2014-07-01

A Model of Children's Acquisition of Grammatical Word Categories Using an Adaptation and Selection Algorithm

Teresa Young
Brigham Young University - Provo

A Model of Children's Acquisition of Grammatical Word Categories Using an Adaptation and Selection Algorithm

Teresa Young

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Ron W. Channell, Chair
Kristine Tanner
Shawn L. Nissen

Department of Communication Disorders

Brigham Young University

June 2014

ABSTRACT

A Model of Children's Acquisition of Grammatical Word Categories Using an Adaptation and Selection Algorithm

Teresa Young
Department of Communication Disorders, BYU
Master of Science

Most children who follow a typical developmental timeline learn the grammatical categories of words in their native language by the time they enter school. Researchers have worked to provide a number of explicit, testable models or algorithms in an attempt to model this language development. These models or algorithms have met with some varying success in terms of determining grammatical word categories from the transcripts of adult input to children. A new model of grammatical category acquisition involving an application of evolutionary computing algorithms may provide further understanding in this area. This model implements aspects of evolutionary biology, such as variation, adaptive change, self-regulation, and inheritance. The current thesis applies this model to six English language corpora. The model created dictionaries based on the words in each corpus and matched the words with their grammatical tags. The dictionaries evolved over 5,000 generations. Four different mutation rates were used in creating offspring dictionaries. The accuracy achieved by the model in correctly matching words with tags reached 90%. Considering this success, further research involving an evolutionary model appears warranted.

Keywords: grammatical word categories, evolutionary programming, language acquisition

ACKNOWLEDGEMENTS

I would like to express my most sincere thanks to my thesis chair, Dr. Ron Channell, for his constant dedication, patient guidance, and encouragement, which turned a seemingly impossible task into an experience of considerable learning and accomplishment. His genuine desire and willingness to assist in each step of this thesis process contributed greatly to my education in countless ways. I would also like to thank the members of my committee, Dr. Nissen and Dr. Tanner, for their helpful input and contributions along the way. My family and friends are also deserving of my earnest appreciation and thanks for their support and prayers on my behalf during this often difficult, but greatly rewarding experience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DESCRIPTION OF CONTENT

This thesis is part of a larger research project, and portions of this thesis may be published as part of articles listing the thesis author as a co-author. The body of this thesis is written as a manuscript suitable for submission to a peer-reviewed journal in speech-language pathology. An annotated bibliography is presented in the Appendix.

Introduction

From a very young age, humans begin to acquire the understanding and use of language. A desire to better understand the processes and mechanisms of language acquisition has prompted the use of explicit, testable algorithms (Pinker, 1988), some of which have been implemented as computer models. These algorithms and computer models typically use transcribed samples of language input to children, apply various procedures to extract patterns from this input, and interpret the resulting changes in the model and its output as representing changes in some aspect of linguistic knowledge. As this methodology has been applied to the problem of how children learn the grammatical word categories of their language, these algorithms and models have met with promising but variable success. The present project describes an alternate algorithm, derived from the principles of adaptation and selection typically associated with the study of evolutionary biology. This algorithm serves as the basis for a testable computer model of grammatical category acquisition.

A number of published studies have investigated hypotheses as to how children learn grammatical categories (also known as the *parts of speech* or syntactic categories). Researchers agree that children are not born with a robust vocabulary and functioning system for syntax (word ordering), but the question of exactly how language is acquired is vigorously debated. Over the past two decades, two major theoretical perspectives have become dominant: the nativist (also know as generativist or Universal Grammar) perspective and the constructivist (also called emergentist, socio-pragmatic, functionalist, or usage-based) perspective (Ambridge & Lieven, 2011). The nativist perspective holds that newborn children already possess an innate knowledge of syntactic categories and are only required to label each new word with an already existing category. In contrast, constructivism holds that language is acquired through input and that children are not born with inherent grammatical categories, needing to somehow extract these categories from the language they hear. Obviously, the theoretical approach that a researcher takes shapes the method and design of any studies and the conclusions that may be drawn.

Models of Grammatical Category Acquisition

In recent years, computational models have been explored as a way to study grammatical category acquisition. Most of the following studies used a distributional approach to language acquisition, wherein the occurrence of a word relative to other words (e.g., following, between, preceding these other words) is noted and inferences drawn. It is important to note that in all of these studies, words were only assigned their most frequent or likely grammatical category.

Cartwright and Brent (1997) used the concept of sentential minimal pairs to describe distributional category acquisition. An example of two sentences that constitute a minimal pair would be *I saw the cat* and *I saw the dog*. *Cat* and *dog* are the only two words that differ in these sentences, and these words are found in the same sentence position. However, most sentences cannot be matched up so succinctly. Generalized forms of minimal pairs must be created by forming templates. The authors hypothesized that when children hear a sentence, they create a new group for each word in the sentence. A new template is formed consisting of the new group sequence found in the sentence. This template is added to a list of templates already formed. A preference list is used in the process of merging similar groups from different (older and new) templates, until the best arrangement is found. A number of experiments carried out by the authors were designed to test their theory in regards to an artificial language, real child-directed speech, and the connection between semantic and syntactic categories. The study's obtained level of accuracy implied that this approach could be useful, though no follow-up studies have been published.

Redington, Chater, and Finch (1998) focused on a distributional approach involving the hierarchical clustering of words into syntactic categories, in which words in the same syntactic category have distributional regularities in common. Redington et al. proposed three stages for using distributional information to determine syntactic category. These stages include analysis of the contexts within which the target word occurs, comparing the contexts or distributions of pairs of words, and grouping words with similar distributions. The computational model forms dendrograms, which visually represent a hierarchical clustering of syntactic categories, with those clusters that are most similar being placed close together. Samples from the Child Language Data Exchange database (CHILDES; MacWhinney, 2000) were used in a series of experiments. Accuracy and completeness were measured based on comparing the benchmark syntactic categories with the derived groupings. Experiments were carried out to measure the effectiveness of distributional information involving position of context words in relation to target word, numbers of context and target words, effectiveness based on word class, corpus size, utterance boundaries, frequency versus occurrence information, removing function words, prior knowledge of other categories, and child-directed versus adult-adult speech. These experiments indicated that analysis of distributional information may contribute to early acquisition of syntactic categories.

Mintz (2003) studied the role of frequent frames in grammatical category acquisition. This distributional approach uses the two words framing (surrounding) a word to determine the word’s grammatical category. Words that are surrounded by the same frequent frame (e.g., the intervening words) are placed in the same grammatical category. For example: The blank in the following phrase is likely to be a verb based on the frame surrounding the blank: *to* ___ *to*. Mintz's study aimed to determine (a) the type of distributional information most informative, (b) the types of distributional cues picked up on and used by children and infants, and (c) how to assemble categories into a system. Six corpora from the CHILDES database (MacWhinney, 2000) were used as input and only the adult utterances were used. The frequent frames for each corpus were analyzed and intervening words were grouped according to frame-based categories. Results showed high accuracy of grammatical category identification. Frequent frames produced categorization accuracy that was equal or more accurate than the distributional methods that were analyzed in previous studies.

Weisleder and Waxman (2010) expanded upon the notion of frequent frames by examining the distributional information available to children learning Spanish. Weisleder and Waxman also expanded on the Mintz (2003) notion of frequent frames by including sequences in which one of the frames is the end of the utterance (referred to as end-frames). This was done due to the prosodic nature of the ends of utterances, which may be more readily acquired by young children. Weisleder and Waxman used six corpora from the CHILDES database (MacWhinney, 2000), three Spanish, three English, and only the adult utterances were analyzed. Accuracy scores were higher for the English samples than the Spanish, and although such things as homophony and noun drop in the Spanish language may create obstacles to the use of distributional evidence, the use of frequent frames as an organizing strategy received further support.

Freudenthal, Pine, and Gobet (2005) refuted the proposed accuracy of the distributional methods supported by co-occurrence statistics in determining syntactic category. Redington et al. (1998) and Mintz (2003), among others, had supported the notion that the observed accuracy of distributional patterns allows insight into models of child grammatical category acquisition. However, Freudenthal et al. suggested that without using these derived syntactic categories to generate language output, accuracy cannot be fully determined. To accomplish this, the authors used the Mosaic software, a computational model designed to simulate syntactic acquisition. Input was fed through the model and co-occurrence statistics were used to generate phrase substitutions that fit the context of the original input. The authors used a new version of Mosaic that classified frequently recurring phrases into chunks, which are treated as one unit, and can consequently only be involved in substitutions as a whole unit, rather than by single word. The authors felt that chunking increased the accuracy of substitutions, decreasing the level of error. Results of the study suggested that measures typically used to determine accuracy of syntactic categories based on co-occurrence statistics might be inadequate. However, the sentences that their model generated were very short (averaging only 3.5 morphemes in length).

The studies described above presented algorithms or computer models that have attempted to describe the process of grammatical category acquisition. Each approach either derived a number of grammatical word categories from the transcribed language input from adults to a child and then compared the derived categories to linguist-coded classifications or – as in the study by Freudenthal et al. (2005) – had humans judge the adequacy of generated sentences. In most of these studies, words were classified as to their single, most likely grammatical word category.

