



Theses and Dissertations

2014-06-06

A Model of Grammatical Category Acquisition Using Adaptation and Selection

Sarah Zitting Cluff
Brigham Young University - Provo

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Communication Sciences and Disorders Commons](#)

BYU ScholarsArchive Citation

Cluff, Sarah Zitting, "A Model of Grammatical Category Acquisition Using Adaptation and Selection" (2014). *Theses and Dissertations*. 4086.
<https://scholarsarchive.byu.edu/etd/4086>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

A Model of Grammatical Category Acquisition Using Adaptation and Selection

Sarah Z. Cluff

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

Copyright © 2014 Sarah Z. Cluff

All Rights Reserved

ABSTRACT

A Model of Grammatical Category Acquisition Using Adaptation and Selection

Sarah Z. Cluff
Department of Communication Disorders, BYU
Master of Science

By the later preschool years, most children have a knowledge of the grammatical categories of their native language and are capable of expanding this knowledge to novel words. To model this accomplishment, researchers have created a variety of explicit, testable models or algorithms. These have had partial but promising success in extracting grammatical word categories from transcriptions of caregiver input to young children. Additional insight into children's learning of the grammatical categories of words might be gained from evolutionary computing algorithms, which apply principles of evolutionary biology such as variation, adaptive change, self-regulation, and inheritance to computational models. The current thesis applied such a model to the language addressed to five children, whose ages ranged from 1;1 to 5;1 (years;months). The model evolved dictionaries linking words to their grammatical tags and was run for 4000 cycles; four different rates of mutation of offspring dictionaries were assessed. The accuracy for coding the words in the corpora of language addressed to the children averaged 92.74%. Directions for further development and evaluation of the model are proposed.

Keywords: grammatical word categories, evolutionary programming, language acquisition

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Channell for all the time and effort he has provided over the past year. Dr. Channell has always been open and willing to answer any questions, give any feedback, and offer direction and support when I needed it. He has not only made the completion of a thesis possible, but also made the process itself an enjoyable learning experience. I would also like to thank my committee, Drs. Nissen and Tanner, for the time they have given to improving this thesis.

I would like to recognize my husband, Josh, as well as my parents. They have constantly supported me in my pursuit of a master's degree and have always seen my potential when I fail to recognize it. I am grateful for their faith in me and their many prayers in my behalf. Without them, I would not have accomplished all that I have thus far.

TABLE OF CONTENTS

| | |
|--|-----|
| LIST OF TABLES | v |
| LIST OF FIGURES | vi |
| DESCRIPTION OF CONTENT | vii |
| Introduction..... | 1 |
| Models of Grammatical Category Acquisition..... | 3 |
| An Algorithm Using Adaptation and Selection..... | 7 |
| Method..... | 9 |
| Participants..... | 9 |
| Instrumentation..... | 11 |
| Procedure..... | 13 |
| Results..... | 13 |
| Discussion..... | 21 |
| Conclusions..... | 25 |
| References..... | 26 |
| Appendix: Annotated Bibliography..... | 29 |

LIST OF TABLES

| Table | Page |
|---|------|
| 1. Mean Accuracy Results of Generations 3800-4000 at Each Mutation Rate for Each Corpus | 14 |

LIST OF FIGURES

| Figure | Page |
|---|------|
| 1. Mean results for each mutation rate across all corpora..... | 15 |
| 2. Results for Eve's corpus at all mutation rates..... | 16 |
| 3. Results for Peter's corpus (sections 1-12) at all mutation rates..... | 16 |
| 4. Results for Peter's corpus (sections 13-20) at all mutation rates..... | 17 |
| 5. Results for Naomi's corpus (sections 1-58) at all mutation rates..... | 17 |
| 6. Results for Naomi's corpus (sections 59-93) at all mutation rates..... | 18 |
| 7. Results for Shem's corpus (sections 1-20) at all mutation rates..... | 18 |
| 8. Results for Shem's corpus (sections 21-40) at all mutation rates..... | 19 |
| 9. Results for Adam's corpus (sections 1-19) at all mutation rates..... | 19 |
| 10. Results for Adam's corpus (sections 20-39) at all mutation rates..... | 20 |
| 11. Results for Adam's corpus (sections 40-55) at all mutation rates..... | 20 |

DESCRIPTION OF CONTENT

This thesis is part of a larger research project, and portions of this thesis may be published as parts 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

A desire to better understand the necessary preconditions and mechanisms of language acquisition has prompted the development of explicit, testable algorithms (Pinker, 1988) and computer models. These algorithms and models typically use transcribed language input to children, apply a set of learning procedures to extract certain patterns from this input, and measure the resulting changes in some aspect of linguistic knowledge. As applied to the problem of how children learn the grammatical word categories of their language, the results of these algorithms and models have been promising but variable in their success. The present project describes an alternate algorithm, derived from the principles of adaptation and selection typically associated with evolutionary biology, which serves as the basis for a testable computer model of grammatical category acquisition.

Past studies have investigated various hypotheses as to how children learn grammatical categories (also known as syntactic categories or parts of speech). Although it is apparent that children are not born with an immediate robust vocabulary and syntax, exactly how language is acquired is still vigorously debated. Two major theoretical perspectives involve the nativist (also known as generativist or Universal Grammar) perspective, and the constructivist (also called emergentist, socio-pragmatic, functionalist, usage-based) perspective (Ambridge & Lieven, 2011). The nativist perspective proposes that all people are born with some innate linguistic knowledge, which probably includes some version of grammatical categories. In comparison, constructivism is built upon the philosophy that language is acquired through input and children are not born with inherent grammatical categories. The theoretical approach a

researcher takes shapes the method and design of his or her study and the conclusions that may be drawn; this shaping is illustrated by several design decisions made in previous models and studies. One decision is whether grammatical categories defined by the community of grammarians are sought in the data or whether categories are formed by clusters in the data. This is similar to the notion of the etic versus emic approaches (Pike, 1967): whether data are fit into the closest categories or whether categories are created which best fit the data. A related design decision is whether the number of grammatical categories is set (and at which level it is set) or whether any number of categories might emerge when supported by data. An additional design decision concerns the number of possible grammatical categories each word might have. DeRose (1988) pointed out that about 11% of word types in English are grammatically ambiguous in that they have more than one possible grammatical category (such as *cook* being either a noun or a verb), but that this 11% of word types are used often enough to account for 40% of the word tokens in running text such as the Brown University Corpus (Francis & Kucera, 1982). In most of the studies described hereafter, each word is allowed to have only a single grammatical category, presumably its most frequently occurring one. A final example as to how these algorithms and models differ is whether a grammatical category is assigned to a word by the word's distribution relative to one or more other words (a "distributional" approach) or whether additional information is used, such as the word's semantic nature (Pinker, 1987) or additional input cues (MacWhinney, Leinbach, Taraban, & McDonald, 1989).

Models of Grammatical Category Acquisition

Previous attempts to answer the question of language acquisition have varied from theories of semantic bootstrapping (Pinker, 1987) to more recent, distributional approaches like Mintz's (2003) frequent frames. 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, or preceding these other words) is noted and inferences are drawn. It is important to note that in all of these studies, words were only assigned their most frequent or likely grammatical category.

One earlier study did not use a distributional approach but was based instead upon various input cues. MacWhinney et al. (1989) used a computational model to investigate how German children learn the declensions (i.e., the agreement rules and patterns) of their language. The authors approached their study with a constructivist perspective and suggested that children learn language through various cues (e.g., phonological, morphological, syntactic, and semantic). Their computational model, which they called the Competition Model, was designed to mirror child language acquisition for the German declension paradigm, chosen specifically due to its highly complex gender, number, and case assignments. With these three parameters applied, there were a resulting six types of articles for the program to choose from when coupling an article with a noun. As nouns were presented one at a time, the program was designed to detect if there were cues associated with those input nouns. Results showed high accuracy in learning and assigning articles to nouns based on cues. A limitation of the study, apart

from using input derived from adult-to-adult speech, was only focusing on only a single grammatical tag per word.

Kiss (1973) was one of the first to introduce a computational model that used a hierarchal analysis to group words based on their distribution relative to other words, introducing an area of computational modeling for language acquisition that has been a study of interest for 40 years. Though based on a relatively small sample of utterances, the model found clusters of words with similar distribution which might represent grammatical categories.

Cartwright and Brent's (1997) distributional study differed in two ways from the Kiss (1973) study. First, the algorithm proposed by Cartwright and Brent resulted in a smaller set of discrete categories and, second, the model operated one sentence at a time, forgetting previous input. The proposed hypothesis was that children create new templates for input they receive and then merge overlapping templates together, based on preferences, until optimal groups and templates are reached. For each of the five experiments, independent variables were manipulated, such as varying the amount of input given to the computer, adding ambiguous words to the input, using child-directed speech, and adding semantic information to the system. Although accuracy and completeness scores showed that distributional input was effective for group merging, the scores were higher when semantic information was combined with distributional information.

A third computational model study based on distributional methods was performed by Redington, Chater, and Finch (1998). The model organized words based on similar co-occurrence patterns to provide a representation of the way a child may

organize the input received. The study used language samples from the Child Language Data Exchange System (CHILDES) database (MacWhinney, 2000) as input. Due to the limitation of the computational algorithm being unable to distinguish between multiple syntactic categories, words were only tagged with their most frequent category. The words were organized into dendrograms, divided at different points, and scored in formulas of accuracy and completeness. Words that were close in proximity in the dendrogram were most likely from the same grammatical category, while words farther apart were least syntactically similar. Results of accuracy and completeness were highest when clusters of both the preceding and succeeding words were analyzed, and gave further support for language acquisition theories based on distribution.

