2016-02-01

A Model of Children's Acquisition of Grammatical Word Categories from Adult Language Input Using an Adaption and Selection Algorithm

Emily Marie Berardi
Brigham Young University

A Model of Children's Acquisition of Grammatical Word Categories from Adult Language Input Using an Adaption and Selection Algorithm

Emily Marie Berardi

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
Martin Fujiki

Department of Communication Disorders

Brigham Young University

February 2016

ABSTRACT

A Model of Children's Acquisition of Grammatical Word Categories from
Adult Language Input Using an Adaption and Selection Algorithm

Emily Marie Berardi
Department of Communication Disorders, BYU
Master of Science

Previous models of language acquisition have had partial success describing the processes that children use to acquire knowledge of the grammatical categories of their native language. The present study used a computer model based on the evolutionary principles of adaptation and selection to gain further insight into children's acquisition of grammatical categories. Transcribed language samples of eight parents or caregivers each conversing with their own child served as the input corpora for the model. The model was tested on each child's language corpus three times: two fixed mutation rates as well as a progressively decreasing mutation rate, which allowed less adaptation over time, were examined. The output data were evaluated by measuring the computer model's ability to correctly identify the grammatical categories in 500 utterances from the language corpus of each child. The model's performance ranged between 78 and 88 percent correct; the highest performance overall was found for a corpus using the progressively decreasing mutation rate, but overall no clear pattern relative to mutation rate was found.

Keywords: grammatical word categories, evolutionary programming, language acquisition

ACKNOWLEDGMENTS

I thank Dr. Channell for his guidance and patience, my committee for their feedback, and my family for their support. I also thank those who participated in the collection of the language samples used in this study as well as the graduate students who formatted them for use in previous theses.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

DESCRIPTION OF THESIS STRUCTURE

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. This thesis is suitable for submission to a peer-reviewed journal in speech-language pathology. An annotated bibliography is presented in the Appendix.

Introduction

Children acquire the understanding and use of grammatical word categories (e.g., noun, verb, adjective) in their language by the end of the preschool years (Berko, 1958; Ratner & Menn, 2000). Researchers agree that children are not born with an innate vocabulary or understanding of grammatical word categories, but debate remains as to how children learn these categories. Theories of how children learn language in general can be applied to this question. Ambridge and Lieven (2011) outline two main theories: innate-centered and learned-centered. According to the innate-centered, or generativist theory, children are born with existing knowledge of grammatical word categories that they then learn to sort vocabulary into. Supporters of the learned-centered, or constructivist theory assert that children need to first form categories based on the speech input they hear and then label words with these categories.

In addition, many have added to the collective understanding of grammatical word category acquisition. MacNamara (1972) proposed that infants learn meaning before syntax but that vocabulary, syntax, and phonology are dependent upon each other and cannot be separated into a neat sequential order. Jusczyk and Aslin (1995) showed that 7.5-month-old infants are able to detect words in fluent speech. This is an important foundation for grammatical word category acquisition because children must first be able to recognize words within the context of a sentence. Shady and Gerken (1999) showed that in order to comprehend speech input, children use multiple sources such as grammatical and caregiver cues. This gives another clue as to how children recognize words within speech input.

All researchers agree that children must have a process for learning and sorting grammatical word categories. A strong contributing theory regarding this process is bootstrapping, which is the ability to solve other problems in an area given partial knowledge in that area. Bootstrapping can be semantic, prosodic, or syntactic. Semantic bootstrapping refers to using the learned meaning of words to infer the grammatical role of that word in sentences (Pinker, 1987). If using prosodic bootstrapping, one uses cues of intonation in spoken language to infer the structure of the sentence (Morgan & Demuth, 1996). With syntactic bootstrapping, infants must use distributional information to infer which grammatical categories words belong to. As semantic and prosodic bootstrapping have yet to be modeled, syntactic bootstrapping is the only type that has been modeled with computer programs. Although some models align more with the generativist theory and some with the constructivist theory, each addresses the question of how children sort words into grammatical word categories.

Cartwright and Brent (1997) showed that a computer model could be a valid approach for studying theories of grammatical word category acquisition. Their model used distributional analysis, and their results showed that this information was useful for merging groups and that accuracy increased with the addition of semantic information. However, the model had a limitation. It could optimize accuracy or completeness, but not both at the same time.

Redington, Chater, and Finch (1998) used a program based on distributional analysis and showed that this information could be used by children learning grammatical word categories. They used corpora from the Child Language Data Exchange System (CHILDES; MacWhinney, 2000) and found that the nearer the context was to the target word, the more information it provided about the word's grammatical word category. This is consistent with other evidence that children and adults are sensitive to local contexts.

Mintz (2002) provided evidence that the brain has ready access to distributional analysis. This suggests that it could be used by children learning grammatical word categories. Adult participants in his study listened to a short sample of an artificial language and then identified phrases that they believed they had heard during the sample. The participants incorrectly recognized category-conforming phrases, suggesting their use of distributional analysis when exposed to a new language.

Mintz (2003) also presented an approach using frequent frames and corpora from the CHILDES database in a computer model. A frequent frame is a pair of words that are commonly found together with one word in between (e.g., it_the). The mean number of word types categorized in the corpora was 440, which is more than twice as many as previous models had categorized. The mean accuracy for standard labeling was 98%, and the mean accuracy for expanded labeling, which includes more specific categorization, was 91%. This study supported the use of distributional analysis to acquire grammatical word categories but included a limitation that each word was only allowed one grammatical word category. Also, this model did not categorize many words because only a small percentage of words in the language fall into the frequent frames.

To offer an alternative that may address the limitations found in bootstrapping models, the present study uses a computer model based on principles of biological evolution. Dennett (1995) argued that Darwin's theory of evolution was so successful because it is essentially an algorithm. Dennett defined an algorithm as something that produces a predictable result when run. Principles of evolutionary biology have previously been applied to computer models and artificial intelligence. They have been used to approach complex problems such as teaching a computer to play checkers (Fogel, 2002) and reading mammograms (Fogel, Watson, Boughton, & Porto, 1998).

Principles of adaptation and selection have also been used to address the question of how children learn grammatical word categories. Channell, Nissen, and Tanner (2014) investigated this application of evolutionary programming. Using the same six corpora from the CHILDES database as Mintz (2003), a computer program created random tags for each word in the sample and evaluated them using adaptation and selection. The most accurate dictionaries were allowed to create new dictionaries, simulating the process of adaptation and selection. The program showed promise, as accuracy increased dramatically through the first 1500 cycles across all corpora. Accuracy averaged 90.65% over 3,000 cycles.

Using a similar computer program, Cluff (2014) evaluated the effectiveness of allowing the assignment of multiple categories to words. Odd numbered utterances were used as input and evaluated using even numbered utterances from the same corpora. Up to three categories were allowed per word, and multiple mutation rates were evaluated over 4,000 cycles. The 1/1200 mutation rate produced the greatest accuracy, and accuracy averaged 92.74%. Stenquist (2015) expanded upon Cluff's findings using adult utterances as input and child utterances for evaluation. She found the 1/2400 mutation rate to produce the greatest accuracy. The program ran for 12,000 evolutionary cycles and accuracy peaked around 80% across the corpora.

These previous studies had limitations. Cluff's (2014) study used adult utterances for both the input and the evaluation, which is not as strong as using child utterances for evaluation. Stenquist's (2015) study only allowed one grammatical word category per word of the sample. The present study addresses the limitations of previous studies. The present computer model allows three possible grammatical word categories for each word but holds the overall number at a specified level. In addition, the mutation rate progressively decreases across the evolutionary cycles. Adult utterances are used as input, and child utterances are used for evaluation.

Method

Participant Samples

Spoken language samples were taken from the CHILDES database (MacWhinney, 2000) and divided into two corpora: adult utterances spoken to the child, and utterances spoken by the child. The samples were collected as audio recordings of adult-child interactions on multiple occasions over several interactions. The adult utterances were used as training utterances. The last 500 non-repeated multiword utterances of each child corpus were used as the test utterances, but other variables differed among corpora such as the child's age, the elicitation technique, and the original purpose of the language sample.

