of possible verb positions—might be sufficient to determine which distributionally defined word class is the verb category. Thus, initially words would be clustered distributionally. Next, the cluster containing nouns would be identified from the semantic tendencies of its members. The location of nouns in utterances would then be used by syntactically constrained mechanisms to guide the labeling of the verb category.

The procedure just outlined for labeling the verb category requires some notion of predicate-argument structure. Interestingly, an effective procedure for labeling verbs exists for the corpora analyzed here that does not require such knowledge. In these corpora, verbs are the largest categories after nouns. If the noun categories are labeled following the manner above, the next largest category could be identified as the verbs. Such an approach would be successful if applied to the consolidated categories as well.

Of course, such a simplistic procedure for identifying the verb categories might not turn out to be viable cross-linguistically, or even in other English corpora. Nevertheless, the results from the present analyses leave open the intriguing possibility that identifying the frequent frames that contain verbs is possible without calling upon knowledge of predicate-argument relations.

There is an alternative to the view that the distributionally defined categories must be linked to syntactic labels and that infants have innate knowledge of syntactic categories. According to Tomasello (2000a; 2000b), children's early lexical categories are not abstract adult categories, like verb. Rather, initial categories are "item-based" and organized around the specific environments in which words occur. The present findings might appear to mesh well with this view: children's early grammar could be constructed around individual (or consolidated) frame-based categories, and only later would these categories take on the more abstract status as in the adult grammar. In that case, perhaps one need not posit that children have an innate verb category that must be associated with the relevant frequent frames. But, eventually, children would have to assign words to the adult category, and something akin to the labeling procedures outlined above would be required. Thus, even if children's initial categories turn out to be more restricted than adults, the issue of how they are eventually integrated into an adult grammar remains.

The previous discussion provided a practical demonstration that a distributional analysis using frequent frames as contexts could successfully group words in accordance with their grammatical category. Approximately half the words in a corpus were categorized by analyzing the distributional contexts of only 5% of the tokens, so frequent frames are efficient categorizing contexts, as well. Together with research suggesting that infants attend to frame-like context, these findings suggest that infants may well be carrying out the kinds of analyses necessary to categorize words using frequent frames. I will now present preliminary findings that show that 12-month-old infants categorize novel words based on the words' distribution within frequent frames, and that they may be especially sensitive to frames that contain *verbs*.

Evidence that Children Use Distributional Information to Categorize Novel Words

This experiment used a version of the Headturn Preference Procedure (HPP, Hirsh-Pasek et al., 1987; Kemler Nelson et al., 1995), similar to the one described by Jusczyk & Aslin (1995). First, infants heard a set of sentences that each contained a nonsense word; they were then tested on whether they categorized the nonsense words based on the frames in which the words occurred. In half the sentences, the nonsense word occurred in a noun frame (henceforth, *nonce nouns*), and in the other half, the nonsense word occurred in a verb frame (henceforth, *nonce verbs*). For example, in *She wants you to deeg it*, the nonce word occurred in the *to__it* verb frame; in *I see the bist in the room*, the nonce word occurred in the *the__in* noun frame. After familiarizing an infant to sentences like these, categorization was assessed by testing for a difference in the infant's listening preference to novel *grammatical* sentence, in which the occurrence of the nonce word was supported by the distributional information (e.g., *I deeg you now!*), versus novel *ungrammatical* sentences (e.g., *I bist you now!*). The ungrammatical and grammatical sentences differed only in the nonce word. A difference in infants' preference for grammatical versus ungrammatical sentences indicated that they could discriminate the two sentence types. The simplest explanation of such a discrimination was that infants categorized the novel words based on the distributional information in the familiarization sentences and noticed whether the nonce word occurred in a position that was consistent with that categorization. Counterbalancing procedures (see below) further ensured that infants' listening behavior was not due simply to an idiosyncratic preference for a specific sentence or sentences, but rather reflected the relation between test and familiarization sentences.

Twenty-four infants averaging 12 months, 2 days (range 11 months, 15 days to 12 months 14 days) participated in the study, and were randomly assigned to two groups A & B, with equal numbers of subjects in each group. Group assignment determined which of two counterbalanced sets of familiarization materials an infant heard.

Familiarization Stimuli

The full set of materials for groups A and B is given in Table 6. Each infant heard two nonce verbs and two nonce nouns in a total of 12 familiarization sentences: six containing verbs and six containing nouns. The six verb sentences were comprised of two sets of "pair-sentences" and two singleton sentences. Paired sentences were identical except they differed in the particular nonce word; for example, one pair in group A was *She wants you to deeg it,* and the other was *She wants you to lonk it*). The four pair-sentences provided the distributional basis for categorizing the two verbs together. The two singleton sentences each contained a different one of the two nonce verbs. The singletons provided a basis for later testing categorization because the "missing pair" corresponds to one of the novel grammatical test items (e.g., the singleton *I lonk you now!* has the test sentence counterpart *I deeg you now!*). Nonce-noun sentences followed the same structure.

An attempt was made to select frames that were found in the prior distributional analyses to be frequent frames across the six analyzed corpora. The prevalence of a given frame across all corpora was taken to be an indicator of its general ubiquity in speech to children, and of the likelihood that it would be recognized by infants in our study. An additional selection criterion was that none of the frames shared the same initial word or the same final word. That is, selection of the frame *you__the* precluded selection of the frame *you__it*.[6] Because of this constraint, it was not possible to select only frames that occurred in all six corpora, since the most frequent and consistent frames overlapped considerably in the framing elements. Nevertheless, two of the verb frames were frequent frames in all six corpora, and one (occurring in a singleton sentence) was a frequent frame in four of the six corpora; one noun frame was a frequent frame in all six corpora, and one (occurring in a singleton sentence) was so in five. The frame surrounding the nonce noun in the remaining singleton noun sentence, *I put his __ on the box*, was not a frequent frame in any corpus because no other frequent frames satisfied the restriction on shared elements. However, that sentence followed the general structure of existing frequent frames in that it there was a determiner to the left and a preposition, *on*, to the right.

