

Reappraising Lexical Specificity in Children’s Early Syntactic Combinations

Stewart M. McCauley (smm424@cornell.edu)

Morten H. Christiansen (christiansen@cornell.edu)

Department of Psychology, Cornell University, Ithaca, NY 14853 USA

Abstract

The flexibility and unbounded expressivity of our linguistic abilities is unparalleled in the biological world. Explaining how children acquire this fundamental aspect of human language is a key challenge for cognitive science. A recent corpus study by Yang (2013) has cast doubt on the lexical specificity of children’s productivity, as hypothesized by usage-based approaches. Focusing on determiner-noun combinations, he suggests that children possess an adult-like determiner category. In this paper, we show that Yang’s results may depend too heavily on an idealized notion of frequency distributions. We propose that these issues may be resolved by sidestepping sampling considerations and directly modeling children’s *actual* language processing. We therefore evaluate the abilities of two computational models to capture children’s productions of determiner-noun combinations. The first model implements a probabilistic context-free grammar, which acquires statistical information incrementally. A second model, the Chunk-based Learner (CBL), provides a simple instantiation of item-based learning. CBL outperforms the rule-based model, successfully producing the vast majority of the determiner-noun combinations in a dense corpus of child speech. The results thus suggest that the case against lexical specificity in children’s early determiner-noun sequences may be overstated.

Keywords: Language Learning; Grammatical Development; Computational Modeling; Usage-based Approaches; Sampling; Syntactic Categories; Lexical Specificity

Introduction

Much of the debate on language learnability has centered on the nature of children’s early productivity. Given the finite and noisy nature of the input, how are children able to generalize to a seemingly unbounded capacity for communicating novel information? The traditional answer invokes a system of words and rules, in which processing is memory-based at the word level, but algorithmic at the multiword level; compositional operations are performed over word classes corresponding to items from a mental lexicon (e.g., Chomsky, 1957; Pinker, 1999). Under this view, children are assumed to possess innate syntactic categories, such as *noun* and *determiner*. While various theoretical approaches differ with respect to the way in which innate word classes are mapped onto words themselves (e.g., Pinker, 1984), they converge on the idea that children’s early language use is—like adult language, under such a perspective—*class-based*. That is, children’s early comprehension and production abilities are governed by computations over their innate syntactic categories.

In recent decades, a number of theoretical alternatives have emerged from the field of cognitive linguistics, such as

construction grammar (e.g., Croft, 2001; Goldberg, 2006). Such approaches dispense with the words and rules framework entirely, holding instead that grammatical processing—and, by extension, children’s grammatical development—is primarily memory-based, driven by stored linguistic units of varying granularity and complexity. That is, knowledge of grammar is inseparable from lexical knowledge; the two can only be distinguished insofar as they constitute polar ends of a spectrum of unit complexity ranging from the level of simple symbols (such as morphemes and simple words) to complex symbols (such as grammatical constructions). While such approaches to explaining linguistic productivity allow syntactic categories to be learned, they do not converge on a single notion of the nature of such categorical knowledge (e.g., Croft, 2001), nor do they seek to directly explain the development of abstract categories themselves. They have, however, inspired developmental research in what has become known as the *usage-based* framework. A number of researchers in the usage-based tradition have identified item-based patterns in children’s early language use, such as verb-island phenomena (e.g., Tomasello, 1992).

With respect to the development of abstract syntactic categories, a number of usage-based corpus studies have focused on the English determiner category as a test case, inspired by early proposals that the categorical knowledge driving children’s early speech is quite limited (e.g., Braine, 1976). In response to work arguing for an early abstract determiner category (e.g., Valian, 1986), Pine and Martindale (1996) analyzed seven corpora of child and child-directed speech. Controlling for the number of multiword utterances in each sample, as well as vocabulary range, the authors found that children in the age range of 1;1 to 2;4 exhibited far less overlap in their determiner use than did their caretakers. Pine and Lieven (1997) extended this general finding to a group of 11 child corpora.

