

The Effect of Word-internal Properties on Syntactic Categorization: A Computational Modeling Approach

Fatmeh Torabi Asr, Afsaneh Fazly, Zohreh Azimifar

Computer Sciences and Engineering

Shiraz University

Shiraz, Iran

{torabi,fazly,azimifar}@cse.shirazu.ac.ir

Abstract

We study the acquisition of abstract syntactic categories of words in children by using a computational model of categorization. Especially, we examine the effect of word-internal properties, such as morphological and phonological cues, on the identification of different categories, such as nouns, verbs, and determiners. To evaluate our model, we use it to determine the syntactic category of actual novel words selected from naturalistic child-directed utterances. We argue that such an evaluation is necessary for a better understanding of the effect of different cues (including word-internal properties and contextual cues) on category acquisition.

Keywords: Computational modeling, Syntactic category learning.

Introduction

Infants have a good understanding of the syntactic categories of words long before attending school. Psychological observations at different stages of child language development have shown the ability of children to recognize abstract (syntactic or semantic) categories, such as verb and noun, countable and uncountable (Brown, 1957; Gelman & Taylor, 1984; Samuelson & Smith, 1999). A variety of proposals exist in the psycholinguistics literature regarding the types of cues that are informative about such word categories, and the way children may use them to learn the categories. Computational modeling has often been used as a powerful tool to shed light on many aspects of language acquisition, including word categorization (Pearl, 2009). In this study, we draw on an existing categorization model in order to achieve a better understanding of the mechanisms and the information sources children use during the acquisition of syntactic categories, such as verbs and nouns.

Syntactic category learning in children has been suggested to be based on several information sources, such as word-external properties including distributional information about neighboring (co-occurring) words, as well as word-internal properties such as phonological and morphological cues (e.g., Brown, 1957; Gerken et al., 2005; Monaghan et al., 2007). Many of the computational studies on syntactic category acquisition focus on showing the relevance of the above properties to the acquisition of adult-like, linguistically-salient categories, such as verbs, nouns, and adjectives. For example, Mintz (2003), Monaghan et al. (2007) and Onnis and Christiansen (2008) present analyses of child-directed data to determine whether there are correspondences between particular syntactic categories and certain types of cues. Moreover, most of the existing computational models of child category

learning lack cognitive plausibility in some respects: The categorization models proposed by Schütze (1993), Redington et al. (1998), and Clark (2003) incorporate a batch (non-incremental) clustering algorithm; The connectionist model of Onnis and Christiansen (2008) is minimally supervised, assumes a fixed number of categories, and can only be used to study words in isolation.

A few studies have introduced cognitively-plausible models for syntactic category learning (Cartwright & Brent, 1997; Parisien et al., 2008; Alishahi & Chrupała, 2009). These incorporate fully-unsupervised incremental algorithms for clustering words as they appear in naturally-occurring utterances. However, these studies have focused solely on the role of context (co-occurring words) for inferring the syntactic category of a target word, and have overlooked the importance of other sources of information, such as phonology and morphology.

In our modeling of syntactic category acquisition, we address some of the aforementioned shortcomings. Specifically, we choose a simple incremental clustering algorithm (one proposed by Alishahi & Chrupała, 2009), which we further modify to increase simplicity. In addition, we examine the role of word-external information sources (namely, word co-occurrence), as well as that of word-internal sources (namely, phonology, and morphology) in order to better understand the interactions among these types of cues on the acquisition of syntactic categories. We use only very simple cues that are known to be accessible by children early in their language development. Finally, we propose and use a novel evaluation framework to examine the role of each type of information in the acquisition of syntactic categories.

Results of our experiments on naturally-occurring English child-directed utterances indicate that different cues are useful for the identification of different classes of words. In particular, we find that the identity of the word is essential to the identification of closed-class words. Open-class words, however, share similarities with respect to other types of cues, both word-external and word-internal. Nonetheless, even among these classes, different categories seem to be identified based on different properties: whereas verbs are better categorized with the help of morphological and phonological properties, co-occurrence information alone is reliable for categorizing nouns.