Mintz (2003) looked at the frequent frames of child-directed speech. Mintz defined frequent frames as two words that often co-occur, separated by one word (e.g., you __ it, the __ one). Mintz used six language samples from the CHILDES database (MacWhinney, 2000), all of which involved children who were 2;6 (years;months) or younger. Analysis of frequent frames in the first experiment was determined by looking at the 45 most frequently occurring frames for each corpus, which met the criteria of occurring frequently enough to be noticeable and contained a range of intervening words. Mintz argued that he looked at a smaller part of the language sample in order to better analyze the extent of which the input the children were receiving was informative. Based on calculations of accuracy, all scores showed significant results, although scores for completeness were relatively low. The second experiment that was conducted was based upon the number of frequency of frames in comparison to the total number of frames per each corpus. Two limitations of this study were: (a) words that belonged to multiple

grammatical categories were only assigned to the most common category, and (b) the percentage of each language sample supporting the analysis and delineation of frequent frames was small.

A study which expanded upon Mintz's (2003) frequent frames partially addressed the second limitation (the small part of the data in each language sample used by the program) previously mentioned. St. Clair, Monaghan, and Christiansen (2010) compared frequent frames to flexible frames, using a computational model to perform the study. Multiple experiments showed the benefits of combining input information from bigrams (e.g., aX, Xb) and trigrams (e.g., aXb) into flexible frames (e.g., aX +Xb), which overcomes the weaknesses of strictly analyzing bigrams or trigrams independently. Accuracy increased, and a larger portion of the language sample was analyzed, suggesting that allowing a less rigid distributional form may provide more information during input for children learning language and acquiring grammar.

A distributional study done by Freudenthal, Pine, and Gobet (2005) also used a computational algorithm that learned syntactic (grammatical) categories via chunking, which was then analyzed using co-occurrence statistics. The authors argued that co-occurrence statistics were more accurate over longer units, decreasing the rate of incorrect substitutions, and was therefore better to analyze results than traditional methods (i.e., accuracy and completeness). The computational model, Mosaic, was used to show syntax acquisition. Chunked units were determined by if the same small phrases frequently occurred together. These phrases were then chunked together as one unit and, in turn, sped up the process of learning syntactic categories. The only variable that was adjusted in this study was the specific threshold of the chunking levels. Freudenthal et al.

found that this chunking mechanism decreased the number of syntactic errors produced by Mosaic.

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 in their computational models and were limited to only one possible grammatical tag per word. The present study approaches the topic of interest from an unexplored method, an adaptation and selection algorithm, and better models real language by assigning multiple grammatical categories to words.

An adaptation and selection algorithm is part of evolutionary computing. Evolutionary computing involves applying the evolutionary biology principles (e.g., variation, adaptive change, self-regulation, and inheritance) to computational models and is comprised of sub-branches consisting of genetic algorithms, evolution strategies, and evolutionary programming (Fogel, 2006). Fogel explains that the branch of evolutionary programming entails a population of solutions that randomly mutates to create offspring from the parent solutions, of which the best-fit offspring are then chosen to become the parents of the next generation. Because the system never receives feedback as to what aspects of the chosen solutions are correct, the correct parts are as likely to mutate in the next generation as are the incorrect elements. However, it is observed that over many selections, reproductions, and mutations, an offspring evolves that represents an adequate solution to the original problem at hand.

These evolutionary algorithms have been widely applied to areas outside of researching an explanation of human evolution. Evolutionary algorithms have been used

to develop seemingly trivial programs, like computers that learn how to play checkers, to more compelling advances, such as systems that provide better interpretation of the radiographic features of mammograms, thus aiding in early breast cancer detection (Fogel, 2002; Fogel, Watson, Boughton, & Porto, 1998). Siegler (1996) based some of his research on the development of human cognitive strategies using evolutionary algorithms as well. He argued that evolutionary biology and cognitive development contain analogous fundamental concerns, one of which is the search for ultimate origins.

The present study applies an adaptation and selection evolutionary algorithm to the problem area of the acquisition of the grammatical categories in language. Previous studies of acquisition have hypothesized and explored, but a satisfactory explanation has yet to be agreed upon. Studies have also limited their analysis to one grammatical tag per word when the English language contains many words that fall into multiple categories. In order to incorporate DeRose's (1988) findings of words having multiple grammatical categories, the current study initially used an automated tagging system to discover the average number of tags per word per corpus. Depending on the corpus, this number ranged anywhere from 1.12 to 1.16 average tags per word. By setting this number in the computational model, the average tags per word provided a control that prevented false positives in accuracy. Due to the varied and often confusing input that children receive day-to-day, the application of an adaptive-selective algorithm might comparatively model the guesswork produced by a child during early language production.

Method

Other researchers collected the language sample corpora for different purposes than were used in this research study. In the present study, the focus for using the corpora was to correctly format and grammatically tag the language samples.

Participants

Five corpora from the CHILDES database (MacWhinney, 2000) were used for input for the adaptation-selection computational algorithm. Three of these corpora are the same samples used in Mintz's (2003) study, while two are additional corpora taken from the database to increase the pool of subjects.

Eve's language samples were recorded over a period of 20 sessions, ranging from 1;6 to 2;3. Samples were taken in Eve's home with her parents and a few other adults. Eve was described as linguistically precocious. From the 20 language samples, there were 14,806 adult utterances spoken in the presence of Eve and 12,114 utterances from Eve herself. Sample collection stopped when her family moved away from the Cambridge area (Brown, 1973). The average number of tags per word was 1.14.

The next child, Peter, was the first-born child in his family. He was from a middle class to upper-middle-class family, who were living in a university community in New York City. Peter's language samples were gathered mostly from play interactions in his home living room with multiple adults. Sampling began when Peter was 1;9 to 3;2. In the 20 language samples collected, 31,738 utterances were spoken by adults in Peter's presence and 29,756 utterances spoken by Peter (Bloom, Hood, & Lightbrown, 1974; Bloom, Lightbrown, & Hood, 1975). The corpus on Peter was broken into two segments, Peter (sections 1-12) and Peter (sections 13-20) respectively. The first section includes

ages 1;9 to 2;4 and contained 16,249 adult utterances. The second section spanned ages 2;5 to 3;2 and had 15,489 adult utterances. The average tags per word for Peter (sections 1-12) was 1.15, while the average tags per word for Peter (sections 13-20) was 1.16.

Naomi's language samples were collected in the home, during parent-child interaction. A total of 93 language samples were collected from 1;1 to 5;1. There were a total of 12,226 utterances spoken by adults in Naomi's presence, and 17,243 utterances were spoken by Naomi (Sachs, 1983). The corpus on Naomi was broken into two segments, Naomi (sections 1-58) and Naomi (sections 59-93). The first section includes ages 1;1 to 2;4 and contained 7,151 adult utterances. The second section spanned ages 2;4 to 5;1 and contained 5,075 adult utterances. The average tags per word for Naomi (sections 1-58) was 1.12, while the average tags per word for Naomi (sections 59-93) was 1.13.

Shem was born to a middle class to upper-middle class professional family in the Palo Alto area in California. He attended a local day care in the mornings as well as occasionally in the afternoons. Language samples were collected in Shem's home from ages 2;2 to 3;2. In the 47 language samples collected, there were 24,097 utterances spoken by adults in Shem's presence and 18,166 utterances spoken by Shem (Clark, 1978). The corpus on Shem was broken into two segments, Shem (sections 1-20) and Shem (sections 21-40). The first section includes ages 2;2 to 2;8 and had 11,409 adult utterances. The second section spanned ages 2;8 to 3;2 and had 12,688 adult utterances. The average tags per word for Shem (sections 1-20) was 1.13, while the average tags per word for Shem (sections 21-40) was 1.12.

Adam was the son of a middle-class family who spoke Standard American English. His parents were described as well educated and were professionally employed as a minister and an elementary school teacher. His language samples started when he was 2;3 and ended when he was 4;10. Samples were taken in Adam's home with his parents and a few other adults. Of the 55 total language samples collected, 26,432 utterances were spoken by adults in Adam's presence and 46,743 utterances were spoken by Adam (Brown, 1973). The corpus on Adam was broken into three different segments, Adam (sections 1-19), Adam (sections 20-39), and Adam (sections 40-55). The first section includes ages 2;3 to 2;11 and had 9,507 adult utterances. The second section spanned ages 3;0 to 3;10 and had 10,277 adult utterances. The third section included ages 3;11 to 4;10 and had 6,648 adult utterances. The average tags per word for Adam (sections 1-19) was 1.12, while the average tags per word for Adam (sections 20-39) was 1.12, and 1.13 for Adam (sections 40-55).

Instrumentation

At the computational model program start-up, the user determines three settings: the number of evolutionary cycles to run, the target level for a dictionary's average number of grammatical tags per word, and the likelihood of dictionary entry mutation (set as a one in X chance). The program then opens a text file of transcribed, grammatically-tagged (coded) utterances which 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]* (etc.) and no punctuation marks. These utterances are stored for re-use in evaluation. An output file, which records results, is opened as well.

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, in such a way that every word in each dictionary has a grammatical tag entry randomly constructed. In order to ensure randomization, the number of tags for the word is randomly set as one, two, or three, and, accordingly, one, two, or three tags are randomly drawn for the word from the grammatical tag options available to the system. The number of dictionaries can be changed in the program code but not at run time.

The adaptation/selection process is then cycled through for the number of generations previously set during the program start-up. The odd-numbered utterances in the input file serve as the basis for the fitness evaluation of each dictionary. This fitness evaluation is set up by 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, a 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. Last, the product of this ratio times the number of correct hits is used as the *fitness* criterion. This fitness number is also written to the screen and the output file, along with the cycle number. While the odd-numbered utterances serve as the basis for the fitness evaluation, the even-numbered utterances in the input file are used as the basis for quantifying the generalization accuracy of the best dictionary. This accuracy level is also written to the screen and to the output file.