Researchers have subsequently criticized the Pine and Martindale (1996) and Pine and Lieven (1997) studies for the sparseness of the data used as well as the inclusion of nouns that children produced with a determiner only once, making it impossible for there to be any overlap (e.g., Valian, Solt, & Stewart, 2009). More recently, Yang (2013) expanded on this criticism by noting that linguistic frequency distributions conform to a Zipfian pattern (Zipf, 1949), in which the frequency of a word is inversely proportional to its rank in a frequency table. Yang argued that such Zipfian patterns have the consequence that, even for adult speech, most nouns appear so infrequently in a corpus that they are unlikely to occur with more than one type of determiner. Yang used calculations based on Zipf’s

law to evaluate a memory-based language model trained on 1.1 million utterances drawn from the CHILDES database (MacWhinney, 2000). Across 1,000 simulations, the model randomly selected determiner-noun sequences from the training data. The amount of determiner overlap in these randomly sampled pairs was then compared to the amount of overlap exhibited by selected target children's determiner-noun productions. Yang's finding was that the memory-based random selection model significantly underpredicts the amount of overlap in children's actual determiner-noun productions, while a class-based calculation using Zipf's law more accurately captures children's productivity. Based on this finding, Yang concluded that previous findings of lexical specificity in children's determiner use are sampling artifacts.

Arguing against the idea that children's item-based patterns in determiner usage are simply artifacts of Zipfian distributions, Pine, Freudenthal, Krajewski, and Gobet (2013) presented a series of corpus analyses. In the first analysis, they compared the overlap of determiners used with nouns appearing in the speech of both children and their caretakers against the overlap of nouns used only by caretakers. They showed that the results of the comparison are sensitive to sample size, and that when this variable is controlled for, caretakers showed more overlap with nouns appearing in child speech than nouns that did not. This led Pine et al. to control for vocabulary range in their second analysis, demonstrating that once size and vocabulary are both controlled for, there were significant differences between children and their caretakers in terms of flexible determiner usage. A third analysis demonstrated increasingly flexible usage of determiners with a fixed set of nouns across two developmental stages.

Pine et al. additionally show that the top 10 nouns in the corpora do not conform to Zipf's law. While this result is informative, the sample is too small to allow any decisive conclusions to be made. In the present paper, we therefore conduct an exhaustive test of Yang's (2013) assumption that nouns in child-directed speech conform to a Zipfian distribution, evaluating the consequences of this analysis for Yang's case against item-based patterns. We then propose an alternative approach that is less susceptible to sampling issues; while corpus analyses have provided great insight into the nature of children's early productivity, it remains for computational studies to explore the psychological mechanisms involved in acquiring adult-like determiner use. As an initial step, we present a computational study that instantiates the principles of item- and class-based learning in two distinct, simple models of language learning and use. The models are evaluated with respect to their ability to capture children's actual determiner-noun combinations through generalization to unseen input.

Experiment 1: Analyzing the Distribution of Nouns in Child-directed Speech

Yang's (2013) claim that previous findings of lexical specificity in children's determiner use are merely sampling

artifacts depends heavily on the assumption that nouns in child-directed speech follow Zipf's law (a power law function). Pine et al. (2013) question this assumption, demonstrating that the frequencies of the top 10 nouns in the corpora used by Yang are different than what would be expected based on Zipf's law. While this result is informative, much larger samples are necessary in order to establish definitively whether frequencies conform to a given distribution. In what follows, we describe statistical tests performed on the frequencies of the *entire* set of nouns in several corpora of child and child-directed speech. Our results suggest that the nouns in each corpus are highly unlikely to be drawn from a power law distribution, and thus do not follow Zipf's law as Yang's analyses assume.

Corpus Selection and Preparation

Previous computational studies on the acquisition of syntactic categories have focused on a variety of publicly available corpora of child-directed speech from the CHILDES database (MacWhinney, 2000). However, these studies have been subject to problems of data sparseness, as they have primarily relied on multiple small corpora that typically account for only 1-2% of the input to and speech of a given child (cf. Maslen, Theakston, Lieven, & Tomasello, 2004). Here, we focus primarily on the linguistic information available to a single child. This is achieved by using a dense corpus of child and child-directed speech, which covers over 10% of the speech of and directed to the target child (the Thomas/Brian corpus; Maslen et al., 2004). This provides an advantage over previous studies that have relied on comparisons across several small sets of data. Nonetheless, for purpose of comparison, we also include the six smaller corpora of child-directed speech analyzed by Yang (2013): the Adam, Eve, Naomi, Nina, Peter, and Sarah corpora from the CHILDES database.