Finally, in one of the noun frames and one of the verb frames, the frame context was modified to allow the nonce word to be the last word in the sentence. This was done in order to heighten the salience of these novel word forms and increase the likelihood

---

[6] A full explanation of the reasons for this restriction would be lengthy and of minimal theoretical interest. Briefly, the restriction eliminates alternative explanations of the predicted results that do not involve categorization.

that infants would successfully segment them when they were sentence-internal. Thus, nonce words in those sentences were not strictly in frames. Nevertheless, the majority of occurrences of a given nonce word in the familiarization materials was within a frequent frame.

Although both nonce nouns and nonce verbs occurred in frequent frames, the frequency of the noun frames and verb frames differed in the six corpora. Pooling the corpora, the verb frames in this study occurred with an overall frequency of approximately 2200, whereas the noun frames occurred approximately 820 times. Hence, the relative difference in frequency between frames types is close to a factor of three. In as much as these differences are indicative of relative differences in frame frequency in the input to a given child, one might predict that the frame environments for nouns and verbs in this study would be differently effective in mediating categorization. Thus, we might expect categorization to be stronger in this study for verbs than for nouns, on the assumption that the more frequent environments might more readily foster categorization.

As can be seen by comparing materials for the two groups, in Table 6, the nonce words were counterbalanced across groups A & B, such that nonce nouns for group A were nonce verbs for group B, and vice-versa.

Test Stimuli

The four singleton frames (2 for nouns, 2 for verbs) provided environments to test category generalization. Table 6 shows the eight test items that were presented to both familiarization groups after the familiarization phase. Each singleton familiarization sentence provided the basis for one novel *grammatical* and one novel *ungrammatical* test sentence. The grammatical sentence was created by replacing the nonce word with the one from the same category that did not occur in that frame during familiarization. The ungrammatical sentence was created by replacing the nonce word with one of the nonce words from the other category. Thus, group A's familiarization sentence, *Can you deeg the room?,* formed the basis for creating the novel grammatical test sentence *Can you lonk the room* and the novel ungrammatical test sentence, *Can you bist the room?* Both groups of infants received the same test items. Due to the counterbalanced design, grammatical test sentences for group A infants were ungrammatical for group B infants, and vice-versa. This ensured that grammaticality was not confounded with particular test sentences.

The familiarization and test sentences were recorded by a trained female native English speaker who was blind to the predictions of the experiment. The speaker was trained to produce the sentences with normal prosody, appropriate for a simple declarative sentence or a question. The spoken materials were then digitized onto the computer that controlled the experiment.

Procedure

The infant sat on the caretaker's lap in a sound attenuated booth, and the experimenter sat in a separate control room. On each wall to the right and left of the infant, approximately at the infant's eye level, a yellow light was mounted, and beneath each light was a loudspeaker. On the wall in front of the infant a red light was mounted, and beneath the light was a small video camera through which the experimenter could view the infant. The presentation of stimuli through the loudspeakers and the activation of the lights were controlled by a computer. The computer also measured and recorded the infant's orientation times to each test stimulus, as indicated by the experimenter.

During the experiment, the parent listened to masking music. through tightly fitting headphones.

To initiate the experiment, the center light flashed until the infant looked center, at which point the center light was extinguished and the familiarization sentences were played through both loudspeakers. While the familiarization sentences played, the lights functioned as in the test phase to help keep the infant alert (see below), however the familiarization materials were played continuously and independently of the infant's head turn behavior and the light activity. The familiarization sentences were presented in six randomized blocks, for a total duration of approximately 90 seconds. Group A subjects heard familiarization sentences from List A, Group B subjects heard sentences from counterbalanced List B.

Following the familiarization phase was a brief *contingency training* phase. The center light was activated until the infant oriented to it for 2 seconds. The light was then extinguished and a randomly selected side light was activated. When the infant looked towards the side light, a 500 Hz pure tone with a duration of one second was repeated through the associated loudspeaker with a 100 msec pause between repetitions. The tone repeated until the infant looked away for two consecutive seconds or until the completion of 15 repetitions. At that point the light was extinguished, the sound stopped, and the center light commenced flashing to initiate a new trial. There were two trials of this type to provide the infant with a demonstration of the contingency between the lights, the auditory stimulus presentation, and the infant's behavior.

Test trials were identical to contingency training trials, except that test sentences replaced the pure tone, and the duration of the pause between repetitions of a test item within a trial was approximately 300 msec. On a given trial, infants heard one of the eight test sentences (four grammatical, four ungrammatical). Trials were presented in two blocks, with trial order randomized within each block. Group A and Group B subjects heard the same test items; but the grammatical sentences for Group A subjects were the ungrammatical sentences for Group B subjects, and vice-versa.

