2015-03-01

External Validity of Grammatical Word Category Classification Using an Adaptation and Selection Model

Michelle Chatterton
Brigham Young University - Provo

External Validity of Grammatical Word Category Classification

Using an Adaptation and Selection Model

Michelle Chatterton

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

March 2015

ABSTRACT

External Validity of Grammatical Word Category Classification
Using an Adaptation and Selection Model

Michelle Chatterton
Department of Communication Disorders, BYU
Master of Science

The process of acquiring language requires children to learn grammatical categories and apply these categories to new words. Researchers have proposed various explanations of this process in the form of algorithms and computational modeling. Recently, adaptation and selection models have been tested and applied as a possible explanation to the process of acquiring grammatical categories. These studies have proven promising, however, the external validity of this approach has not been examined by grammatically coding samples outside the training corpus. The current thesis applies an adaptation and selection model, which pauses the evolution of dictionaries after every thousand cycles to allow the tagging of 30 outside samples, which are then checked for tagging accuracy. The accuracy across the five training corpora by the six thousandth cycle averaged 76.75%. Additional research is needed to explore the effects of altering the parameters in the model.

Keywords: grammatical word categories, evolutionary programming, language acquisition

ACKNOWLEDGEMENTS

I would like to thank the researchers who donated child language samples to the CHILDES archive as well as the families whose data were sampled. I would especially like to thank my thesis chair, Dr. Ron Channell, for his guidance, support, and encouragement throughout this process. He successfully "de-awful-ized" writing this thesis. I appreciate all of his hard work and expertise. I would also like to thank my committee members, Drs. Kristine Tanner and Shawn Nissen, for their guidance and support.

My deepest gratitude goes to my parents, Michael and Julie Chatterton, for their love and support throughout the entirety of my education. Thank you for helping me believe that "I am smart, I am confident, and I can do this."

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

DESCRIPTION OF THESIS CONTENT AND 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. 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

Language acquisition is a complex, multifaceted process. One essential part of language acquisition is organizing words into utterances. Considering the large number of words available in any natural language, this organizing task would be an insurmountable feat without extracting the commonalities of word relations, grouping conventional word usages into patterns, and grouping these patterns into categories. There is evidence suggesting that preschool children who are typically developing have acquired a knowledge of grammatical categories in that they are capable of expanding this knowledge to novel words (e.g., Berko, 1958; Ratner & Menn, 2000).

As the study of language acquisition has developed, two major theoretical perspectives have emerged. First, the nativist perspective, also known as generativist or Universal Grammar perspective, asserts that grammatical categories are innately present in a child and therefore the child's remaining task is to place each word of the language into one or more of the grammatical categories. In contrast, the constructivist perspective, also known as emergentist, socio-pragmatic, usage-based, or functionalist perspective, assumes that grammatical categories are not innately present in the child, but need to be extracted or created by detecting commonalities in the distributional use of words in the language (Ambridge & Lieven, 2011). Both the nativists and the constructivists usually compare their efforts toward explanation or explicit, computational modeling of language acquisition to grammatically coded text samples to gauge progress. Often an additional goal for constructivists is to process language input data in such a way as to allow their model or program to produce new, acceptable utterances in the target language.

During the past decades several studies have investigated how children learn to identify or construct grammatical categories. The method and design of these studies has been shaped by the researcher's chosen theoretical perspective. Algorithms, accompanied by computer models to implement these algorithms, have been created to describe and perhaps explain how children learn to identify or construct grammatical categories.

Models of Grammatical Category Acquisition

Berko (1958) investigated whether children possess morphological rules. Children ranging in age from four to seven years were presented with 27 nonsense words and asked to produce English plurals, verb tenses, possessives, derivations, and compounds of those 27 words. Nonsense words were used to ensure that the child's answer was not due to memorization, but a reflection of the child's knowledge of morphological rules. The children's answers were rated according to the answers of twelve adults previously shown the nonsense words. Results suggested that children within this range operated with clearly delimited morphological rules.

Pinker (1987) discussed the problem facing a child who begins to learn language by having only small hints or cues, a process termed *bootstrapping*, and uses this information to start forming correct grammatical rules. Pinker reviewed four proposed solutions; the first possible solution was correlational bootstrapping, which suggests that children might be sensitive to a set of distributional properties among words, such as a certain type of word always following or preceding another type of word. Pinker felt that this hypothesis was weak because not all distributional properties of words are noticeable and many of these patterns are merely incidental and thus useless for figuring out the system. Next, Pinker investigated prosodic bootstrapping. This hypothesis suggests that children retain data on the intonational contour, stress pattern, relative timing of elements, and pauses found in input sentences. However, Pinker contended that this hypothesis did not have enough evidence to support or reject it and thus would be hard to evaluate. The third hypothesis Pinker examined was syntactic bootstrapping, which postulates that possible grammars are subject to several innate constraints, making it possible for a small amount of distributional analysis to yield correct categorization of linguistic elements. This argument appeared to Pinker to be unable to solve the bootstrapping problem alone. Finally, Pinker discussed semantic bootstrapping. This hypothesis claims that children use the presence of semantic entities (such as human-ness or object-ness) to infer that the input contains tokens of the corresponding syntactic substantive universals (i.e., grammatical word categories such as nouns). This hypothesis asserts that these semantic elements are adequate conditions for learning the use of syntactic symbols. This hypothesis also faces several problems including: the semantic regularities are not completely universal, many non-basic sentences do not hold syntax-semantic correspondences, child semantic interpretation of parental input is fallible, and individual learning varies between children. Pinker suggested that while none of these bootstrapping hypotheses alone may appear viable, each might contribute to the overall solution to the language acquisition problem.

Cartwright and Brent (1997) proposed an explicit, incremental strategy as an effective way for children to categorize words. The authors discussed hierarchical cluster analysis (HCA), which is a computational model that groups words together based on similar distributional patterns. Their proposed theory has two distinct differences from HCA. First, Cartwright and Brent's strategy results in a set of discrete categories of words instead of a large number of nested categories. Second, the strategy is incremental, operating one sentence at a time and forgetting all previous sentences. This strategy uses templates, which are the sequences of word classes describing an utterance, to form what they termed generalized minimal pairs, which can then be used to group or merge words together based on distributional information. The authors proposed that before merging any words into groups, children follow a particular set of preferences. These preferences are as follows: minimize the number of templates, minimize the sum of the lengths of the templates, create templates with the highest possible frequency, minimize the total number of groups, put all instances of a word type together, minimize the number of types whose instances are divided among different groups, minimize the number of words in each group, minimize the number of groups consisting of more than one type, maximize the frequency of word types within their groups, and use large groups (in terms of numbers of types) in the templates. The authors adopted the Minimum Description Length paradigm, which provides an analytic framework to analyze inductive learning problems as optimization problems. This approach involves formulating a hypothesis and derivation. The authors hypothesized that merging would continue until no improvement of the arrangement of groups can be made.

Cartwright and Brent (1997) conducted five experiments using computer simulations to test whether this strategy would be useful in learning grammatical categories. Independent variables were manipulated during each of the five experiments. These manipulations included varying the amount of input, adding ambiguous words to the input, using child-directed speech, and adding semantic information. The accuracy and completeness scores averaged together about 50%, demonstrating that distributional input was effective for group merging; however, the scores were higher with semantic information combined with distributional information.