Adam. Adam (Brown, 1973) lived in a middle-class family with parents that were described as well educated and employed as an elementary school teacher and a minister. The language samples were taken from age 2;3 to 4;10 (years;months) in Adam's home with his parents and a few other adults. The total number of samples taken was 55. These samples were taken as part of a study to investigate psychological and linguistic aspects of preschool-age language development. Of the utterances spoken by adults, 19,301 were used as training utterances.

Anne. Anne (Sawyer, 1997) was age 3;5 when language samples were taken in her preschool class during unstructured play time. Unfortunately, socioeconomic status data and parental employment data on Anne's family were not available. The samples were taken as part of a study examining how the interactional dynamics of unstructured children's play affects development of conversational skills, social skills, and creativity. Of the utterances spoken by adults, 25,551 were used as training utterances.

Aran. Aran (Theakston, Lieven, Pine, & Rowland, 2001) was the oldest child in his middle class family. The interactions used for the samples took place in his home twice every three weeks over one year. These samples were taken as part of a study of children's acquisition of verb-argument structure. Of the utterances spoken by adults, 20,192 were used as training utterances.

David. David (Henry, 1995) was age 2;0 to 4;2 while these informal samples were taken in his family's or the investigator's home once a month for approximately one hour. He was the oldest child in his family. These samples were taken as part of a study looking at how children acquiring English in Belfast, Northern Ireland, use variable subject-verb agreement, which is common in Belfast English. Of the utterances spoken by adults, 9,933 were used as training utterances.

Naomi. Naomi's (Sachs, 1983) language samples were taken in her home during parent-child interactions from age 1;1 to 5;1. Her mother was the author of the study these samples were taken for. This study was about the acquisition of displaced reference, or talking about subjects outside of the current time and setting. Of the 93 samples taken, 12,034 adult utterances were used as training utterances.

Nina. Nina's (Suppes, 1974) samples were taken in 1972 and 1973. Nina was age 1;11 to 3;3 during this time. These samples were taken for a study of the semantics in children's speech. Of the utterances spoken by adults, 35,381 were used as training utterances.

Peter. The oldest child of his middle to upper-middle class family, Peter lived in a university community in New York City (Bloom, Hood, & Lightbown, 1974; Bloom, Lightbown & Hood, 1975). His 20 language samples were gathered from play interactions with multiple adults from when he was age 1;9 to 3;2. These samples were taken as part of a study examining the role of imitation in the acquisition of a first language. There were 20,827 adult utterances used.

Sarah. Sarah, (Brown, 1973) a child of a working class family, was age 2;3 to 5;1 during the time her language samples were taken. These samples were taken as part of the same study as Adam's were, to investigate psychological and linguistic aspects of preschool-age language development. Of the utterances spoken by adults, 48,205 were used as training utterances.

Computer Model

The computer model used was a program called EB_EV (Version 1.0; Channell, 2015), which was based on the NS-EV program used by Stenquist (2015). Like NS-EV, EB_EV uses principles of adaptation and selection to evolve sets of grammatical tags (e.g., noun, adjective, verb) for words. This mimics a "guess and check" procedure that infants may use when learning grammatical word categories and assigning vocabulary words to these categories. A dictionary is defined as the words of a corpus and their randomly assigned tags.

At the program's start, 500 dictionaries are populated such that each word is assigned a randomly chosen grammatical word category. The program then runs for a selected number of cycles. During each cycle, a section of the training corpus (5,000 utterances) is coded by each dictionary. The tags are compared with previously assigned tags of the same utterances. The 20 most accurate dictionaries as compared to the grammatical categories of the words in the training text are then allowed to produce new or "child" dictionaries. These child dictionaries are identical to the previous dictionaries except for random differences due to a selected mutation rate. This mutation causes random reassignment of grammatical tags with no relation to whether or not the tag was correct. The child dictionaries then compete by coding the training corpus and the process is repeated for the selected number of cycles.

At 100-cycle intervals, the best dictionary codes the test corpus of the child’s utterances. This tagging is compared to the tagging of the same utterances, and the percent of accuracy is written to a file. This information does not affect the evolving dictionaries.

Procedures

Each utterance was separated so that each line contained one independent utterance. Punctuation, coding, and notations were removed. All words were in lowercase except for the pronoun I and proper nouns. Words in the corpora were tagged automatically and manually checked for accuracy prior to their introduction to the computer model. The adult training corpora and the child test corpora were then processed by the computer model. The program was set to allow up to three tags per word but to keep the average number of tags per word at a specified level corresponding with that of the training corpus’s manual tags. The mutation rate was programmed to change across the evolutionary cycles from 1/800 to 1/2400 to mimic how a child’s categorization of words may be more flexible at first and become more rigid over time. For comparison, the model was also run at a 1/800 mutation rate and a 1/2400 mutation rate. After each 100 cycles, the computer program reported the agreement between the best dictionary's coding of the child corpus and the manually checked coding.

The model mimics the procedure the children may have used while learning grammatical word categories. The computer model uses the adaptation and selection algorithm described above to learn the grammatical categories of words in the parent or caregiver corpora, and the accuracy of this learning is assessed by examining how well it identifies the grammatical categories in the children's corpora.

Results

The model was run for each of the eight corpora three times: at a 1/800 mutation rate, at a 1/2400 mutation rate, and at a progressively decreasing mutation rate that started at 1/800 and ended at 1/2400. Table 1 represents the evolutionary cycle at which the dictionaries for each corpus reached the targeted average number tags per word, which had been set based on the actual tags for the corpus. It can be seen in Table 1 that one child's corpus (Sarah) never reached the targeted average tags per word when run at the 1/800 mutation rate. In general, the model tended to reach the average tags per word target earlier when run on the progressively decreasing mutation rate and the 1/2400 mutation rate than when run on the 1/800 mutation rate.

One of the corpora (Sarah) never reached the set average tags per word when run at the 1/800 mutation rate. In general, the model tended to reach the average tags per word value earlier when run on the progressively decreasing mutation rate and the 1/2400 mutation rate than when run on the 1/800 mutation rate.

Table 1

Cycle At Which Each Program Reached The Set Average Tags Per Word

Child	1/800	1/2400	Progressively Decreasing
Adam	1900	2100	1900
Anne	1600	1900	1600
Aran	1800	2000	1800
David	1900	2100	1800
Naomi	1700	2000	1700
Nina	2200	2300	2000
Peter	1500	1900	3100
Sarah	LNA	3100	3200

Table 2 shows the set average number of tags for each sample and the Mean Length of Utterance (MLU) values of the children. The MLU values may be inflated and may not represent the child's actual MLU at the time of recording because the samples used did not include any one-word utterances and did not separate compound sentences into independent clauses. LNA: Level not achieved.

Table 2

MLU Values and Average Number of Tags

Child	MLU	Average Tags
Adam	5.16	1.15
Anne	4.10	1.13
Aran	4.32	1.12
David	4.52	1.13
Naomi	5.45	1.14
Nina	4.87	1.13
Peter	4.63	1.14
Sarah	4.72	1.14

Table 3 shows the peak accuracy values for each language sample at each mutation rate. Four children's samples reached peak accuracy with the 1/2400 mutation rate, three reached peak accuracy at the progressively decreasing mutation rate, and one reached peak accuracy at the 1/800 mutation rate.

Table 3

Peak Averages at Each Mutation Rate

Child	1/800	1/2400	Progressively Decreasing
Adam	81.69	83.01	82.54
Anne	87.76	85.32	88.25
Aran	78.54	84.46	80.67
David	79.22	79.57	82.47
Naomi	80.87	81.57	81.64
Nina	79.79	82.09	80.82
Peter	87.10	81.57	81.22
Sarah	79.65	81.22	80.29

The results for Adam's corpus are shown in Figure 1, which shows the growth curves for the program when run at each mutation rate. The 1/800 mutation rate was more accurate in the earlier cycles but reached a plateau sooner. It experienced a temporary dip in accuracy around cycle 8000. The 1/2400 mutation rate increased in accuracy slower but reached a plateau later and did not experience any dips in accuracy. The progressively decreasing mutation rate followed the accuracy of the 1/800 mutation rate for the early cycles and the 1/2400 for the later cycles. At the last cycle the most accurate mutation rate was 1/2400.