Tags and codes were removed from each corpus, leaving only the speaker identifier and the original sequence of words. Nouns and determiner-noun sequences were then identified and extracted using TreeTagger (Schmid, 1994).

Methods

To evaluate the hypothesis that the noun frequency data from the corpora follow a power law distribution, we use the Kolmogorov-Smirnov (KS) goodness-of-fit test (Press, Teukolsky, Vetterling, & Flannery, 1992), with corresponding *p*-values for the power law fit calculated according to the method described by Clauset, Shalizi, and Newman (2007). The KS test evaluates the null hypothesis that a sample is drawn from a given distribution (in this case, a power law). We also compare the power law fit to alternative fits of lognormal and exponential distributions—both appropriate candidates for frequency data with a long tail—using likelihood ratio testing (cf. Clauset et al., 2007).

Results and Discussion

The results of the KS test for the distribution of nouns across the entire dense corpus strongly suggest that the noun

frequencies do not conform to a power law distribution ($D = 0.19, p < 0.001$)¹. The same pattern followed for the six smaller corpora originally used by Yang (2013) (with D statistics ranging from 0.11 to 0.21, all p 's < 0.001).

Comparison to alternative distributions using likelihood ratio testing confirmed that while the power law distribution provided a better fit to the dense corpus noun data than the exponential distribution ($R = 25.34, p < 0.001$), the lognormal distribution was a far better fit than either the power law distribution ($R = 17.8, p < 0.001$) or the exponential distribution ($R = 28.75, p < 0.001$). A complementary cumulative distribution function (CCDF) plot comparing the noun data from the dense corpus to the three distributions is shown in Figure 1. The data from the six smaller corpora followed the same pattern in each case (all p 's < 0.001).

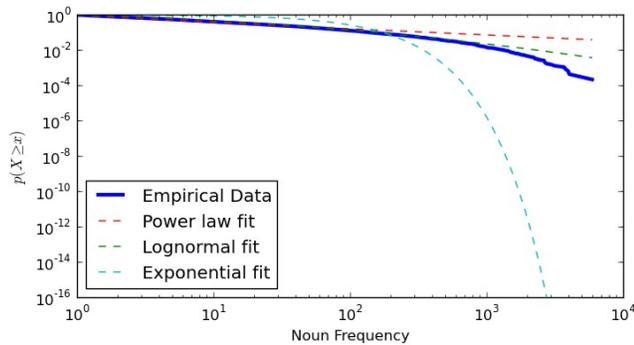


Figure 1: CCDF plot depicting the distribution of nouns in the dense corpus fit to power law, lognormal, and exponential distributions.

Our results strongly suggest that the distribution of nouns in the selected corpora do not follow Zipf’s law. As our analysis covers not only the largest currently available corpus of English child-directed speech, but all of the corpora used by Yang (2013), it suggests that Yang’s calculations based on Zipf’s law depend on a highly idealized notion of the distribution of noun frequencies, and mischaracterize the degree of determiner-noun overlap that would result from following the actual distributions of nouns in corpora of child-directed speech.

Rather than attempting to control for sampling considerations (as in Pine et al., 2013), we propose an alternative approach that more directly evaluates the nature of children’s early syntactic combinations. Specifically, we suggest that to resolve these issues we need to move beyond corpus analyses to the explicit modeling of the mechanisms children are hypothesized to use in acquisition, and test how well they account for children’s actual linguistic behavior. In what follows, we take an initial step towards modeling children’s actual comprehension and production processes, focusing on determiner-noun combinations.

¹ A separate test performed on only those nouns produced by the child in a determiner-noun combination met with similar results ($D = 0.15, p < 0.001$).