Redington, Chater, and Finch (1998) studied a potential solution to the problem of grammatical category acquisition, which involved looking at distributional information. Words clustered together with similar co-occurrence patterns were used to model how children might process incoming data. Redington et al. used the CHILDES database (MacWhinney, 2000) to obtain a large corpus of adult utterances collected from adult-child language conversation samples. The words in the corpus were each classified according to their most likely grammatical category, and then these hypothesized categories were compared to conventional grammatical categories. The more similar the words, the more tightly clustered they became. The clusters were highly accurate when compared to conventional grammatical categories. Several experiments further demonstrated that clustering is an informative strategy. However, it should be noted that although the obtained clusters were highly accurate, only a small portion of the corpus could be clustered.

Mintz (2003) investigated the use of frequent frames, which consist of two jointly occurring words with one word in between. The words inside a particular frame were categorized together as representing a single grammatical word category. Six corpora from the CHILDES database provided input for the analysis. Only utterances from adults to children 2;6 or younger were analyzed. Forty-five frequent frames were analyzed to determine if the words contained in each frame were from a single grammatical category. Mintz concluded that frequent frames may yield high accuracy and are effective at categorizing words. Although this study demonstrated that frequent frames produced accurate categories, Freudenthal, Pine, and Gobet (2004) criticized Mintz for using a small number of grammatical tags in the labeling scheme. Mintz used two labeling schemes, a standard labeling scheme, which only included noun, verb, adjective, preposition, adverb, determiner, wh-word, “not”, conjunction, and interjection, and an expanded labeling scheme, which further divided the noun category into pronouns and nouns, as well as the verb category into verbs, auxiliaries, and the copula. Despite using this expanded labeling scheme, Freudenthal et al. pointed out that the scheme didn’t adequately divide grammatical categories enough to make them useful or accurate in actual production.

Freudenthal, Gobet, and Pine (2004) proposed a chunking mechanism to resolve ambiguities in the extraction of syntactic categories. A generative link was created between words using a program named MOSAIC, a computational model of syntax acquisition, which reduced overall error rates and prevented the substitution of similar words in incorrect contexts by treating multi-word phrases as one unit. A simulation was conducted using data from two corpora. The corpus was used to train the model and then generate 500 utterances with a mean length of utterance (MLU) of 3.5. These utterances were scored based on syntactic errors by two human raters, which resulted in reduced error rates. While this model reduced error by preventing some incorrect substitutions, it also led to different types of substitutions and errors.

Much like Mintz (2003), Weisleder and Waxman (2010) investigated frequent frames. Their research, however, looked at frequent frames for English and Spanish across sentence positions and across different grammatical forms. Weisleder and Waxman used six parent-child corpora, three in English and three in Spanish, from the CHILDES database. After analyzing the corpora, Weisleder and Waxman reported that across both languages frequent frames proved to provide strong cues for categorizing grammatical forms. Compared to English, frequent frames for Spanish had compromised accuracy due to homophony and noun-drop. In addition, while frequent frames provide strong cues for the grammatical forms noun and verb, these frames provided weaker cues for adjectives. Findings showed that the distributional information was more accurate for mid-frames compared to end-frames.

St. Clair, Monaghan, and Christiansen (2010) proposed flexible frames as an alternate model of grammatical acquisition. The authors note that while frequent frames can result in very accurate grammatical classification, frequent frames also result in sparse data, they are difficult to use, and they only categorize words that are routinely surrounded by other frequent words. In contrast, flexible frames combine input from bigrams (e.g., aX, Xb) and trigrams (e.g., aXb). These flexible frames (e.g., aX + Xb) were tested in several experiments and showed increased accuracy, suggesting that flexible frames are more effective for learning grammatical categories than single bigrams or fixed frame trigrams alone. The results of these experiments suggest that flexible frames, with their less rigid distributional form, might provide more information for children acquiring grammatical categories.

An Algorithm Using Adaptation and Selection

In addition to the approaches discussed above, a new approach referred to as evolutionary computing has been suggested to solve the problem of learning the grammatical categories of words. Evolutionary computing applies algorithms for adaptation and selection, borrowed from the biology of evolution. Concepts such as genetics, fitness, survival, reproduction, and mutation are applied. Evolutionary programming (EP), a sub-branch of evolutionary computing, makes a population of random solutions to a problem, the fitness of this population is assessed, and the most fit member or members reproduce to form the next population, with slight variation in each "child" (Fogel, 2006). The correctness of the separate components or individual items of the solution is never given, therefore, correct and incorrect details are equally likely to change or mutate. The fitness levels of the best solutions or "children" tend to rise and over time and provide a suitable, although imperfect, solution to the problem of learning grammatical categories of words.

These evolutionary algorithms have well-known applications in attempts to explain of the emergence of species (Dennett, 1995). In addition, these algorithms have offered successful models for short-term adaptations of organisms to the environment, the actions and reactions of viruses and the human immune system, and human cognitive strategies and development (Siegler, 1996). EP has been applied to a diverse range of problem areas including the interpretation of mammograms (Fogel, 1996), programs that have the capability to play checkers or chess at a level of expertise matching humans, the design of antennae, etc. With the vast array of applications for EP, it seems logical to apply it to the problem of learning the grammatical categories of words.

Applying an Adaptation and Selection Model to Grammatical Category Acquisition

In recent studies, Cluff (2014), Judd (2014), and Young (2014) applied an adaptation and selection approach to modeling the acquisition of grammatical word categories. Using a smaller set of 16 categories (Cluff), a larger set of 85 categories (Young), or using the categories of Spanish words (Judd), the approach was found to present a rapid improvement over the first 1,500-2,000 evolutionary cycles, with slower improvement up to about 2,500 cycles, and a plateauing there after at between 80% and 90%. However, all three of these initial studies looked at the acquisition of the categories in alternate utterances drawn from the same corpora used for training. The external validity of this approach has not yet been examined by coding language samples collected outside of the training corpus. Such was the focus of the present study.

Method

Training Corpora

Five corpora (Anne, Aran, David, Nina, and Sarah) from the CHILDES database (MacWhinney, 2000) were separately used to train the computer model in the present study. Each corpus had been collected and transcribed as part of previous studies of child language acquisition. The background information available on each participant varied as it was detailed by the study's author.

Anne. Anne was one of 12 children (six boys and six girls) whose samples were collected as part of a longitudinal study of English-speaking children by Theakson, Lieven, Pine, and Rowland (2001). The children, who were mainly from middle-class families, were between the age range of 1;8 to 2;0 (years;months), with mean length of utterance (MLU) scores ranging from 1.06-2.27 at the beginning of the study. The language sampling consisted of two one-hour visits every 3 weeks for the period of one year. The samples were collected in the home of the child while they engaged in normal play activities with their mothers. The child played with their own toys during the first 30 minutes and with toys provided by the experimenter for the last 30 minutes. Anne was monolingual, the first-born child, and cared for primarily by her mother. This corpus contained 25,551 child-directed utterances.

Aran. Aran was also one of the children who were part of the Theakson et al. (2001) study, and the participant parameters and sampling procedures were the same as for Anne. Aran was monolingual and the firstborn in his family. This corpus contained 20,192 child-directed utterances (Theakson et al., 2001).

David. David was one of eight children (four boys and four girls) from upper working class families whose samples were collected over a period of 3 years, from 1995 and 1998. Each child who sampled was mainly cared for at home and was either the oldest or only child in their family. David was born on September 21, 1993, and language samples were taken between ages 2;0 and 4;2. The samples were collected in his home once a month for approximately one hour and in general, the adult-child recordings were informal. Toys were provided and role-plays with props were encouraged. This corpus contained 9,933 child-directed utterances (Henry, 1995; Wilson & Henry, 1998).