Algorithm 1: Incremental word clustering

```

1: initialize set of clusters  $\mathcal{K} = \emptyset$ 
2: for every frame  $f$  do
3:    $C_M = \operatorname{argmax}_{C \in \mathcal{K}} \operatorname{Sim}(f, C)$ 
4:   if  $\operatorname{Sim}(f, C_M) \geq \theta$  then
5:     Add frame  $f$  to cluster  $C_M$ 
6:   else
7:     Construct a new cluster for frame  $f$ 
8:   end if
9: end for

```

(This algorithm is a modification of the one proposed by Alishahi & Chrupała, 2009).

Modeling the acquisition of syntactic categories

Our goal is to build a computational model of syntactic categorization that is cognitively plausible, i.e., we make as few assumptions as possible about the type of cues accessible to young children, and about the mechanisms children might use for categorization. We thus use an adaptation of a simple incremental algorithm proposed by Alishahi and Chrupała (2009), which forms categories simply by drawing on the similarity among words to be categorized. Here, we present an overview of our adaptation of the algorithm, and a description of three types of cues we use for categorization.

The categorization algorithm

The unsupervised clustering algorithm proposed by Alishahi and Chrupała (2009) works based on contextual similarities among words. The algorithm is incremental in that it processes words one by one, discarding each word after clustering. For each newly-observed *frame* (a target *head-word* along with its neighboring words from left and right), if the similarity to all of the already-shaped clusters is less than a predefined threshold, a new cluster is constructed. Otherwise, the word is assigned to the most similar cluster. We modify this algorithm in two ways: (i) the original algorithm of Alishahi and Chrupała includes a phase in which clusters are merged if they are sufficiently similar. To keep the algorithm simple, we removed this step; (ii) our frames are composed of three different types of *features* (five features in total besides the head-word content; see next subsection for details). We thus need to slightly modify the similarity score calculation in order to accommodate for more than one set of features. The similarity between a frame and a cluster (a group of frames) is calculated as in:

$$\operatorname{Sim}(f, C) = \sum_{i \in \mathcal{F}} \omega_i * \operatorname{Sim}_i(f, C) \in \mathcal{F} \quad (1)$$

where f is a frame, C is a cluster, i is a feature, \mathcal{F} is the set of all features, $\operatorname{Sim}_i(f, C)$ is the similarity of frame f to cluster C with respect to the i^{th} feature, and ω_i determines the weight of the contribution of feature i in determining the overall similarity. Weights for all features need to sum to 1, i.e., $\sum_i \omega_i = 1$. The modified version of the algorithm is shown in Algorithm 1.

Cues used for categorization

As previously mentioned, children are known to group words into syntactic categories by drawing on a number of different information sources. In our work, we include three different sources of information, and five types of cues (features) in total, as explained below:¹

- Distributional information about word co-occurrences: This kind of information has been reported to be reliable and very important in syntactic categorization (Schütze, 1993; Redington et al., 1998; Mintz, 2003; Clark, 2000; Parisien et al., 2008; Alishahi & Chrupała, 2009). We take one word from each side of a target head-word as its co-occurrence features, because in many of the above studies words closer to a word have been shown to be more informative about its category. For example, considering sentences, such as “There is a cat in the basket”, and “We need a table in our kitchen”, “A cat is in the basket”, and “A table is in the kitchen.” provides a clue to the model to group *cat* and *table* together since they share similar co-occurrence features. In our framework, each co-occurring word is considered as an independent feature when determining similarity between a word (frame) and a cluster (as in many previous studies, and in contrast to representations such as “frequent frames” of Mintz, 2003). For example, even if the two tokens *cat* and *table* did not share the preposition *in*, they would still be considered as similar because of the preceding determiner *a* they have in common.