Experiment 2: Modeling Children’s Production of Determiner-Noun Sequences

As an initial step toward modeling children’s actual comprehension and production processes during learning, we evaluate the ability of a simple, developmentally motivated model of item-based learning processes, the Chunk-based Learner (based on McCauley & Christiansen, 2011), to account for children’s determiner-noun combinations. This ability is compared to that of a class-based model with built-in grammatical categories, based on a standard probabilistic context-free grammar model (PCFG; cf. Manning & Schütze, 1999). Unlike most computational approaches to acquisition, both models are designed to capture the *incremental* nature of the task facing the learner: each is trained and evaluated in an incremental rather than batch fashion, and is *only able to draw upon what has been learned from previously encountered input*. After describing the models, we compare their ability to capture the determiner-noun combinations of the target children in corpora of child-directed speech, using the same seven corpora described above (the dense corpus and the six corpora used by Yang, 2013). Unlike the approach described by Yang, both models are evaluated on their ability to generalize to previously unseen input.

Modeling Children’s Determiner Productivity Using Grammatical Categories

The class-based model involves a developmentally motivated modification to the standard PCFG language model; statistical information tied to each rewrite rule is acquired incrementally, during a single pass through a corpus. This allows the language model to maintain the generative capacity of the traditional PCFG through pre-established word classes and rewrite rules while also simulating a gradual buildup of lexical information, as is necessary even under nativist accounts of language acquisition (e.g., Pinker, 1999). For the current simulation, we focus on a single fragment of the PCFG, corresponding to two syntactic categories and a single rewrite rule:

NP \rightarrow DET + N
DET: {the, a, an}
N: {set of nouns encountered thus far in the corpus}

Thus, we focus only on simple noun phrases involving definite or indefinite nouns (as in Yang, 2013). The simulation involves two simultaneous tasks: 1) comprehension, in which distributional information tied to determiners is acquired, and 2) production, in which noun phrases are produced stochastically according to the information gleaned during comprehension up to the given point during the simulation at which a production attempt is made. Lexical knowledge in the model contains only two categories. The determiner category is pre-established, as depicted above, while the noun category is gradually built up on the basis of the input.

Each time an adult utterance is encountered, the model engages in the comprehension task. During comprehension, frequency information tied to each word type is incremented. This allows the probability of a given word type to be calculated as the number of tokens of that type normalized by the total number of tokens encountered in a given category. Each time a child utterance of the form DET+N is encountered, the model engages in the production task. In this task, the PCFG is used in an active rather than passive fashion; given the target noun, the model stochastically produces a new DET+N sequence by selecting one of the available determiners probabilistically, according to the probability of each terminal (which is updated incrementally during learning). The determiner is then concatenated with the noun from the utterance, thus directly implementing Yang's (2013: p. 6324) assertion that "very young children's language is consistent with a grammar that independently combines linguistic units (...)."

The model was scored according to the number of correctly produced determiner-noun combinations. The total number of correctly produced noun phrases was normalized by the total number of attempted DET+N productions, yielding a production accuracy score (percentage). As the model is stochastic, 100 separate iterations were performed on each input corpus. The mean score across all 100 simulations was then taken as the final score.

Chunk-based Learner (CBL)

Language learning in CBL involves improving the model's ability to perform two tasks: "comprehension" of child-directed speech, through the statistical discovery and use of chunks as building blocks, and "production," which utilizes the same chunks and statistics involved in comprehension. Comprehension is approximated in terms of the model's ability to segment a corpus into phrasal units, and production is approximated in terms of the model's ability to reconstruct utterances produced by the child. While comprehension and production in the model are two sides of the same coin, we describe them separately for simplicity.