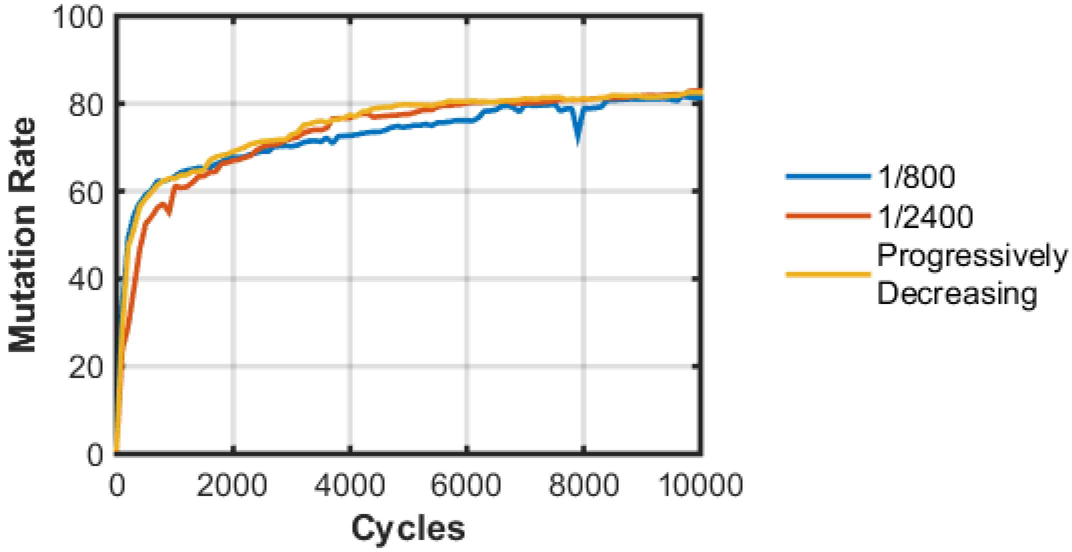

Figure 1. Results for Adam's corpus at all mutation rates.

Figure 2, which shows the growth curves for the program when run at each mutation rate, shows the results for Anne's corpus. The 1/800 mutation rate was the most accurate across all cycles until the end, when the progressively decreasing mutation rate surpassed it. The 1/2400 mutation rate was the least accurate across all cycles except for just before cycle 2000 when the progressively decreasing mutation rate temporarily decreased in accuracy.

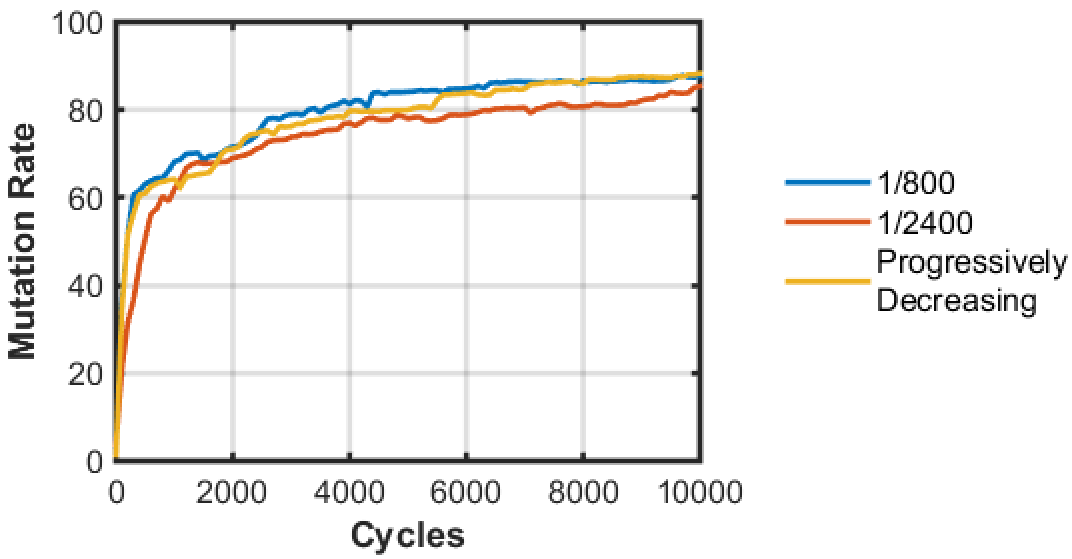

Figure 2. Results for Anne's corpus at all mutation rates.

The results for Aran's corpus are shown in Figure 3, which shows the growth curves for the program when run at each mutation rate. The 1/800 mutation rate was more accurate in the earlier cycles but reached a plateau sooner. The 1/2400 mutation rate increased in accuracy slower but reached a plateau later. The progressively decreasing mutation rate followed the accuracy of the 1/800 mutation rate for the early cycles and reached a final accuracy greater than that of 1/800 but less than that of 1/2400.

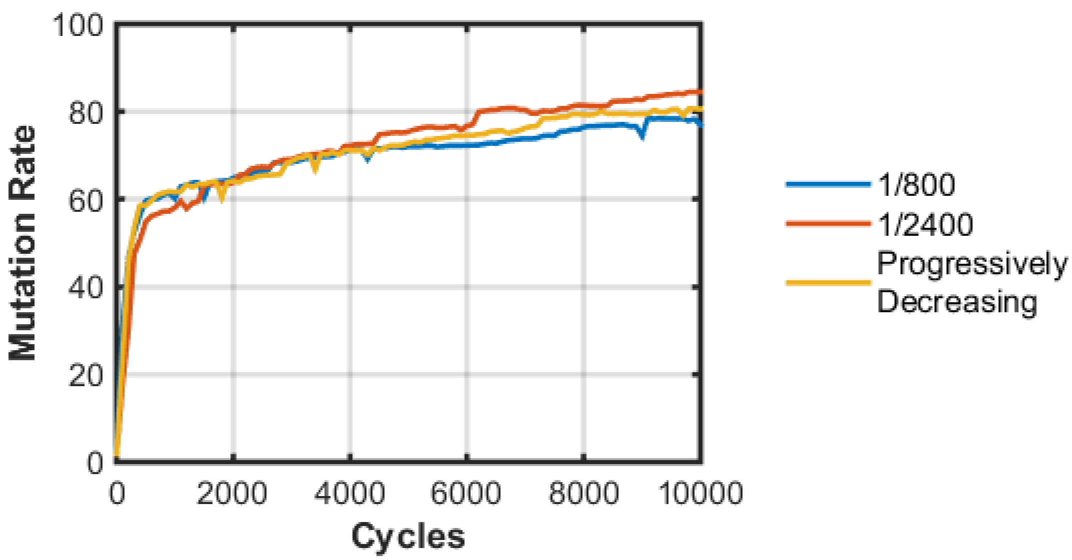

Figure 3. Results for Aran's corpus at all mutation rates.

Figure 4, which shows the growth curves for the program when run at each mutation rate, shows the results for David's corpus. The 1/800 mutation rate was more accurate in the earlier cycles but reached a plateau sooner. It experienced temporary decreases in accuracy before cycle 6000 and around cycle 7000. The 1/2400 mutation rate increased in accuracy slower but reached a plateau later. However, both ended with a lower accuracy than that of the progressively decreasing mutation rate, which surpassed the others by cycle 4000.