Selection of the stimulus presentation side on a given trial was random, but constrained such that the same side would not be selected in more than three consecutive trials

Results and Discussion

Overall, infants listened longer to ungrammatical over grammatical strings. Infant's mean listening time to grammatical and ungrammatical sentences was 7.5 seconds (SE=.35) and 8.2 (SE=.33) seconds, respectively. The difference was significant by a two-tailed t-test ($t(23) = -3.38$, $p < .005$), and by a Wilcoxon matched-pairs signed-ranks test ($p < .005$). Eighteen out of 24 infants showed longer looking times to ungrammatical sentences. Since the only systematic difference between the two sentence types was the distributional category of nonce words, infants' sensitivity to the difference strongly suggests that they categorized the nonce words based on distributional information.

Next, the grammaticality effect was tested separately for noun frame and verb frame sentences. Recall that the corpus frequencies of the noun frames and verb frames differed such that the verb frames were nearly three times as frequent as the noun frames. It was of interest, then, to see if there was a correlated difference in the degree to which the noun and verb frame contexts revealed the grammaticality effect. Separate analyses of listening times to test sentences in which the nonce word was in a verb frame and those in which the nonce word was in a noun frame revealed that the grammaticality effect was due entirely to verb-frame sentences. For noun-frame sentences, listening times to grammatical and ungrammatical items were not significantly different (7.5s, SE=.42, and 7.8s, SE=.46, respectively; $t(23)=-.581$, n.s.), whereas listening times to grammatical and ungrammatical verb-frame items were (7.5s, SE=.41, and 8.6s, SE=.26, respectively; $t(23)=-4.32$, $p<.001$, two-tailed). Figure 1 graphs mean listening times to grammatical and ungrammatical sentences for each frame type individually.

The grammaticality effect provides support for the hypothesis that 12-month-olds categorized novel verbs based on distributional information in the familiarization sentences. Thus, at an age when infants are just starting to produce their first words, they apparently are using sequential information to group words into classes. The failure to find this effect in the noun frame items could have been due to a variety of factors, and should be interpreted with caution. However, an intriguing possibility is that the difference in the effect for noun frames and verb frames is related to their different relative frequencies in children's input. As previously mentioned, the noun frames were, overall, less frequent in the corpus analyses than the verb frames. These and other factors could have resulted in the verb frames being noticed more, or used more effectively for categorization.

It is interesting to note that in a study with older, 15-month-old learners of German, Hohle, Weissenborn, Kiefer, Schulz, & Schmitz (2004) found evidence that infants use distributional information to categorize novel *nouns*, but not novel verbs. Using bigram contexts, they showed that novel words following a determiner were categorized as nouns but novel words following a subject pronoun were not categorized as verbs. They attributed this difference to the fact that determiners are more predictive of nouns than pronouns are of verbs in child-directed German. In other words, they, too, linked categorization to distributional properties in the input. There are other differences between the present study and Hohle et al.'s could account for differences in the results. But in a broader sense, the findings are similar: both studies suggest that categorization is driven by the informativeness of distributional contexts.

An additional point is that in this study, *detecting* categorization of the nonce words depended on different environments than initially categorizing them did. Infants may have categorized the nonce nouns together based on the familiarization pair-sentences (e.g., *I see the gorp in the room!* and *I see the bist in the room!*), but the test sentences may not have provided a sufficient cue to trigger recognition of the category. To address these issues, we are carrying out further studies in our lab, a) to replicate the finding with verbs, and b) to explore conditions under which categorization of noun frames might also be obtained.

Whatever the explanation for the advantage for verb categorization turns out to be, the verb/noun discrepancy suggests further that infants' categorization of verbs was not solely due to the distributional information contained within the familiarization material itself. If a grammaticality effect had resulted for both verb frames and noun frames, one could not with certainty conclude that categorization was due to the frames being frames *of English*. That is, the patterning of the nonce words in the set of familiarization sentences could provide evidence to a distributional learner who lacked experience with English that certain words "belonged together" (Mintz, 2002). Evidence of strictly experiment-internal distributional learning would certainly be interesting, however, the present pattern of results suggests that this is not the critical process at work. The immediate distributional environments of the nonce nouns and nonce verbs are, formally, equally informative. But they differ is in the range of situations in which a given infant has heard those words and environments prior to the study. Infants apparently brought their experience with English to bear on how they analyzed the nonce words in this study.

A final note of caution is warranted in interpreting the results as evidence of infants' categorization from *frames*, as such. Although the frequent frames were carefully selected for the experimental materials, in accord with the corpus analyses presented earlier, there is no guarantee that the categorization cues infants used were the frames themselves. There could be other informative distributional cues that infants exploited, including bigram information from the word preceding (or following) the nonce word. However, much research that I am aware of on bigram contexts suggests that bigram cues aid categorization only when correlated cues are provided from other sources, for example, phonological cues, semantic cues, or other types of distributional cues (Braine, 1987; Frigo & McDonald, 1998; Gerken, Gómez, & Nurmsoo, 1999; Gómez & Lakusta, 2004; Mintz, 2002; Smith, 1966; Wilson, 2000; Wilson, Gerken, & Nicol, 2000). In normal circumstances, correlated cues certainly are available (see Christiansen & Monaghan, this volume), and it is reasonable to suppose that learners will recruit converging sources of information in acquisition. However, in the present study it is unlikely that any other source of information was available, apart from distributional patterns.[7] Future work can make use of this experimental method to elucidate what specific distributional cues infants respond to. For example, by manipulating the distributional contexts such that frame-based accounts would predict degraded categorization performance but other accounts (e.g., bigram models) would not predict a change, one can see whether infants' behavior follows one or the other prediction in a consistent manner.