Comprehension Although the model's comprehension performance is not directly assessed in the current study, it drives the model's ability to create utterances during production, including the determiner-noun combinations that are the focus of this paper. Comprehension begins with the tracking of simple distributional statistics: As the model processes utterances word-by-word, it tracks frequency information for words and word-pairs, which is used on-line to track the backward transition probability (BTP) between words and maintain a running average BTP for previously encountered pairs. When the model calculates a BTP that is greater than expected, based on the running average, it groups the word together with the previous word(s). When the calculated BTP falls below the running average, a boundary is placed and the chunk thereby created (consisting of one or more words to the left of the inserted boundary) is added to the *chunkatory*, the model's inventory of single- and multi-word units. Importantly, the model

maintains frequency information for each chunk in the chunkatory. The model also uses the chunkatory to make on-line predictions for which words will form a chunk, based on previously learned chunks. Each time a word-pair is encountered, it is checked against the chunkatory; if it has occurred before as a complete chunk or as part of a larger chunk, the words are grouped together and the model moves on to the next word. If the word-pair is not found in the chunkatory, the BTP is compared to the running average, with the same consequences as before. Because there are no a priori limits on the number or size of the multi-word building blocks that can be learned, the resulting chunkatory will contain a mix of words and multi-word chunks.

For example, consider the following scenario in which the model encounters the phrase *the blue ball* for the first time and its chunkatory includes *the blue car* and *blue ball*. When processing *the* and *blue*, the model will not place a boundary between these two words because the word-pair is already represented in the chunkatory (as in *the blue car*). Instead, it predicts that this bigram will form part of a chunk. Next, when processing *blue* and *ball*, the model reacts similarly, as this bigram is also represented in the chunkatory. The model thereby combines its knowledge of two chunks to discover a new, third building block, *the blue ball*, which is added to the chunkatory, and the model then goes on to process the next word in the utterance.

Thus, the model gradually creates an inventory of building blocks and uses these to segment the corpus into phrasal units—akin to shallow parsing—favoring sequential information. This shallow processing approach was adopted because it is consistent with evidence on the relatively underspecified nature of human sentence comprehension (e.g., Frank & Bod, 2011; Sanford & Sturt, 2002) and provides a mechanistic approximation of the item-based way in which children are hypothesized to process sentences by usage-based theories (cf. Tomasello, 2003). CBL's ability to do phrasal segmentation compares well with off-the-shelf shallow parsers in English, German and French (see McCauley & Christiansen, 2011, for details).

Determiner-Noun Production Each time the model encounters a multi-word child utterance featuring a determiner-noun combination, it is required to produce its own determiner-noun combination using the corresponding noun. The chunkatory is searched for chunks featuring the target noun with *a*, *the*, or *an*, and the chunk with the highest frequency count is output. This provides a lexically-specific analogue to the PCFG production task, and thus scoring is identical: the determiner-noun sequence must match the child's.

In summary, as the model is exposed to a corpus, one word at a time, it 1) builds a chunkatory—an inventory of single- and multi-word building blocks—and uses these to segment and learn from the incoming input, and 2) uses the same chunks to attempt to reproduce the child's determiner-noun sequences as it comes across them in the corpus.

Sentence Production As an initial step towards capturing the semantic dimension of children's determiner-noun

combinations within a more comprehensive item-based model of production, we report an additional set of simulations involving the full version of CBL (cf. McCauley & Christiansen, 2011). This version differs from that described above only with respect to the production task: each time a multi-word child utterance is encountered, the model attempts to reproduce the entire utterance using only building blocks discovered in the previously encountered input. Following Chang, Lieven and Tomasello (2008), we assume that the overall message, which the child wants to convey, can be approximated by treating the utterance as a randomly-ordered set of words: a “bag-of-words.” The task for the model, then, is to output these words in the correct order (as originally produced by the child). Following usage-based approaches, the model utilizes building blocks from its chunkatory to reconstruct the child’s utterances. In order to model retrieval of stored chunks during production, word combinations from the utterance that are represented as multi-word chunks in the chunkatory will be placed in the bag-of-words instead of the individual words that make up those chunks. E.g., consider a scenario in which the model encounters the child utterance *the dog chased a cat* and has both *the dog* and *a cat* as chunks in its chunkatory. These two chunks would then be placed in the bag along with *chased*, and the order of these three chunks is randomized. The model then has to reproduce the child’s utterance using the unordered chunks in the bag. We model this as an incremental, chunk-to-chunk process rather than one of whole-sentence optimization. Thus, the model begins by removing from the bag the chunk with the highest BTP given the # tag (which marks the beginning of each utterance in the corpus), and outputs it as the start of its new utterance. Next, the remaining chunk with the highest BTP given the most recently produced chunk is removed from the bag and output as the next part of the utterance. In this manner, the model uses chunk-to-chunk BTPs to incrementally produce the utterance, outputting chunks one-by-one until the bag is empty. Using this method, CBL is able to produce the majority of utterances produced by children in 24 different Old World languages (McCauley & Christiansen, 2011). For the present study, the overall percentage of correctly produced determiner-noun sequences (i.e., those that are identical to the determiner-noun sequences in the target child’s original utterance) is evaluated.