An Algorithm Using Adaptation and Selection

Although these past studies have provided insight into the topic of grammatical category acquisition, these studies were largely based on distributional methods. The present study approaches the acquisition of grammatical word categories from a very different direction by using an adaptation and selection algorithm.

An adaptation and selection algorithm is a contribution of evolutionary computing. Evolutionary computing involves applying the principles of evolutionary biology (e.g., variation, adaptive change, self-regulation, and inheritance) to computational models and is comprised of the sub-branches of genetic algorithms, evolution strategies, and evolutionary programming (Fogel, 2006). According to Fogel, the branch called evolutionary programming starts with a population of solutions that randomly mutates to create offspring from the parent solutions, of which the most-fit become the parents of the next generation. The system never receives feedback as to what aspects of the chosen solutions are correct, and thus the correct parts of the solution are as likely to mutate in the next generation as are the incorrect parts. However, over many selections, reproductions, and mutations, offspring evolve that represent an adequate solution to the problem to be solved.

Evolutionary algorithms have been applied to areas other than suggesting an explanation of human evolution. Evolutionary algorithms have been used to develop solutions to seemingly trivial programs, like computers that learn how to play games such as checkers, to more important solutions, such as systems that assist in the interpretation of mammograms (Fogel, 2002; Fogel, Watson, Boughton, & Porto, 1998). An interesting alternate application of evolutionary algorithms was made by Siegler (1996), whose research on the development of human cognitive strategies was inspired by evolutionary algorithms. Siegler asserted that evolutionary biology and cognitive development shared analogous fundamental concerns, one of which was a search for ultimate origins.

The present study applies an adaptation and selection algorithm to the problem of the acquisition of the grammatical categories in language. Specifically, changes in type accuracy (different words) and token accuracy (total words) associated with different mutation rates will be observed across evolutionary cycles. Given the varied and often contradictory language input that children receive, the application of an adaptation-selection algorithm might reflect the sometimes random attempts and false starts made by a child during early language production.

Method

Other researchers collected the longitudinal language sample corpora that were used in the present research study. In the present study, the task was to correctly format and grammatically tag the language samples, to run the corpora through the evolutionary algorithm modeling program, and to tabulate and present the results.

Participants

Six corpora from the Child Language Data Exchange System (CHILDES; MacWhinney, 2000) were used as the source of input for the adaptation-selection computational algorithm. These corpora are the same samples that were used in Mintz's (2003) study.

Anne. (Theakston, Lieven, Pine, & Rowland, 2001) Anne was from a middle-class, English-speaking family. She was the oldest child in her family. Anne was audiotaped in her home during normal interaction for one hour, two times every three weeks for a period of one year. Anne's mother provided the input in the sample. The number of child-directed utterances in the corpus was 25,551.

Aran. (Theakston et al., 2001) Aran was from a middle-class, English-speaking family. As a participant in the same study as Anne (Theakston et al.), Aran was also an oldest child. Interactions for the sample took place in his home for one hour, two times every three weeks for a period of one year. The number of child-directed utterances in the corpus was 20,192.

Eve. (Brown, 1973) Eve was age 1;6 to 2;3 during the time this sample was collected and quite advanced in her language development. Twenty sessions were completed before Eve's family moved from the area. During the study, she demonstrated a great increase in her speech abilities. The number of child-directed utterances in the corpus was 14,543.

Naomi. (Sachs, 1983) Naomi was age 1;1 to 5;1 during the time this sample was taken by her mother, Jacqueline Sachs, who completed a longitudinal study of her daughter. The number of child-directed utterances in the corpus was 7,053.

Nina. (Suppes, 1974) Nina was age 1;11 to 3;3 during the time the sample was collected, from 1972-1973. The number of child-directed utterances in the corpus was 14,543.

Peter. (Bloom, Hood, & Lightbown, 1974; Bloom, Lightbown, & Hood, 1975) Peter was age 1;9 to 3;2 during the time the sample was collected. He was the first born in his family. Peter was from an upper-middle class family and his parents were college-educated. They lived in a university community in New York City during the time of the study. The number of child-directed utterances in the corpus was 18,004.

Instrumentation

At start-up the program implementing the algorithm opens the file of utterances that were spoken in the context of a child. This input file has one utterance per line, with the format of "word <tag word <tag word <tag..." and no punctuation marks. These utterances are stored for use in evaluation. An output file, which records the results, is opened as well. The mutation rate to be used by the program is set as a "one in X chance," where X is set by the program operator.

Next, a list of the grammatical tags used in the input file is initialized. These tags are used solely for evaluation purposes, not for training the program. A list is made of the words used in the file. This list is the basis of the dictionaries (of words and their possible grammatical tags), which will be evolved as the core task of the program. A population of 100 dictionaries is created wherein every word in each dictionary has a grammatical tag entry randomly chosen.

The adaptation-selection process is then cycled through for 5,000 generations. The odd-numbered utterances in the input file are the basis for the fitness evaluation of each dictionary. This fitness evaluation consists of examining the odd-numbered input file utterances to see if a particular dictionary contains the correct tag for each word in the utterance; if it does, the tally of the number correct is increased by one. Then, the number of tags used in the dictionary is divided by the number of words in the dictionary; the product of this ratio times the number of correct hits is used as the fitness criterion. The odd-numbered utterances serve as the basis for the fitness evaluation, and the even-numbered utterances in the input file are used as the basis for quantifying the generalization accuracy of the dictionary. These accuracy levels are written to the screen and to the output file.

After all candidate dictionaries have been evaluated, the one with the highest fitness score serves as the starting point for populating the next generation. Each word entry in this dictionary has one in X chance of having its grammatical tag entry replaced with a randomly chosen one as part of making an offspring dictionary, regardless of if that particular entry was actually correct or not. The X level was set at program start-up. By this process a population of new offspring dictionaries is created, which are then evaluated, and the fittest one becomes the basis for future dictionaries. After the specified number of evolutionary cycles has been reached, final data and the best-evolved, most fit dictionary are written to the output file for reference.

Procedure

The corpus (i.e., the set of language samples) for each child was formatted and grammatically coded before being run through the adaptation-selection computational model. As there was a random component in creating the initial program dictionaries, each corpus was run through the program three times at each mutation likelihood, and the data points presented are the average of the three runs. The number of generations of adaptation and selection allowed was 5,000. The mutation-variation rate for each dictionary tag was set to one of four chances of mutation, independent of whether the tag was correct or not. The program was designed to grammatically tag the words from the language sample corpus input with the most likely grammatical category per word. Typically, tagging each word in a corpus with the most likely tag for the word results in about 92% of the words being tagged correctly; the exact maximum level for each corpus was calculated and the goal was to see how close the adaptation and selection algorithm could approach that level.

Results

The computational model analyzed the effects of two independent variables, the number of evolutionary cycles and the mutation likelihood, on the two dependent variables, the token accuracy and type accuracy of the program output. Token accuracy refers to the accuracy attained by the model in correctly tagging the words in the even utterances in each corpus, following the selection of the most accurate dictionary for each generation. Type accuracy refers to the accuracy of the model in correctly tagging different words in each corpus. Each language corpus was run through the adaptation selection model three times for 5,000 generations each and at different mutation likelihoods. The token and type accuracy varied according to number of evolutionary cycles or generations (1-5,000) produced by the evolutionary program. Chance of mutation was set at four different likelihoods: 1/800, 1/1200, 1/1600, and 1/2400, and accuracy was observed at each of these rates.

Table 1 shows the mean token accuracy percentages for generations 4801-5000 at each mutation likelihood rate for each corpus. It may be seen in Table 1 that higher rates of mutation resulted in increased accuracy of token means. Table 2 displays percentages for the mean type accuracies for the same evolutionary cycles. The mutation likelihood of 1/1200 produced in general slightly higher type accuracy overall.

Table 1

Mean Token Accuracy Results of Generations 4801-5000 at Each Mutation Likelihood for Each Corpus

Corpus	1/800	1/1200	1/1600	1/2400
Anne	88.81	88.18	86.23	83.73
Aran	86.80	86.19	84.68	78.27
Eve	90.11	89.25	86.99	82.62
Naomi	88.59	87.54	84.62	80.07
Nina	90.08	89.39	88.79	80.56
Peter	89.77	88.65	88.48	78.85

Table 2

Mean Type Accuracy Results of Generations 4801-5000 at Each Mutation Likelihood for Each Corpus

Corpus	1/800	1/1200	1/1600	1/2400
Anne	60.43	61.34	60.36	57.77
Aran	52.33	54.59	51.98	52.80
Eve	65.47	66.23	63.81	60.21
Naomi	65.68	65.23	63.60	59.19
Nina	66.50	67.44	66.30	61.44
Peter	64.78	64.98	63.76	59.80

The growth curves shown in Figures 1 and 2 illustrate the token and type means for all corpora for 5,000 generations at each of the four mutation rates. In Figure 1, the computational model produced the most rapid increase in mean token accuracy from generations 1 to 2,000 (for the mutation likelihood of 1/800) and then slowed in rate of increase until reaching its maximum at approximately cycle 4,500. Variation in results based on mutation likelihood is also illustrated in Figure 1, with a 1/800 chance of mutation yielding slightly the highest accuracy and 1/2400 chance of mutation resulting in a significant decrease in accuracy. Figure 2 shows a more gradual incline in accuracy of type means across number of generations. There is less variation associated with mutation rate levels in the findings for mean type accuracy.