General Discussion

The first part of this chapter motivated, on logical and empirical grounds, the benefits of frequent frames in providing distributional information from which learners could determine the category membership of novel words. The second part of the chapter presented a preliminary step in investigating whether very young children indeed categorize a novel word based on its distribution within frames. The initial results suggest that, as early as 12 months of age, infants categorize novel words based on some type of distributional information, and, tentatively that infants might be especially sensitive to the distributional information relevant to verbs. Although the results of the study with infants are consistent with the interpretation that infants categorized based on

[7] The results also leave open the possibility that infants employed structure-dependent distributional learning, as outlined by Pinker (1984). That is, infants could have used knowledge of English phrase structure, as opposed to frequent frames, to guide their distributional analyses. Again, further study is needed to address the plausibility of this and other alternative distributional explanations. However, since these infants were only a year old, it is not clear whether the representations necessary for structure-dependent learning would be in place.

frames, many details remain to be worked out as to precisely which distributional patterns were and are the relevant ones. Nevertheless, the fact that infants used distributional information to categorize words after very little exposure, and that categorization is measurable with this experimental technique, suggests that further research will shed light on these remaining questions.

Cross-Linguistic Viability

The distributional analyses explored here rely on word order patterns to form categories. As noted above, the patterns corresponding to simple transitive sentences might make frequent frames especially useful for categorizing many verbs. Given the fact that word order patterns differ cross-linguistically, and that in many languages word order is much freer than in English, it is important to consider whether this approach to categorization is universally viable. Two points about this issue are worth noting. First, in a language in which word order is relatively free – grammatical relations being marked by inflectional morphology – it may turn out that there is nevertheless enough consistency in word orders that informative frequent frames would result. This is clearly an empirical question, however Slobin & Bever (1982) show that both children and adults have preferred canonical patterns for sequencing nouns and verbs, even in free word order languages (e.g. Turkish). Perhaps these canonical patterns will turn out to yield informative frequent frames. Second, even if canonical patterns are not informative, the essential properties of frequent frames might nevertheless be relevant for categorization in heavily inflected, free word order languages. A fundamental property of a frequent frame is that it is a relatively local context defined by frequently co-occurring units. In the procedures explored here, the units were words and the frame contexts were defined by words that frequently co-occur. In heavily inflectional languages, frequently co-occurring units are likely to be the inflectional morphemes, which are limited in number and are extremely frequent (much like the pronouns and other closed-class words in the frames here). Hence, if the learning mechanisms that notice non-adjacent dependencies (Gómez, 2002) are flexible as to the level of granularity of the entities involved (i.e., words or affixes), such mechanisms would potentially identify frames in which the framing elements are inflections.[8] Indeed, the morphosyntactic dependencies that infants in Santelmann & Jusczyk's (1998) study represented involved the affix -*ing*. It is an empirical question whether *morphological* frames would provide useful category information for the relevant languages, or whether other types of distributional information is best suited for analysis at this level; what type of information is relevant for infant learners of these languages is yet another empirical question. Clearly, further research into typologically different languages is necessary to determine whether a frame-based approach is universal applicability. Of course, cross-linguistic issues equally impact other types of distributional information. With additional flexibility of the type just described, frequent frames are amenable, in principle, to categorization in typologically different languages.

---

[8] Concerns about the perceptibility of affixes, and therefore their reliability in early acquisition are mitigated when one considers that for languages for which this level of analysis would be most fruitful, the affixes are phonologically more prominent (Gleitman & Wanner, 1982).

Summary and Conclusions

In summary, as with other recent studies, the frequent frames analyses described here demonstrate that many of the hypothesized problems for bootstrapping into word categories from distributional information turn out not to be relevant when actual corpora are analyzed. Furthermore, the non-adjacent dependencies involved in frequent frames form distributional patterns to which infants and young children have been shown to be sensitive. The distributional analyses presented here show that these patterns would be especially useful for categorizing verbs: frames containing verbs constituted the plurality of frequent frames, and when overlapping frame-based categories were joined together, the largest category contained the verbs. An account was provided as to how distributionally defined categories could become syntactic. The proposal was that the distributional information provides a bootstrap into a pre-existing linguistic system in which grammatical form-class categories—e.g., noun, verb, adjective—are distinguished[9]: Frequent frames provide a means of initially categorizing words, then the distributional category that contains the nouns can be identified by the tendency of its members to refer to concrete objects. Several possible procedures for identifying the distributional category containing the verbs were then discussed: Either by simple descriptive properties (e.g., the largest category that is not the nouns), or more linguistically informed procedures that involve structural notions. Just as cross-linguistic research is necessary to empirically test the viability of frequent frames as a universal approach to initial word classification, cross-linguistic research is also necessary to

[9] See Pinker (1984, p. 43) for an excellent discussion of what this means.

determine what kinds of mechanisms might be universally appropriate for labeling the verb category, but several plausible options are available.

Finally, the accompanying behavioral study was an initial investigation into infants' use of frames in categorizing novel words. The results demonstrate that infants as young as 12 months of age categorize novel words based on distributional information. After hearing novel words used in very limited contexts, infants categorized the words on distributional grounds. Moreover, there is some indication, although preliminary, that 12-month-olds might be selectively attentive to frames containing verbs.

One might wonder whether this advantage for verbs puts in question the proposal outline above, in which labeling verb categories depends on first labeling nouns. How would this work if infants are first sensitive to verb distributions? Recall that categorization and labeling are separable processes, and the infant study reported here concerned only categorization. It is conceivable that the order in which words are grouped together differs from the order in which the resulting groups are associated with syntactic categories (i.e., labled).

The research on frames and frame-like contexts in early syntax acquisition is just beginning, and many questions remain open. Nevertheless, the computational and behavioral studies presented here provide converging evidence that young learners initially use distributional information to find the verbs.