After all of the candidate dictionaries have been evaluated, the one with the highest fitness score is used as the starting point for populating the next generation. Due to the mutation setting at program set-up, each word entry in this dictionary has a mutation likelihood chance of having its grammatical tag entry replaced with a randomly chosen one as part of making an offspring dictionary, regardless of whether that particular entry was actually correct or not. By this process a population of new "baby" dictionaries is created, which are then evaluated, and the best one becomes basis for future dictionaries. After the specified number of evolutionary cycles has been reached, final data 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. The number of generations of adaptation and selection allowed was 4,000. The program was designed to grammatically tag the words from the language sample corpus input with the predetermined average number of tags per word. Each of the 11 corpora was run through the program four times, once for each level of the mutation-variation rate. The levels of the mutation-variation rate were a 1/400 chance, a 1/800 chance, a 1/1200 chance, and a 1/1800 chance. The average number of tags allowed per word was based on the average tags per word for each individual corpora (e.g., 1.12, 1.13, 1.14, 1.15, or 1.16 tags per word), as determined by a tabulation program.

Results

The computational model was run for each of the 11 corpora for 4,000 cycles to see the effects of each mutation rate (1/400, 1/800, 1/1200, 1/1800) on the accuracy of

word coding. Generally, accuracy increased across cycles as the mutation rate was decreased. Table 1 illustrates the mean percentage of accuracy for generations 3800-4000 at the four mutation rates for each corpus. It can be seen in Table 1 that the lowest accuracy rate in each corpora was always at the 1/400 mutation rate, while the other three (e.g., 1/800, 1/1200, and 1/1800) varied as to which produced the highest accuracy. As the program was allowed fewer mutation opportunities, it was better able to correctly tag words in the corpus as to their grammatical categories.

Table 1

Mean Accuracy Results of Generations 3800-4000 at Each Mutation Rate for Each Corpus

| Corpus | 1/400 | 1/800 | 1/1200 | 1/1800 |
|-------------|-------|-------|--------|--------|
| Eve | 90.72 | 93.64 | 94.13 | 94.54 |
| Peter 1-12 | 91.91 | 94.71 | 94.63 | 94.06 |
| Peter 13-20 | 89.98 | 93.81 | 93.60 | 93.75 |
| Naomi 1-58 | 89.52 | 92.34 | 92.93 | 92.41 |
| Naomi 59-93 | 86.39 | 90.27 | 90.85 | 89.47 |
| Shem 1-20 | 84.90 | 92.14 | 92.86 | 92.37 |
| Shem 21-40 | 83.48 | 87.64 | 88.27 | 89.16 |
| Adam 1-19 | 89.20 | 93.54 | 93.66 | 92.93 |
| Adam 20-39 | 85.61 | 90.93 | 91.92 | 92.14 |
| Adam 40-55 | 88.11 | 90.92 | 91.78 | 91.06 |

The results for the mean percentage at each mutation rate across all corpora can be seen in Figure 1. Growth curves for each child's corpus at each of the established mutation rates can be found in Figures 2-11. As illustrated, the adaptation-selection computational model improves rapidly within the first 500 generations and then gradually increases through the 4000th generation. In general, the highest mutation rate (e.g., 1/400) is the one that increases the quickest, but then fails to continue to maintain the accuracy and the other mutation rates surpass it by later generations, as seen in Figure 1. The mutation rate that produced the highest accuracy across the 4000 generations was 1/1200, although the 1/1800 rate was just slightly below it.

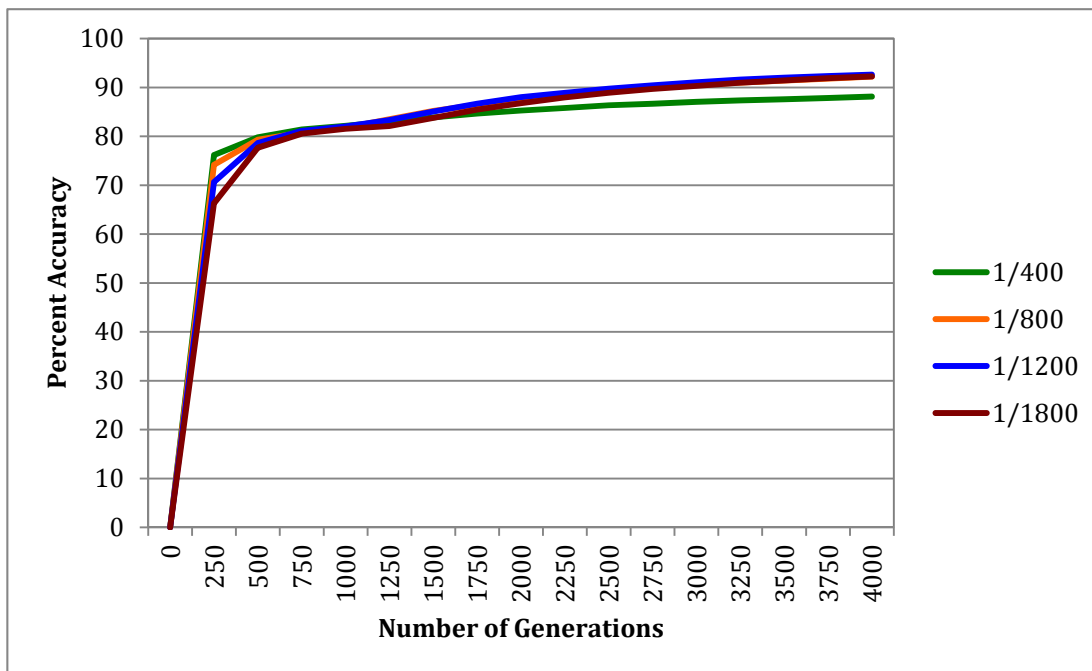


Figure 1. Mean results for each mutation rate across all corpora.

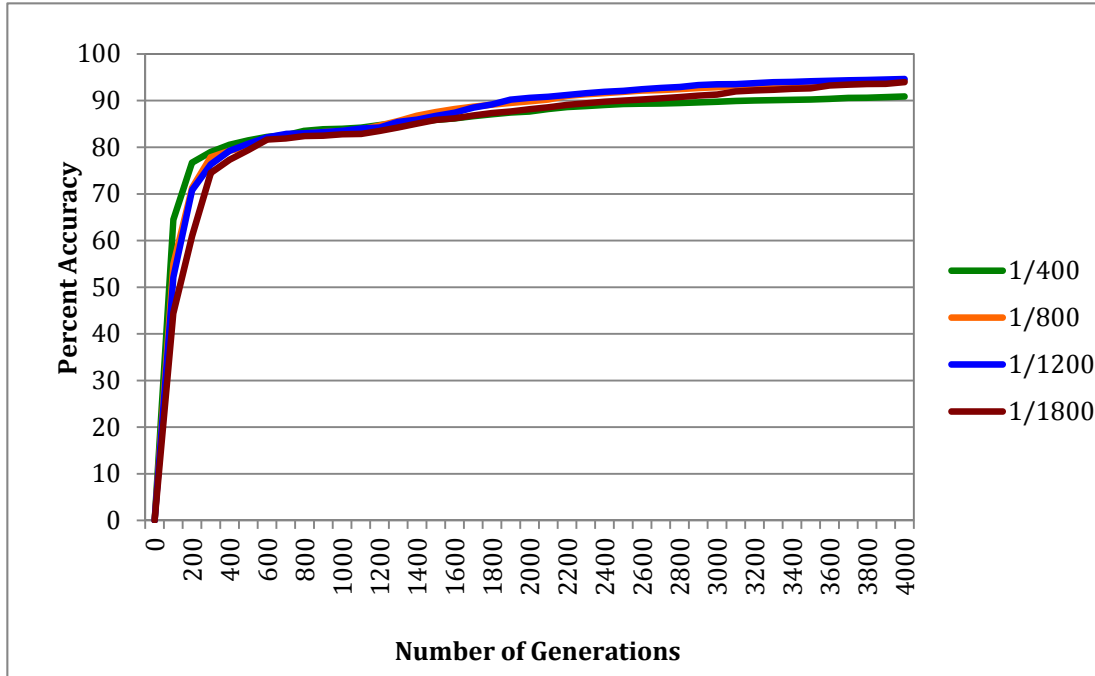


Figure 2. Results for Eve's corpus at all mutation rates.

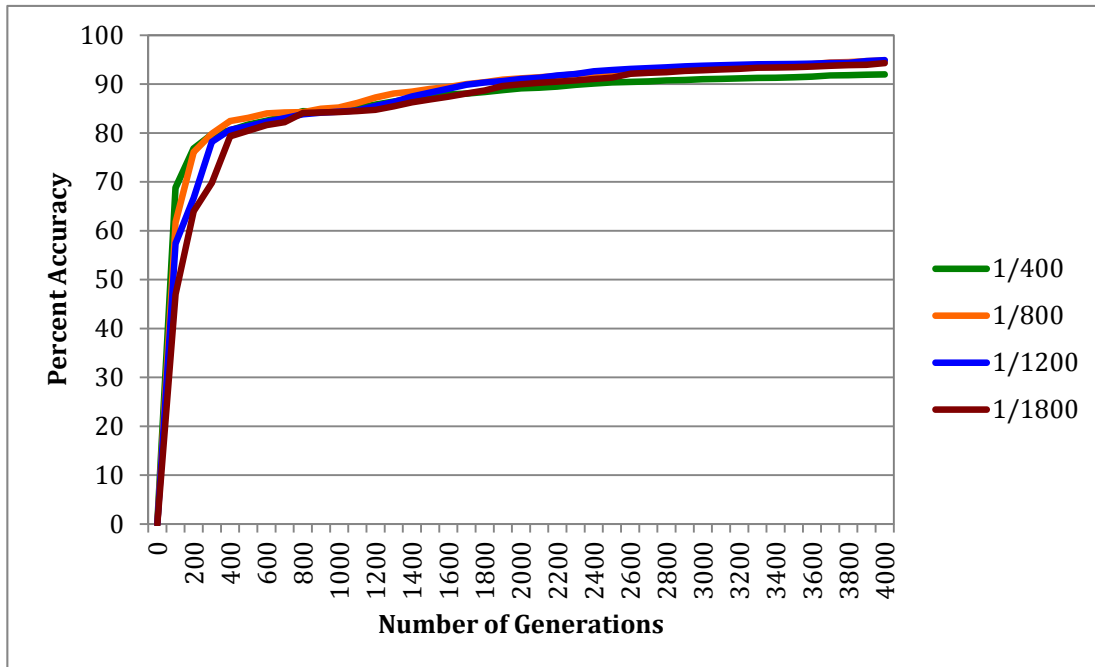


Figure 3. Results for Peter's corpus (sections 1-12) at all mutation rates.