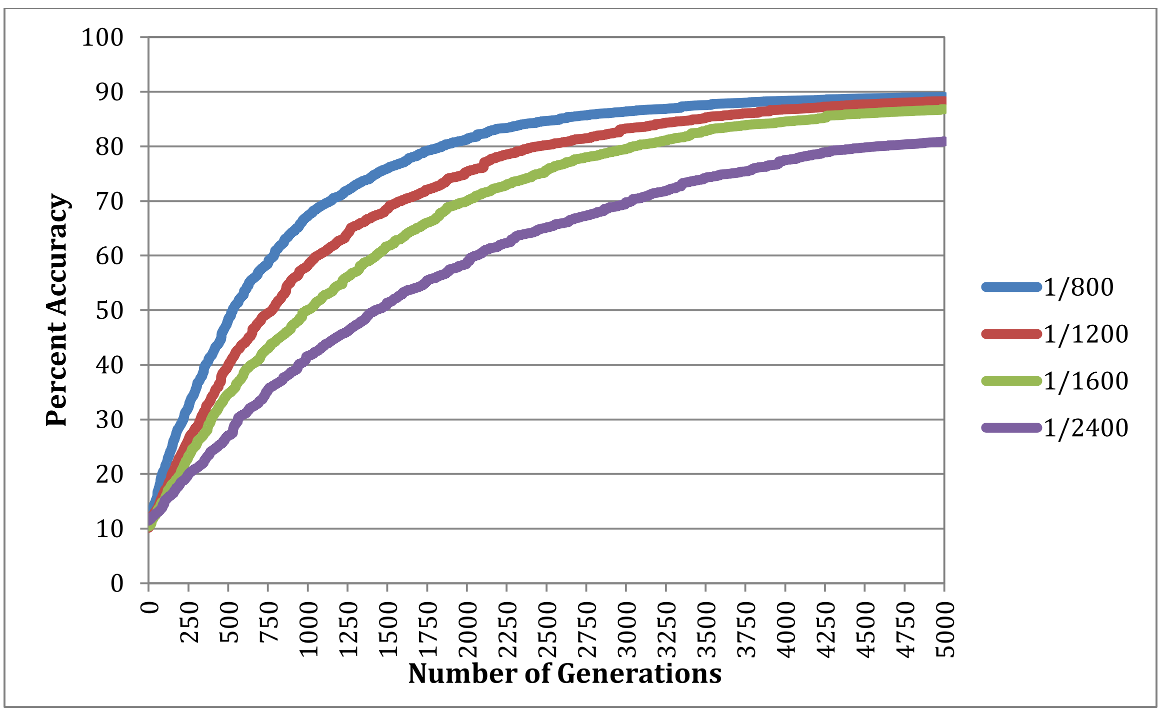

Figure 1. Token means for each mutation rate across all corpora.

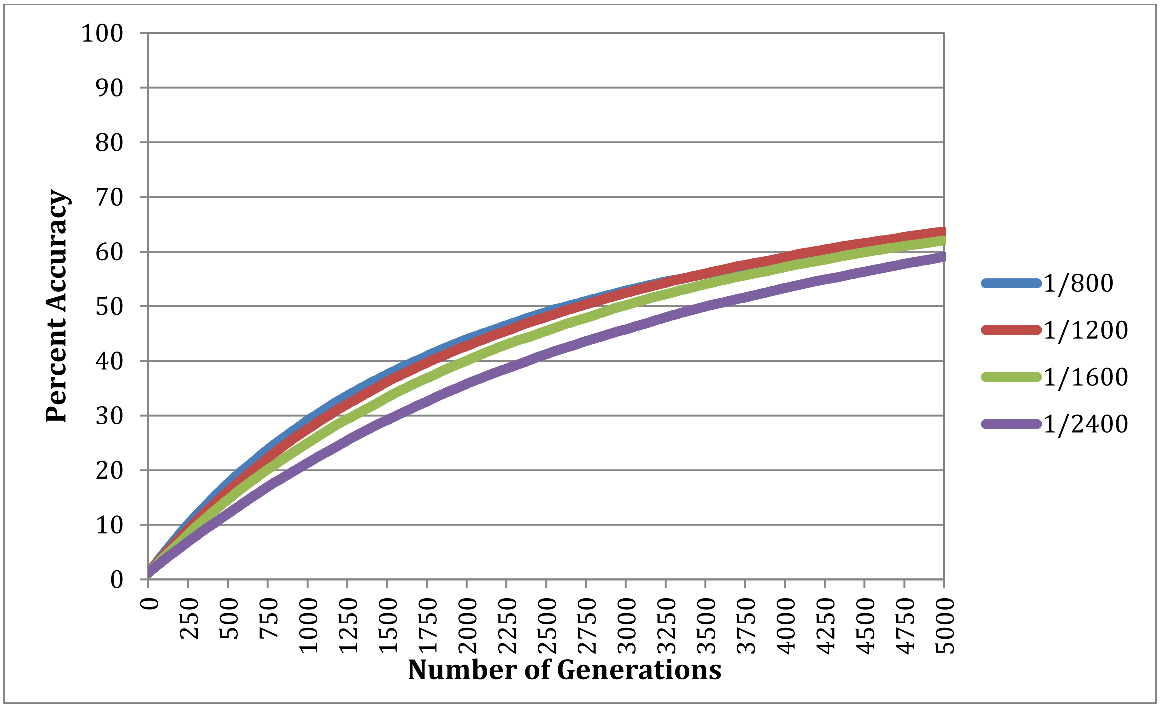

Figure 2. Type means for each mutation rate across all corpora.

Figures 3 through 8 show the growth curves for accuracy of token and type means for each corpus. Similar patterns are found in terms of accuracy rates according to mutation likelihood and cycle number when comparing these figures to the figures described above. Again, the higher mutation rates resulted in increased accuracy of token means. A 1/1200 mutation rate for type means produced in general a slightly higher accuracy overall. There is a range of 4% in maximum token accuracy across corpora at a 1/800 mutation rate and a range of 15% in maximum type accuracy across corpora at the same mutation rate.

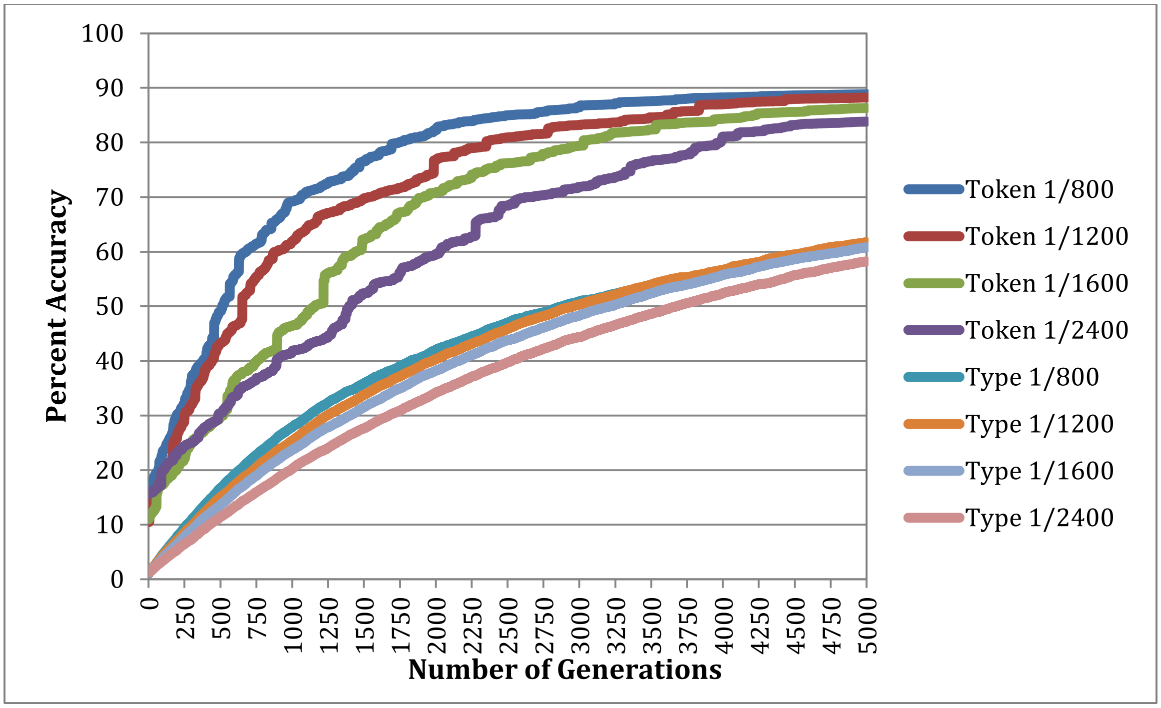

Figure 3. Token and type means for Anne's corpus at all mutation rates.