Table 1. Several frame-categories derived from the Peter corpus, including frequency counts.

Peter

you__it			*I__it*		*the__one*
put (52)	move (3)	squeeze (1)	see (18)	knock (1)	other (21)
see (28)	hold (3)	showing (1)	put (12)	knew (1)	red (11)
do (27)	give (3)	show (1)	think (9)	get (1)	yellow (8)
did (25)	fixing (3)	said (1)	got (8)	fixed (1)	green (8)
want (23)	drive (3)	rip (1)	thought (5)	finished (1)	orange (6)
fix (13)	close (3)	read (1)	have (5)	close (1)	big (6)
turned (12)	catch (3)	reach (1)	found (5)	build (1)	blue (5)
get (12)	threw (2)	pushed (1)	do (4)	bet (1)	right (4)
got (11)	taking (2)	push (1)	take (3)		small (3)
turn (10)	screw (2)	play (1)	open (3)		little (3)
throw (10)	say (2)	pick (1)	fix (3)		wrong (1)
closed (10)	ride (2)	parking (1)	did (3)		top (1)
think (9)	pushing (2)	made (1)	closed (3)		round (1)
leave (9)	hit (2)	love (1)	use (2)		only (1)
take (8)	hiding (2)	left (1)	tie (2)		light (1)
open (8)	had (2)	knock (1)	tear (2)		empty (1)
find (8)	eat (2)	knew (1)	need (2)		black (1)
bring (8)	carry (2)	hid (1)	know (2)		
took (7)	build (2)	flush (1)	hear (2)		
like (6)	brought (2)	finished (1)	guess (2)		
knocked (6)	write (1)	expected (1)	give (2)		
putting (5)	wiping (1)	dropped (1)	doubt (2)		
pull (5)	wipe (1)	drop (1)	wear (1)		
found (5)	wind (1)	draw (1)	took (1)		
make (4)	unzipped (1)	covered (1)	throw (1)		
have (4)	underneath (1)	closing (1)	threw (1)		
fixed (4)	turning (1)	call (1)	saw (1)		
finish (4)	touching (1)	broke (1)	read (1)		
try (3)	tore (1)	blow (1)	pushed 1)		
swallow (3)	tie (1)		pick (1)		
opened (3)	tear (1)		move (1)		
need (3)	swallowed (1)		leave (1)		

Table 2. Several frame-categories derived from the Aran corpus, including frequency counts.

Aran

you__it

put (28)	gave (2)	move (1)
want (15)	found (2)	manage (1)
do (10)	fit (2)	make (1)
see (7)	enjoy (2)	load (1)
take (6)	eat (2)	liked (1)
turn (5)	chose (2)	lift (1)
taking (5)	catch (2)	licking (1)
said (5)	with (1)	let (1)
sure (4)	wind (1)	left (1)
lost (4)	wear (1)	hit (1)
like (4)	use (1)	hear (1)
leave (4)	took (1)	give (1)
got (4)	told (1)	flapped (1)
find (4)	throwing (1)	fix (1)
throw (3)	stick (1)	finished (1)
threw (3)	share (1)	drop (1)
think (3)	sang (1)	driving (1)
sing (3)	roll (1)	done (1)
reach (3)	ride (1)	did (1)
picked (3)	recognise (1)	cut (1)
get (3)	reading (1)	crashed (1)
dropped (3)	ran (1)	change (1)
seen (2)	pulled (1)	calling (1)
lose (2)	pull (1)	bring (1)
know (2)	press (1)	break (1)
knocked (2)	pouring (1)	because (1)
hold (2)	pick (1)	banged (1)
help (2)	on (1)	
had (2)	need (1)	

put__in

it (49)	dolly (2)	panda (1)
them (14)	yourself (1)	her (1)
him (11)	you (1)	Pingu (1)
things (6)	what (1)	
that (5)	this (1)	
those (4)	these (1)	
teddy (2)	some (1)	

the__and

tractor (5)	ignition (1)
horse (4)	hut (1)
shark (3)	holes (1)
back (3)	hippo (1)
zoo (2)	hens (1)
top (2)	ham (1) (1)
tiger (2)	floor (1)
roof (2)	fire+engine (1)
leg (2)	eye (1)
grass (2)	entrance (1)
garage (2)	elephant (1)
window (1)	dolly (1)
wellingtons (1)	doctor (1)
water (1)	cups (1)
video (1)	cows (1)
train (1)	controls (1)
sun (1)	carts (1)
station (1)	carpark (1)
stars (1)	cake (1)
shop (1)	bus (1)
shirt (1)	bull (1)
sand (1)	brush (1)
round (1)	box (1)
rain (1)	bottom (1)
pussycat (1)	book (1)
postbox (1)	blue (1)
panda (1)	bits (1)
nuts (1)	bank (1)
mother (1)	bananas (1)
monkey (1)	animals (1)
lion (1)	air (1)
kite (1)	

Table 3. Number of frequent frames underlying the largest consolidated category, by corpus. Proportions represent the number of verbs (tokens and types) in the largest consolidated category out of the total number of words in that category (percentages in parentheses).