Nina. Nina's samples were collected as part of a single child longitudinal study. The samples were collected between 1972 and 1973 when she was age 1;11 to 3;3. This set of corpus contained 35,881 child-directed utterances (Suppes, 1974).

Sarah. Sarah's sampling began when she was age 2;3 and ended at age 5;1, resulting in a corpus of 139 samples. Brown (1973) and his students collected the samples between 1962 and 1966. This corpus contained 48,205 child-directed utterances. Sarah was from a working class family.

Test Corpora

The set of 30 child language samples used in the Channell and Johnson (1999) study was used as the set of test samples. These samples had previously been collected by three graduate students in speech-language pathology. The samples were collected as typically developing children, ages 2;6 to 7;11, interacted with the graduate students. There were 3 children within each 6-month interval between ages 2;6 and 6;11. There were 3 children between ages 7;0 and 8;0. Unintelligible utterances were not tagged, and utterances with one or more unintelligible words were excluded from the sample. Approximately 200 intelligible child utterances were contained in each child's sample. All adult utterances were excluded. The samples were then manually tagged as part of Channell and Johnson's study.

Instrumentation

The mc_ev program (Channell, 2015) was used in the present study. To start the program, the user sets several program parameters, the two most important being the number of evolutionary cycles that will be run and the likelihood of dictionary entry mutation, which is set as a one in X chance (e.g., 1/2,400). For its training, the program uses a text file of transcribed and grammatically coded adult utterances that were spoken in the context of a child. This input file contains one utterance per line. Each line has the format of “word <tag word <tag word <tag (etc.)” with punctuation marks removed. An output file is opened, to which the results will be written. The words used in the training corpus file are compiled into a list that will serve as the basis of the dictionaries consisting of words and their possible grammatical tags, which will evolve as the primary task of the program. Five hundred dictionaries are created with each word in each dictionary having a grammatical tag entry initially chosen at random.

The adaptation and selection process is cycled through for the number of generations that was set at program start-up. A fitness evaluation for each dictionary is completed using the tags of the words in the utterances of the input training corpus file. If that dictionary's tag for a word matches the tag for the word in that utterance of the training corpus, the score for that dictionary is incremented. After all of the candidate dictionaries have undergone this evaluation process, the 20 determined to have the highest fitness score serve as the foundation for populating the next generation. There is a mutation-likelihood-sized chance for each word in the dictionary that the grammatical tag entry for that word will be replaced with a randomly chosen one as part of creating the next generation, regardless of whether that tag entry is correct or incorrect. This mutation likelihood chance was set at program start-up. This process continues to propagate new dictionary offspring, which are then evaluated to determine the most fit dictionaries that will in turn serve as the foundation for future dictionaries.

At user-set intervals, the evolutionary process described above is temporarily paused and the best dictionary of the just-finished cycle is used to grammatically tag each word of each of the 30 test files. These tags are compared to the manually-assigned tags for the same words, and the percentage of agreement is written for each test file.

Procedure

The main task of the present study was to prepare and run the training corpora and run the test samples. This involved collating multiple language samples into a single corpus, removing the child utterances and any extraneous coding, formatting the corpus for grammatical tagging, and then checking the accuracy of the grammatical coding. The set of test language samples had been previously grammatically hand-coded, but needed to be formatted to run through the mc_ev software.

Results

The five training corpora each served as the basis for dictionaries, which evolved for 6,000 cycles each. This process was carried out using two different mutation rates, 1/1,600 and 1/2,400. In all comparisons, the level of agreement between the previously tagged training, or test corpora, and the new tagging done by an evolving dictionary, was used as the basis for evaluating accuracy, with the previous tagging used as the gold standard of correctness.

Figure 1 illustrates the mean percentage of accuracy using the 1/1,600 mutation rate for each training corpus at the first cycle and each subsequent 1,000 cycles. Generally, the accuracy of each corpus increased rapidly between the first cycle and 2,000 cycles, with continued, however less rapid growth from 2,000 to 3,000 cycles. Accuracy then very gradually improved until the final, 6,000th cycle. It can be noted, however, that the accuracy with which the test samples were tagged when the model was trained with David's corpus made notable growth between 2,000 and 3,000 cycles as well as between 1,000 and 2,000 cycles.

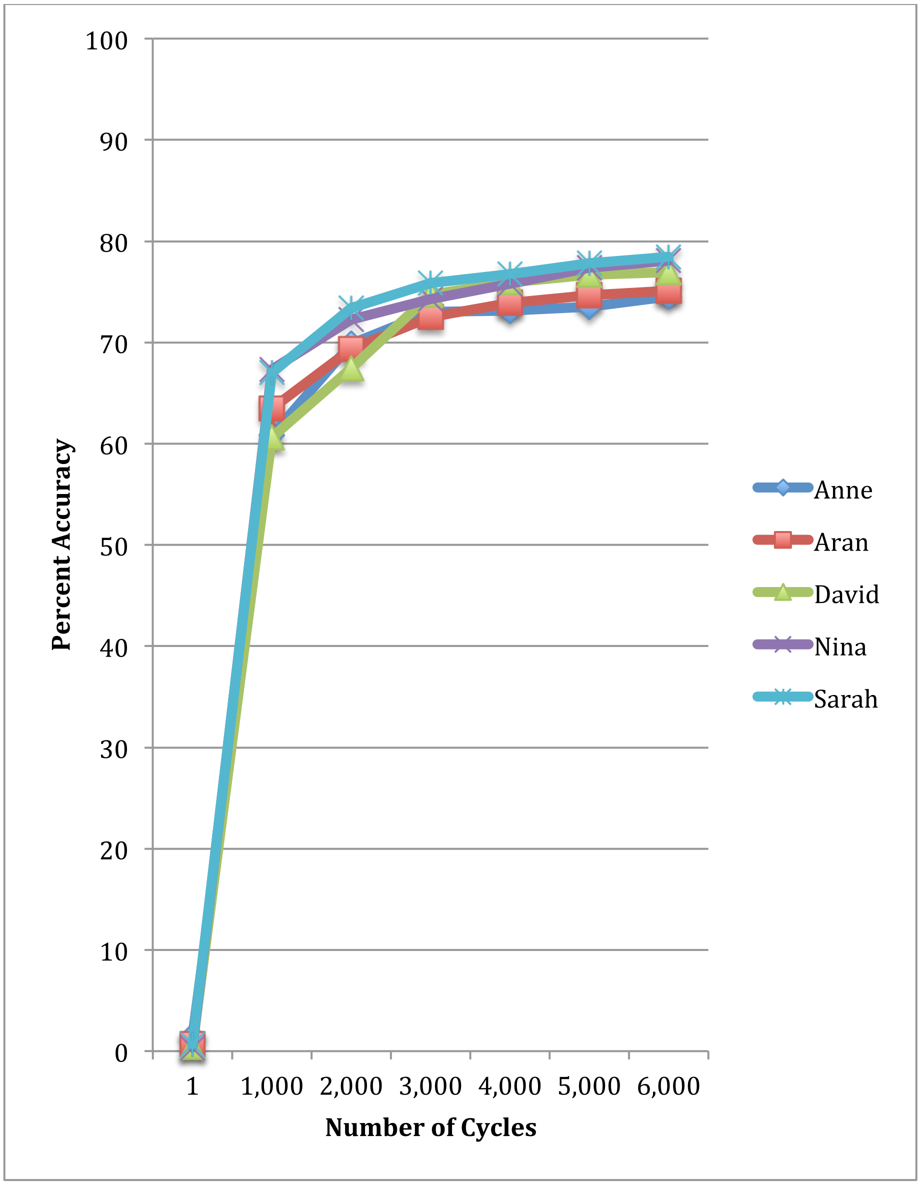

Figure 1. Mean results for each training corpus at the 1/1,600 mutation rate.