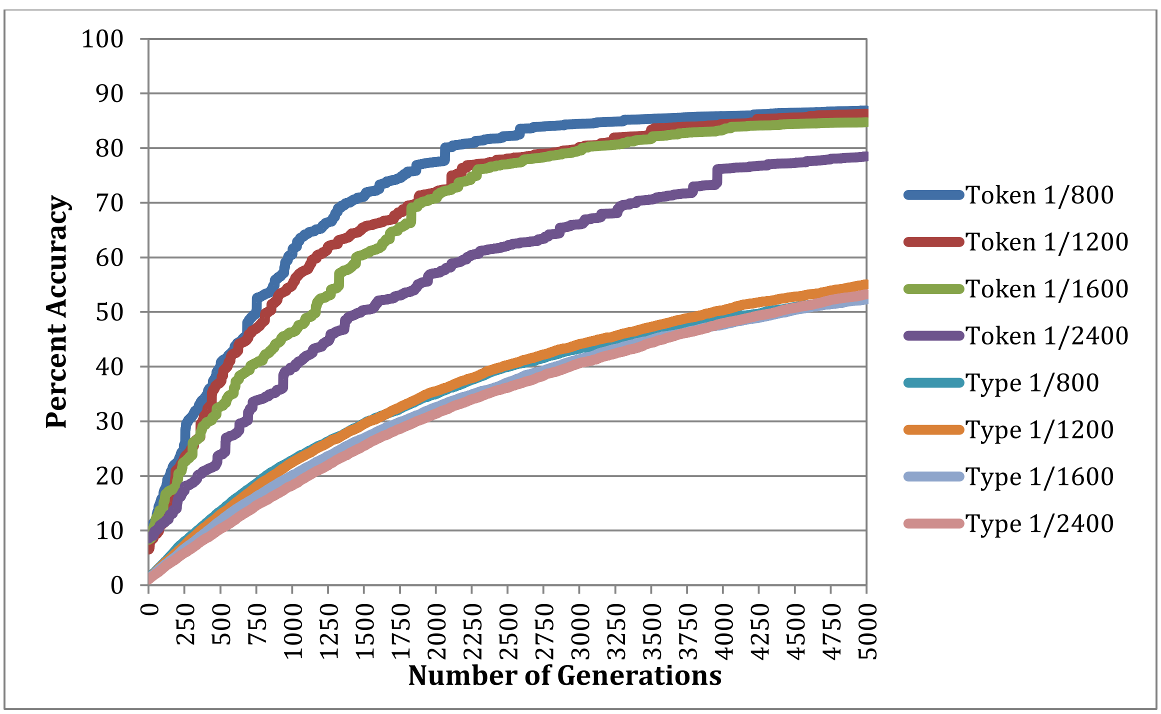

Figure 4. Token and type means for Aran's corpus at all mutation rates.

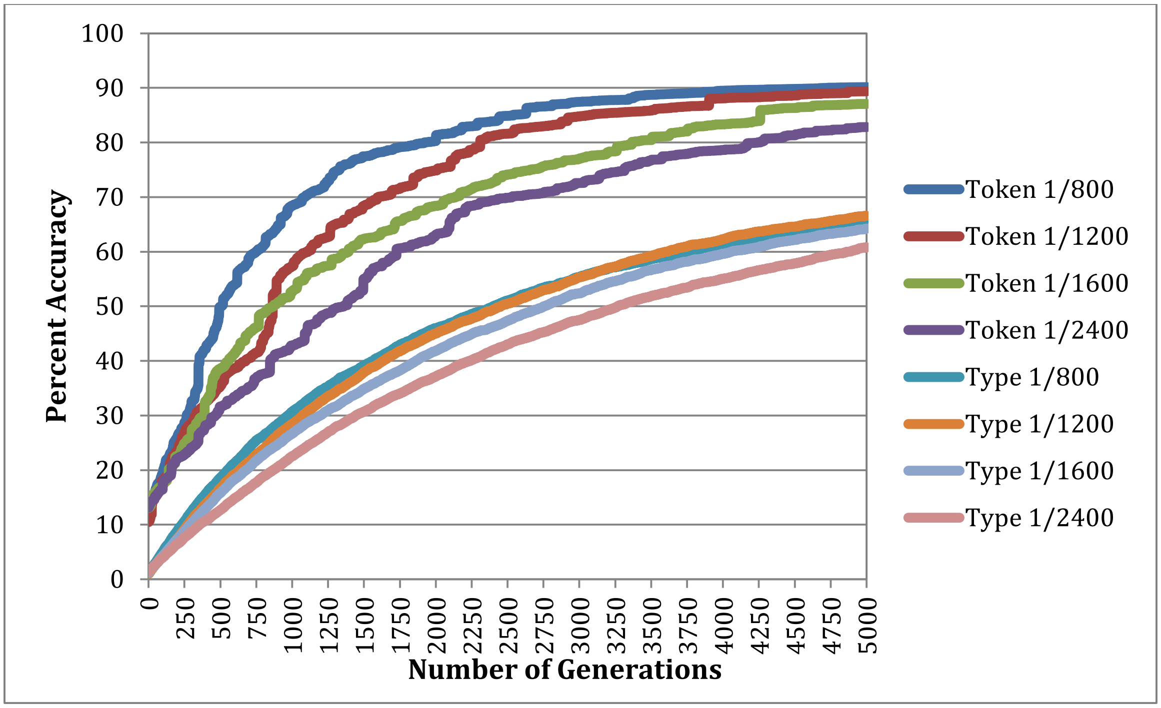

Figure 5. Token and type means for Eve's corpus at all mutation rates.

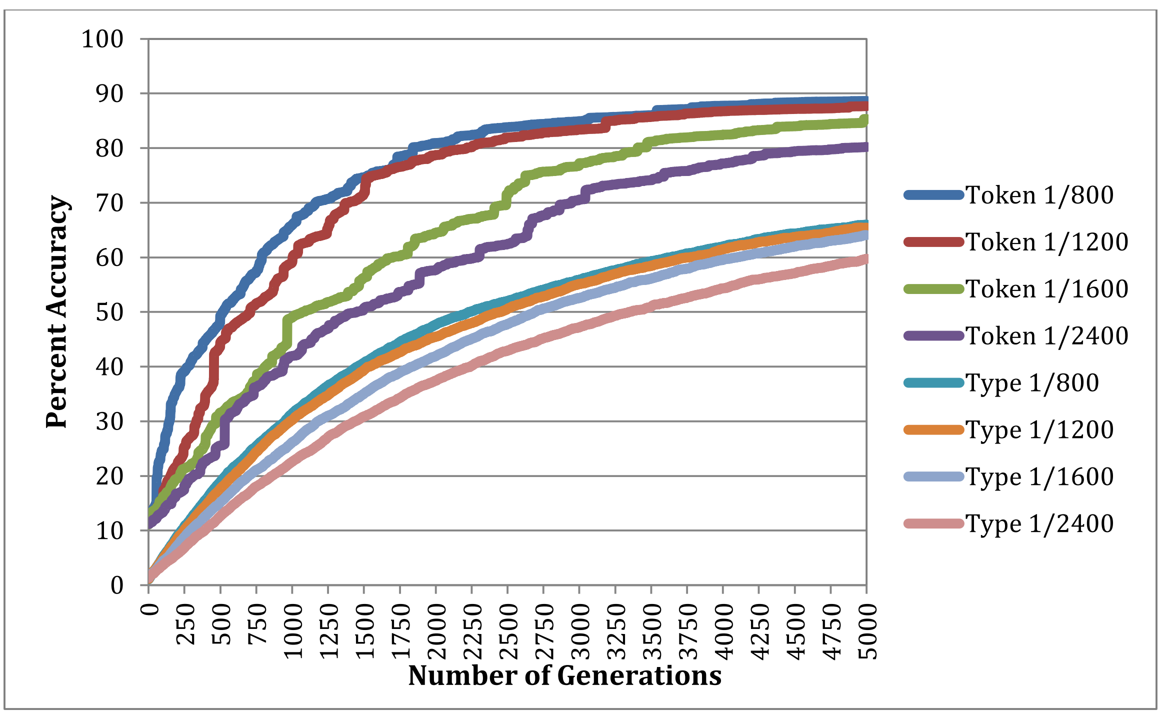

Figure 6. Token and type means for Naomi's corpus at all mutation rates.