Corpus	Number of frames consolidated into largest (verb) category	Verb tokens ----------- (%) Total tokens	Verb types ----------- (%) Total types
Peter	24	2904/3191 (91%)	254/283 (90%)
Eve	16	1449/1468 (99%)	180/194 (93%)
Nina	10	1316/1333 (99%)	155/167 (93%)
Naomi	18	690/700 (99%)	135/144 (94%)
Anne	13	1445/1474 (98%)	169/182 (93%)
Aran	9	1721/1760 (98%)	199/224 (89%)
MEAN:	**15**	**(97%)**	**(92%)**

Table 4. Largest consolidated category for the Peter corpus. Words are ranked by frequency in the contributing frames (listed in parentheses), and asterisks indicate words that did not adhere to the predominant verb tendency of the category.

Peter

put (307)
want (262)
have (182)
see (143)
*in (131)
think (96)
like (92)
get (83)
take (76)
need (75)
fix (73)
*on (72)
open (70)
do (66)
is (64)
turn (56)
make (40)
throw (38)
got (38)
close (35)
leave (31)
hold (30)
did (30)
thought (29)
ride (28)
give (28)
bring (25)
*to (24)
pull (24)
read (21)
hit (21)
do (21)
push (20)
going (19)
find (19)
use (16)
turned (15)
try (15)
took (15)
finish (15)
closed (15)
move (14)
hear (13)
found (13)
eat (13)
wind (12)
trying (12)
show (12)
putting (12)
play (12)
draw (12)
touch (11)
wanted (10)
tell (10)
tape (10)
taking (10)
had (10)
tear (9)
know (9)
knocked (9)
help (9)
catch (9)
unscrew (8)
made (8)
guess (8)
*about (8)
write (7)
screw (7)
knock (7)
fixed (7)
cut (7)
say (6)
making (6)
keep (6)
drive (6)
doing (6)
does (6)
change (6)
build (6)
bang (6)
*under (5)
talking (5)
pick (5)
mean (5)
let (5)
learn (5)
fixing (5)
*down (5)
carry (5)
bet (5)
wipe (4)
were (4)
unwind (4)
told (4)
tie (4)
swallow (4)
roll (4)
riding (4)
pour (4)
*from (4)
eating (4)
changed (4)
break (4)
*at (4)
writing (3)
was (3)
turning (3)
*through (3)
threw (3)
stir (3)
stick (3)
squeeze (3)
spill (3)
smell (3)
said (3)
rip (3)
pushing (3)
pushed (3)
patting (3)
opened (3)
left (3)
kick (3)
*into (3)
*how (3)
hiding (3)
*for (3)
feel (3)
dump (3)
drink (3)
brought (3)
bringing (3)
bite (3)
*behind (3)
*all (3)
*with (2)
were (2)
wearing (2)
wash (2)
used (2)
unwrap (2)
understand (2)
throwing (2)
talk (2)
start (2)
showing (2)
shaking (2)
saw (2)
repeat (2)
remember (2)
lose (2)
learned (2)
knew (2)
*inside (2)
hurt (2)
heard (2)
hand (2)
goin(g) (2)
go (2)
giving (2)
finished (2)
empty (2)
drew (2)
doubt (2)
call (2)
broke (2)
bought (2)
are (2)
answer (2)
work (1)
wish (1)
wiping (1)
*who (1)
*what (1)
wear (1)
waving (1)
watching (1)
watch (1)
walk (1)
using (1)
unzipped (1)
untie (1)
understood (1)
underneath (1)
under (1)
tying (1)
tryin(g) (1)
tried (1)
touching (1)
tore (1)
tip (1)
thougt (1)
taught (1)
taping (1)
swallowed (1)
supposed (1)
stuff (1)
spread (1)
sit (1)
singing (1)
shut (1)
sharpening (1)
set (1)
scared (1)
scare (1)
saying (1)
s(u)pposed (1)
run (1)
referring (1)
*ready (1)
reach (1)
promise (1)
practiced (1)
poured (1)
pointing (1)
persuade (1)
pay (1)
passed (1)
parking (1)
pack (1)
*over (1)
opening (1)
*off (1)
*of (1)
*not (1)
*near (1)
moving (1)
misunderstood (1)
missed (1)
mess (1)
marked (1)
love (1)
lost (1)
losing (1)
lock (1)
listening (1)
lifted (1)
lick (1)
kissed (1)
juggle (1)
hug (1)
hole (1)
hid (1)
having (1)
havin(g) (1)
has (1)
*happy (1)
grabbed (1)
goes (1)
getting (1)
gave (1)
forgot (1)
fold (1)
flush (1)
fitting (1)
finding (1)
feed (1)
fasten (1)
expected (1)
erase (1)
emptying (1)
emptied (1)
dropped (1)
drop (1)
drawing (1)
decorate (1)
covered (1)
cover (1)
cook (1)
come (1)
closing (1)
clicking (1)
chew (1)
*by (1)
buy (1)
bouncing (1)
blow (1)
bend (1)
be (1)
*back (1)
attach (1)
ask (1)
*and (1)
Pat (1)

Table 5. Largest consolidated category for the Eve corpus. Words are ranked by frequency (listed in parentheses), and asterisks indicate words that did not adhere to the predominant verb tendency of the category.