Models Summary

To summarize, we test the following models: 1) an incrementally trained PCFG (with built-in classes and rewrite rules) which stochastically selects a determiner for each target noun, and thus provides a straightforward implementation of Yang’s (2013) claim that children combine linguistic units independently, 2) the CBL model, which produces determiner-noun combinations based on the frequencies of lexically-specific chunks learned and stored in its chunk inventory, and 3) a more comprehensive version of CBL which roughly approximates the overall message

the child wishes to convey (using a bag-of-words approach) to incrementally produce entire utterances based on chunks and transition probabilities learned previously (the determiner-noun sequences in the utterance are then compared to those of the child).

Results and Discussion

In all cases (see Figure 2), CBL outperformed the PCFG by a wide margin. For the dense corpus, CBL successfully produced 70% of the child’s determiner-noun sequences (94.3% on the sentence production task), while the PCFG achieved a performance score of just 49.2%. Across the six smaller corpora used by Yang (2013), CBL attained a mean score of 69.2% (87.3% on the alternative production task), while the PCFG achieved a mean score of 51.6%².

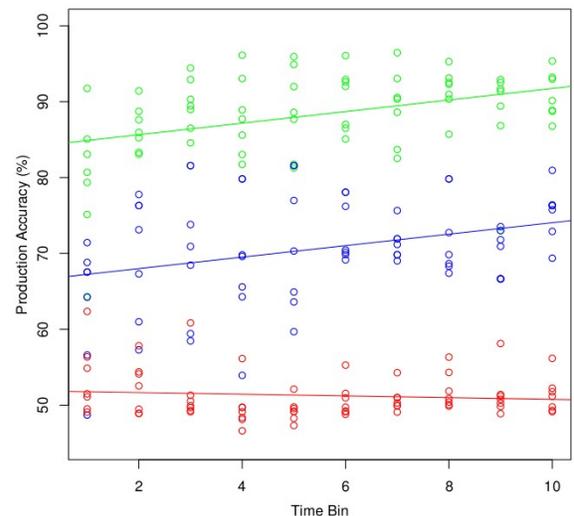


Figure 2: Production accuracy as a function of time: CBL during full sentence production task (green), CBL during determiner-noun task (blue), PCFG (red). Trend lines derived from linear regression.

The incremental nature of both models allows us to further compare the development of model performance over time. After dividing each individual simulation into ten bins of equal size (with the first bin representing the first tenth of the model’s pass through the corpus, the second bin representing the second tenth, and so forth), we examined the trajectory of model performance using bin as a temporal dimension to predict production scores. This yielded a small but reliable correlation between performance and time bin for CBL across all simulations ($R^2 = 0.1$, $F_{1,68} = 7.46$, $p < 0.01$), with a mean score of 63.6 during the first phase and 75.4 during the last. The correlation was also present for the sentence production task ($R^2 = 0.15$, $F_{1,68} = 11.76$, $p < 0.01$), with a mean score of 79.9 during the first phase and 90.9 during the last. However, the PCFG did not exhibit a significant difference in performance as a function of time

² To counter the potential objection of a lack of phonological constraints, we re-ran the PCFG simulations treating both *a* and *an* as a single indefinite article. The mean production accuracy across all 7 corpora improved by less than one percentage point.

($R^2 = 0.01$, $F_{1,68} = 0.76$, $p = 0.38$), with a mean score of 53.5 during the first phase and 51.4 during the last.