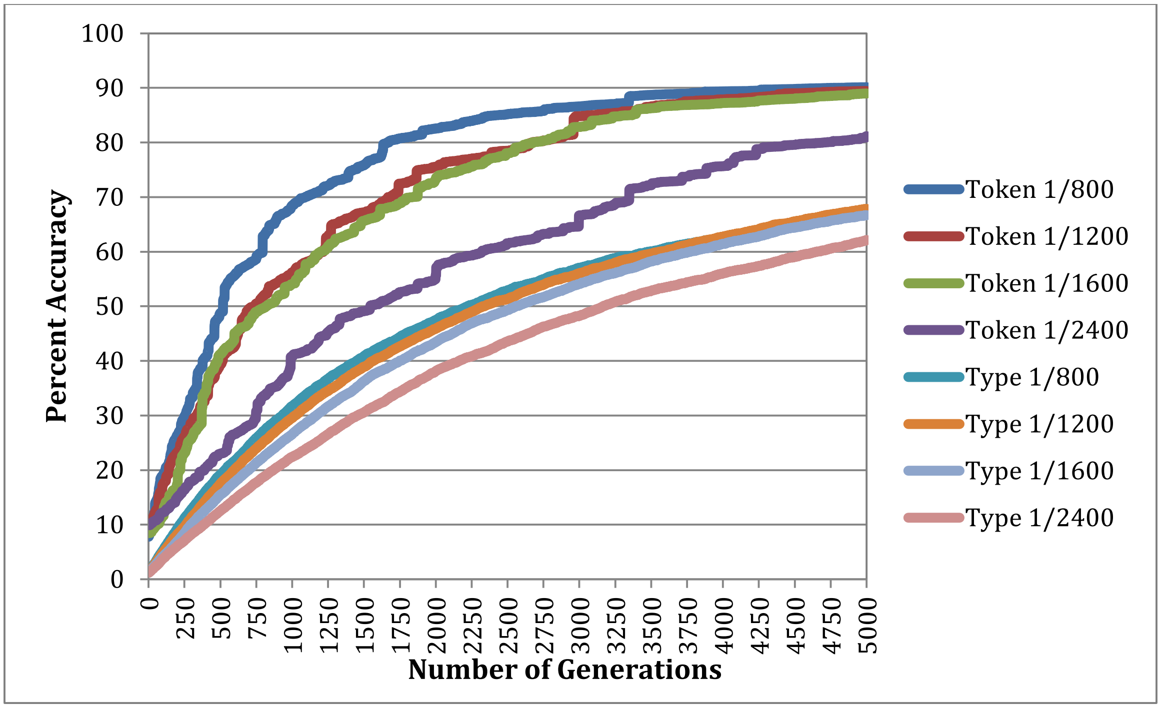

Figure 7. Token and type means for Nina's corpus at all mutation rates.

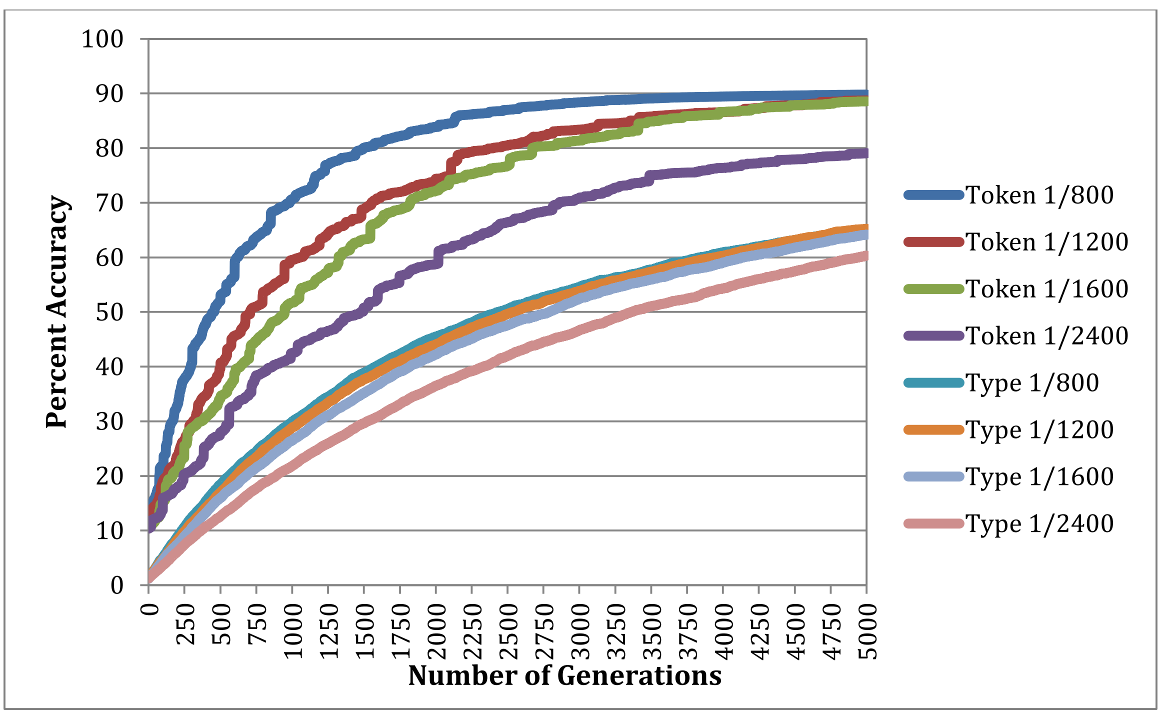

Figure 8. Token and type means for Peter's corpus at all mutation rates.

Discussion

The purpose of this study was to determine the accuracy of an adaptation-selection computational model in establishing the grammatical categories of words from six language corpora. Accuracy was determined for both word tokens (all the words in the corpora) and word types (all different words used in the corpora). The findings of this study suggest that the algorithm is able to produce high accuracy in its ability to correctly label grammatical categories of child-directed input. The number of evolutionary cycles used and the mutation likelihood had a direct impact on the accuracy of token and type means. For word tokens across all corpora, accuracy of results increased at an even rate at a 1/800 mutation likelihood for the first 2,000 evolutionary cycles, and then increased at a more gradual rate until reaching 90% at 5,000 cycles. For word types across all corpora, a more gradual rise in accuracy was found. Type accuracy reached its maximum of only 63% at a mutation likelihood of 1/1200 by the 5,000th cycle. Lower mutation rates resulted in lower accuracy of word tokens, while the mutation rate of 1/1200 produced in general a slightly higher accuracy for word types. The corpora for both Eve and Nina resulted in the highest accuracy of token means, while the Nina corpus produced the highest accuracy among type means.

Pilot tests were performed to determine the impact of the random component inherent in the adaptation-selection computational model. Due to the possibility of fluctuation in corpora accuracy scores, the Naomi corpus was run through the program eleven times to determine the extent of output variation. The scores resulted in a less than ±0.5% difference. Pilot studies also explored the effects on accuracy of manually correcting the grammatical category of each word in the language samples after running them through the grammatical tagging software vs. using only the grammatical tagging software. Portions of two corpora, Peter and Anne, were manually tagged following the autotagging process. This involved checking autotags for accuracy and correcting words that were incorrectly tagged. Results between the two tagging processes were essentially the same and no further study was done in the area.

The results of the current study offer a slightly different contribution to the area of grammatical category acquisition when compared with previous studies of a similar nature. A recent study by Cluff (2014) also explored the use of an adaptation-selection model in learning grammatical categories of words. While the current study focused on using only the single most likely tag for each word (what has commonly been done in previous studies) and used a total of 85 possible tags, the study by Cluff used a total of 16 tags, with each word having up to three possible tag options. The resulting growth curves of the Cluff study illustrate a much faster learning rate, with the accuracies reaching approximately 80% by the 500th generation. This accelerated learning rate, which differs from the gradual learning rate of the present study, is most likely due to the variation in number of total tags. Another difference between studies was found in the effect of predetermined mutation rates on the program output. The study by Cluff set the chances of mutation at 1/400, 1/800, 1/1200, and 1/1800. The mutation rates that resulted in a less likely chance of mutation (1/800, 1/1200) produced the highest accuracy. In other words, as the number of opportunities for mutation went down, accuracy generally went up. In the present study, the mutation likelihood variable had the opposite effect on output – the greater the chance of random mutation, the higher the accuracy. It is unclear what might have caused this difference; however, there are several variations between studies that could be considered. As previously mentioned, the study by Cluff used multiple possible tag options per word and considered the effect of a different set of mutation rates. Additionally, Cluff's language corpora went through only 4,000 generations, rather than the 5,000 used in the current study. Any number of causes may have contributed to the multiple differences in outcome.

As discussed earlier, a number of contributions to increasing understanding in the area of determining child grammatical category acquisition through computation models have been published. The present study, along with the study by Cluff (2014), uses an approach that differs significantly from previous models. Because of this, direct comparison to previous studies is impossible. However, each study has made its own unique contribution to this area. The study by Cartwright and Brent (1997) used the concept of sentential minimal pairs to describe distributional category acquisition. In their distributional approach, Redington et al. (1998) used a computational model to form dendrograms, which visually represent a hierarchical clustering of syntactic categories. Mintz (2003) used frequent frames as a means of explaining an additional distributional method of grammatical category acquisition. Freudenthal et al. (2005) refuted the proposed accuracy of the distributional methods supported by co-occurrence statistics in determining syntactic category, instead suggesting that without using these derived syntactic categories to generate language output, accuracy cannot be fully determined. These studies, including the present analysis which examines an adaptation-selection algorithm, are just a few of the studies that have made an impact on our current understanding of grammatical category acquisition in children.