- Phonological information: Words belonging to the same syntactic category tend to have common phonological properties. For example, looking at child-directed utterances, (Monaghan et al., 2007) show that verbs and nouns are different with respect to several phonological features, including the number of syllables. The study of Monaghan et al. focuses on the relevance of syntactic categories and a large number of word-level, syllable-level, and phoneme-level phonological properties. We focus here on two of the simplest word-level phonological properties that we assume are readily accessible by young children, namely the length of a word in terms of number of syllables and phonemes (we use the number of letters to approximate the number of phonemes in a word).
- Morphological information: It has been shown that English affixes, such as *-ing* in verbs, can provide strong clues to the identification of syntactic categories, and that such information is abundant in child-directed speech (Onnis & Christiansen, 2008). Nonetheless, it is not clear whether we can assume that children have access to such accurate morphological knowledge about words and categories prior to syntactic category learning. Inspired by the work

¹In this study, we do not consider one other important source of information for learning of syntactic categories, namely, semantic information about words. This type of information requires making assumptions about what *meaning* is and how children may represent it, and hence is outside the scope of this study.

of Onnis and Christiansen (2008), here we use the last phoneme (ending) of the words as an approximation of the morphological affixes.²

Overall, we include six different features (cues) in our categorization: two Cooc features, Head word, two Phon features, and one Morph feature. The Cooc cues are considered as properties external to the word (properties of the context the word appears in), whereas the rest are related to the word itself and hence are considered as word-internal cues. In our experiments, we examine the effect of each different type of cue on categorization, and also consider the role of word-internal cues versus external ones.

Experimental Setup

Corpus

We extract our input data (both for training and testing) from the Manchester corpus (Theakston et al., 2001), one of the English subsets in the CHILDES database (MacWhinney, 2000). The Manchester corpus contains conversations of parents/caregivers with 12 British children between the ages of 1;8 (years;months) and 3;0.³ For training, we choose around 10000 child-directed utterances from the conversations of all 12 children, such that the chronological order of the utterances is maintained, and the utterances contain only words selected from a limited vocabulary of 500 words. When selecting the 500 words, we make sure that their distribution in the corpus matches a Zipfian distribution, so that our results are not biased towards words from certain frequency ranges. We limit the size of vocabulary because some feature values need to be determined manually. In addition, in one experimental task, we need access to actual novel words not previously seen in the training corpus, as opposed to made-up novel words used in many psychological experiments.

We use two different test corpora, one for each experimental task (as explained in the Evaluation subsection below). The first set of test data (used in the Word Category Prediction Task) is selected exactly as the training data, though from a non-overlapping portion of the original (Manchester) corpus. The other test data (used in the Novel Word Categorization Task) is selected such that the target words to be categorized are a novel word not in the vocabulary of 500 words. This second test set is similar to the training data in all other aspects. Each test corpus contains 2000 word usages (tokens to be categorized).

Feature Extraction

From each utterance (in the training or test data), we extract a number of frames to be clustered. As explained previously,

²We also included the first phoneme (beginning) of a word as also done by Onnis and Christiansen (2008). However, in our initial evaluations we found that the inclusion of this feature did not affect the results, and hence removed it from our set of features.

³Thanks to Chris Parisien for providing us with a preprocessed version of this corpus.

Head:	<i>table</i>	Cooc:	<i>a, in</i>
Phon:	2, 5	Morph:	<i>l</i>

Figure 1: Sample frame extracted for the target word *table* from the utterance “We need a *table* in the kitchen”.

each frame contains a head word (the target word to be categorized), as well as several other features (two Cooc, two Phon, and one Morph features). A sample frame is shown in Figure 1. The head word and the Cooc features can be directly extracted from the utterance. If any of the Cooc features are missing (i.e., the target word is the first or the last word of the utterance), that feature is set to “Unknown”. For the two other types of features (Phon and Morph) we need to have access to a phonemic representation of words and other phonological features. We extract two of these features (the ending phoneme, and the number of syllables) from the MRC Psycholinguistic Database, a publicly available resource built for use in studies on child language (Wilson, 1988).⁴ If a word is not found in MRC, we set the values of the above features manually. For the third feature, the number of phonemes in a word, we use the number of letters as an approximation.