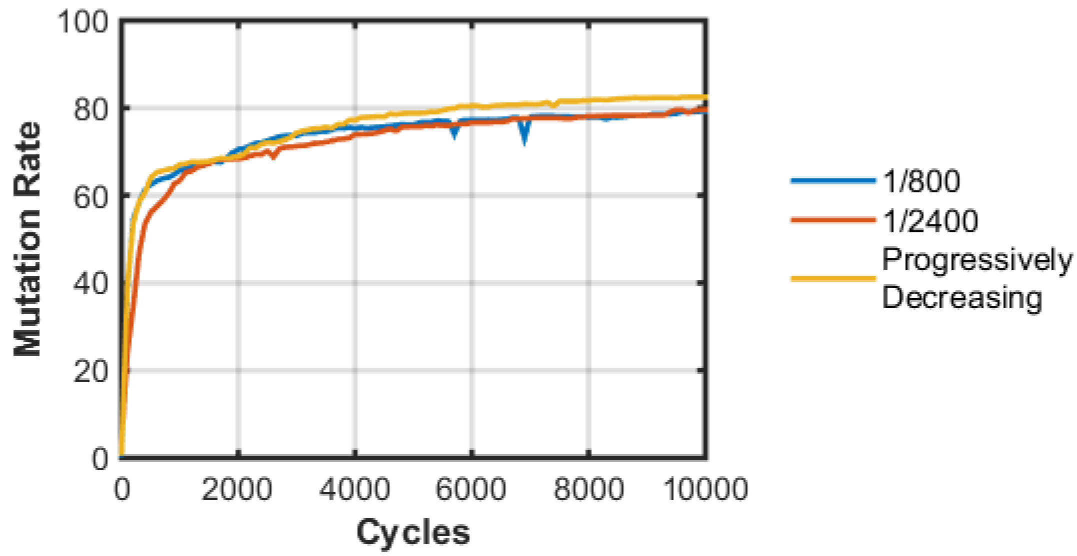

Figure 4. Results for David's corpus at all mutation rates.

The results for Naomi's corpus are shown in Figure 5, which shows the growth curves for the program when run at each mutation rate. All three mutation rates had similar growth. The 1/800 mutation rate plateaued before the others, and the progressively decreasing mutation rate showed the highest accuracy at the last cycle.

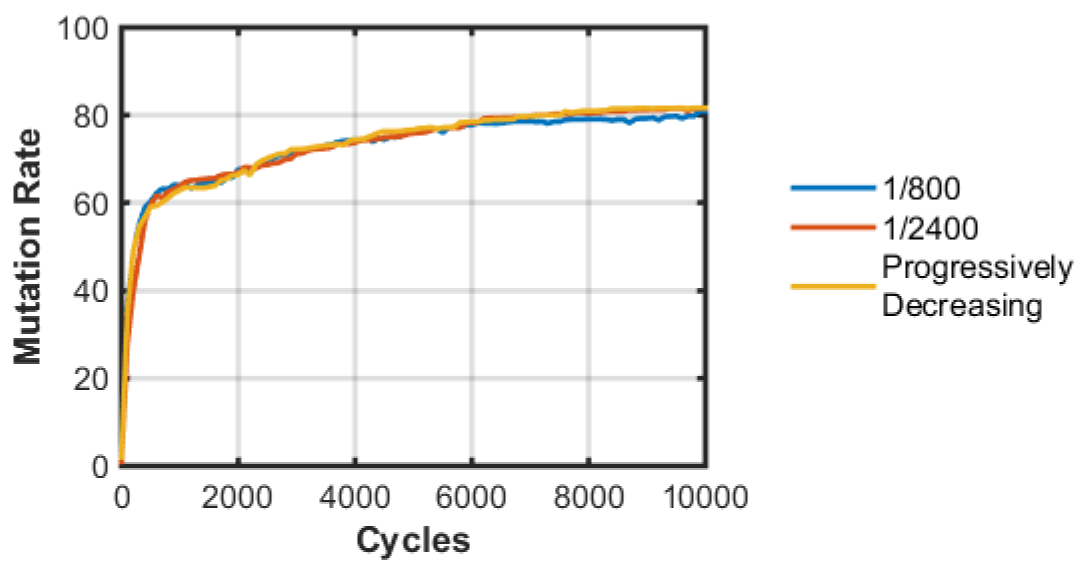

Figure 5. Results for Naomi's corpus at all mutation rates.

Figure 6, which shows the growth curves for the program when run at each mutation rate, shows the results for Nina's corpus. The 1/800 mutation rate was more accurate than the 1/2400 mutation rate in the earlier cycles but reached a plateau sooner. The 1/2400 mutation rate increased in accuracy slower but reached a plateau later. The progressively decreasing mutation rate followed the accuracy of the 1/800 mutation rate for the early cycles and the 1/2400 for the later cycles. At the last cycle the most accurate mutation rate was 1/2400.

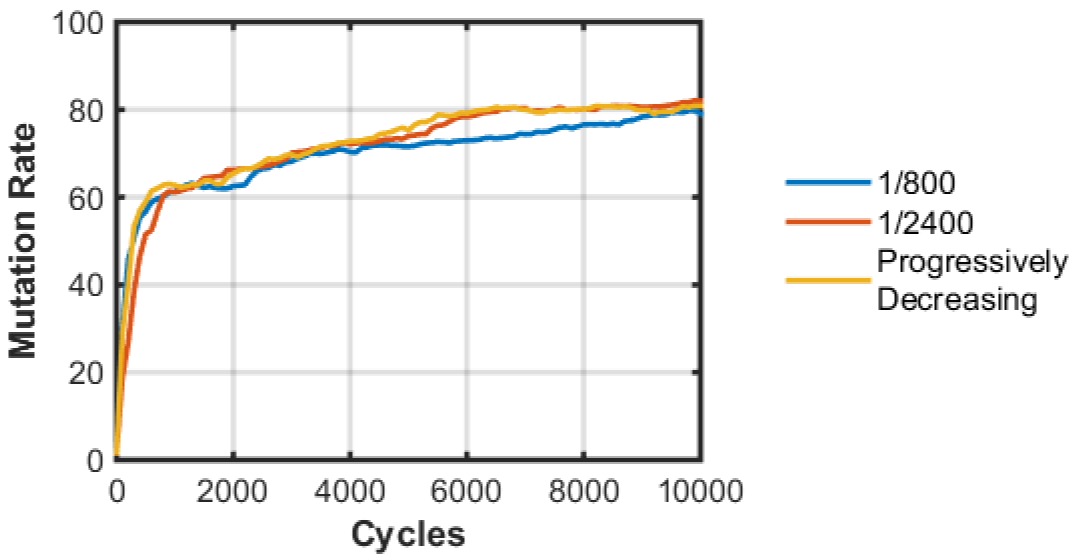

Figure 6. Results for Nina's corpus at all mutation rates.

The results for Peter's corpus are shown in Figure 7, which shows the growth curves for the program when run at each mutation rate. The 1/800 mutation rate was more accurate in the earlier cycles and was the most accurate at the last cycle. The program experienced temporary decreases in accuracy around cycles 4000 and 8000. The 1/2400 mutation rate increased in accuracy slower and ended with the lowest but did not experience any decreases in accuracy. The progressively decreasing mutation rate follows the 1/800 mutation rate for the early cycles and ends with an accuracy higher than the 1/2400 mutation rate but lower than the 1/800 mutation rate.