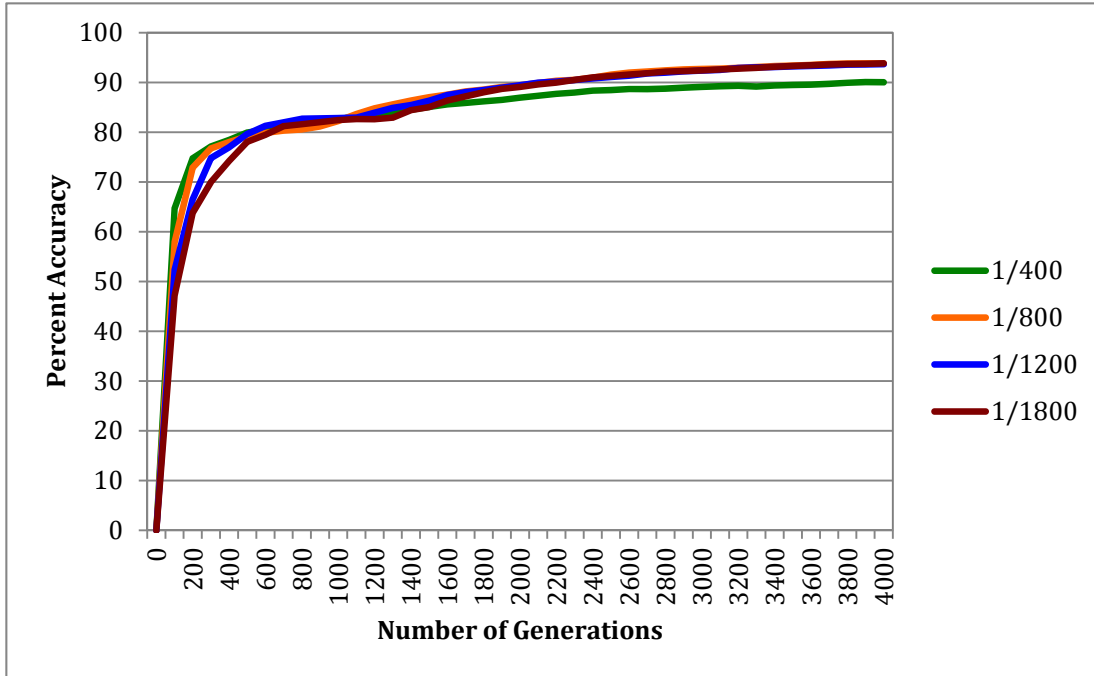


Figure 4. Results for Peter's corpus (sections 13-20) at all mutation rates.

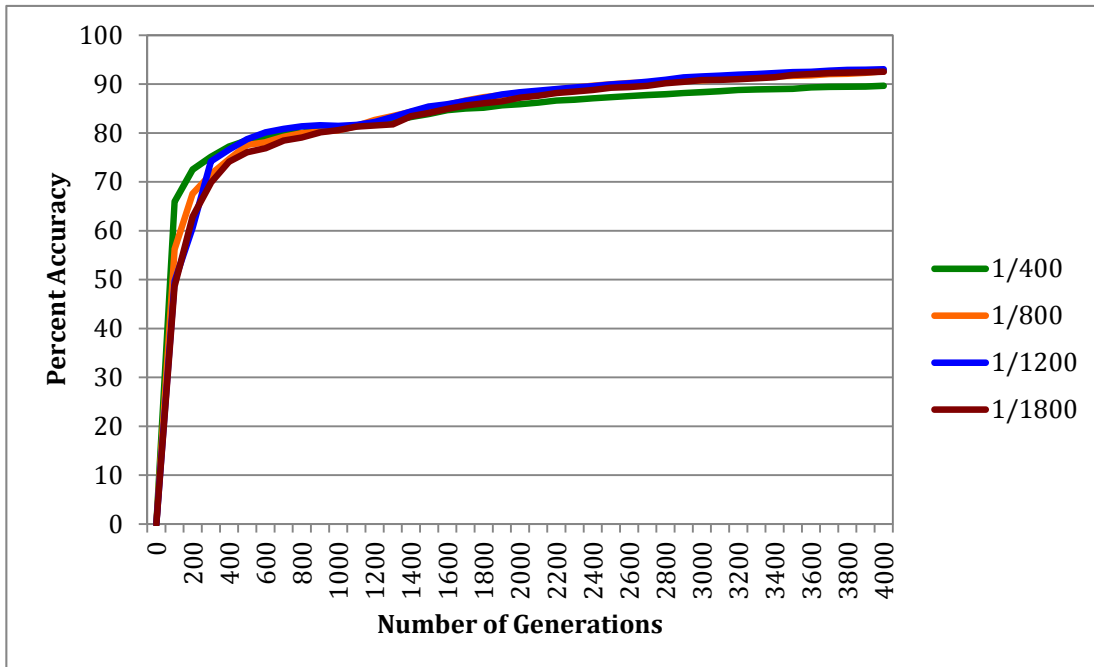


Figure 5. Results for Naomi's corpus (sections 1-58) at all mutation rates.

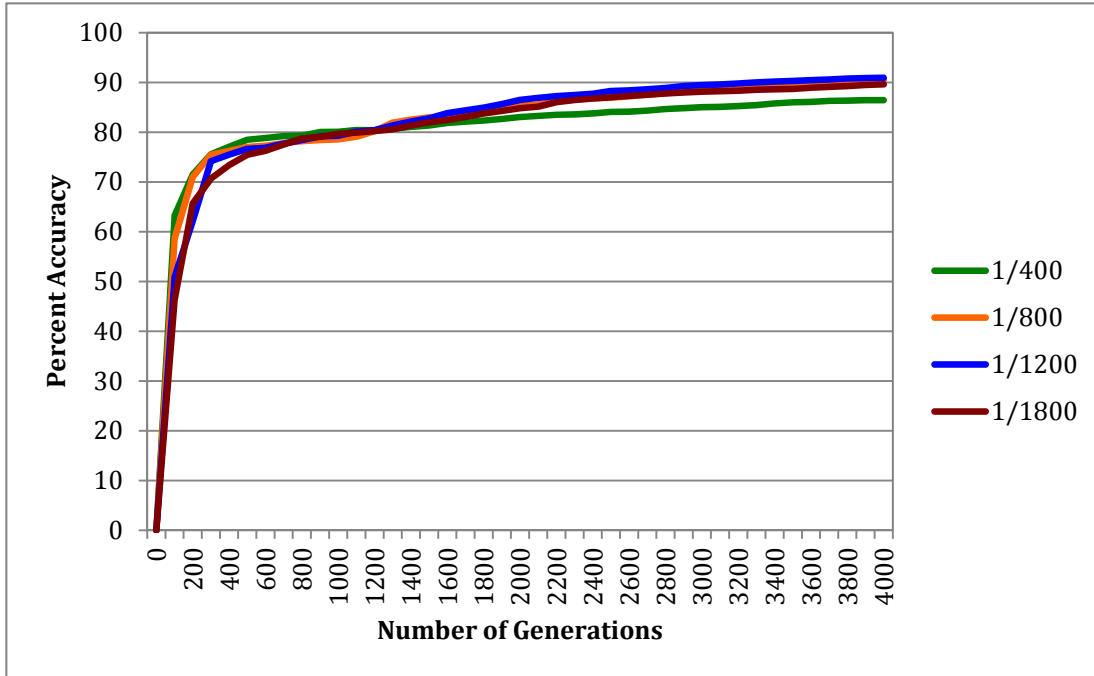


Figure 6. Results for Naomi's corpus (sections 59-93) at all mutation rates.

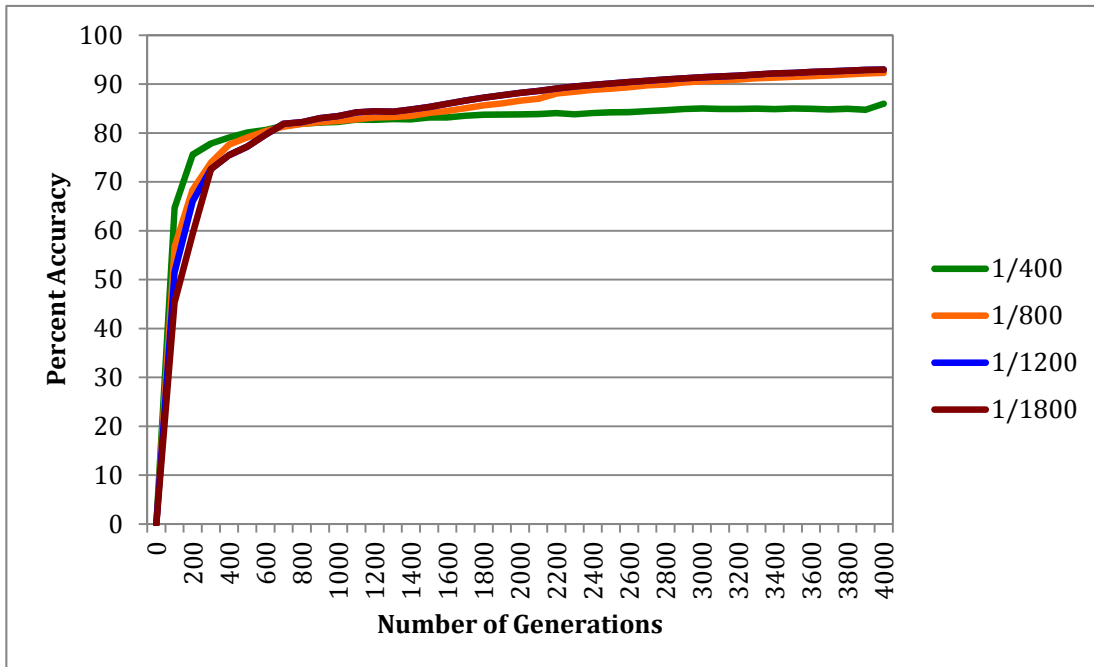


Figure 7. Results for Shem's corpus (sections 1-20) at all mutation rates.