Evaluation

To examine the contribution of different types of cues on syntactic categorization, we evaluate the effectiveness of clusters resulting from one or a combination of features in two tasks. Specifically, we train our model (on the training corpus) in three different conditions, that is, using one of the following feature combinations: Head+Cooc, Head+Cooc+Morph, Head+Cooc+Phon. We then determine the effectiveness of the resulting clusters in each condition by examining the performance of the model on inferring the category of a number of test words. Note that the model does not create any new clusters during the test phase, but assigns each word to one of the clusters formed in the training phase.

We evaluate our model using two experimental tasks: one is to predict the syntactic category of a word whose identity is known to the model/learner; the other one is to infer the syntactic category of a novel (previously-unseen) word. In the word category prediction task (Experiment 1) the Head of a frame is considered as a feature, whereas it is not included in the task of novel word categorization (Experiment 2). More details on each of these tasks is given in the following section.

Note that the resulting categories do not necessarily need to match the conventional adult-like categories put forth by linguists. Nonetheless, as a first-line evaluation, here we compare the categories learned by our model to a gold-standard set of syntactic categories. To measure test performance, we must compare the ‘true’ syntactic category of each test word (according to the gold-standard) to the label of its associated cluster. We thus need to label each cluster with a syntactic category. Words in the Manchester corpus are tagged with their parts of speech according to a fine-grained tag set. For

⁴<http://www.psych.rl.ac.uk/>

our evaluation, we use a coarse-grained version of this original tagging (also used by Parisien et al., 2008), including 11 tags, namely: Noun, Verb, Adjective, Adverb, Determiner, Negation, Infinitive, Auxiliary, Conjunction, Preposition, and Others. Each cluster is assigned the majority label among all its members. E.g., a cluster containing 30 nouns, 90 verbs, and 20 adjectives is labeled as Verb.

Test performance is measured using *Accuracy*: the proportion of test words assigned to their correct category. We also look into the accuracy for different groups of words, such as Verbs and Nouns, as well as open-class and closed-class words.

Model Parameters

Our model contains two sets of parameters: the weights ω_i used for measuring the similarity of a frame to a cluster (in Eqn 1), and a similarity threshold θ used for deciding whether to create a new cluster for a given frame. We set the weights ω_i uniformly, giving equal weights to all features. The value of θ affects the number of generated clusters: a low value increases the likelihood of grouping words, hence decreasing the total number of clusters. We set this parameter to different values for different experimental conditions (i.e., different combinations of features), so that we maintain the total number of clusters generated in each condition within a desired range.

We use two different ways of measuring $Sim_i(f, C)$ in Eqn 1 depending on feature i . For categorical features (Head, Cooc, Morph) we use the cosine of the vectors (widely used for similar clustering algorithms). A vector representing a categorical feature such as Head is of the size of word types in the corpus. E.g., for a sample frame f this vector includes 0 in all elements except where the value of Head in that frame is presented. For numerical features (Phon) we use the Euclidean distance.