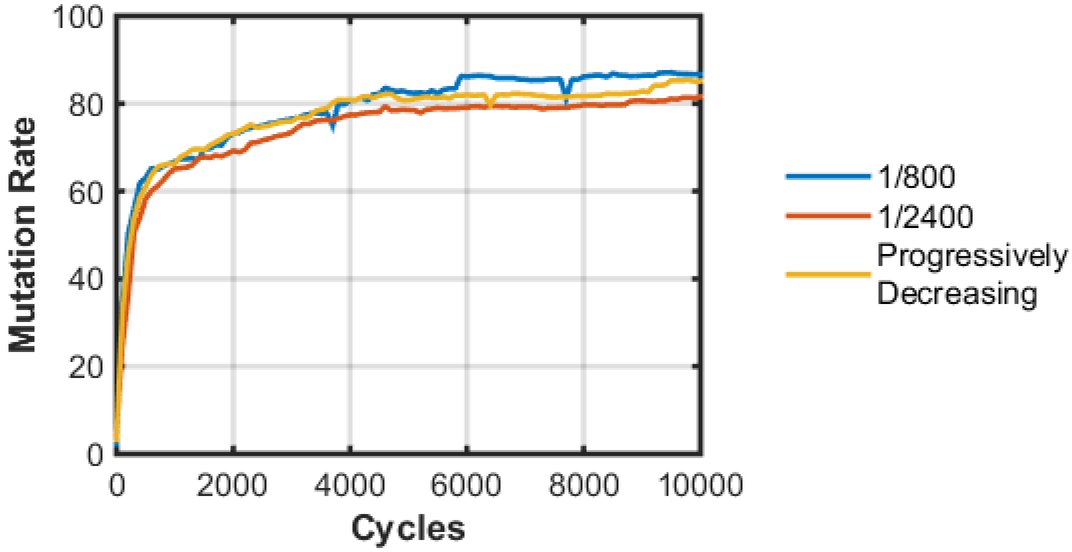

Figure 7. Results for Peter's corpus at all mutation rates.

Figure 8, which shows the growth curves for the program when run at each mutation rate, shows the results for Sarah's corpus. The 1/800 mutation rate was more accurate in the earlier cycles but reached a plateau sooner. The 1/2400 mutation rate increased in accuracy slower but reached a plateau later. The progressively decreasing mutation rate follows the 1/800 mutation rate in the early cycles but does not reach as high of a plateau. At the last cycle the most accurate mutation rate 1/2400.

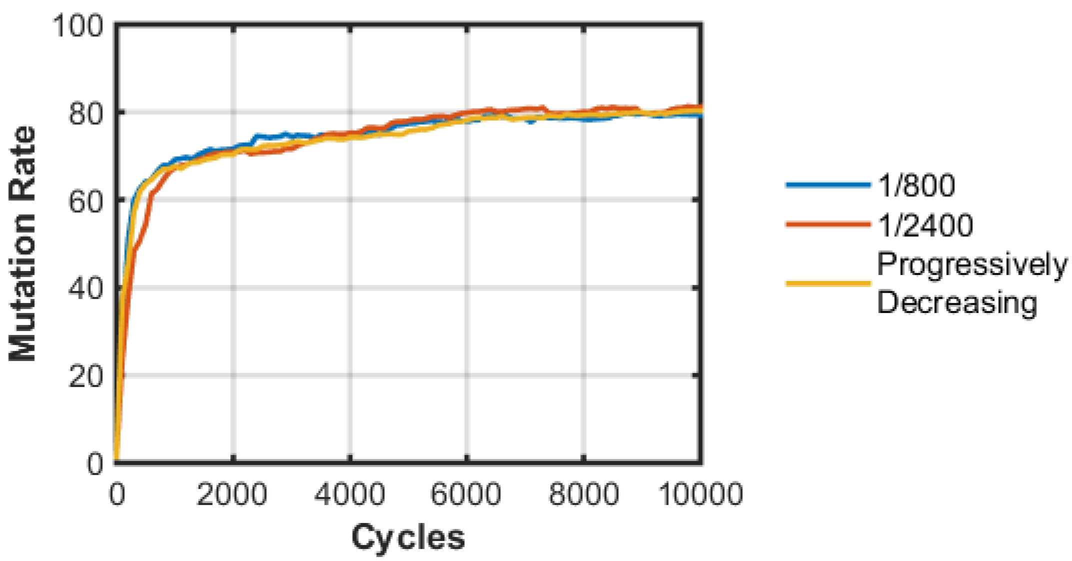

Figure 8. Results for Sarah's corpus at all mutation rates.

Figure 9, which shows the growth curves for the program when run at each mutation rate, shows the average results for all corpora. On average, the 1/800 mutation rate increased in accuracy sooner but plateaued earlier. The 1/2400 mutation rate increased in accuracy later but reached a higher final accuracy by plateauing earlier. The progressively decreasing mutation rate followed the 1/800 mutation rate for the early cycles and the 1/2400 mutation rate for the later cycles, ending with the highest accuracy at the last cycle.

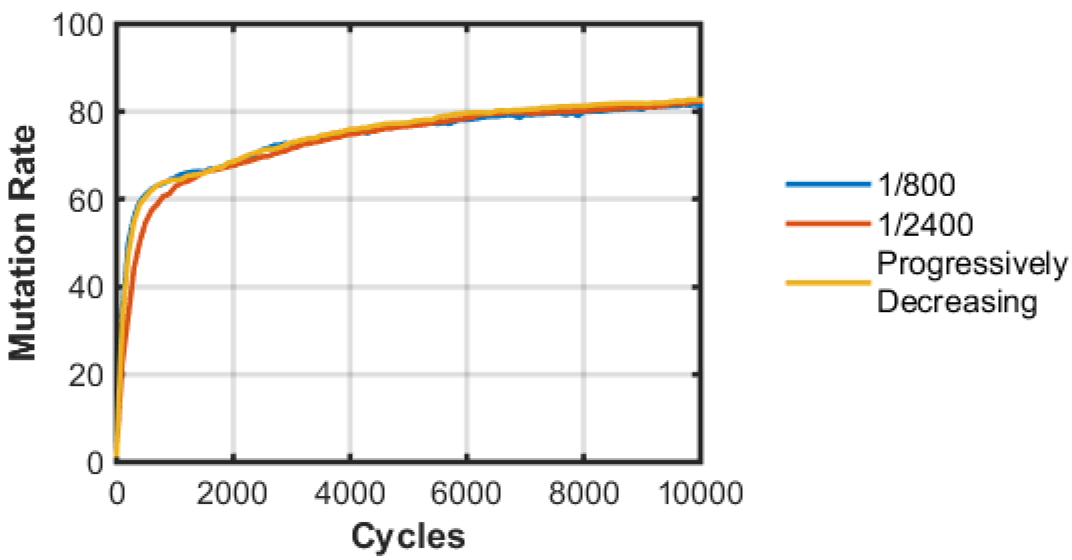

Figure 9. Average results for all corpora at all mutation rates.

Discussion

The purpose of this study was to evaluate accuracy with which the EB_EV computer model could learn grammatical categories from adult language input and use this learning to label the grammatical categories in the child corpora. EB_EV differed from a previous version (NS_EV) in two ways: it allowed each word to have multiple grammatical category options, and it allowed the use of a progressively decreasing mutation rate as well as the fixed mutation rates used in NS_EV. The effects of these two additional features on the accuracy of grammatical category coding using the model were analyzed. When run with the progressively decreasing and

1/2400 mutation rates, the model was able to reach the targeted average tags per word value sooner. It is unknown why one of the corpora (Sarah) never reached the targeted average tag per word when run at the 1/800 mutation rate. The majority of corpora reached their peak accuracy at the progressively decreasing or 1/2400 mutation rate, with only one corpus finding the most accuracy with the 1/800 mutation rate. Although the progressively decreasing mutation rate was designed to mimic how a child's word categorization may become less flexible over time, it is unclear whether this addition was a significant improvement.

It is unknown why there were differences between children or between the mutation rates. The 1/800 mutation rate showed a few dips in accuracy, possibly because accurate tags were randomly replaced with inaccurate tags. On average the 1/800 mutation rate tended to have higher accuracy early in the program, but the 1/2400 continued to improve longer. Overall the progressively decreasing mutation rate was the most accurate, but only by a small margin. One corpus (David) had significantly fewer child-directed utterances than the other corpora, and the progressively decreasing mutation rate was the most accurate mutation rate for that sample. This could suggest that the progressively decreasing mutation rate is better than the other mutation rates when utterances are limited. One corpus (Sarah) had significantly more utterances than the other corpora, and all mutation rates had comparable results. This suggests that the type of mutation rate used is less important when more utterances are available. Further studies could investigate this correlation.