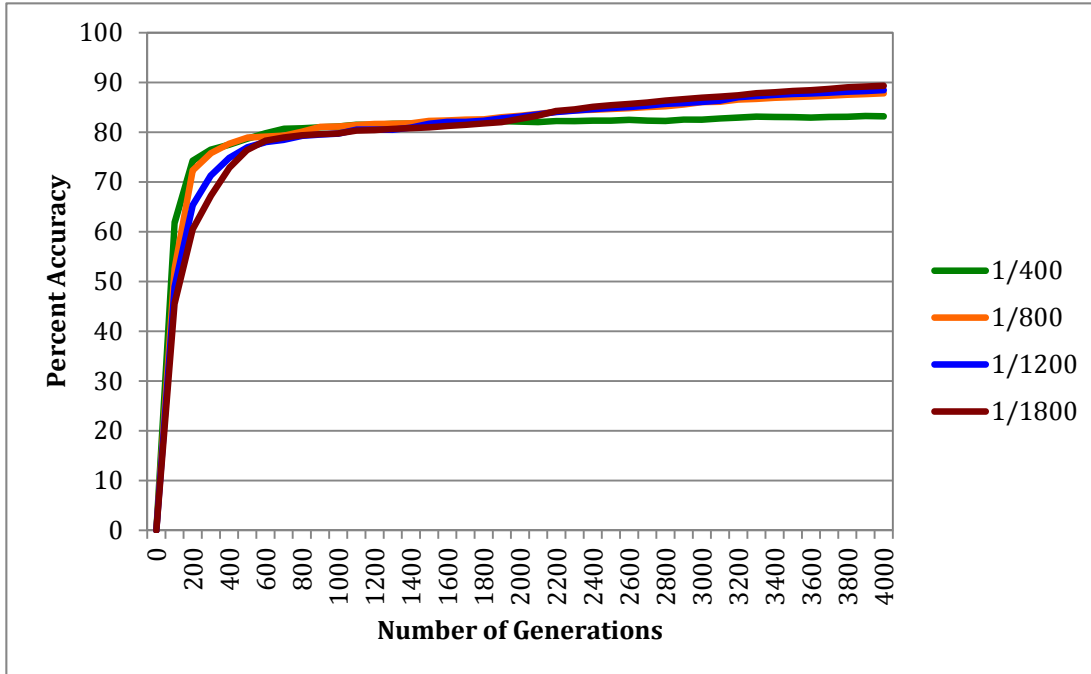


Figure 8. Results for Shem's corpus (sections 21-40) at all mutation rates.

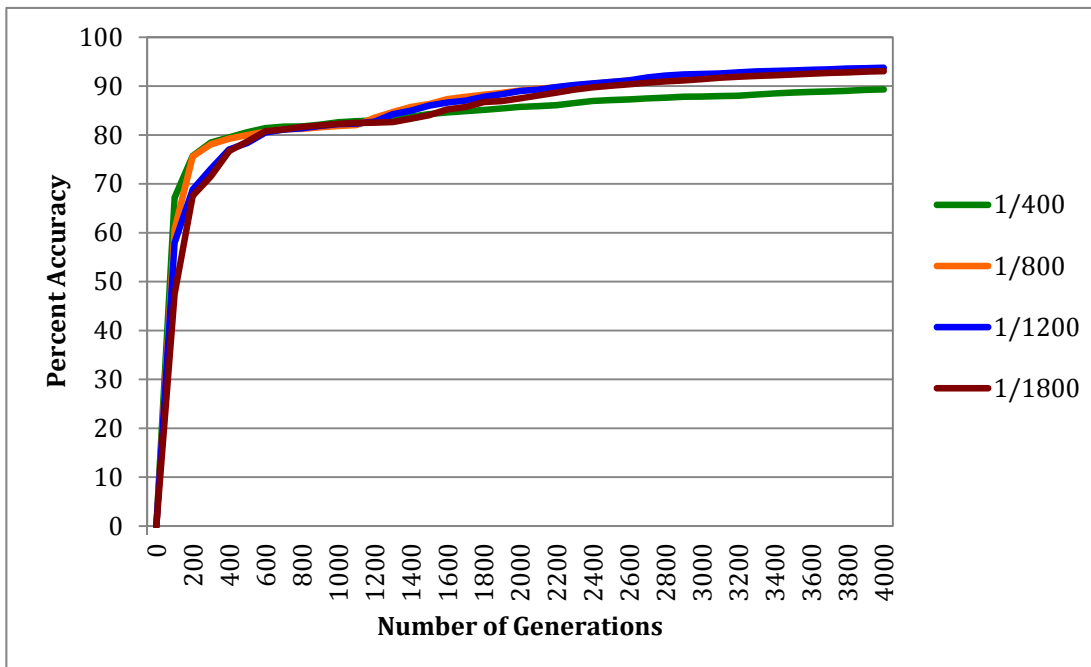


Figure 9. Results for Adam's corpus (sections 1-19) at all mutation rates.

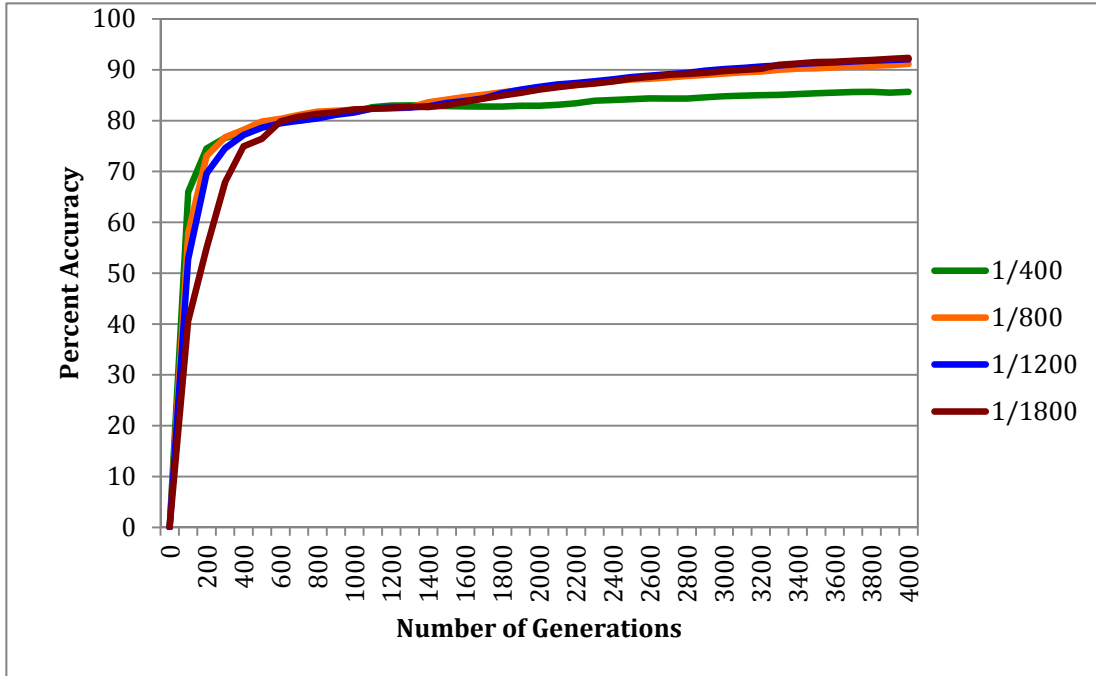


Figure 10. Results for Adam's corpus (sections 20-39) at all mutation rates.