Figure 2 illustrates the mean percentage of accuracy with the 1/2,400 mutation rate for each corpus at the first cycle and each subsequent 1,000 cycles. Similar to the results of the 1/1,600 mutation rate, the most rapid improvement in accuracy was noted between the first cycle and 2,000 cycles. Notable improvement, although less rapid, was noted between 2,000 and 3,000 cycles, with gradual improvement in accuracy continuing through the 6,000th cycle.

Comparing the accuracy of both mutation rates at 1,000 cycles, the accuracy for the 1/2,400 mutation rate is slightly lower. However, accuracy for the 1/2,400 mutation rate improves as the cycles progress and is comparable, in fact rises slightly above, the accuracy for the 1/1,600 mutation rate by the 2,000th cycle. From that point on, accuracy for the 1/2,400 mutation rate remains slightly elevated compared to the accuracy of the 1/1,600 mutation rate.

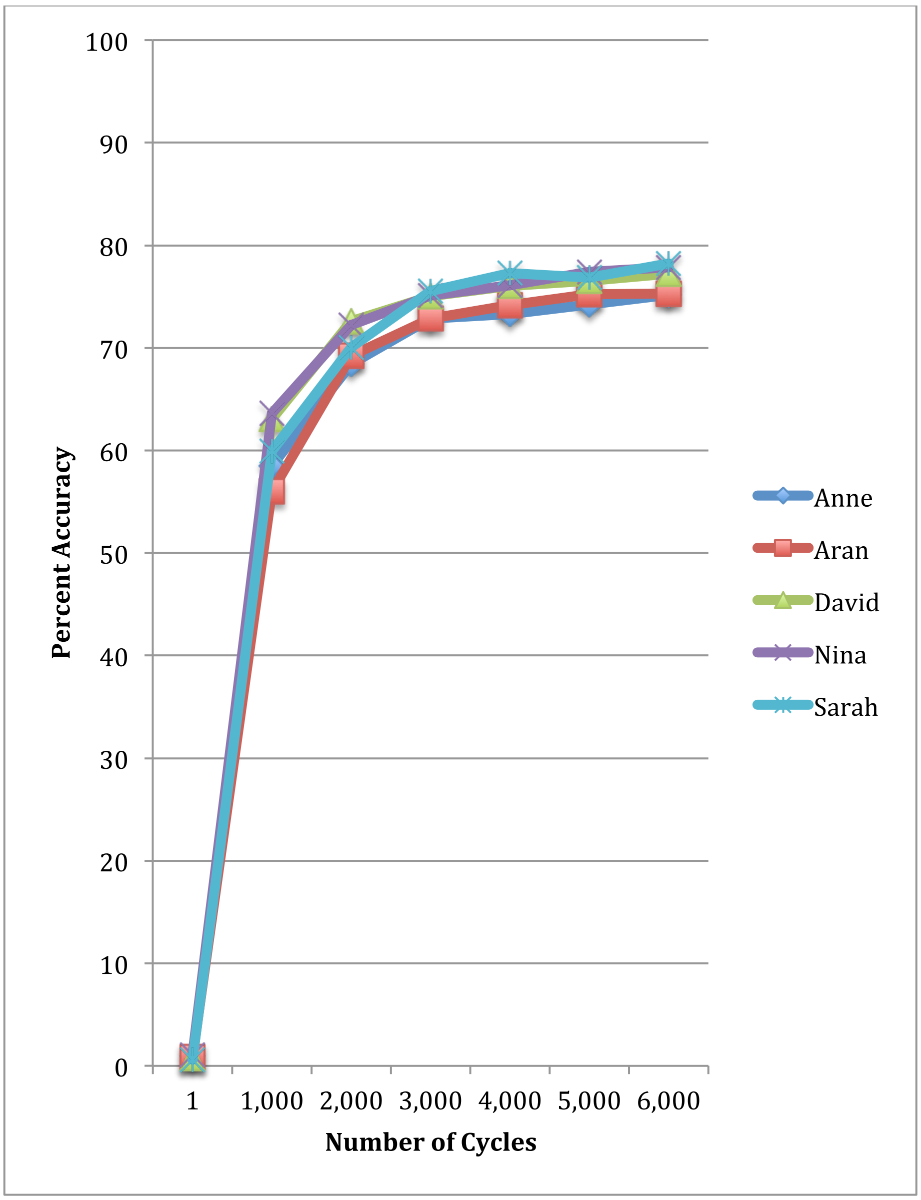

Figure 2. Mean results for each training corpus at the 1/2,400 mutation rate.

The mean accuracy levels across all five corpora for each of the 30 Provo samples with the 1/1,600 and 1/2,400 mutation rate are contained in Table 1. There is higher variability in accuracy at 1,000 cycles among the Provo samples, with accuracy becoming more comparable with subsequent cycles. The decreasing range among the Provo samples as the number of cycles increases illustrates this. The range of accuracy at 1,000, 3,000, and 6,000 cycles with the 1/1,600 mutation rate is 10.12, 7.57, and 6.91 respectively. The range of mean accuracy at 1,000, 3,000, and 6,000 cycles with the 1/2,400 mutation rate is 9.69, 7.69, and 7.00 respectively. This is comparable to the range of the 1/1,600 mutation rate mean accuracy.

Pearson's correlations between the accuracy observed at the last (6,000th) cycle and the test child's MLU or the number of utterances in their sample, whether for mutation rates of 1/1,600 or 1/2,400, were not significant.

Discussion

The purpose of this study was to investigate the accuracy of the evolutionary algorithm software program by comparing it to real-world data. This goal was achieved by looking at separately tagged corpora every thousand cycles up to six thousand cycles. In general, the trend across the five training corpora, regardless of mutation rate, was to make rapid gains in accuracy initially between the first cycle and the 2,000th cycle. Accuracy continued to improve between 2,000 and 3,000 cycles, however, less rapidly than the gains made between the first cycle and 2,000 cycles. Accuracy between 3,000 and 6,000 cycles improved very gradually. A 12,000 cycle trial was implemented for one training corpus, however, improvement between 6,000 and 12,000 cycles was minimal. Initially, mean accuracy for the 1/2,400 mutation rate was slightly decreased compared to the mean accuracy for the 1/1,600 mutation rate. However, accuracy for

Table 1

Mean Accuracy Results of Each Sample at 1,000, 3,000, and 6,000 cycles at the 1/1,600 and 1/2,400 Mutation Rates