Experimental Results

Experiment 1: Word Category Prediction

Recall that to determine the effect of different types of cues (Head, Cooc, Phon, Morph) in the acquisition of syntactic categories, we train our model in three conditions (i.e., using three combinations of features, namely Head+Cooc, Head+Cooc+Phon, and Head+Cooc+Morph). In Experiment 1, we measure the accuracy of category prediction over a test data containing 2000 known words. Comparing the accuracy of the categorization model across these conditions is fair and meaningful only if the number of clusters are relatively close for all conditions. Generally, allowing a larger number of clusters makes the categorization more conservative (i.e., by forming too many small clusters each containing one or a few word types that are highly similar). Based on our observation, this implicitly affects the test accuracy. Hence, in the training phase for each of the three above-mentioned conditions, we use different values for the similarity threshold θ to obtain approximately similar number of final clusters (i.e.,

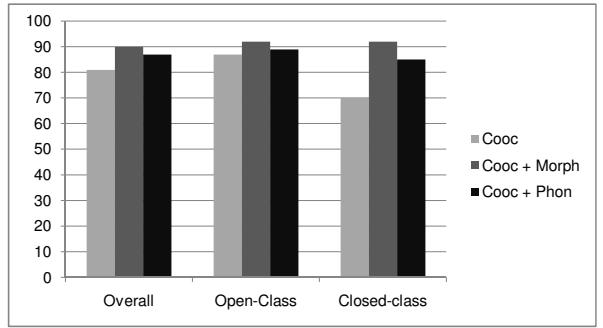


Figure 2: %Accuracy of known-word category prediction in three conditions; the total number of clusters constructed during training phase is in the range 258–288.

between 258–288).⁵ This way we maintain one factor (number of clusters) constant, allowing us to focus on the effect of different features involved in categorization.

Results are presented in Figure 2. In each condition, we measure accuracy on all 2000 words (displayed in the figure as the Overall accuracy), as well as for open-class and for closed-class words separately. Since Head is used as a feature in all conditions, for the ease of exposition, the figure refers to the conditions as Cooc, Cooc+Phon, and Cooc+Morph.

Figure 2 shows that the overall categorization accuracy of the model is improved by adding morphological or phonological information, reinforcing that word-internal features are indeed informative about a word’s syntactic category. The best performance is achieved by combining Cooc and Morph features, suggesting that our morphological feature might be more indicative of syntactic category than the phonological features.

Comparing the accuracy on open-class words and on Closed-class words, we can see that in two out of the three conditions (i.e., Cooc and Cooc+Phon), open-class words are better categorized in comparison to closed-class words. This is expected because it is more likely that the word co-occurrence information (which is the main source of information in all conditions) reveals the similarity among open-class (content) words more easily than for closed-class (function) words. As an example, we expect nouns to often appear after a small set of determiner types (e.g., *a*, *an*, *the*), whereas determiners may precede many different nouns, sharing fewer context features.

Previous studies have shown a strong effect for the Head feature in determining a word’s syntactic category (e.g., Chang et al., 2006). It is thus reasonable to compare the over-

⁵We have performed experiments with different ranges of cluster numbers, and found that the general patterns in results are similar. As noted before, we prefer fewer clusters (fewer than our vocabulary size) to allow for generalization. Indeed, we observe that even with 258–288 clusters, the generalization of the model is reasonably good. Since more than 55% of the clusters contain three or more word types.

all performance of our model in the three conditions with that of a simple category learner that uses only the Head feature, which we refer to as the *lex-stat* learner following Chang et al. (2006). For the performance of our model and that of the *lex-stat* learner to be comparable, we must set the similarity threshold so that we end up with around 500 clusters for all conditions (since the *lex-stat* learner constructs a separate cluster for each word type in the vocabulary). Indeed, we find that the overall performance of *lex-stat* (92%; not shown in Figure 2) is better than for Cooc (89%), and is comparable to the other two categorization conditions, Cooc+Phon (92%), and Cooc+Morph (92%). This raises an important question: whether the positive effect we observe here for the addition of Phon and Morph features is a true effect. In other words, since both Phon and Morph features are word-internal, it is possible that their inclusion in categorization increases the contribution of the Head feature in calculating similarity, implicitly giving more weight to the Head feature.