This study's accuracy findings were similar to those of Stenquist (2015) but lower than those of Cluff (2014). Stenquist's model and the present study's model both used adult utterances as input and child utterances as evaluation. Cluff's model used adult utterances for both input and evaluation, but it and the present study's model both allowed for multiple tags per word. Both Stenquist and Cluff also found that a lower mutation rate leads to lower accuracy. The present study was the first to use a progressively decreasing mutation rate. Stenquist's model approached or exceeded 80% accuracy. The present study achieved slightly higher accuracy on average, probably due to the allowance of multiple tags per word. Cluff's model approached or exceeded 90% accuracy. This is likely because Cluff's study used even numbered utterances of the corpus for training and odd numbered utterances for evaluation and Stenquist's model and the current model both evaluate only a limited window of the corpus at a time. Cluff found a relationship between higher average tags per word and higher accuracy, but the present study included a larger sample size and found no such correlation.

The methodology of this study is different from three previous attempts to address the question of grammatical category acquisition, but it contributes to the same body of knowledge. Cartwright and Brent (1997) proved that a computer model could be a valuable tool in approaching this question. Like Mintz (2003), Redington et al. (1998) also used distributional analysis. Redington et al. found that the context of a word provides important information about grammatical category of that word. Mintz's model utilized frequent frames and reached accuracy above 90%. The present study is unique in its use of evolutionary programming and of other features that further attempt to mimic the process possibly used by young children. Its model uses adult utterances as input and child utterances as evaluation, it allows for multiple tags per word, and its mutation rate becomes less frequent over time.

A limitation of this and similar studies is the size of the training corpora available in public databases. An average child in a professional family would hear almost 45 million words by age four, and an average child in a family of a lower socioeconomic status level would hear almost 15 million (Hart & Risley, 1995). Therefore, the average child will be exposed to many more words than exist in the largest training corpus used in this study (Sarah), which consisted of 193,891 words. The limited size of available corpora may restrict the amount of learning that can be modeled from these data. Alternate sources of realistic data for simulation modeling should be investigated.

Further research is warranted in a few aspects of this study. Although the sample size of the current study is larger than those of previous studies, it is still not representative of the general population, as many of the children came from similar backgrounds. This is a limitation of what longitudinal language sample data are available in databases. Also, the optimal mutation rate is still unclear, as the rate that achieved the highest accuracy varied by sample. Further parameters have yet to be fully explored.

However, this study did contribute to the question of how children acquire grammatical word categories. It showed that on average a progressively decreasing mutation rate does produce the highest level of accuracy and that allowing for multiple grammatical tags per word increases the accuracy of the computer model. Also, it suggested that a progressively decreasing mutation rate could be significantly more accurate when the number of utterances in a corpus is limited. These findings show promise in the use of adaptation and selection algorithms in contributing to understanding of language acquisition and disorders.

References

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

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

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*, 120-170.

Channell, R. W. (2015). EB_EV [Computer software]. Provo, UT: Brigham Young University.

Channell, R., Nissen, S., & Tanner, K. (2014). *Modeling children's acquisition of grammatical word categories using an adaptation and selection algorithm*. Paper presented at the meeting of the American Speech-Language Hearing Association, Orlando, FL.

Cluff, S. Z. (2014). *A model of grammatical category acquisition using adaptation and selection*. (Master's thesis). Retrieved from http://scholarsarchive.byu.edu/etd/4086/

Dennett, D. C. (1995). Darwin's dangerous idea: Evolution and the meanings of life, 48-60. New York, NY: Simon & Schuster.

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

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*, 317-326.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes Publishing.

Henry, A. (1995) *Belfast English and Standard English: Dialect variation and parameter setting*. New York, NY: Oxford University Press.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology, 29*, 1-23.

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

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

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

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

Morgan, J. L., & Demuth, K. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.) *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ratner, N. B., & Menn, L. (2000). In the beginning was the wug: Forty years of language elicitation studies. In L. Menn & N. B. Ratner (Eds.) *Methods for studying language production.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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,* 1-28*. Children's language*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Sawyer, K. (1997). *Pretend play as improvisation*. Mahwah, NJ: Lawrence Erlbaum Associates.

Shady, M., & Gerken, L. (1999). Grammatical and caregiver cues in early sentence comprehension. *Journal of Child Language, 26*, 163-176.

Stenquist, N. A. (2015). *Modeling children's acquisition of grammatical word categories from adult input using an adaptation and selection algorithm.* (Unpublished Master's thesis) Brigham Young University, Provo, UT.

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.

Appendix: Annotated Bibliography

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

The introduction of this book describes various theories of child language acquisition. These include innate-centered approaches (e.g., nativist, generativist, universal grammar) and learned-centered approaches (e.g., constructivist, emergentist, socio-pragmatic, functionalist, usage-based). Chapter six discusses how children learning word-order languages learn the rules of syntax, including lexical categories. Previous findings of various theories are presented. These include semantic bootstrapping, prosodic bootstrapping, distributional analysis (i.e., frequent frames, chunking, merging templates), and the use of phonological cues. Limitations are also presented, suggesting that more understanding is needed in this area. The present study will contribute to that goal.

Cartwright, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition, 63*, 120-170.

After reviewing previous relevant studies, Cartwright and Brent propose a theory of how children use distributional regularity to learn syntax. They assert that this theory is better than previous models because, among other advantages, it processes sentences individually instead of collectively, allows more than one category per word, and uses other sources of information such as semantics. The authors suggest that children use sentence templates based on generalized minimal pairs and merge these templates based on certain preferences. Five experiments were conducted to test this idea. The first experiment tested whether this strategy could effectively learn an artificial language, and

the second added ambiguous words to the lexicon to test the aspect of the theory of grouping words into more than one category. The other experiments used child-directed speech, larger input files, and the addition of semantic information. The accuracy and completeness scores showed that the distributional input was effective for merging groups and that results were improved when semantic information was added. These results show that the proposed distributional model was effective at categorizing words, suggesting that computational models are a valid approach for studying theories of syntactic category acquisition, which is the premise of the current study.

Channell, R., Nissen, S., & Tanner, K. (2014). Modeling children's acquisition of grammatical word categories using an adaptation and selection algorithm. Paper presented at the meeting of the American Speech-Language Hearing Association, Orlando, FL.

The purpose of this study was to investigate the use of evolutionary programming in explaining how children learn grammatical categories of words. The same six language sample corpora used by Mintz from the CHILDES database were used. The input to the computer model was the utterances from the adults. The program tags each word in the sample, and the tags are evaluated. Each word is randomly assigned a grammatical tag, and dictionaries are produced that contain lists of the tags used. Next an evaluation process uses adaptation and selection cycles according to the number of generations selected. There is a mutation rate assigned which corresponds with how often a grammatical tag will be randomly replaced. The dictionaries with the highest number of accurate tags are used to create the following dictionaries. New dictionaries are then created and evaluated. All corpora showed major improvement in the first 1500 cycles, with all corpora showing over 85% accuracy. Between 1500 and 3000 cycles, improvement averaged only about 2%, and afterwards there was very little progress. These results show promise in using evolutionary programming to model grammatical category acquisition. The current study evaluates the modification of parameters of the computer model used in this study.

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(3)*, 296-406.

This study further investigates the findings of Mintz's 2003 study regarding the use of frequent frames and distributional analysis to model grammatical category acquisition. This follow-up study included three experiments. The first applied the use of frequent frames to French. Because French relies less on closed-class words, there is more potential for erroneous generalization. A French corpus from the CHILDES database was used. As in Mintz's previous study, scores of accuracy and completeness were computed. As more groups of syntactic categories were taken into account, accuracy remained high but completeness decreased. However, the results showed that this approach was valid for French, despite the presence of more ambiguous function words. The second experiment compared results using front contexts, which categorize following words, and back contexts, which categorize preceding words. The same French corpus was used alongside an English corpus. The authors found that front contexts yielded slightly better results in both French and English. The third experiment attempted to use recursive frames and asymptotic manipulation, but neither were found to improve results of frequent frame categorization. Overall, this study showed that the theory of frequent frames is viable across languages and that front contexts are more useful for distributional analysis.