	1/1,600 Mutation Rate			1/2,400 Mutation Rate		
Samples	1,000	3,000	6,000	1,000	3,000	6,000
Aaron	65.75	76.03	78.06	62.86	75.26	78.29
Aimee	63.14	73.08	76.29	59.49	74.15	76.56
Alisha	63.88	74.81	76.54	60.96	74.79	76.69
Amber	64.93	76.90	79.22	60.45	77.02	79.49
Ambree	62.02	72.00	74.65	58.88	72.23	75.19
Andrus	68.03	76.80	79.46	65.52	77.08	79.30
Ashley	62.91	74.75	75.83	59.88	75.17	75.96
Ashley B	62.81	74.10	77.24	61.80	74.60	77.46
BJ	66.06	75.23	77.20	61.69	75.09	77.80
Christine	67.15	76.14	78.51	60.12	76.92	78.49
Clarissa	62.41	72.98	76.04	57.23	73.17	76.42
Cody	61.61	70.94	73.39	57.15	71.81	73.89
Elizabeth	63.20	73.29	75.29	58.83	73.48	75.05
Heather B	64.93	74.19	76.85	59.51	73.83	76.91
Heather C	62.45	72.01	74.00	59.37	71.94	73.80
Jack	61.14	72.85	75.71	58.10	72.94	76.48
Jarom	66.36	74.96	76.58	61.06	74.68	76.30
Katie	64.70	74.94	77.15	61.60	75.25	77.49
Kevin	64.89	74.75	76.59	62.75	74.62	76.48
Kile	59.41	70.26	73.78	57.22	70.85	73.84
Michael	63.17	72.92	76.67	59.44	73.18	76.51
Patrick	66.95	75.29	77.80	63.36	75.51	78.41
Rebecca	57.91	69.58	73.03	55.84	69.50	73.61
Rebekah	66.35	77.16	79.93	62.73	77.19	80.60
Sarah	63.33	74.50	77.72	58.45	74.80	77.68
Scott	62.73	73.98	77.28	60.93	75.11	77.03
Talon	61.60	72.67	74.72	58.67	73.22	74.46
Tavida	66.11	75.39	78.90	61.64	75.77	78.73
Tiffany	63.48	73.47	75.68	58.48	73.05	75.57
Toinette	65.99	76.27	78.45	62.73	76.24	78.08

the 1/2,400 mutation rate was slightly increased compared to the accuracy for the 1/1,600 mutation rate for the remaining cycles.

Recent studies have explored the use of an evolutionary algorithm in the area of grammatical category acquisition. Cluff (2014) examined the effects of mutation rate on the accuracy achieved by an adaptation and selection computational model. Cluff was able to achieve accuracy scores in the 90% range and three possible tags were allowed per utterance. Young (2014) also used an adaptation and selection computational model, which investigated the effects of evolutionary cycles and mutation likelihood on token and type accuracy. Similar to Cluff, accuracy scores reached the 90% range. Judd (2014) applied an adaptation and selection computational model to Spanish language corpora. Accuracy rates reached about 85% for word tokens and about 72% for word types. A limitation of these studies is that odd-numbered utterances within the same corpora were used for fitness evaluation, with the even-numbered utterances being used to quantify the accuracy of the dictionary as it generalized. In contrast, the current study used an external corpus to determine accuracy of the evolved dictionaries, which established a stricter standard for grading.

While the current study is similar to previous research aimed to investigate the process of acquiring grammatical categories, several fundamental differences are present. As discussed earlier, Cartwright and Brent (1997) explored a computational model, which grouped words with similar distributional patterns together. Applying an incremental strategy, templates were used to form generalized minimal pairs that could be applied to group or merge words together. This strategy resulted in discrete categories of words. Similarly, Reddington et al. (1998) applied a model based on distributional information, which clustered words together with similar co-occurrence patterns. However, only a small portion of the corpus could be clustered. Mintz

(2003) proposed the use of frequent frames, which tagged words based on their distributional pattern between two jointly occurring words with one word in between. As with Reddington et al., Mintz only applied frequent frames to a small portion of the corpus. In contrast, the current study used each corpus in its entirety. Ultimately, it is difficult to compare previous research efforts to the current study due to the uniqueness of using an adaptation and selection computational model. This variety of approaches and methods illustrate that investigating the process of acquiring grammatical categories is an area of interest and an endeavor of many researchers.

The success of this adaptation and selection computational model to investigate the acquisition of grammatical categories suggests a number of questions meriting inquiry. Some samples resulted in higher accuracy rates compared to others. Perhaps this might be due to some training corpora incidentally sharing a larger number of words with the test samples, or perhaps it might be due to the developmental level of the child spoken to in the training corpus; these are questions for further research. Superficial differences, such as having the training corpora drawn from various English dialects, did not seem to affect the training, however the effect of other variables remains unexplored at this time.

There are several limitations of the present study. Only five training corpora were used and only two mutation rates, 1/1,600 and 1/2,400, were tested. The model was trained on a successively moving window of 5,000 utterances at a time from a training corpus, limiting our knowledge of the effect of using a larger or smaller window. The current study implemented a model setting of 500 evolving dictionaries and allowed the 20 most fit dictionaries to reproduce. The effect of allowing more or fewer dictionaries to evolve and reproduce are unknown at this time. The parameter settings implemented in the current study leave several areas unexplored and thus needing further research.

Nevertheless, the present study offers useful information about the validity of applying an adaptation and selection approach to modeling the acquisition of grammatical categories. It suggests that this approach is robust enough to accurately code independent samples, and it implies that an adaptation and selection computational model of grammatical category acquisition warrants further development and exploration.

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.

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

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.

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

Channell, R. W., & Johnson, B. W. (1999). Automated grammatical tagging of child language samples. *Journal of Speech, Language, and Hearing Research, 42,* 727-734.

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

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

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

Fogel, D. B., & Fogel, L. J. (1996). An introduction to evolutionary programming. *Proceedings of Evolution Artificielle 95*, Springer-Verlag, Berlin, 21-33.

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.

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

Judd, C. L. (2014). A model of grammatical category acquisition in the Spanish language using adaptation and selection. (Unpublished master's thesis). Brigham Young University, Provo, UT.

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

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

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

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: Erlbaum.

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

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.

Suppes, P. (1974). The semantics of children's language. *American Psychologist*, *29*, 103-114.

Theakson, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: an alternative account. *Journal of Child Language*, *28*, 127-152.

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

Young, T. (2014). A model of children's acquisition of grammatical word categories using an adaptation and selection algorithm. (Unpublished master’s thesis). Brigham Young University, Provo, UT.

Appendix: Annotated Bibliography

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

This book reviews the major theoretical approaches to language acquisition. On one side there is the generativist approach, which asserts that some important aspects of linguistic knowledge are innate in children. On the other side is the constructivist approach, which asserts that the ability to learn language is innate, but children must generalize across speech they hear to acquire grammatical categories. In this book, chapter six specifically addresses how children learn the word-order 'rules' of their language. While there are strengths and limitations to each approach, each offer valid components and interesting perspective.

The generativist approach suggests that children use semantic bootstrapping, meaning that children innately have a set of linking rules to connect syntactic categories and semantic categories. Another hypothesis is prosodic bootstrapping, which suggests that prosodic information such as pauses, syllable lengthening, and pitch contours are used to divide an utterance into clauses and then into syntactic phrases. Additionally, there is the idea of distributional analysis, which asserts that children group words together that appear in similar distributional contexts.

The constructivist approach includes semantic assimilation, which states that children's earliest grammar is entirely semantic and that events and entities that are not true actions or agents will be assimilated into the correct category based on their similarity to real actions and agents. In addition, the constructivist approach includes distributional learning. This model uses a clustering algorithm, which groups words together based on the number of contexts, meaning the preceding and following word, that they share.

While these two approaches differ significantly from each other, the use of phonological cues is one potential learning strategy that is plausible with both. This strategy cites that words from different syntactic categories generally differ in the number of phonological properties. Learning how children assign words to syntactic categories is challenging, therefore, it is important to understand both theoretical approaches and the evidence that drives each.