Note that the *lex-stat* learner is a very conservative model with no generalization abilities (since each word type is in its own cluster). Such a model thus fails to properly categorize novel (previously unseen) words. In contrast to such a learner, children have the ability to categorize novel words (even meaningless artificial words made up for experimental purposes), by the help of the context, or based on their morphological properties (Brown, 1957). We thus argue that for a categorization model to reveal the true effect of features such as morphology or phonology, it should be able to generalize well on unseen words. In the second Experiment, we use our three categorization models to determine the category of novel words. We consider actual novel words in this task because we want to draw on word-internal features, e.g., phonological and morphological properties of words.

Experiment 2: Novel Word Categorization

In this task, we use our model (in the three conditions) to categorize 2000 novel words. In such cases, the Head feature is not informative (since test words have not been seen during training), and hence the model has to utilize other sources of information to determine the category of a word. Results are presented in Figure 3. Comparing performance on this Experiment with those on Experiment 1 (Figure 2) shows a substantial decrease in the overall categorization accuracy (note that here Head feature is taken out of consideration). We especially observe a significant drop in performance for closed-class words. This decrease in performance emphasizes the importance of the Head feature for word categorization, particularly in determining the category of closed-class words. This is again an expected result, given our discussion presented in the previous subsection about the weakness of co-occurrence features in categorizing closed-class words.

Comparing results for the conditions shown in Figure 3 reveals that, as in Experiment 1, the use of Morph features does not improve the overall accuracy of categorization. These results are in contrast to the findings of Onnis and Christiansen (2008), who claim that featuring words solely based on their

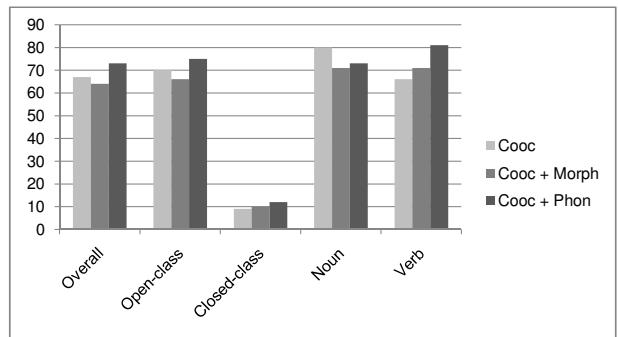


Figure 3: %Accuracy of novel word categorization in three conditions; the total number of clusters constructed during training phase is in the range 258–288.

(beginning and) ending phonemes results in good categorization. Their approach differs from ours in that they perform a batch processing over child-directed utterances, which allows their model to more easily learn the correspondences between a certain category, e.g., verbs, and endings shared by words from this category, such as *-ing* in *finishing*, *playing*, *reading*. Our model has to learn such correspondences incrementally, and hence is prone to making errors when calculating similarity between a word form such as “finishing” (a verb ending in the suffix *-ing*) and one such as “string” (a noun with a similar ending which is not a suffix but part of the word itself). Such errors in early stages may cause the algorithm to form incoherent clusters in later stages.

Figure 3 also includes the performance of our model (in all three conditions) separately shown for Nouns and Verbs. Although the use of Morph features does not help the overall categorization accuracy, it does seem to be particularly helpful in identifying Verbs. Interestingly, using Cooc features alone results in a better detection of novel nouns, whereas for verbs, other types of information (Morph and Phon) are helpful. Hence, even among open-class words, discovering different categories seems to rely on different types of information. This is supported by the observation that, typically, context words such as determiners mark the appearance of nouns; in contrast, verbs particularly share morphological and phonological properties. Related statistical analysis, such as that of (Monaghan et al., 2007; Clark, 2003) suggest such a complementary contribution of different cues; and moreover, some psychological studies implicitly take this into account when designing their experiments on children (Brown, 1957).

Conclusions

We have used an adaptation of a categorization algorithm proposed by Alishahi and Chrupała (2009) to model the acquisition of syntactic categories (e.g., verbs and nouns) in children, and to examine the effect of different types of cues on this task.