want (143)	fold (8)	turned (3)	buy (2)	shake (1)
have (118)	did (8)	stir (3)	bed (2)	scratch (1)
put (91)	thought (7)	spill (3)	ask (2)	rocking (1)
like (91)	finish (7)	sing (3)	are (2)	reading (1)
get (63)	dropped (7)	shoot (3)	wrote (1)	reach (1)
going (54)	crack (7)	putting (3)	wiped (1)	ran (1)
see (52)	be (7)	jump (3)	wear (1)	push (1)
do (52)	wipe (6)	hit (3)	washed (1)	pour (1)
take (35)	spilled (6)	had (3)	wanted (1)	poke (1)
eat (34)	need (6)	goin(g) (3)	wan(t) (1)	pointing (1)
know (30)	move (6)	drop (3)	use (1)	playing (1)
say (26)	help (6)	catch (3)	untied (1)	ought (1)
play (25)	do (6)	bit (3)	twist (1)	must (1)
think (24)	cook (6)	watch (2)	turning (1)	loving (1)
read (24)	were (5)	wash (2)	trying (1)	look (1)
write (21)	taste (5)	untie (2)	try (1)	liked (1)
turn (16)	peel (5)	tie (2)	top (1)	lick (1)
make (16)	open (5)	swim (2)	throwing (1)	left (1)
find (15)	cut (5)	standing (2)	talk (1)	learn (1)
tell (14)	chew (5)	saying (2)	swallow (1)	kiss (1)
go (14)	blow (5)	riding (2)	suck (1)	*just (1)
doing (14)	bite (5)	*on (2)	stick (1)	*it (1)
throw (13)	*with (4)	made (2)	step (1)	*how (1)
bring (13)	went (4)	leave (2)	stay (1)	hope (1)
sit (12)	took (4)	knit (2)	stand (1)	heard (1)
give (12)	shut (4)	keep (2)	spit (1)	*head (1)
drink (12)	said (4)	fixed (2)	*some (1)	having (1)
hear (11)	pull (4)	fall (2)	snap (1)	guess (1)
fix (11)	pee+pee (4)	eating (2)	slipped (1)	*glad (1)
show (10)	*in (4)	cracking (2)	slept (1)	gave (1)
hold (10)	hurt (4)	cool (2)	sleep (1)	found (1)
touch (9)	draw (4)	come (2)	sitting (1)	*for (1)
had (8)	climb (4)	color (2)	shook (1)	fly (1)
got (8)	wish (3)	close (2)	sharing (1)	fixed (1)
forgot (8)	wind (3)	carry (2)	share (1)	finished (1)
				drinking (1)
				drew (1)
				drank (1)

Table 6. Familiarization sentences, group A & B, and test sentences, both groups.

GROUP A	
Verb Frame Familiarization Sentences	*Noun Frame Familiarization Sentences*
She wants <u>to deeg it</u>.	I see <u>the gorp in</u> the room.
She wants <u>to lonk it</u>.	I see <u>the bist in</u> the room.
You <u>can deeg.</u>	That's <u>your gorp.</u>
You <u>can lonk.</u>	That's <u>your bist.</u>
Can <u>you deeg the</u> room?	I put <u>his gorp on</u> the box.
<u>I lonk you</u> now!	Here's <u>a bist of</u> a dog.
GROUP B	
Verb Frame Familiarization Sentences	*Noun Frame Familiarization Sentences*
She wants <u>to gorp it</u>.	I see <u>the deeg in</u> the room.
She wants <u>to bist it</u>.	I see <u>the lonk in</u> the room.
You <u>can gorp.</u>	That's <u>your deeg.</u>
You <u>can bist.</u>	That's <u>your lonk.</u>
Can <u>you gorp the</u> room?	I put <u>his deeg on</u> the box.
<u>I bist you</u> now!	Here's <u>a lonk of</u> a dog.
TEST ITEMS	
Grammatical-A, Ungrammatical-B	*Ungrammatical-A, Grammatical-B*
Can <u>you lonk the</u> room?	Can <u>you bist the</u> room?
<u>I deeg you</u> now!	<u>I gorp you</u> now!
I put <u>his bist on</u> the box.	I put <u>his lonk on</u> the box
Here's <u>a gorp of</u> a dog.	Here's <u>a deeg of</u> a dog.

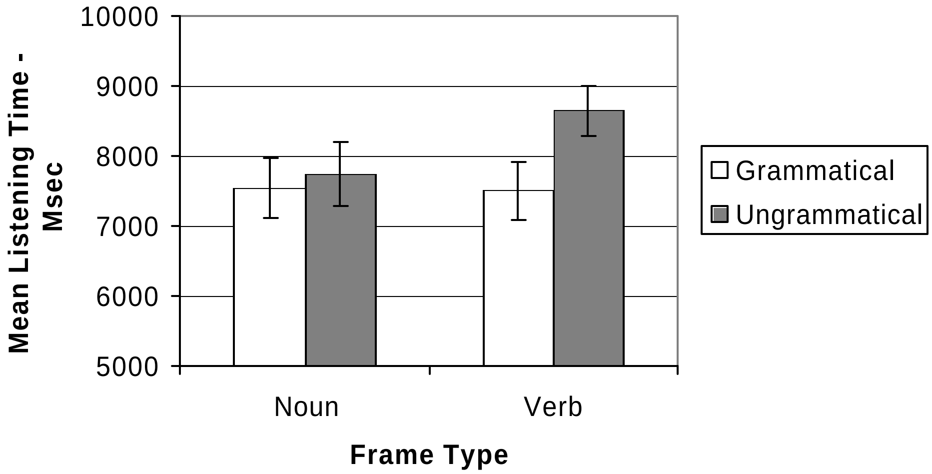

Figure 1. Mean listening times to grammatical and ungrammatical sentences, by frame type.

References

Bloom, L., Hood, L., & Lightbown, P. (1974). Imitation in language development: If, when, and why. *Cognitive Psychology*, *6*(3), 380-420.

Bloom, L., Lightbrown, P., & Hood, L. (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development*, *40*(2, Mono), 1-97.

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

Brown, R. (1957). Linguistic determinism and the part of speech. *Journal of Abnormal & Social Psychology*, *55*, 1-5.