Despite its contributions, the current study has a few limitations. One of these limitations is the program's ability to use only one grammatical tag per word. Although the most likely tag per word was chosen, many words can be placed in multiple categories based on their usage. For example, the word *bark* can be either a noun or a verb. As previously mentioned, the version of the adaptation-selection model implemented by Cluff (2014) provided for a choice of three possible tags per word, which allowed for the program to achieve a higher chance of correct tag choice. Future studies using this model could allow for multiple tag options per word, which might change the program's potential for achieving greater accuracy. Further research might also shed light on the impacts of other variables in the study, such as optimal mutation rate, number of grammatical categories in the set, and the number of evolutionary cycles. The work done by Cluff is one example of a very similar study in which these variables were different. The adjustments of these variables resulted in a broad spectrum of disparities in the output, which were discussed previously.

An additional limitation of this study is the restricted number of participants. The six individual language corpora used in this study were taken from MacWhinney's (2000) CHILDES database and were the same used by Mintz in his 2003 study. Future studies could make use of a larger number of corpora, which might contribute to increased accuracy. Moreover, the nature of the language samples used in the current work was fixed and invariable because they were obtained from an online source. Such things as child age, socioeconomic status, setting, parent interaction style or language input, and corpus size were determined far in advance of the present study. This resulted in less control over the variables found in language input that might have impacted results of the study. Future researchers have the option to manipulate these variables in ways that might impact the success of their studies.

While this adaptation-selection program helps researchers in the field understand one possible method of language development, it does not aim to posit a direct model of how children actually learn grammatical categories. In terms of clinical application, Siegler (1996) suggested the interesting possibility that concepts of evolution that have help scientists understand how species evolve might also be used by cognitive developmentalists in understanding how children develop the ability to think. This theory, while relevant, still needs quite a bit of development before actual application can be made to explaining aspects of children's language development.

Despite the limitations, this evolutionary model of grammatical category acquisition has made a notable contribution to a growing understanding of one foundational aspect of child language learning. The fairly high accuracy of results obtained by the adaptation-selection computational model suggests that the program functions well and can be viewed as one possible representation or tentative explanation of grammatical category acquisition. The positive nature of the findings warrants further studies involving the evolutionary model. Such research can continue to answer the questions regarding the actual patterns and methods of learning the grammatical categories of words by children.

References

Ambridge, B., & Lieven, E. V. M. (2011). *Child language acquisition: Contrasting theoretical approaches* (pp. 1-3, 191-209). New York, NY: Cambridge U Press.

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

Bloom, L., Lightbown, P., & Hood, L. (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development, 40* (Serial No. 160).

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

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

Cluff, S. Z. (2014). A model of grammatical category acquisition using adaptation and selection. (Unpublished master's thesis). Brigham Young University, Provo UT.

Fogel, D. B. (2002). *Blondie24: Playing at the edge of AI.* San Francisco, CA: Morgan Kaufmann.

Fogel, D. B. (2006). *Evolutionary computation: Toward a new philosophy of machine intelligence (3rd ed.)*. Hoboken, NJ: IEEE/Wiley.

Fogel, D. B., Watson, E. C., Boughton, E. M., & Porto, V. W. (1998). Evolving artificial neural networks for screening features from mammograms. *Artificial Intelligence in Medicine, 14*(3), 317-326.

Freudenthal, D., Pine, J. M., & Gobet, F. (2005). On the resolution of ambiguities in the extraction of syntactic categories through chunking. *Cognitive Systems Research, 6,* 17-25.

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

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*(1), 91-117. doi:10.1016/S0010-2077(03)00140-9

Pinker, S. (1988). Learnability theory and the acquisition of a first language. In F. Kessel (Ed.), *The development of language and of language researchers: Papers presented to Roger Brown.* Hillsdale, NJ: Erlbaum.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22*, 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.), *4. Children's language*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking.* New York, NY: Oxford University Press.

Suppes, P. (1974). The semantics of children'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*, 127-152.

Weisleder, A., & Waxman, S. R. (2010). What's in the input? Frequent frames in child-directed speech offer distributional cues to grammatical categories in Spanish and English. *Journal of Child Language, 37,* 1089-1108. doi: 10.1017/S030500090999006

APPENDIX: ANNOTATED BIBLIOGRAPHY

Ambridge, B., & Lieven, E. V. M. (2011). *Child language acquisition: Contrasting theoretical approaches* (pp. 1-3, 191-209). New York, NY: Cambridge University Press.

Ambridge and Lieven described two major approaches to child language acquisition, the generativist and the constructivist approaches. The generativist (or nativist) view supports the idea that children are born with an innate knowledge of grammar, its rules and categories. They are required to label each new word with an already existing category. The constructivist (also known as emergentist, functional or usage-based) perspective proposes that children have no innate knowledge of grammar, and acquisition is input-based (knowledge based on characteristics of input). Children must acquire categories and new words, placing new words into created categories appropriately.

The authors presented major approaches under the generativist perspective. One theory describes linking rules, which link syntactic and semantic categories; however, the notion of innate linking rules is being abandoned for other theories. Similarly, the evidence for prosodic bootstrapping is weak.

Distributional analysis approaches that apply the generativist perspective include Mintz' (2003) frequent frames approach. This has had more success than other distributional approaches. Problems with the generativist approach are found when it comes to connecting categories that are defined through distribution with categories that are innately specified.

The authors stated that the constructivist perspective supports the idea that grammatical categories are not innately formed groups, but labels used to classify words that behave similarly. One approach is semantic assimilation, which involves the use of semantic similarities early on, with distributional similarities used later to determine more difficult category members. The authors included Redington, Chater, and Finch's (1998) distributional approach using clusters, as well as Cartwright and Brent's (1997) study involving templates and sentential minimal pairs in the constructivist category for distributional approaches. The authors cited Freudenthal, Pine, and Gobet (2005), who pointed out the limitations of these studies by suggesting that producing output generated using the random words from the newly formed grammatical categories was required in order to truly measure accuracy. Although Freudenthal et al. attempted to address this issue, the overall challenge with constructivist distributional approaches is that the processes of comprehension and production of grammatical categories need to be more clearly shown.

The authors concluded their summary by outlining grammatical category acquisition through using a phonological approach (e.g., syllable stress), which fits under both generativist and constructivist perspectives.

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*, 150-177.

Berko used nonsense materials to test subjects for acquisition of morphological rules. Nonsense words were necessary because, if known words were used, the subjects would have been tested for their ability to memorize words, rather than a system of rules for English morphology. Berko first examined actual vocabulary of 1st graders to determine what aspects of morphology were present in children's utterances. Nonsense words and picture cards were then created to examine children's ability to use plural and possessive nouns, third person singular, as well as past and progressive tense verbs, and comparative and superlative adjectives. The words were tested first on 12 adult subjects who were all college graduates. Their answers were considered correct; the children's answers were compared against the answers of the adult subjects. Eighty children participated in the experiment, preschoolers and 1st graders. Results indicated that children are able to use morphological rules correctly, as shown by their accuracy in placing correct endings on nonsense words. The children performed most successfully on words with frequently used endings.

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

Cartwright and Brent used the concept of sentential minimal pairs to describe distributional category acquisition. An example of two sentences that constitute a minimal pair is: *I saw the cat. I saw the dog. Cat* and *dog* are the only two words that differ in these sentences, and they are found in the same sentence position. However, most sentences cannot be matched up so succinctly. Generalized forms of minimal pairs must be created by forming templates. The authors hypothesized that when children hear a sentence, they create a new group for each word in the sentence. A new template is formed consisting of the new group sequence found in the sentence. This template is added to a list of templates already formed. A preference list is used in the process of merging similar groups from different (older and new) templates, until the best arrangement is found. In experiment 1 computer simulations tested the theory in regards to an artificial language. Experiment 2 analyzed how words can be grouped into more than one category, using the same artificial language as in experiment 1. Experiment 3 and 4 tested the authors' theories using English samples of child-directed speech. Experiment 4 differed from 3 in that the analysis was done incrementally. Experiment 5 explored the connection between semantic and syntactic categories.

Benefits of this strategy include: a set of discrete categories can be obtained; a whole sentence is used as context (rather than one or two words immediately preceding or succeeding the target word) so that phrase boundaries are kept intact. The authors theorized that the use of minimal pair sentences in their analysis would allow for groupings of words into more consistent grammatical categories. The study's high level of accuracy inferred that this approach could be useful.

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

Chang et al. studied the use of several algorithms whose purpose is to use distributional information in learning syntactic categories. They used as the basis for their study a prediction model, which allowed them to use corpora in several different languages to test theories of syntax acquisition and the production of adult utterances. The authors used several typologically-different languages, allowing the study to be free from biases toward particular languages, such as English, or toward languages with specific typologies. This aspect of the study is unique because tagged corpora are typically not available in different languages. The authors used an evaluation measure called Word Order Prediction Accuracy (WOPA). WOPA works by beginning with a candidate set (this includes the unordered set of words from the utterance that is to be predicted) and then trying to predict the order of words one word at a time. Each predicted utterance is given a score based on its accuracy when compared with the original utterance. Using WOPA, six syntax acquisition algorithms were compared to see which one best matched the knowledge used in producing the corpora. The study also aimed to show that WOPA can be used as a method for comparing computational approaches. Results indicated that words are better characterized using specific categories, rather than broad categories.