Our novel word categorization task provides a suitable

framework to evaluate the helpfulness of word-external (e.g., context) as well as word-internal features (e.g., morphological and phonological properties), independently from the identity of the word being categorized (head word). For example, our results indicate that categorizing closed-class words strongly relies on the head word. Specifically, these classes of words do not share intra-category similarities (neither contextual nor morpho/phonological similarities), and hence cannot be categorized well only by drawing on such properties. In contrast, open-class words can be successfully categorized based on a combination of word-internal and word-external properties, even without considering the head word.

In a more detailed investigation of the roles of word-external versus word-internal features, we find that verbs are better recognized when phonological and morphological properties are taken into account in addition to the context (co-occurring words). Note that we do not assume a full knowledge of morphology, but instead use word endings as an approximation to word suffixes (as suggested by Onnis & Christiansen, 2008). Interestingly, for nouns, considering only the information about the co-occurring words results in a more accurate categorization. This finding is in contrast to that of Onnis and Christiansen (2008). We argue this difference to be due to the incremental nature of our model.

Evaluating the effect of different cues in word categorization models needs much care. Studies such as those of Parisien et al. (2008) and Alishahi and Chrupala (2009) have reported the capability of co-occurrence information in categorizing words. They include, however, the head word itself as part of their features used for categorization. These studies evaluated the performance of their models on various tasks, such as noun/verb disambiguation, and semantic feature prediction. But they did not provide a comparison between their models and a categorization model that only uses the head word. As shown in our experiments, it is possible to achieve a high accuracy on a task by using such a simple conservative model. The task of novel word categorization that we propose is appropriate for evaluating the ability of a set of categories generated by a model to make generalizations.

In this study, we have shown that different types of cues, e.g., contextual or word-internal properties, provide children with complementary information, each helping with the categorization of a particular group of words. However, our framework is general and can be extended to incorporate other similar features (e.g., other morphological or phonological cues), as well as information about the semantic properties of words.

References

Alishahi, A., & Chrupala, G. (2009). Lexical category acquisition as an incremental process. In *Proceedings of the CogSci-2009 workshop on psychocomputational models of human language acquisition*. Amsterdam.

Brown, R. W. (1957). Linguistic determinism and the part of speech. , 55(1).

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

Chang, F., Lieven, E., & Tomasello, M. (2006). Using child utterances to evaluate syntax acquisition algorithms. In *In Proceedings of the 28th Annual Conference of the Cognitive Science Society*.

Clark, A. (2000). Inducing syntactic categories by context distribution clustering. In *Proceedings of the CoNLL-2000 and LLL-2000* (pp. 91–94).

Clark, A. (2003). Combining distributional and morphological information for part of speech induction. In *Proceedings of the 10th annual meeting of the European Association for Computational Linguistics* (pp. 59–66).

Gelman, S., & Taylor, M. (1984). How two-year-old children interpret proper and common names for unfamiliar objects. *Child Development*, 55(4), 1535–1540.

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, 32(2), 249–268.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed., Vol. 2: The Database). MahWah, NJ: Lawrence Erlbaum Associates.

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

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, 55(4), 259–305.

Onnis, L., & Christiansen, M. H. (2008). Lexical categories at the edge of the word. *Cognitive Science*, 32(1), 184–221.

Parisien, C., Fazly, A., & Stevenson, S. (2008). An incremental bayesian model for learning syntactic categories. In *Proceedings of the CoNLL-2008*.

Pearl, L. (2009). Using computational modeling in language acquisition research. *To appear in Experimental Methods in Language Acquisition Research*.

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

Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, 73(1), 1–33.

Schütze, H. (1993). Part of speech induction from scratch. In *Proceedings of the ACL-1993* (pp. 252–258).

Theakston, A. L., Lieven, E. V., 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, 127–152.

Wilson, M. (1988). *MRC Psycholinguistic Database: Machine-readable Dictionary, version 2.00*.