Brown, R. (1973). *A first language; the early stages*. Cambridge: Mass. Harvard University Press.

Cartwright, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition, 63*(2), 121-170.

Childers, J. B., & Tomasello, M. (2001). The role of pronouns in young children's acquisition of the English transitive construction. *Developmental Psychology, 37*(6), 739-748.

Fisher, C., Gleitman, H., & Gleitman, L. R. (1991). On the Semantic Content of Subcategorization Frames. *Cognitive Psychology, 23*(3), 331-392.

Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. R. (1994). When It Is Better to Receive Than to Give - Syntactic and Conceptual Constraints on Vocabulary Growth. *Lingua, 92*(1-4), 333-375.

Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language, 39*(2), 218-245.

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

Gibson, E., & Wexler, K. (1994). Triggers. *Linguistic Inquiry, 25*(3), 407-454.

Gillette, J., Gleitman, H., Gleitman, L. R., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*(2), 135-176.

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

Gleitman, L. R., & Gleitman, H. (1997). What is a language made out of? *Lingua, 100*(1-4), 29-55.

Gleitman, L. R., & Wanner, E. (1982). Language Acquisition: The State of the Art. In E. Wanner & L. R. Gleitman (Eds.), *Language Acquisition: The State of the Art*. Cambridge: Cambridge University Press.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431-436.

Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science, 7*(5), 567-580.

Grimshaw, J. (1981). Form, Function, and the Language Acquisition Device. In C. L. Baker & J. J. McCarthy (Eds.), *The Logical Problem of Language Acquisition*: MIT Press, Cambridge.

Harris, Z. S. (1951). *Structural Linguistics*. Chicago, IL: University of Chicago Press.

Hirsh-Pasek, K., Nelson, D. G. K., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26*(3), 269-286.

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

Jusczyk, P. W., & Aslin, R. N. (1995). Infants Detection of the Sound Patterns of Words in Fluent Speech. *Cognitive Psychology, 29*(1), 1-23.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development, 18*(1), 111-116.

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

MacNamara, J. (1972). Cognitive basis of language learning in infants. *Psychological Review, 79*, 1-13.

MacWhinney, B. (2000). *The CHILDES project : tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Maratsos, M. P., & Chalkley, M. A. (1980). The Internal Language of Children's Syntax: The Ontogenesis and Representation of Syntactic Categories. In K. E. Nelson (Ed.), *Children's Language, Volume 2* (pp. 127-214): Gardner, New York.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition, 30*(5), 678-686.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*(1), 91-117.

Mintz, T. H., Newport, E. L., & Bever, T. G. (1995). Distributional regularities of grammatical categories in speech to infants. In J. Beckman (Ed.), *Proceedings of the 25th Annual Meeting of the North Eastern Linguistics Society.* Amherst, MA: GLSA.

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science, 26*(4), 393-424.

Naigles, L. R. (1990). Children Use Syntax to Learn Verb Meanings. *Journal of Child Language, 17*(2), 357-374.
Naigles, L. R., & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. *Journal of Child Language, 25*(1), 95-120.
Naigles, L. R., & Kako, E. T. (1993). 1st Contact in Verb Acquisition - Defining a Role for Syntax. *Child Development, 64*(6), 1665-1687.
Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127-162.
Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance - II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology, 49*(2), 85-117.
Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Pinker, S. (1994). How Could a Child Use Verb Syntax to Learn Verb Semantics. *Lingua, 92*(1-4), 377-410.
Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes, 13*(2-3), 129-191.
Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22*(4), 425-469.
Sachs, J. (1983). Talking about the there and then: the emergence of displaced reference in parent-child discourse. In K. E. Nelson (Ed.), *Children's language* (Vol. 4). Hillsdale, NJ: Lawrence Erlbaum Associates.
Saffran, J. R. (2001). Words in a sea of sounds: the output of infant statistical learning. *Cognition, 81*(2), 149-169.
Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: evidence for limitations in processing space. *Cognition, 69*(2), 105-134.
Slobin, D. I., & Bever, T. G. (1982). Children Use Canonical Sentence Schemas - a Crosslinguistic Study of Word Order and Inflections. *Cognition, 12*(3), 229-265.
Smith, K. H. (1966). Grammatical intrusions in the recall of structured letter pairs: Mediated transfer or position learning? *Journal of Experimental Psychology, 72*(4), 580-588.
Snedeker, J., Brent, M. R., & Gleitman, L. R. (submitted). The changing character of the mental lexicon: An information-based account of early word learning.
Snedeker, J., Gleitman, L. R., & Brent, M. R. (1999). The Successes and Failures of Word-to-World Mapping. In A. Greenhill, H. Littlefield & C. Tano (Eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development* (pp. 654-665): Cascadilla, Somerville, MA.
Suppes, P. (1974). The semantics of childnre's language. *American Psychologist, 29*, 103-114.
Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: an alternative account. *Journal of Child Language, 28*(1), 127-152.

Tomasello, M. (2000a). Do young children have adult syntactic competence? *Cognition, 74*(3), 209-253.

Tomasello, M. (2000b). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences, 4*(4), 156-163.

Wexler, K., & Culicover, P. W. (1980). *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.

Wilson, R. (2000). *Category learning in second language acquisition: what artificial grammars can tell us.* Unpublished masters thesis, University of Arizona.

Wilson, R., Gerken, L. A., & Nicol, J. (2000, November). *Artificial grammar research extends to natural language: Implicit learning of categories.* Paper presented at the annual meeting of the Psychonomic Society, New Orleans.