Chang, F., Lieven, E., & Tomasello, M. (2008). Automatic evaluation of syntactic learners in typologically-different languages. *Cognitive Systems Research, 9,* 198-213.

Chang et al. studied syntax acquisition through a specific computational model, which was unique because of its ability to compare results of several different languages. Most computer models that evaluate syntax acquisition are limited by their ability to focus on only one language. The authors' evaluation measure incorporated a bag-of-words incremental generation (BIG) approach with an automatic sentence prediction accuracy (SPA) measure. This BIG-SPA measure was used to evaluate several learners using n-gram statistics, as well as to evaluate an Adjacency-Prominence learner (based on a psychological account of syntax acquisition, which considers multiple aspects of syntactic learning). These models allowed for comparison of theories of syntax acquisition in several languages in terms of how well they gain syntactic knowledge from the input. The bag-of-words measure considers a group of words that have previously formed an utterance and uses one word at a time to try to predict the utterance. The sentence prediction accuracy is the percentage of complete utterances correctly predicted in the corpus. While the Adjacency-Prominence learner resulted in the greatest accuracy of sentence prediction, the BIG-SPA tasks allowed for the comparison of syntactic learners in different languages.

Chemla, E., Mintz, T. H., Bernal, S., & Christophe, A. (2009). Categorizing words using 'frequent frames': What cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental Science, 12,* 396-406. doi: 10.1111/j.1467-7687.2009.00825.x

The study by Chemla et al. focused on the accuracy of Mintz' (2003) frequent frames model when applied to the French language. This distributional model by Mintz analyzed the role of frequent frames (the words surrounding or "framing" a target word) in contributing to the process of grammatical category acquisition. This model is quite successful in English, and the first purpose of the present study was to determine if similar levels of accuracy could be obtained in French. Despite doubts that the nature of the French language would support a frequent frame model, the results yielded high accuracy in determining grammatical categories of content words. The authors' second purpose was to explore the effects of discontinuity on the accuracy of the model. Instead of using the framing method (A x B), two other contextual environments were used, that of front contexts (A B x) and back contexts (x A B). The authors found that, while front contexts produced higher accuracy than back contexts, neither of them resulted in the high level of accuracy produced with the frequent frames model. The third experiment carried out by the authors involved using the categories of the frequent frames themselves, instead of the specific words found in the frames, to predict accuracy of the target word. The outcomes of this experiment indicated poor accuracy in comparison to the normal version of the frequent frames model.

Cluff, S. Z. (2014). A model of grammatical category acquisition using adaptation and selection. (Unpublished master's thesis). Brigham Young University, Provo UT.

Cluff studied a new computational model of grammatical category acquisition involving a theory of adaptation and selection. An evolutionary algorithm was used to analyze the child-directed input contained in five language corpora from the CHILDES database (MacWhinney, 2000). The language samples were transcribed and grammatically coded using tagging software. In running the corpora through the adaptation-selection model, dictionaries are created, which then reproduce based on accuracy ratings. Each sample was run through the program for 4,000 generations (or evolutionary cycles) with the likelihood of mutation set at a one in X chance. While most previous studies have adopted a single, most-likely tag scheme, Cluff's study allowed a possibility of multiple grammatical tags per word. The program produced high accuracy of grammatical tags. The 1/400 mutation rate (the highest of the four possible rates) produced the lowest accuracy. The other possibilities were 1/800, 1/1200, and 1/1800, (lower likelihoods of mutation) which all produced higher accuracies. Additionally, words with more possible tags resulted in better accuracy. The results of this study were promising and warrant further research in this area.

Freudenthal, D., Pine, J. M., & Gobet, F. (2005). On the resolution of ambiguities in the extraction of syntactic categories through chunking. *Cognitive Systems Research, 6,* 17-25.

Freudenthal et al. refuted the proposed accuracy of the distributional methods supported by co-occurrence statistics in determining syntactic category. Redington, Chater, and Finch (1998) and Mintz (2003), among others, supported the notion that the observed accuracy of such distributional patterns allows insight into models of child grammatical category acquisition. However, Freudenthal et al. suggested that without using these derived syntactic categories to generate language output, accuracy cannot be fully determined. To accomplish this, Freudenthal et al. used the software Mosaic, a computational model designed to simulate syntactic acquisition. Input was fed through the model and co-occurrence statistics were used to generate phrase substitutions that fit the context of the original input. The authors used a new version of Mosaic that classified frequently recurring phrases into chunks, which are treated as one unit, and can consequently only be involved in substitutions as a whole unit, rather than by single word. Freudenthal et al. felt that chunking increased the accuracy of substitutions, decreasing the level of error. Results of the study indicated that measures typically used to determine accuracy of syntactic categories based on co-occurrence statistics are inadequate.

Hills, T. (2013). The company that words keep: Comparing the statistical structure of child- versus adult-directed language. *Journal of Child Language, 40,* 586-604. doi:10.1017/S0305000912000165

Hills studied the differences between adult-directed language and child-directed language in terms of how each might influence a person's ability to learn words. Past studies showed that child- and adult-directed language do differ in several ways. This study involved the analysis of several different aspects of language input, including associative structure (the chances of a word occurring with its free associates), contextual diversity, word repetitions, and frequency, and how these aspects of input might affect the ease or difficulty with which language is learned. Longitudinal studies of six language corpora were completed, with four of the corpora being child-directed input (children were ages 1;0 to 5;0) and two of the corpora adult-directed input. The findings of this study indicated that child-directed language, in comparison to adult-directed language, is more associative, shows more repetition, demonstrates the use of greater word frequencies, and relies on more usage consistency. This is more true for words learned earlier (e.g., the words learned by age 1;0) than for words learned at later ages. These results suggest that adults structure their language differently when directing it toward young children, and this difference in structure does positively affect children's ability to learn language.

Kiss, G. R. (1973). Grammatical word classes: A learning process and its simulation. *Psychology of Learning and Motivation, 7*, 1-41.

In this study, Kiss aimed to explore the system of language acquisition involving word classes. He explained this process as incorporating several components, including associative learning, making use of semantic and syntactic information, adjusting label of word class based on further learning, etc. Kiss focused specifically on associative learning and classification (distributional information). He used computer models to demonstrate the mechanisms by which this learning occurs. The program used to demonstrate associative learning "reads an input corpus and establishes associative links between the words and their contexts" (p. 21). The context taken into consideration here is the word preceding the target word. The program used to demonstrate classification learning is meant to "establish internal representations for groups of words which are distributionally similar to each other" (p. 22). Child-directed input was used in determining the effectiveness of these models. The models were shown to be effective in determining word classes, including features of nouns, verbs, adjectives, articles, and adjectives.

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

MacNamara discussed the acquisition of language in terms of its relationship with meaning. He proposed that, in analyzing input, meaning is first determined, and then language is discovered. The study focused more specifically on cognitive strategies used by children in comprehending language. MacNamara asserted that infants form thoughts before they place a linguistic code on those thoughts in order for them to be communicated. Infants may understand the meaning of something first. They then discover the name for that referent, and then use the new vocabulary in forming an idea of syntactic structure. An example of this is given in the sentences: *The boy struck the girl* and *The girl struck the boy*. The child must rely on comprehension of what is happening before he can apply the proper syntactic form to that meaning. Additionally, grammatical devices or forms that are concerned with meaning are developed prior to syntactic forms that have less to do with meaning. MacNamara also supported the idea that phonological forms in language are also determined based on meaning. He explained that it is important to remember that, while there may be some chronology to this acquisition process, each process has an interdependent relationship with the others for development.

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

In this study, Mintz analyzed the role of distributional information in aiding the acquisition of grammatical categories in an artificial language. Participants included 49 undergraduate students, nine of whom were excluded from the study due to error caused by faulty equipment or the examiner. Participants were given training sets and test sets of sentences to listen to. After 6 minutes of listening to the training set of sentences (consisting of different combinations of words in the artificial language), participants were asked questions regarding what they could remember about the sentences. They were asked whether specific sentences had been heard during the training set and how confident they were in their answer. The results of this study were positive, indicating that students used the distributional information contained in the artificial language to determine grammatical categories of the words. These results differ from previous studies of a similar nature, indicating that the larger amount of distributional information available contributed to the outcome. Limitations of this study include the simple nature of the training and test sentences. Real languages are much more complex than the artificial language used here. Additionally, this study tests the abilities of adults, rather than children. While this information is helpful, future studies will need to look at how children use distributional information.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*, 91-117. doi:10.1016/S0010-2077(03)00140-9

Mintz studied the role of frequent frames in grammatical category acquisition. This distributional approach uses the two words framing (surrounding) a word to determine the word's grammatical category. These framing words that frequently occur together are referred to as frequent frames. Words that are surrounded by the same frequent frame (intervening words) are placed in the same grammatical category. Example: The blank in the following phrase is likely to be a verb based on the frequent frame surrounding the blank: *to* _____ *to*.

The purpose of the study was to determine type of distributional information most informative; types of distributional cues picked up on and used by children and infants; and how to assemble categories into a system. The purpose was not to provide a model for grammatical category acquisition by language learners, but to explore the assumptions that could be formed about a model.