Berko, J. (1958). The Child's Learning of English Morphology. *Word*, *14*, 150-177.

This study investigates the question: do children possess morphological rules? Children ranging in age from four to seven years were presented with 27 nonsense words and asked to produce English plurals, verb tenses, possessives, derivations and compounds of those words. Nonsense words were used to ensure that the child's answer was not due to memorization, but it reflected inner knowledge of morphological rules. The children's answers were rated according to the answers of twelve adults previously shown the nonsense words.

Result showed children within this age range operate with clearly delimited morphological rules. Both sexes did equally well, but there were differences between preschoolers and first graders. Because both groups were not qualitatively different in that they employed the same simplified morphological rules, the groups were combined for this analysis.

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

This article proposes an explicit, incremental strategy as an effective way for children to categorize words. The authors discuss hierarchical cluster analysis (HCA), which is a computational model that groups words together based on similar distributional patterns. Their proposed theory has two distinct differences from HCA. First, Cartwright's strategy results in a set of discrete categories of words instead of a large number of nested categories. Second, their strategy is incremental, operating one sentence at a time and forgetting all previous sentences. This strategy uses templates, which are the sequences of word classes describing an utterance, to form generalized minimal pairs, which can then be used to group or merge words together based on distributional information. The authors propose that before merging any words into groups, children follow a particular set of preferences. These preferences are as follows: minimize the number of templates, minimize the sum of the lengths of the templates, create templates with the highest possible frequency, minimize the total number of groups, put all instances of a word type together, minimize the number of types whose instances are divided among different groups, minimize the number of words in each group, minimize the number of groups consisting of more than one type, maximize the frequency of word types within their groups, and use large groups (in terms of numbers of types) in the templates. The authors adopt the Minimum Description Length (MDL) paradigm, which provides an analytic framework to analyze inductive learning problems as optimization problems. This approach involves formulating a hypothesis and derivation. The authors hypothesize that merging will continue until no improvement of the arrangement of groups can be made.

Five experiments were conducted using computer simulations to test whether this strategy would be useful in learning grammatical categories. The first experiment used artificial language based on a simple template grammar. Results suggested that this learning strategy is effective at categorizing grammatical categories from a small sample in a short amount of time. The second experiment added ambiguous words, which proved the strategy performs well on simple, distributionally defined artificial language input even with ambiguity. Experiment three used spontaneous, child-directed English, which resulted in a better categorization job than on either baseline program. Experiment four was the same as the previous experiment, with the exception of using substantially longer input files. Results were slightly more conservative than on experiment three. Experiment five tested whether semantic information would improve performance. Results were improved, suggesting that even a small amount of semantic information can lead to more correct group merges and improved performance. The authors admit additional experiments are warranted to expand the evidence cited by these experiments.

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

This article presents a method that uses a bag-of-words incremental generation (BIG) task and sentence prediction accuracy (SPA) to evaluate how well syntactic learners have acquired syntactic knowledge from the input. This method is applied to several formally-specified learners and to a new learner, the Adjacency-Prominence learner. The BAG task involves putting the words from the utterance into an unordered bag-of-words. The learner attempts to predict the utterance incrementally, one word at a time. The SPA measure is the percentage of complete utterances correctly produced.

Twelve corpora from the CHILDES database, each in a different language, were selected, as well as two larger English and German-Dense corpora. All corpora involved interactions between a child and at least one adult. Five learners were tested in adult-child and adult-adult situations including bigram, trigram, bigram+trigram, unigram+bigram+trigram, and backed-off trigram. The backed off trigram learner would use trigram statistics if available, the back off to bigram statistics if trigrams weren't available, and then back off to unigram statistics if the other two were not available. Results showed that trigrams are more informative than bigrams and the bigram+trigram learner is better able to use sparse data than just the bigram learner. Surprisingly, results suggest that unigram frequencies in certain language typologies may have a negative affect on word ordering processes, as the unigram+bigram+trigram learner was numerically worse than the bigram+trigram learner.

Finally, this study used the Adjacency-Prominence-type syntactic learners. This model uses dual-pathways, the sequencing and meaning pathways. The model learned different types of information in each pathway and this information was integrated in production. Results showed that the Adjacency-Prominence learner worked better than the bigram learner across all 14 corpora. The SPA results suggested that prominence statistics are biased for analytic languages as compared to synthetic languages.

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

This article compares five algorithms against real corpora of children's speech. The authors implemented an evaluation measure called Word Order Prediction Accuracy (WOPA), which was used to compare six syntax acquisition algorithms. Twelve typologically diverse corpora were used, as well as two larger English and German corpora. Each algorithm used a dual-pathway model including a sequence system, which produced a context statistic, and a meaning system, which produced an access statistic. Results showed lexically specific knowledge to be more advantageous than knowledge of broad-based categories for predicting word order. However, none of the learners acquired syntactic categories (e.g., nouns, verbs, adjectives, determiners), leading the authors to propose that a combination of broad and specific categories might lead to improved results.

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*, 296-406.

This study is an extension of the work of Mintz (2003), who looked at the use of frequent frames in the English language. This study applies frequent frames to the French language to examine the effects of a language, which does not rely heavily on closed-class words and also has a greater number of determiners. Three experiments were implemented using corpora from the CHILDES database. Completeness and accuracy score were analyzed. The first experiment revealed that frequent frames produced accurate groups of content words in the French language. It was noted that, although French has more varied and ambiguous function words, frequent frames were able to categorize content words. The second experiment analyzed the effect of front contexts and back contexts as opposed to frames. While front contexts produced slightly better results compared to back contexts, frames were found to have far better results than either of the alternate contexts. The third experiment tested the effects of the recursive application of frequent frames. Results showed that this procedure did not lead to any sort of benefit.

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

In this selection Dennett outlines how natural selection is an algorithmic process. Dennett explains that an algorithm is a formal process that can logically be depended on to lead to a particular result. He includes substrate neutrality, underlying mindlessness, and guaranteed results as characteristics of an algorithm. Additional examples of evolutionary algorithms, such as tournaments and the process of annealing metal, are given. Dennett asserts that although controversial, Darwin introduced the idea of evolution stemming from an algorithmic process.

Dickinson, M., & Jochim, C. (2009). Categorizing local contexts as a step in grammatical category induction. *Proceedings of the EACL 2009 Workshop on Cognitive Aspects of Computational Language Acquisition* (pp. 34-41), Athens. Association for Computational Linguistics.

This article investigates treating contexts, or frames, as categories, attempting to explain what role contexts contribute on their own in defining grammatical categories. This study used the Wall Street Journal portion of the Penn Treebank with 36 tags. Defining frequent frames as those occurring at least 200 times, there was 79.5% token precision, showing that frames are not fully able to disambiguate corpus categories.

In addition, this study uses lexical information to combine contexts, while preserving the intended category. Overlapping the word sets of frames that contain the same word resulted in more categories. Token-based overlap was used, which is consistently more accurate than type-basked overlap. Results confirmed that merging frames resulted in higher precision and the resulting categories extended to a large portion of the corpus. To test this method on a language with freer word order and more morphological complexity, the German TIGER corpus was used. The results were on par with English, although the study acknowledges this might be due to the majority of the frames being singular or mass noun. This study did not investigate splitting frames, which merits further research. Additional experiments on more corpora, with different tagsets, and for different languages are warranted to further verify these findings.

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.