Thus, the item-based model improved with exposure to more input, which is further underscored by the slightly better overall performance for the dense corpus than the smaller corpora. This trend is significant despite the target children's increasingly flexible use of determiners over time (cf. Pine et al., 2013), suggesting that purely item-based processes continue to play a role even as children's grammatical categories appear to grow more abstract. This idea resonates with recent psycholinguistic evidence for item-based processing in adults (e.g., Arnon & Snider, 2010), and is consistent with usage-based theory more generally.

General Discussion

The aims of the present study were twofold: firstly, to evaluate the claims of Yang (2013) that item-based patterns in children's determiner-noun combinations are merely artifacts of sampling from a Zipfian distribution, and secondly, to offer a further computational approach to studying children's early productivity—complementary to previous corpus analyses—based on modeling the mechanisms involved in children's incremental language learning and use.

Our statistical tests of the distributions of nouns from each of the corpora used by Yang, in addition to the currently largest available corpus of English child/child-directed speech, strongly suggest that the nouns in each case do not conform to Zipf's law. Consequently, Yang's calculations, which are based on the assumption of a Zipfian distribution, likely underestimate the degree of determiner-noun overlap that would be expected based on the distribution of nouns alone. This would mean that lexically specific patterns found in previous corpus analyses (e.g., Pine & Lieven, 1997; Pine & Martindale, 1996) are more than mere sampling artifacts.

In this context, we argue that corpus analyses should be complemented by an approach that sidesteps sampling considerations, focusing instead on modeling the mechanisms involved in language acquisition according to the particular theoretical approaches being evaluated. Our simulations provide an initial step in this direction: we report a simple, developmentally motivated model of item-based language learning and use which successfully captures a large proportion of the *actual* determiner-noun combinations made by the target child of a dense corpus. That this simple approach dramatically outperforms a class-based model in which determiners and nouns are combined independently (a notion key to Yang's approach) lends support to usage-based approaches to children's early syntactic combinations. The finding that the production attempts of the item-based model improved over the course of the simulations, despite the increasingly flexible use of determiners by the target child as a function of age (Pine et al., 2013), resonates with the idea that item-based processing continues to play a role throughout development,

even as grammars grow more abstract, consistent with theoretical proposals emerging from cognitive linguistics (e.g., Croft, 2001; Goldberg, 2006).

Acknowledgments

Thanks to Julia Ying for helpful comments. This study was partially supported by BSF grant 2011107, awarded to MHC.

References

- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62, 67-82.
- Braine, M. D. (1976). Children's first word combinations. *Monographs of the Society for Research in Child Development*, 41, 1-104.
- Chang, F., Lieven, E., & Tomasello, M. (2008). Automatic evaluation of syntactic learners in typologically-different languages. *Cognitive Systems Research*, 9, 198-213.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. New York: Oxford University Press.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk, Volume II: The Database*. Lawrence Erlbaum.
- Manning, C. D. & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge: MIT press.
- Maslen, R. J., Theakston, A. L., Lieven, E. V. ., & Tomasello, M. (2004). A dense corpus study of past tense and plural overregularization in English. *Journal of Speech, Language, and Hearing Research*, 47, 1319-1333.
- McCauley, S.M. & Christiansen, M.H. (2011). Learning simple statistics for language comprehension and production: The CAPPUCINO model. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1619-1624). Austin, TX: Cognitive Science Society.
- Pine, J. M., Freudenthal, D., Krajewski, G., & Gobet, F. (2013). Do young children have adult-like syntactic categories? Zipf's law and the case of the determiner. *Cognition*, 127, 345-360.
- Pine, J. M., & Lieven, E. (1997). Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics*, 18, 123-138.
- Pine, J. M., & Martindale, H. (1996). Syntactic categories in the speech of young children: The case of the determiner. *Journal of Child Language*, 23, 369-395.
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: Harper Collins.
- Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge University Press.
- Valian, V. (1986). Syntactic categories in the speech of young children. *Developmental Psychology*, 22, 562-579.
- Valian, V., Solt, S., & Stewart, J. (2009). Abstract categories or limited- scope formulae? The case of children's determiners. *Journal of Child Language*, 36, 743-778.
- Yang, C. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences*, 110, 6324-6327.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Reading, MA: Addison-Wesley.