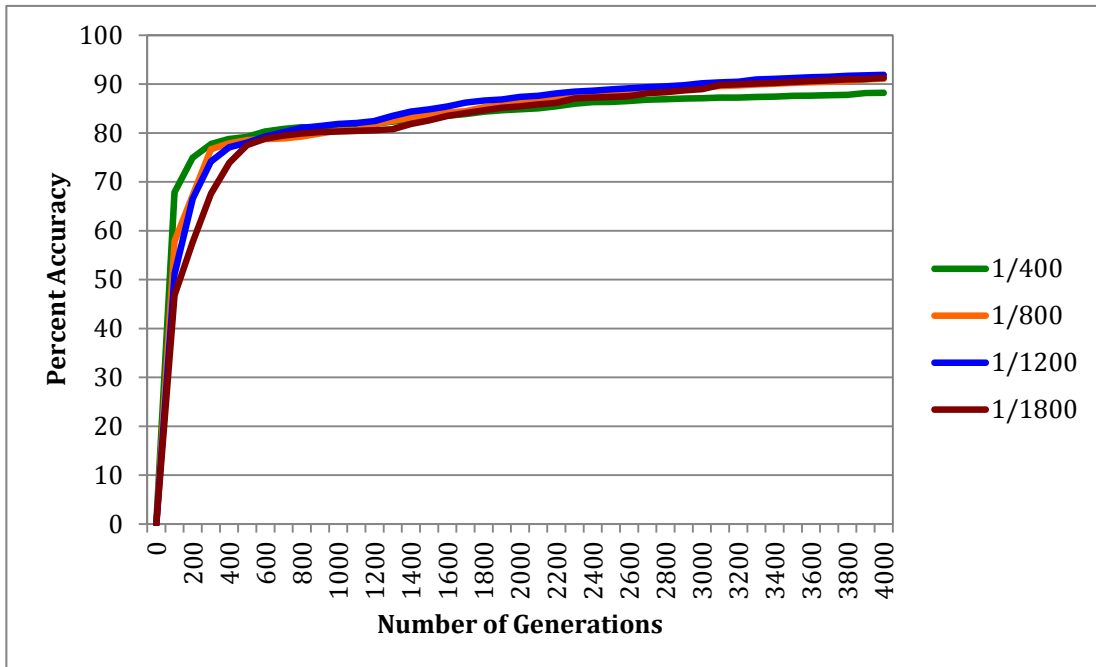


Figure 11. Results for Adam's corpus (sections 40-55) at all mutation rates.

Discussion

The purpose of this study was to investigate the effects that the mutation rate had upon the accuracy of an adaptation-selection computational model while controlling the average number of tags per word. The study found that lower mutation rates yielded better program accuracy. It is easy to see from all of the figures presented that the model increased in accuracy very quickly during the first 500 or so generations. Because the model was able to produce high accuracy early on, but still had many generations left to run, higher mutation rates would replace correct tags for incorrect tags more often than the lower mutation rates.

One of the advantages to dividing the some of the participants' set of samples into smaller corpora was that it allowed a preliminary examination as to how age affected accuracy. For example, Adam's sample was divided into three corpora, all of which had the highest accuracies at the 1/1200 mutation rate. However, as displayed in Table 1, accuracy did decrease somewhat with the increase of age. This outcome was also consistent for Naomi, Shem, and Peter. This finding could suggest that the difference between the language addressed to the children when they were younger, compared to the language addressed to them when they were older, might make a difference in the acquisition of grammatical categories. Of course, a closer analysis of the syntactic and lexical characteristics of the language input in these corpora would be necessary to more firmly support this relationship. However, this correlation might suggest why other researchers have generally used samples from children younger than about 2;6. Further research is needed to examine if this observed trend holds true with other longitudinal child language corpora.

In terms of the setting of the average number of tags per word, a relationship was observed between the three highest average tags per word and the highest percent accuracy (see Table 1) though the sample size was too small to allow statistical corroboration. Eve's corpus and both of Peter's corpora produced the highest accuracy rates in the study, as well as having the highest average number of tags per word (e.g., 1.14, 1.15, 1.16) of all the corpora. Perhaps increasing the number of tags per word even slightly allows the computational model more opportunities to select the correct tag and thus increase the overall attained accuracy. Again, analysis of more corpora would provide insight into the potential reasons for this admittedly possible effect.

Pilot tests were performed to examine the effects of increasing the amount of generations run in total. One child's corpus was randomly selected and run at 8,000 generations as opposed to the normal 4,000. The resulting data showed that the accuracy plateaued around the 4,000th generation. There was no visible improvement during the rest of the generational cycles, as the model reached a point where it was as likely to replace correct tags for words as it was to replace incorrect tags for words. This pilot testing for generational cycles suggests that future studies could use the 4,000 generations setting as a constant, at least with similar corpora.

Pilot testing was also performed to explore the results of changing the number of dictionaries produced at each generational cycle. The program used in the study produced 100 dictionaries each cycle. When the number of dictionaries produced was changed from 100 to 500, accuracy results improved for the corpora tested, though the program ran more slowly. Perhaps allowing the program increased opportunities to produce more accurate dictionaries early on could be the reason for this finding; however,

the use of 1,000 dictionaries on a corpus did not appear to produce better accuracy than 500 dictionaries. Additional testing would need to be performed to investigate if accuracy was consistently enhanced with a 500-dictionary version of this computational model or if some other number of dictionaries might yield consistently better results.

Although this study of grammatical category acquisition is similar to previous research in that the foundational area of interest is the same, the difference in methodology and outcome variables makes a direct comparison to past studies difficult. As mentioned earlier, several studies have made significant contributions to the questions surrounding grammatical category acquisition. For example, Cartwright and Brent (1997) investigated how children may create and merge familiar templates when learning grammatical categories. Redington et al. (1998) modeled grammatical category acquisition using dendrograms to show distributional information, and Mintz (2003) introduced the idea of frequent frames. These approaches all contributed knowledge to the question regarding grammatical category acquisition, but all took very different approaches to this problem. The current study is a new and different approach as well. With that being said, the comparison of the present findings with those of the Mintz study might be of limited value, though three of the same corpora were used.

Token accuracies for Peter, Eve, and Naomi in the second experiment of the Mintz (2003) study were reported as 0.98, 0.98, and 0.96. These accuracies were higher than results for the same children in the current study. Although accuracy in the Mintz study was higher for Peter, Eve, and Naomi, it is important to consider the amount of data analyzed by Mintz. When determining frequent frames, Mintz selected the 45 most frequent frames from a child's sample and then analyzed the rest of the children's corpora

using these 45 frames. This method unfortunately left quite a bit of input unanalyzed, which makes a true comparison of the accuracies between the present study and Mintz impossible. Perhaps if all the data were analyzed in each corpus using frequent frames, the token accuracies might have decreased to some degree.

As suggested above, one of the limitations of the present study was the number of corpora analyzed. Corpora from five children were selected from MacWhinney's (2000) CHILDES database, but perhaps increasing the number of corpora used would provide a better display of accuracy growth than obtained in the present study. It would be desirable to expand the diversity of those samples as well, in order to provide a more representative display of language input. However, it is difficult to acquire such data due to the limited number of longitudinal studies performed and available for public use.

As in any study of grammar acquisition, multiple languages will need to be tested in order to increase the validity to this new acquisition model due to the variability of syntactic and lexical differences affecting word order in these languages. The results of the present study, which used only English corpora, will need corroboration from other languages.

Generally speaking, it is hard to say if children really do acquire grammar using some sort of adaptation-selection algorithm. Siegler (1996) suggested that the cognitive strategies used by humans, including children, mirror the elements displayed in an evolutionary algorithm, like the one used at the core of this study. However, further research and theorizing in the area of adaptation and selection models of human cognition and its development would be both necessary and desirable.

Conclusions

Despite its limitations, this study has made several contributions to research studying grammar acquisition. It is the first model done in this area of research to use an adaptation-selection algorithm, and thus it has provided an additional, novel approach and computer model implementation to the question of grammatical category acquisition. This model also grammatically tagged entire corpora of child language, instead of only tagging defined sections, thus showing a more holistic analysis. High accuracy resulted with lower mutation rates and a tentative correlation between average number of tags per word and accuracy was seen. These findings are encouraging and suggest that further research in this area is warranted. Other areas outside of the learning of the grammatical categories of words might also be explored using an adaptation-selection model.

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., & Lightbrown, P. (1974). Imitation in language development: If, when and why. *Cognitive Psychology*, 6, 380-420.
- Bloom, L., Lightbrown, P., & Hood, L. (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development*, 40 (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.
doi: 10.1016/S0010-0277(96)00793-7
- Clark, E. V. (1978). Awareness of language: Some evidence from what children say and do. In R. J. A. Sinclair & W. Levelt (Eds.), *The child's conception of language*. Berlin: Springer Verlag.
- DeRose, S. J. (1988). Grammatical category disambiguation by statistical optimization. *Computational Linguistics*, 14, 31-39.
- 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.
- Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston, MA: Houghton Mifflin.
- 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. doi: 10.1016/j.cogsys.2004.09.003
- Kiss, G. R. (1973). Grammatical word classes: A learning process and its simulation. *Psychology of Learning and Motivation, 7*, 1-41.
- MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language, 28*(3), 255.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.), Vol. 2: The Database. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*(1), 91-117. doi:10.1016/S0010-2077(03)00140-9
- Pike, K. L. (1967). *Language in relation to a unified theory of structure of human behavior* (2nd ed.). The Hague, Netherlands: Mouton.
- Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.) *Mechanisms of language acquisition*. Hillsdale, NJ: LEA.
- 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.
- 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

Appendix: Annotated Bibliography

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

The introduction begins by outlining the different theoretical approaches to child language acquisition before going more in-depth with the studies that have been done. Although there are various names given, the main approaches fall into the nativist or constructivist categories. Chapter six focuses specifically on how children learning word-order languages learn the rules governing the word order of that language. The author's explore Pinker's theory of semantic bootstrapping and point out the inherent problems with it that have led many new researchers to abandon it as a possibility. The chapter then reviews multiple distributional approaches (i.e., frequent frames, chunking, merging templates), as well as the possible approach of language acquisition via phonological cues. The research support for each theory, as well as weaknesses in the theory, is presented in such a way as to show that although much has been learned, the mystery of child language acquisition has yet to be solved.

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

This study investigated how children acquire language and their knowledge of morphological rules by teaching them nonsense words. Morphological areas tested were the plural and two possessives of a noun, the third person singular form of a verb, the progressive and past tenses, and the comparative and superlative of adjectives. The author inventoried 1000 most frequent words in a first grader's vocabulary in order to appropriately assess which morphemes and grammatical rules to test. A total of 56 child participants, 4-7 years old, were asked to inflect, derive, compound, and analyze compound words, in order to test the extent they extended morphological rules to new words. A total of 28 questions were developed analyzing these rules and administered to each child. In general, results showed that gender did not make a difference in the ability to produce morphology, but that there was a difference based on age, with first graders performing better than the preschool children. Overall, the study showed that children generally know how to use morphological rules.