Cluff, S. Z. (2014). *A model of grammatical category acquisition using adaptation and selection*. (Master's thesis). Retrieved from http://scholarsarchive.byu.edu/etd/4086/

The purpose of this study was to evaluate the effectiveness of using a computer model based on adaptation and selection that assigned multiple categories to words. This model is designed to represent the guesswork that children use when acquiring knowledge of grammatical categories. This particular thesis study evaluated the effect of various mutation rates on grammatical tagging accuracy. The five corpora used in this study were from the CHILDES database, and three were samples used by Mintz (2003). In the computer model, the user determines the number of evolutionary cycles to run, the average number of grammatical tags per word, and the chance of mutation, which is when a tag is randomly replaced. Up to three tags can be chosen for each word. The words are randomly tagged, then the adaptation and selection process is cycled for as many generations as the user chose. The tagging accuracy is evaluated by comparing the random tags in the even numbered utterances to the correct tags in the odd numbered utterances. In this study the program ran 4,000 evolutionary cycles. The mutation rate levels were 1/400, 1/800, 1/1200, and 1/1800. The average number of tags allowed per word corresponded to the average for each individual corpora. The 1/400 mutation rate produced the lowest accuracy. The accuracy improved dramatically over the first 500 cycles and continued to improve through the 4,000th cycle. The 1/1200 mutation rate produced the greatest accuracy over 4,000 cycles. The current study will also evaluate modifications to parameters of the same computer model.

Dennett, D. C. (1995). *Darwin's dangerous idea: Evolution and the meanings of life*, 48-60. New York, NY: Simon & Schuster.

In this chapter Dennett argues that Darwin's theory of evolution was so influential because it was based on the power of an algorithm. He defines an algorithm as something that produces a predictable result when run. Three features of an algorithm are substrate neutrality, underlying mindlessness, and guaranteed results. Substrate neutrality refers to the principle that the results of an algorithm are due to logical structure, not materials. Underlying mindlessness is defined as simple steps throughout the process. An algorithm also produces guaranteed results every time. According to Dennett, Darwin's theory of evolution is an algorithm. In the present study, principles of evolution are used in an algorithm applied to the question of how children learn syntactic categories.

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

In the introduction to this book, an evolutionary programming process used by a computer to learn to beat humans at checkers is described. Evolutionary programming mimics human cognition by finding the optimal solution to a problem. Using a criterion that is defined as success, it chooses one stimulus over another until it finds the one that best fits the criterion. Random variation is used in which the offspring (dictionaries, in the present study) are similar to the parents apart from a mutation. These mutations help the program to explore other stimuli that may better fit the criterion. The principle of survival of the fittest is applied by only allowing the best fitting offspring to reproduce new offspring. In the present study, only the most accurate dictionaries will produce new dictionaries.

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*, 317-326.

In this study, artificial neural networks (ANNs) were used to analyze radiographic features in mammograms. This was motivated by variability and inconsistencies in human analysis of mammograms. The ANNs were trained using evolutionary programming as they analyzed 216 suspicious mammograms. Their conclusions were either proven or disproven with surgical biopsy. Overall the ANNs were successful in diagnosing breast cancer. The results of this study indicate that artificial intelligence can be successful in solving complex problems. The computer program in the present study also uses evolutionary programming to solve a complex problem.

Freudenthal, D., Pine, J., & Gobet, F. (2005). On the resolution of ambiguities in the extraction of syntactic categories through chunking. *Cognitive Systems Research, 6*, 17-25. doi:10.1016/j.cogsys.2004.09.003

The purpose of this study was to investigate the result of using a chunking strategy with distributional analysis to improve accuracy and quality of syntactic category formation by a computer model called MOSAIC. This addresses the issue caused by some words belonging to multiple grammatical categories. Two corpora of adult speech, Anne and Becky, from the CHILDES database were run through the program with varying levels of chunking. These corpora included about 33,000 and 24,000 utterances. The chunking procedure works by treating frequent phrases as one unit. MOSAIC then produces novel utterances by linking phrases that occur in similar contexts. Two raters coded 500 utterances of each output file for the presence of syntactic errors. Nearly all of the errors involved the incorrect use of a word that belongs to more than one syntactic category, and the authors note that these are the errors likely not to be recognized in previous studies of distributional analysis. Error rates were lower for the models that used chunking. ($p<.001$ for Anne, $p<.05$ for Becky) The results demonstrate that chunking is a useful addition to distributional analysis in modeling syntactic category acquisition by restricting the contexts in which words may be incorrectly substituted. The current study also accounts for the fact that some words belong to multiple syntactic categories by allowing multiple tags per word.

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

The purpose of this study was to determine whether children memorize language or if they have an underlying understanding of morphological rules. The participants, children ages four to seven, were presented with nonsense words and asked to derive different forms of these words based on morphological rules (i.e., plural, verb tense, possessive, etc). The results showed that children of these ages have an understanding of morphological rules, can sort words into grammatical categories, and can apply this knowledge to new words.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology, 29*, 1-23.

The purpose of this study was to investigate whether infants can recognize familiar words in the context of a sentence. This study consisted of four experiments and used two groups of infants. This study utilized the infant behavior such as head turn to signify when the participants recognized words they were previously attuned to. In the first experiment, 7 ½ month old infants were familiarized with two monosyllabic words and then exposed to passages that included these words. The infants detected the familiar words, shown by their increased attention to the passages that included the target words. Experiment 2 was comparable but with 6 month old infants, who did not show any indication that they recognized the target words. Experiment 3 was similar but used nonwords that were similar to real words in the passages with the 7 ½ month old infants and found that familiarization with the similar nonwords did not increase their attention to the words in the passages. Experiment 4 involved familiarization in sentence contexts rather than in isolation and the infants were able to recognize the target words during the test phase. Overall, this study showed that 7 ½ year old infants are able to detect words in fluent speech, which is an important foundation for grammatical category acquisition because children must first be able to recognize words within the context of a sentence.

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

This study represents the first computer model of grammatical category acquisition. The language samples were taken from seven families with children between 1 and 3 years old. Words belonging to the same grammatical categories were found to have stronger connections when fed through the computer program. This study uses mechanisms that are neurologically plausible, analyzing frequency and distribution. The output of the program is grammatically appropriate and comparable to the grammatical categories of young children. Although this study uses a relatively small sample, it shows that the use of a computer model is valid in exploring this topic. The present study builds upon this idea.

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

In this article MacNamara discusses the nature of language acquisition, specifically of semantics, syntax, and phonology. He proposes that infants learn meaning before applying the meaning to work out the syntax of the language. He describes the strategies they use to do this and presents evidence from vocabulary, syntax, and phonology. MacNamara concludes by conceding that the acquisition of these elements of language are dependent upon each other and cannot be separated into a neat sequential order. Although an incomplete portrayal, this article adds to the area of language acquisition.

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

The purpose of this study was to investigate the role of distributional information in learning grammatical categories. This was done by introducing an artificial language to adults and evaluating whether they used distributional analysis based on a memory task performed afterwards. The participants included 49 undergraduate students. They listened to simple sentences in an artificial language for six minutes and were told that afterwards they would be tested on what they heard. They then listened to test sentences and were asked if they had heard them during the training phase. Included in the test sentences were control sentences, sentences repeated from the training phase, and novel category-conforming sentences. Participants' incorrect recognition of these novel sentences would suggest the use of distributional analysis. Participants gave "yes" responses more often to repeated sentences than category-conforming sentences, but more often to category-conforming sentences than to the control sentences. This suggests that the participants did

use distributional information to form categories. Because the participants were only exposed to the artificial language for a short time, Mintz concluded that distributional analysis is available for the brain to use readily. Although one limitation of this study is that the participants were adults, the results provide support for the theory that children could use distributional analysis when learning grammatical categories and also shows that further research is necessary in this area.