Six corpora from the CHILDES database (MacWhinney, 2000) were used as input – adult utterances only were used. Frames were identified and frequent frames were selected based on certain guidelines including frames that occurred often enough that they were noticeable and those frames which surrounded a variety of words that could be categorized together. The frequent frames for each corpus were analyzed and intervening words were grouped according to frame-based categories. Accuracy and completeness of hits, false alarms, and misses was analyzed.

Results showed high accuracy of grammatical category identification. Accuracy was slightly lower for identification of an extended set of grammatical categories. Additional experimentation was done to control for frame selection method according to specific corpus and to confirm that the high accuracy found in experiment 1 did not occur due to small categories with few member types. Accuracy and completeness outcomes were very similar between the two experiments.

Frequent frames are effective in categorizing words in children's input. Frequent frames produced categorization accuracy that was equal or more accurate than the distributional methods that were analyzed in previous studies. Additional studies are necessary to determine whether children actually use frame information in categorizing words.

Monaghan, P., Christiansen, M., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, 55, 259-305. doi: 10.1016/j.cogpsych.2006.12.001

The study by Monaghan et al. aimed to show the role of both phonological cues and distributional information in the determination of grammatical categories. Their hypothesis is called the Phonological-Distributional Coherence Hypothesis (PDCH). The PDCH was tested by measuring the contribution of phonological and distributional information in differentiating between open and closed class words, as well as noun from verbs in the languages of Dutch, English, French, and Japanese. The authors predicted that words which were not placed in their appropriate grammatical category through distributional means, would be correctly classified through phonological cues. Five different experiments were designed: Experiment one evaluated the success of phonological cues in determining correct grammatical category in the four languages; Experiment 2 assessed the role of distributional information in determining category; Experiment 3 determined the relative contribution of both types of cues for determining category; Experiment 4 was similar to 3, but looked at words with multiple tags; and Experiment 5 measured PDCH's accuracy when morphological elements were removed from the language. Results of the study suggest that the authors' hypothesis was correct. Additionally, use of both forms of cues resulted in generally higher accuracy than the use of only either phonological or distributional cues.

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.) *Mechanisms of language acquisition.* (pp. 399-441). Hillsdale, NJ: LEA.

Pinker emphasized the importance for a language acquisition theory to take into account the difficulties of language acquisition, such as grammar rules and specific/unique language constraints, and how children overcome them. The semantic bootstrapping hypothesis states that children use semantic aspects of language in the process of acquiring syntactic or grammatical categories. Once initial rules are learned, children are able to apply those rules to other situations/contexts and build up their language base. Bootstrapping hypothesis differs from distributional analysis in that the child applies structural rules as a whole rather than in definitive, specific ways. This allows for fewer labeling errors to occur. Additional advantages of the bootstrapping hypothesis are a greater efficiency in learning new rules and universal conditions that apply to all rules and sentences. Disadvantages include limitations of this theory when it comes to application to certain languages, confusion in regards to nonbasic sentences, semantic misinterpretation of parental input, and so on.

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

Redington et al. focused on a distributional approach involving the hierarchical clustering of words into syntactic categories. Words in the same syntactic category have distributional regularities in common. Three stages were proposed in using distributional information to determine syntactic category. These stages include analysis of the contexts within which the target word occurs, comparing the contexts or distributions of pairs of words, and grouping words with similar distributions.

Stage 1 – context includes distribution of words near target word. Distributions were mapped with context vectors, which showed target word positions in relation to context word positions. Stage 2 – a context vector is a point in a space of possible distributions of contexts. Target words in the same syntactic category should have similar distributions. In terms of vectors, this would indicate vectors that lie near each other in a space. Vector similarity measures were taken. Similar context vectors (i.e., vectors that are near each other in space) represent a similar syntactic category. Stage 3 – hierarchical clustering was done when an algorithm combined items closest together, and then those groups could be clustered, etc. Clusters are formed at different scales and a dendrogram is formed. Nodes in the dendrogram represent clusters, with the clusters on the right being the most similar.

Samples from the CHILDES database (MacWhinney, 2000) were used in a series of experiments. Accuracy and completeness were measured based on comparing the benchmark syntactic categories with the derived groups. Different experiments were carried out to measure the effectiveness of distributional information involving position of context words in relation to target word, numbers of context and target words, effectiveness based on word class, corpus size, utterance boundaries, frequency versus occurrence information, removing function words, prior knowledge of other categories, and child-directed versus adult-adult speech. The results of these experiments indicate that distributional information may contribute to early acquisition of syntactic categories.

St. Clair, M. C., Monaghan, P., & Christiansen, M. H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition, 116,* 341-360. doi: 10.1016/j.cognition.2010.05.012

St. Clair et al. described a new distributional approach, based on the frequent frames study by Mintz (2003). While Mintz used a trigram approach, looking at three words in a sequence, and focusing only on the rigid structure of the frequent frame, the present study focused on an extension of the frequent frame model, called a flexible frame. The frequent frame model provides great accuracy, but limited coverage (only the 45 most frequent frames). Other models provide good coverage, but produce limited accuracy. St. Clair et al. attempted to provide a model that makes use of the trigram method based on bigram distributional information. In their flexible frames model, the preceding and succeeding word provide information about word category. A visual example of the different frame models follows: fixed or frequent frames (aXb), bigram frames (Xb) and (aX), and flexible frames (aX + Xb). The flexible frame model produced the highest accuracy, and the frequent frame or fixed frame model produced the lowest accuracy. Flexible frames are able to incorporate the advantages of both bigram models and trigram models in demonstrating a method of grammatical category acquisition.

Stumper, B., Bannard, C., Lieven, E., & Tomasello, M. (2011). "Frequent frames" in German child-directed speech: A limited cue to grammatical categories. *Cognitive Science, 35*, 1190-1205. doi: 10.1111/j.1551-6709.2011.01187.x

Stumper et al. based their study of grammatical category acquisition with German child-directed speech on the study done by Mintz (2003). The study by Mintz defined a frequent frame as the two words surrounding a target word. Mintz found considerable success using frequent frames to identify word category. His study was repeated in French (Chemla, Mintz, Bernal, and Christophe, 2009) and Dutch (Erkelens, 2009). The purpose of the present study by Stumper et al. was to determine the role of frequent frames in grammatical category acquisition in the German language. The authors did not find high levels of accuracy in comparison to frequent frames studies done in English and French. This may be due to the less restricted word order of German and the more frequent use of certain lexical forms. The study did yield results that were more accurate than what was found with random trials. There was a significant difference between token accuracy scores and type accuracy scores. The authors pointed out one limitation of the study being the use of only one child-directed speech corpus, rather than multiple corpora, which would have provided a greater variety of linguistic information from which to pull.

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*, 127-152.

The purpose of the study by Theakston et al. was to replicate a study completed by Valian (1999) in the area of verb-argument structure acquisition. Valian's study claimed that children are able to produce intransitive frames earlier than transitive frames due to their lack of direct object argument. Valian supported the idea of an abstract knowledge of grammar and performance limitations keeping children from producing adult-like grammatical utterances. The present study aimed to test Valian's claims through replication of her study. The authors used nine children whose mean length of utterances matched those of Valian's subjects and who participated in a longitudinal study in which language samples were collected on a regular basis for one year. The resulting corpora were transcribed and verb usage was recorded with verbs being categorized according to transitive, intransitive, and mixed verbs. The present study confirmed Valian's findings that children use more direct object arguments with mixed verbs (whose argument is optional) in later developmental stages and also proportionally use more transitive verbs compared to early developmental stages. However, the authors found that children are more likely to use those verb structures that are found more frequently in adult input, rather than rely more completely on performance limitations for production.

Weisleder, A., & Waxman, S. R. (2010). What's in the input? Frequent frames in child-directed speech offer distributional cues to grammatical categories in Spanish and English. *Journal of Child Language, 37,* 1089-1108. doi: 10.1017/S030500090999006

Recent studies have supported the notion that children can acquire grammatical categories based on distributional patterns. The frequent frames approach is a distributional approach focusing specifically on the two words that surround or frame an intervening word. To this point, studies have been conducted in this area that focus on the distributional patterns of English. However, there is a lack of information on distributional information found in other languages. Grammatical forms and categories are marked differently in different languages and an identifying pattern that is central to one language may not be central for another language. This study focused on the distributional information available to children learning Spanish as compared to those learning English and analyzed the accuracy by which frequent frames describe grammatical categories. This study also expanded on the Mintz (2003) notion of frequent frames by including in the analysis sequences in which one of the frames is the end of the utterance (referred to as end-frames). This was done due to the prosodic nature of the ends of utterances, which may be more readily acquired by young children. Six corpora were used from the CHILDES database (MacWhinney, 2000), three Spanish, three English, and the adult utterances were analyzed. The English samples were the same used by Mintz. Results indicated that frequent frames may contain distributional information that is helpful for young children learning the grammatical categories of language. Accuracy scores were higher for the English samples than the Spanish, and although such things as homophony and noun drop in the Spanish language may create obstacles to the use of distributional evidence, the results were still supported.