Cartwright, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition, 63*, 120-170. doi: 10.1016/S0010-0277(96)00793-7

Cartwright and Brent show that while there have been computational models done in past studies, none of the studies, with one exception, were using computational data to recommend a theory of syntactic category acquisition. The authors review past theories and studies, which have explored acquisition based on distributional patterns and syntactic input. Cartwright and Brent's study differs in two ways on a computation model study that was done by Kiss in 1973. First, the proposed theory

results in a smaller set of discrete categories and, second, the model operates one sentence at a time, forgetting previous analysis as it moves from sentence to sentence. The proposed hypothesis was that children create new templates for input they receive and then merge overlapping templates together, based on preferences, until optimal groups and templates are reached. Five experiments were organized to test this hypothesis, varying the amount of input given to the computer, adding ambiguous words to the input, using child-directed speech, and adding semantic information to the system. Although accuracy and completeness scores showed that distributional input was effective for group merging, the scores were higher when semantic information was combined with distributional information. The work done by Cartwright and Brent show that computational model studies are a valid approach to explore theories of categorical language acquisition in children and provide further hypotheses as to the processes that occur during language learning.

Chang, F., Lieven, E., & Tomasello, M. (2008). Automatic evaluation of syntactic learners in typologically-different languages. *Cognitive Systems Research, 9*, 198-213. doi: 10.1016/j.cogsys.2007.10.002

This article explored a method of evaluating syntactic learners across multiple languages. The method used a “bag-of-words incremental generation” (BIG) task and a “sentence prediction accuracy” (SPA) measure for evaluating the data. One of the problems of assessing multiple languages that was pointed out by the authors is that evaluating corpora with a particular tagging scheme (e.g., HMM-TS2) is inherently biased towards the English language, in that it performed significantly better than when tagging the grammar of other languages. This leads to the reason why the authors chose to use the BIG task to evaluate the different corpora due to the fact that this bias is overcome. In order to truly have a typologically varied set of corpora, 12 different languages were chosen from the CHILDES database, as well as two more from the Max Planck Institute for Evolutionary Anthropology. Adult-to-child utterances were used in analysis, as well as adult-to-adult utterances, to increase the complexity of input. An analysis of paired *t*-tests were performed on the SPA results to examine the results of the various learners (e.g., Chance, Bi-gram, Tri-gram). The Chance learner was found to be statistically lower than the other learners. The authors also investigated how the BIG task and SPA analysis performed in regards to analyzing synthetic languages (e.g., French, German, Japanese) versus analytic languages (e.g., Cantonese, English). The study also found that the programs’ Adjacency-Prominence Learner performed significantly better when predicting word order than the Adjacency Learner or the Prominence Learner, suggesting that the adjacency and prominence statistics work together without interfering with each other. This research is beneficial due to system of analysis established to analyze syntax acquisition across multiple 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*(3), 296-406.

This study builds off of Mintz’s 2003 analysis of frequent frames by looking at the distributional analysis in the French language. French was chosen because it is a language that does not rely heavily on closed-class words as English does, thus allowing for errors in generalizations (e.g., due to the similarity between clitic object pronouns and determiners). French also has more flexibility for structure organization, which defies the ‘frequent frame’ structure. Language corpora were chosen from the CHILDES database and analyzed based on levels of completeness and accuracy, as seen in Mintz’s earlier study. Completeness was found to be quite low as the number of groups increases, but accuracy remained high. Results suggested that although French syntax varies from English syntax in multiple ways, a frequent frame theory of grammar acquisition continues to prove effective in either language. For further analysis, another experiment was composed to investigate the difference between front contexts (i.e., word order pairs that categorize following words) and back contexts (i.e., word order pairs which categorize preceding words). Results showed that front contexts were more accurate than back contexts. A third experiment probed the possibility of using the grouped frequent frames produce utterances, essentially a recursive application. This experiment proved to be not beneficial. Overall, frequent frames proved to be a possible theory to grammar acquisition across multiple languages and front contexts provide a richer linguistic environment for a frame analysis.

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

This chapter in Dennett’s book focused solely on Darwin’s definition of natural selection and how it can be viewed as an algorithmic process. Dennett’s persuasion is that Darwin’s natural selection was really an algorithm, by definition. “An algorithm is a certain sort of formal process that can be counted on – logically – to yield a certain sort of result whenever it is ‘run’ or instantiated” (p. 50). He goes on to say that the characteristics of an algorithm consist of substrate neutrality, underlying mindlessness, and guaranteed results. Dennett then gives multiple examples of, specifically, evolutionary algorithms and ends with how Darwin’s dangerous idea was that looking at evolution of species in the context of an algorithm is really the basis of his theory.

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.

Due to the inter- and intra-observer disagreement or discrepancies in mammogram interpretation, artificial neural networks (ANNs) have been used to better interpret the mammogram film screens. An ANN is a pattern recognition algorithm which, in this study, read the radiographic input of the mammogram and then produced an

output of a decision concerning with the likelihood of malignancy. This particular study was conducted in Hawaii and, of those that participated in the study, they identified 216 as being suspicious and confirmed through biopsy, which were then classified as being either malignant or benign. An evolutionary programming algorithm was set at 200 generations and given certain input information (e.g., the patient's age, mass size, calcification density). One of the surprising outcomes from the study was the discovery that the input element of mass size did not affect the outcome results from the ANN. This may mean that the amount of variables including in the input could be reduced so that the program can focus on the more influential variables, thus leading to more accurate outcomes.

Freudenthal, D., Pine, J., & Gobet, F. (2005). On the resolution of ambiguities in the extraction of syntactic categories through chunking. *Cognitive Systems Research*, 6(1), 17-25. doi:10.1016/j.cogsys.2004.09.003

The authors make the argument that past studies of grammatical category acquisition have used inherently flawed measures of accuracy that are not suitable for the purpose of the study. This study suggested that co-occurrence statistics are more accurate over longer units, decreasing the rate of incorrect substitutions. Specifically, Freudenthal et al. put forth that an alternative method of chunking should be used instead to measure well-formed utterances. A computational model, Mosaic, is used to show syntax acquisition. Chunked units are determined by if the same small phrases frequently occur together. These phrases are then chunked together as one unit and, in turn, speeds up the process of learning syntactic categories. The only variable that was changed in this study was the threshold of the chunking levels. It was found that this chunking mechanism decreased the number of syntactic errors produced by Mosaic. This research is beneficial because it shows that computational models have been used in past research to explore the way children acquire syntactic categories and produce language.

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

This article mainly focused on how learners acquire nonadjacent dependencies (e.g., pairing auxiliary "is" with the present progressive "-ing"). Gómez and his students tested this area of language by creating two artificial languages that could only be learned by understanding the nonadjacent dependencies. Participants were students from John Hopkins University who volunteered. The participants were trained for 18 minutes using the two artificial languages and then were tested. Results showed that there was no evidence that the learners were using first-order dependencies in higher-order ones. A second experiment was conducted involving infant participants, with auditory training lasting 3 minutes. Testing was administered via the head-turn preference procedure. The results were the same as with the adult participations. Overall, the study showed that learners, adult and infant, are sensitive to structure when acquiring a new language.

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 authors investigated verb acquisition during interactions between 57 children and their mothers. Each mother and child was videotaped in the home at the beginning of the study and then 10 weeks later. Verbs like *go*, *come*, *fall*, *drop*, and *see* were analyzed for a frequency count from the mother's language output and were then again analyzed for a frequency count in the child's language output after the 10-week period. Multiple measures (e.g., diary, checklist, and elicited production) were taken to record when the child produced each of the predetermined verbs. The authors were concerned with if certain verb characteristics affected the acquisition rate as well. The three characteristics they focused on were total frequency of the verb, the frequency in the final position of an utterance, and the variety of syntactic environments in which the verbs appeared. The authors found that the frequency of a particular verb and diversity of syntactic environments were positive predictors of usage by the child, but that verbs in the final position were a negative predictor of usage. Generally, this study showed that the type of child input influences language acquisition.

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

This study ran four experiments relating to the familiarity of a word for infants to the duration of their attention to that word in a given sentence. The authors built their research on past evidence showing that infants halfway through their first year become attuned to the sound structure and prosody of their language. Attention span was measured using the head turn preference procedure, meaning that if the infant turned his or her head and maintained it, it was due to the recognize of a previously exposed word. Experiment 1 tested 24 American infants, all of who aged 7 ½ months old. Four familiar words were trained in isolation and then subsequently presented via loudspeaker in the context of a sentence. The same was done with unfamiliar words. Results showed that 19 of the 24 infants had longer average listening times to sentences containing the familiar words than the unfamiliar words. The second experiment was conducted identically to the first experiment with the exception of using infants who were 6 months old. Results showed that 13 out of 24 infants listened longer to the sentences containing familiar words, showing a decrease of sensitivity to sound patterns in relation to the 7 ½ month old comparison group. The other two experiments were carried out with nonwords and with a reverse familiarization process (e.g., bombardment of the word embedded within a sentence, instead of isolation). Overall, the study showed that 7 ½ month old infants do have some capacity for identifying sound patterns of familiar words in 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.

The computational model created by Kiss was an information processing formulation. The model would read certain input and then analyze how strong the links were between words, giving further support to grammatical category acquisition based upon distribution. A small study was outlined to test the model using a computer program. The study involved seven families being recorded as mothers played with their children, all of which were between the ages of 1 and 3. The child-directed language resulted in a corpus of 15,000 words, a relatively small sample of what a typical child would hear in one day. When this corpus was fed through the computer program, it showed that there were strong connections between certain words (e.g., “the” and “a,” “horse” and “house,” “horse” and “cow”). These results show that words in the same grammatical category would often have the strongest connections, as the computer program recognized their frequency and distribution from the input. This study was one of the first to introduce the idea of using computational models to investigate grammatical category acquisition.