This study applied an evolutionary programming algorithm to mammogram interpretation. The use of artificial neural networks (ANNs) offers benefits to radiologists, because there is inter- and intra-observer disagreements or inconsistencies when interpreting mammograms. The ANN is an algorithm that uses pattern recognition as it reads radiographic input from the mammogram and considers age of the patient to determine the output decision, or malignancy likelihood. This study, which was conducted in Hawaii, identified 216 cases, which presented suspicious characteristics. These cases were then analyzed and malignancy was confirmed in 111 cases by surgical biopsy. The evolutionary program was trained and run for 200 generations. A higher threshold value resulted in lower false positive rates and decreased sensitivity. In contrast, a lower threshold value resulted in higher false positive rates, however sensitivity was increased. The study found that mass size was not relevant to diagnosis and therefore, the authors suggest that with improved emphasis on critical features, the use of ANNs will increase accurate outcomes.

Freudenthal, D., Gobet, F., & Pine, J.M. (2004). On the resolution of ambiguities in the extraction of syntactic categories through chunking. *Cognitive Systems Research, 6,* 17-25.

This article proposes a chunking mechanism to resolve ambiguities in the extraction of syntactic categories. A generative link is created between words using MOSAIC, a computational model of syntax acquisition, which reduces overall error rates and prevents the substitution of similar words in incorrect contexts by treating multi-word phrases as one unit.

A simulation was conducted using data from two corpora. The corpus was used to train the model and then it generated 500 utterances with an MLU of 3.5. These utterances were scored based on syntactic errors by two human raters, which resulted in reduced error rates.

While this model reduces error by preventing incorrect substitutions, it also leads to different types of substitutions and errors. The strengths of MOSAIC as a model of language acquisition are the use of realistic child-directed speech and the opportunity to compare production of utterances with child speech.

The goal of this article is to propose a possible solution to the difficulty of co-occurrence statistics resulting in utterances that deviate substantially from child speech. This model offers a valid proposal and also reminds researchers to use caution when relying on measures of accuracy to evaluate the quality of syntactic categories because the addition of production may display reduced quality of the categories.

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

This study investigated how nonadjacent dependencies are acquired. Nonadjacent dependencies pose a challenge by requiring learners to form relations with irrelevant material in the middle. Participants of the first experiment included 48 university students. The experiment required each participant to complete training by listening to one of two artificial languages, which could only be learned by understanding nonadjacent dependencies, for approximately 18 minutes. Participants were informed that these auditory strings followed word order rules and that they would later hear 12 strings, half which would follow the same word order of the training strings and half that would violate that set of rules governing word order. The participants were asked to indicate whether the string complied with or violated the rules governing word order. Results showed that a great proportion of learners who heard the 24 set size demonstrated perfect discrimination. Evidence suggested that learners were not embedding first-order dependencies in higher-order dependencies. A second experiment, similar to the first was conducted with infant participants. Auditory training was conducted for 3 minutes and a head-turn preference procedure was implemented. Results showed that the infants, as well as the adults, had increased discrimination for the largest set size and that there was no evidence of embedding first-order dependencies in higher-order ones. Results of both experiments demonstrate that human learners are sensitive to and guided by statistical structure.

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 investigates whether infants can recognize words that occur in various sentential contexts. The first experiment specifically looked at whether infants, who were familiarized with a certain word in isolation, demonstrated a tendency to listen longer to sentences with that word compared to sentences without it. Participants consisted of 24 American infants, approximately 7 ½ months old. The infants were trained by listening to two words in isolation. During the test phases, the infants listened to four six-sentence passages. A modified version of the headturn preference procedure was implemented. Results showed that 19 of the 24 infants listened for significantly longer average times to passages with familiar words than unfamiliar words. Experiment two was a replica of the first experiment with the exception that participants were approximately 6 months old. Results showed that 13 of the 24 infants listened for longer average times to passages with familiar words compared to unfamiliar words, however results were not found to be significant. This indicates that distinguishing familiar words in fluent speech may be an ability that develops after 6 months of age. Experiments 3 involved infants 7 ½ months old, however nonwords that were minimal pairs of the original test words were used during the training phase. The test phase consisted of the original passages. Results showed that 11 of the 24 infants listened for longer average times to the passages with the "familiar" words (the words that were minimal pairs of the trained words) compared to the "unfamiliar" nonwords, which was not found to be significant. This result indicated that infants were not simply matching vowels or syllable rimes. Experiment 4 used a reverse familiarization process, meaning training consisted of familiarizing the infant with passages and testing the infant using isolated words. Results showed that 18 of the 24 infants had longer average listening times for the words that were from familiar passages compared to unfamiliar words. This result was found to be significant. Overall, these experiments show that 7 ½ month old infants have some ability to detect the sound pattern of some words in the context of fluent speech.

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

This study examines the application of a computational model involving information processing formulation to explain the early learning of word classes by children. Input corpora were collected from recordings taken of the interaction between 7 mothers and their children. The children were between 8 and 35 months old and came from middle class families. This resulted in a 15,000 word corpus, which provided a small sample of the type of speech a child would hear. This corpus provided input for the computer program, which uses associative learning, which is forming associative links between words and their contexts, and classification learning, which is establishing internal representations for groups of words based on the similarity of distribution. Results of the study showed that the learning process proposed in this study is capable of identifying word class features from input corpora. The results of this study agree with other findings of early language acquisition in terms of the order word classes emerge.

MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language, 28*, 255-277.

This study investigated the application of a computational model created to model language acquisition of German children. This approach encompasses a connectionist perspective, which asserts that language is acquired through the use of various cues. The focus was on the German declension paradigm, which is due to complexities involved in assigning gender, number, and case in German. The authors detail several types of cues, including morphological, phonological, semantic, and syntactic cues, that are associated with assigning gender, number, and case. The model was designed to look at nouns and the article that accompanies the noun. The nouns were presented to the model one at a time and if a cue was detected for a given noun, it was activated, if a cue was not present, it remained off. Model 1 resulted in high accuracy, which learned and assigned articles to given nouns based on cues. Two main criticisms of Model 1 resulted in two additional models. Model 2 was a replica of Model 1 with one exception of excluding the 11 arbitrary features, which were used in Model 1 to decrease confusion between lexical items. Without these features included, performance dropped significantly. Model 3 did not include the 11 arbitrary features and also depended on the raw phonological form of the stem. Results were improved compared to both of the previous models in terms of training and generalization. This study is an example of the use of algorithms to further understand the acquisition of particular aspects of language.

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

This article discusses frequent frames, two jointly occurring words with one word in between, that Mintz asserts are possible units that language learners pay attention to. The words that occur inside a particular frame are categorized together. Mintz conducted two experiments to test the claim that these units might be used to categorize words. Six corpora from the CHILDES database provided input for the analysis. Only utterances from adults to children 2;6 or younger were analyzed. Forty-five frequent frames were analyzed to determine if the words contained in each frame were from one grammatical category. The findings of these studies concluded that frequent frames yield high accuracy and are effective at categorizing words.

Although this study demonstrated that frequent frames produce accurate categories, for a given grammatical category there was often more than one frame-based category. In addition, there were some frames belonging to different grammatical categories that were grouped together. This study states that further research is warranted to determine if children actually use frame-like information to categorize words.

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.