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

In this study Mintz introduces the idea of using frequent frames in a software model tagging grammatical categories. A frequent frame is a pair of words that are commonly found together with one word in between (e.g., it_the). Mintz's theory is that requiring such a context will reduce errors due to accidental context labeling. He proposes that this model could be valuable due to previous evidence that children and adults are both sensitive to frequent frames in language (e.g., verb acquisition based on verb argument structure). In both of Mintz's experiments, he used six corpora from the CHILDES database. In the distributional analysis procedure, a tally of all the frames was recorded and a list of the frequent frames was made. In Mintz's first experiment, he categorized words that appeared in the 45 most frequent frames of each sample. In the second experiment, the frame selection method correlated with the frequency of frames relative to the total number of frames. Accuracy was measured by the proportion of words that were correctly grouped together into categories. The mean number of word types categorized in the corpora was 440, which is more than twice as many as previous models categorized. The mean accuracy for standard labeling was 98 percent, and the mean

accuracy for expanded labeling, which includes more specific categorization, was 91 percent. Although multiple categories were formed for the same grammatical categories, Mintz's findings showed that frequent frames are very effective at categorizing words although only a small portion of each corpora was analyzed (6% in the first experiment, 5% in the second). He mentions that a theoretical problem with previous distributional models is that they don't account for the possibility that one word may belong to multiple categories. Although his own study shares this limitation, it lays a foundation for further research of distributional analysis used to determine grammatical categories. The current study addresses this limitation by allowing more than one grammatical tag per word.

Naigles, L., & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effects of frequency and structure on children's early verb use. *Journal of Child Language, 25*, 95-120.

The purpose of this study was to investigate whether the type of verb input that a child is exposed to will affect the ease at which a child learns a verb. The sample included 57 mothers and their children. An analysis of the children's verb acquisition revealed three variables that showed significant correlation with order of acquisition. These factors were total frequency, final position frequency, and diversity of syntactic environments. The conclusion that diversity of verb context affects the child's learning provides support for the idea of syntactic bootstrapping, that the surrounding context of a word will provide information about a word's meaning.

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

The purpose of this study was to evaluate whether distributional information plays a significant role in the acquisition of syntactic categories. The corpora used came from the CHILDES database, and they were not altered beyond the stripping of coding information, capitalization, and punctuation. Only adult speech was analyzed. The 1,000 most frequent words were used as target words, and the 150 most frequent words were used as context words. Each word was given only its most common syntactic category as determined by the Collins Cobuild lexical database. Results of the analysis were scored with three measurements: accuracy, completeness, and informativeness. Redington, Chater, and Finch found that the nearer the context was to the target word, the more information it provided about the word's syntactic category. This is consistent with evidence that children and adults are sensitive to local contexts. Preceding context is more useful than succeeding context, but the combination of both types of context was the most informative. They also found that performance was better with a larger number of target words, that distributional information is the most useful for nouns and verbs, that results were best for a corpus size of 100,000 words or more, that utterance boundary information is not necessary, that frequency information is important, that the removal of function words had a positive influence on results, that the grouping of one class affects the successful grouping of another, and that "motherese" does not necessarily influence learning. Redington, Chater, and Finch contributed knowledge about multiple factors influencing syntactic category acquisition. Their results show that distributional information of syntactic contexts is relevant to grammatical category acquisition.

Shady, M., & Gerken, L. (1999). Grammatical and caregiver cues in early sentence comprehension. *Journal of Child Language, 26*, 163-176.

This study's purpose was to further investigate the idea that children use cues from multiple sources to identify linguistic units in speech input. The participants included 60 children age 2;0 to 2;2. The first experiment examined the roles of prosody and grammar. A robot spoke to the child in grammatical sentences, agrammatical sentences, sentences with nonsense words, sentences with natural pauses, and sentences with unnatural pauses. The children then identified pictures based on what the robot said. Results showed that grammaticality was significant, that prosody was significant, but that there was not a significant interaction between the two. In the second experiment the authors investigated the effects of grammatical cues and of caregiver cues such as utterance length and placement of the key word in the utterance-final position. This experiment used the same procedure as the first. Similarly, the results showed that grammaticality was significant, that caregiver cues were significant, but that the two did not have a significant interaction. This study showed that children do use multiple sources as cues for comprehending speech input.

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.

The authors propose that children use flexible frames (trigrams) based on distributional patterns (bigrams) to learn grammatical categories. After reviewing previous research, they suggest that flexible frames (aX + Xb) are more advantageous than fixed frames (aXb). Six experiments were performed using computer models to evaluate the use of flexible frames and fixed frames. The results showed that trigram information may be useful for category acquisition and that accuracy increases when frames are flexible. This study is further evidence that children may use distributional information in grammatical category acquisition.

Stenquist, N. A. (2015). *Modeling children's acquisition of grammatical word categories from adult input using an adaptation and selection algorithm.* (Unpublished Master's thesis) Brigham Young University, Provo, UT.

This thesis study addresses the question of how children learn grammatical categories by using an adaptation and selection algorithm and input corpora of adult speech. The corpora originated from the CHILDES database. The last 500 non-repeated multiword utterances of the child corpora were used as test utterances. The program randomly populated 5000 dictionaries and ran for 12,000 evolutionary cycles. These dictionaries are compared to the automatic tagging and the 20 most accurate dictionaries reproduce new dictionaries at a set rate of mutation. The adult training corpora and the child corpora were then processed. At 500-cycle increments, the program's coding of the child corpora was compared with a manual coding. In all corpora, the greatest increase in accuracy occurred in generations 1 through 1500, continued to improve through 3000, and showed only gradual improvement through generation 12,000. The highest percent accuracy occurred at a mutation rate of 1/2400. This study shows the success and accuracy of a model that uses a training corpus of the adult speech and tests on the child speech. The current study uses a similar procedure and program but allows multiple categories per word.

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.

This study expanded the frequent frames theory in 2003 Mintz study to the German language. Previous studies shows that this theory was viable for French but that results were not as accurate for Dutch. The corpus used was the largest sample of child-directed speech in German. The procedures used in Mintz’s (2003) study were replicated. The results were similar to the results of the Dutch study, with low accuracy compared to English and French. The authors concluded that frames do carry some information of categories but that there is great variability between frames in German. This shows that the question of how children acquire syntactic categories remains fully answered.

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/S0305000909990067

This study built upon Mintz’s work by analyzing the accuracy of frequent frames in providing clues for grammatical category assessment in different languages, sentence positions, and grammatical categories. Three English language samples and three Spanish samples were used. In addition to Mintz’s frequent frames, this model also used end frames, which use the context at the end of the sentence. The accuracy and completeness scores showed significant results. The Spanish samples yielded less accurate results due to homophones and noun-drop. The authors suggested that some languages such as Spanish may rely on other cues such as prosody to make up for this. Mid-frames were found to be more useful than end-frames, and frequent frames were found to give better

cues for nouns and verbs than for adjectives. Overall this study further advanced the idea that frequent frames could provide distributional information useful for grammatical category acquisition.

Wintner, S. (2010). Computational models of language acquisition. In A. Gelbukh (Ed.) *Computational Linguistics and Intelligent Text Processing,* 86-99. Switzerland: Springer. This article begins by outlining the two main theories of language acquisition. These include the nativist theory, which is innate-centered, and the emergentist theory, which is learned-centered. Then the language learning task is discussed and applied to both children and computer programs. The learner is presented with raw data and uses it to form a model, or a grammar. The success of this model is tested on new data, or utterances, by either assigning structure to utterances or generating new utterances. This parallels the methods of the present study. Next, methods of evaluation are discussed, and current approaches to the question of child language acquisition are presented. These include approaches that attempt to explain the phenomenon from a cognitive perspective, approaches that attempt to create a program to create a grammar from raw data, and the few that attempt to combine these two concepts. The author ends by suggesting that future work in this area should combine understanding of early language acquisition but also be "rigorously defined and robustly evaluated."