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

The article is built on the basis that infants use meaning as a foundation for language, not vice versa. One of the first examples given to support this thesis is that children who are deaf, although do not hear structured language as other infants will, do form thoughts and ideas about the world around them without the support of language. The author continues to discuss how context and language are thus associated with each other and come to be learned. Children will first learn how to name objects before they learn the object’s attributes. In order to make an argument for syntax learning in young children, the author included examples that highlighted points of normal development, but may or may not be fully due to his prediction of meaning prior to language organization (e.g., “telegraphese”). Overall, MacNamara’s hypothesis is lacking a bit in support, but give some insight to the developmental process of acquiring language.

MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language*, 28(3), 255.

The authors approached their study with a connectionist perspective, submitting that children do not acquire language via grammar rules, but rather through various cues. The study includes a computational model that is designed to mirror child language acquisition for the German language, focusing on the German declension paradigm due to the complex gender, number, and case assignments. With these three different parameters applied, there are a resulting six types of articles to choose from when coupled with a noun. The authors then go into detail describing the phonological, morphological, syntactic, and semantic cues that are associated with

the three parameters. For the method, the program was designed to detect if there were cues associated with the nouns, which were presented to the program one at a time. If it detected accurate cues, it was activated, and if not, remained off. The results of Model 1 showed high accuracy in learning and assigning the articles to nouns based on cues. Models 2 and 3 were performed to address limitations found in Model 1, with continued high accuracy. One limitation of all the models was that the authors lacked the input of German-speaking children and had to select words from a corpus of German-speaking adults. Even though this study focuses solely on acquisition of German articles and not grammatical categories, it shows language learning algorithms have been used in past studies to better understand a specific part of language acquisition.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90(1), 91-117. doi:10.1016/S0010-0277(03)00140-9

Mintz makes a research-supported argument that children and adults may learn grammatical categories through distributional patterns of speech. Mintz chose to focus specifically on frequent frames. Frequent frames are two words that often co-occur, separated by one word (e.g., you __ it, the __ one, etc.). Mintz submitted that these frames aid language learners in determining grammatical categories. In his study, he obtained six language samples from the CHILDES (MacWhinney & Snow, 1985) database, all of children who were 2;6 or younger. He looked at this smaller language sample number in order to better analyze the extent of which the input the children were receiving was informative. Analysis of frequent frames in the first experiment was determined by looking at the 45 most frequently occurring frames for each corpus, which met the criteria of occurring frequently enough to be noticeable and continued a range of intervening words. Based on calculations of accuracy, all scores showed significant results, although scores for completeness were relatively low. One limitation was that the first experiment confined the results to only 45 frequent frames. In order to overcome this particular limitation, a second experiment was conducted that was based upon the frequency of frames in comparison to the total number of frames per each corpus. Another limitation of the study was that words that belonged to multiple grammatical categories were only assigned to the most common one (e.g., no discrimination between nouns and pronouns), which was shown in the low completeness results. Despite this limitation, this research significantly advances the knowledge of how certain parts of language is acquired and is widely accepted in field. Mintz's publication lays a strong foundation that grammatical categories can be gleaned from distributional information, which can be built upon in further studies.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Bioessays*, 22(4), 425.

This research presented past theories of how language learners acquire syntactic categories, and then presented the possibility of acquisition via distributional information. Distributional information has been studied before in linguistics, but only to describe structural elements. However, the authors present neural research that shows that language in the same linguistic category is topographically similar. The study used language samples from the CHILDES (MacWhinney & Snow, 1985) database and described how words would be grouped together, using one classification. Due to the fact that the current method could not distinguish between multiple syntactic categories, words were assigned their most frequent category. The words were then organized into dendrograms and divided at different points, then scored in formulas of accuracy and completeness. Nine different experiments were then conducted to analyze various aspects of the data. A strong case was made that distributional information does aid in the acquisition of syntactic categories.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3), 210-229.

Two general methods to computer-learning programs are a neural-net approach and a network designed only to learn very specific things. The first approach is built on a punishment-rewards system while the second, which is much more efficient and is the method of choice in this article, requires reprogramming for each new application. Because machine learning can be more easily studied in a game form, the game of checkers was chosen as the system's vehicle for learning. The computer's process of learning involves a "look ahead" procedure in which the computer looks at the possibilities a few moves ahead at a time, creating a "tree" of possible moves. It then calculates and weighs options, determining the best move. The amount of look-ahead is called a ply. The article goes on to describe various ways that the program is given a sense of direction to win, how information is stored, rote-learning versus generalization procedures, etc. One of the conclusions that was found from the tests with this checker computer program is that it is possible to create programs that will outperform the average person. This article is important because it adds support to the fact that learning and acquisition has been shown in other areas of knowledge outside of syntax.

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.

St. Clair, Monaghan, and Christiansen (2010) compared frequent frames to flexible frames, using a computational model to perform the study. Multiple experiments showed the benefits of combining input information from bigrams (e.g., aX, Xb) and trigrams (e.g., aXb) into flexible frames (e.g., aX +Xb), which overcomes the weaknesses of strictly analyzing bigrams or trigrams independently. Accuracy

increased, and a larger amount of the amount of the language sample was analyzed, suggesting that allowing a less rigid distributional form may provide more information during input for children learning language and acquiring grammar.

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

Past research shows that children may use multiple cues to understand speech input (e.g., grammatical morphemes, prosody, utterance length). This study was specifically focusing on the influence of grammar and prosody had on a child's comprehension. Toddlers between the ages of 2;0 and 2;2 were chosen to participate in the study. A robot was programmed to have grammatical sentences, agrammatical sentences, sentences with nonsense words, and sentences where a pause was placed in various spots, some natural and some unnatural. The child then had to identify the correct picture based on what the robot said. Results showed that prosody was significant, grammar was significant, but the prosody and grammar interaction was not significant. This lack of interaction shows that children use both cues in comprehending language. A second experiment was conducted, this time investigating utterance position and length. Similarly, the authors found the individual factors to be significant, but their interaction not significant. Overall, this research supports the fact that children are able to draw from multiple sources to more fully understand language input.

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

This study is an extension of the Mintz (2003) research that investigated frequent frames as a means to acquire grammatical categories. A previous study (Chemla et al., 2009) showed that these frequent frames were a successful way of modeling acquisition not only in English, but also in French. The present study attempted to examine if accuracy and completeness remained high in the German language as well. German was chosen because it has a less restricted word order than French or English. The authors used German corpora from CHILDES database and replicated the Mintz 2003 study. The results showed that accuracy was significantly lower than found in the studies using English or French corpora. The distribution of words from the same category seemed random across different frames, thus showing that a single frame is insufficient in reliability to offering cues as to what category a word is a part of. The study concluded that children may use multiple and all available cues to acquire grammar.

Tettamonti, M., & Perani, D. (2012). Structure-dependency in natural language grammar. In M. Faust (Ed.) *The handbook of the neuropsychology of language* (pp. 229-242). Hoboken, NJ: Wiley-Blackwell.

This chapter in *The Handbook of the Neuropsychology of Language* discusses how the brain organizes and produces grammatical categories. The authors talk about the initial challenge that babies have in segmenting continuous input and cite past research about the use of event related potentials (ERP) and fMRIs which have showed that even young infants become quickly attune to the prosodic and phonemic discriminations of their language. As to how children acquire grammatical categories, the authors cite many past studies and also include their own opinion that children had a neurobiological predisposition or some other kind of learning ability in order to achieve full acquisition without direct instruction. The authors cited research that offers a wide variety of hypotheses as to what areas of the brain are underdeveloped or damaged, thus creating problems with the acquisition of syntax (e.g., problems with the anterior left-hemispheric brain maturation processes, lack of myelination of the arcuate fasciculus). The rest of the studies, which were given, associated hierarchical rule learning with Broca's area of the brain, which was determined to play a key role in grammatical category and syntax acquisition.

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(5), 1089-1108.
doi:10.1017/S0305000909990067

This publication expanded upon Mintz's work on frequent frames. It mainly differed by also analyzing frequent frames for Spanish samples, as well as English. Another area that was explored was that end frames were examined. End frames are defined in the study as sequences at the end of a phrase, where the end is considered the other part of the frame (e.g., *that ____*). Research supported that children are sensitive to the prosodic cues that often accompany end phrases and associate those cues as serving as boundaries. The study took 6 parent-child samples from the CHILDES (MacWhinney & Snow, 1985) archive, 3 English and 3 Spanish. The authors then identified and analyzed 45 frequent frames and 45 end frames, patterned after Mintz's study in 2003. The accuracy and completeness scores showed highly significant results overall. Homophony among function words and noun-drop did compromise some of the accuracy for the Spanish samples. This study further supports the acquisition of grammatical categories via frequent frames and end frames, while showing that this process occurs across different languages.

Wilson, J. & Henry, A. (1998). Parameter setting within a socially realistic linguistics. *Language in Society*, 27, 1-21.

This article approaches the debate of grammar acquisition from what is called the Principles and Parameters approach of Universal Grammar. That is to say that however grammar is acquired it must be done while taking into account the varying principles true to all grammar acquisition (e.g., varying grammar rules per language, varying rules when learning multiple languages, and varying rules per dialects within the same language). The article went into more in-depth examples of varying grammar, like inversion structures or strong imperative morphemes, in different dialects (e.g., Standard English versus Belfast English). The authors suggest that instead of make the distinctions of individual grammar (e.g., core linguistics) versus community grammar (e.g., sociolinguistics), as has been done in the past by Chomsky and Labov, they do, in fact, co-exist and both must be taken into account. Overall, variation can be understood by both perspectives by adopting the principles and parameters theory.