This study investigates how children acquire verbs. Specifically, this study examines the hypothesis that verbs that are frequently found in child input, often in utterance-final position, and appearing in diverse syntactic frames, are used more frequently and in more diverse ways. Participants included 57 children and their mothers. Interactions between the child and mother at home were videotaped twice, once during an initial visit and then 10 weeks later. Frequency of verbs spoken by the mothers were analyzed after the initial language collection and then a frequency count from the child's language output was collected following the 10 week interim. Results showed that verbs that were frequently used in the input were, in turn, used more frequently and with more flexibility by the child after 10 weeks. Additionally, verbs that were used in diverse syntactic frames in the input were found to positively predict more frequent and syntactically divers use by the child after the 10-week interim. However, findings revealed that utterance-final position was a negative predictor of a child's subsequent flexibility of use and did not predict the frequency of use by the child. Ultimately, this study demonstrated that the input a child receives has an impact on their subsequent language acquisition.

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

This chapter discusses the problems facing bootstrapping in language acquisition. The author asserts that a viable theory of language acquisition must show how children acquire the rules of language. The author reviews several proposed solutions to the bootstrapping problem.

First, correlational bootstrapping, which suggests that children are sensitive to a set of distributional properties, was discussed. This hypothesis presents weakness because not all distributional properties are noticeable and many of them are useless. Next, prosodic bootstrapping was investigated. This hypothesis suggests that children record the intonation contour, stress pattern, relative timing of elements, and pauses found in input sentences. However, this hypothesis does not have enough evidence and it is hard to evaluate. The third hypothesis examined was syntactic bootstrapping, which states that possible grammars are subject to several innate constraints making it possible for a small amount of distributional analysis to yield correct categorization of linguistic elements. This argument appears to be unable to solve the bootstrapping problem alone.

Finally, semantic bootstrapping was discussed. This hypothesis claims that children use the presence of semantic entities to infer that the input contains tokens of the corresponding syntactic substantive universals. It asserts that these semantic elements are adequate conditions for the use of syntactic symbols. This hypothesis also faces several problems including: the semantic regularities are not completely universal, many non-basic sentences do not hold syntax-semantic correspondences, child semantic interpretation of parental input is fallible, and individual learning varies between children. The author suggests that while none of these hypotheses may appear viable, each contribute to the overall solution to the bootstrapping problem.

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

The aim of this article is to demonstrate that the distributional properties of words can be highly informative of syntactic categories. The distributional approach used in this study includes three stages. The first stage measures the distribution of contexts where each word occurs, which means collecting information about the distribution of words that occur near the target word. In the second stage they compare the distributions of pairs of words, finding that this serves as a cue as to whether two words belong to the same syntactic category. In the third stage they look at grouping together words with similar distributions.

Adult speech taken from the CHILDES database was analyzed. This large corpus from 6,000 speakers contained several million words of speech. One thousand target words were used and these were classified to their most common syntactic category. Scoring was measured according to accuracy and completeness, with overall performance level dependent on achieving both simultaneously. Results indicate that distributional information provides powerful cues for learning syntactic category membership.

A number of additional experiments were conducted to gather additional information about this model. Findings show that best results were obtained when combining the preceding and succeeding context. This method was effective even with a small vocabulary of target and context words and is most effective for learning nouns, then verbs, and least effective for function words. This method proved to work with or without explicitly marked utterance boundaries, but it benefits when frequency information is included. When function words are removed the method works slightly less well, but interestingly, this method is as effective with adult-adult speech as with the child-directed adult speech taken from the corpus. While keeping a neutral stance as to the nativist and empiricist approaches to language acquisition, this model asserts that distributional information is potentially a cue for acquiring syntactic categories.

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

This article investigates computer-learning programs. The author begins by contrasting two general methods used, a neural-net approach, which uses a punishment-rewards system, and a network created to learn very specific things, which is more efficient and becomes the focus of this study. The computer program is given the task of learning and playing checkers. The learning occurs as the program looks ahead and complies a “tree” of possible future moves. The program then weighs various options and calculates the best move. The distance of the look-ahead, also known as the ply, is dynamic and varies according to situation. The article outlines other specifics of the program design. It also discusses two different types of learning, rote-learning and generalization procedures. Both types of learning showed benefit in certain situations. The results of these experiments demonstrated that it is possible to develop learning schemes that can outplay the average person. The authors propose that learning schemes, like the one discussed in their article, may prove applicable to real-life problems in the future.

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

This study examined the effect of prosody and grammar on the comprehension of children. Children, ages 2;0 to 2;2 heard sentences by a robot programmed to read grammatical, ungrammatical, and nonsense sentences, with early natural, late natural, and unnatural prosody. The child was then asked to identify the correct picture based on the sentence read by the robot. Results showed that children are sensitive to prosody and grammar, however the interaction between prosody and grammar was not found to be significant. A second experiment, similar to the first, was conducted to determine the effect of utterance position and length. Results showed that these caregiver cues play an important role in comprehension, with utterance-final targets and shorter utterance length resulting in better comprehension. Additionally, results of both experiments suggest that prosody, grammar, and caregiver cues all contribute to the increased comprehension of children, and that the use of one type of cue does not minimize the use of another type of cue.

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.

This study proposes flexible frames as a new model for grammatical acquisition. The authors note that frequent frames can result in very accurate grammatical classification, however, they assert that frequent frames result in sparse data, they are difficult to use, and they only categorize words that are routinely surrounded by other frequent words. Flexible frames combine input from bigrams (e.g., aX, Xb) and trigrams (e.g., aXb). These flexible frames (e.g., aX + Xb) were tested in several experiments and showed increased accuracy, suggesting that flexible frames are more effective for learning grammatical categories than single bigrams or fixed frame trigrams alone. The results of these experiments suggest that flexible frames, with their less rigid distributional form, provide more information for children acquiring grammatical categories.

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 aim of this study was to determine the accuracy of applying frequent frames to German as a means of acquiring grammatical categories. Frequent frames were previously implemented by Mintz (2003) in English and later successfully extended to French by Chmela, Mintz, Bernal, & Christophe (2009). However, frequent frames were applied to Dutch by Erkelens (2009), which did not enable accurate lexical categorization. This motivated the authors to look at German, which is similar to Dutch in several ways. Compared to English and French, German is characterized by a less restricted word order. This study used a longitudinal corpus of child-directed German speech from the CHILDES database. Results revealed that although all scores differed significantly from chance categorization, distribution of words from the same category seemed arbitrary across frames. As with previous studies, the authors concluded that a single frame is not adequate to provide a reliable cue to one category.

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.

This article explores the distributional information contained in frequent frames for English and Spanish, across sentence positions, and across different grammatical forms. This study used six parent-child corpora, three in English and three in Spanish, from the CHILDES database. After analyzing the corpora it was reported that across both languages frequent frames proved to be strong cues for categorizing grammatical forms.

Compared to English, frequent frames for Spanish had compromised accuracy due to homophony and noun-drop. In addition, while frequent frames provide strong cues for

the grammatical forms noun and verb, it provided weaker cues for adjectives. Findings showed that the distributional information was more accurate for mid-frames compared to end-frames. This article advances the work of Mintz and offers additional information regarding frequent frames.

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

This article proposes that a possible resolution to the debate between core linguistics and sociolinguistics, supported by Chomsky and Labov respectively, is the Principles and Parameters approach of Universal Grammar. This approach suggests that grammar acquisition must account for varying principles inherent to acquiring grammar. These varying principles include varying grammar rules, varying rules when acquiring more than one language, and varying rules between dialects of the same language. This article cites many examples of these variances. The authors suggest that parameters provide a method, which is both formal and syntactic, to consider variation between and within languages. Additionally, the authors suggest that parameters provide a foundation for the union of core linguistics and sociolinguistics.