2014-07-02

A Model of Grammatical Category Acquisition in the Spanish Language Using Adaptation and Selection

Camille Lorraine Judd
Brigham Young University - Provo

A Model of Grammatical Category Acquisition in the Spanish Language

Using Adaptation and Selection

Camille L. Judd

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 Nissen

Department of Communication Disorders

Brigham Young University

June 2014

ABSTRACT

A Model of Grammatical Category Acquisition in the Spanish Language
Using Adaptation and Selection

Camille L. Judd
Department of Communication Disorders, BYU
Master of Science

Most typically developing children have achieved a knowledge of the grammatical categories of the words in their native language by school age. To model this achievement, researchers have developed a variety of explicit, testable models or algorithms which have had partial but promising success in extracting the grammatical word categories from the transcriptions of caregiver input to children. Additional insight into children's learning of the grammatical categories of words might be obtained from an application of evolutionary computing algorithms, which simulate principles of evolutionary biology such as variation, adaptive change, self-regulation, and inheritance. Thus far, however, this approach has only been applied to English language corpora. The current thesis applied such a model to corpora of language addressed to five Spanish-speaking children, whose ages ranged from 0;11 to 4;8 (years; months). The model evolved dictionaries which linked words to their grammatical tags and was run for 5000 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 peaked at about 85%. Directions for further development and evaluation of the model and its application to Spanish language corpora are suggested.

Keywords: grammatical word categories, evolutionary programming, language acquisition

ACKNOWLEDGEMENTS

I am grateful for the opportunity that I have had to study at Brigham Young University. My sincere gratitude goes out to Dr. Ron Channell without whose help I could not have accomplished what I did. I thank him for his kindness and patience with each of his thesis students. I would also like to thank my family for all of their love and support getting through this process; it was so appreciated. Thank you to all of my friends who cheered me on each step of the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

DESCRIPTION OF CONTENT

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

Introduction

By the end of the preschool years, typically developing children demonstrate a knowledge of grammatical word categories (e.g., noun, verb, preposition, conjunction) and can extend this knowledge to novel words (Berko, 1958; Ratner & Menn, 2000). Both the constructivist and the innatist perspectives on language development view the issue of grammatical category acquisition as important yet offer vastly different explanations for it (Ambridge & Lieven, 2011). Desires to test and improve these theoretical explanations have led to the development of explicit, testable algorithms (Pinker, 1988) which have often been implemented as computer models. These algorithms and models use transcriptions of the language input to children and extract patterns from this input in order to measure changes in some aspect representing linguistic knowledge. As implemented by various researchers, these algorithms and models have yielded insights but have found limited success in modeling the process of grammatical word category acquisition. 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 computer model of grammatical category acquisition.

Models of Grammatical Category Acquisition

Studies of models of grammatical category acquisition have used what is called a distributional approach (Harris, 1951) that examines positional regularity in the occurrence of a word relative to other words, whether preceding, between, or following these other words. Words which have similar distribution patterns are posited to be in the same grammatical category. Previous models have attempted to assign words to their most likely grammatical category based on extracting distributional patterns and thus co-occurrence statistics from the naturalistic data of adults talking to young children, as a brief review of existing studies of algorithms and models will illustrate.

Kiss (1973) introduced perhaps the earliest computational model for language acquisition. Kiss used a hierarchical analysis to gather and group words based on the similarity of their distributions relative to other words. Though based on a relatively small sample of utterances, the isolation of these clusters could be seen as representing grammatical categories.

Cartwright and Brent (1997) proposed a distributional strategy that would allow children to group words with similar syntax into more distinct categories; this model was based partly on the idea of sentential minimal pairs. An example of such minimal pair sentences would be *I saw a rabbit* and *I saw a cat. Rabbit* and *cat* are the only two words that differ between these sentences, which thus would be called minimal pair sentences. Most sentences are not matched up so perfectly, however, and therefore more generalized forms of minimal pairs could be created by forming templates. Cartwright and Brent claimed that upon hearing a sentence, children create a new group for each word in the sentence. A new template is then formed that consists of the new group sequences found in the previous sentences. Each template is added to a list of already formed templates. The best arrangement of sentences is found by using a preference list that merges similar groups from other templates. This strategy uses categorization cues from other domains such as semantics and phonology as well as computer simulations to illustrate the value of these strategies for children's grammatical category learning. However, additional work on this model has not been published.

Redington, Chater, and Finch (1998) used hierarchical clusters to demonstrate that distributional information does in fact aid children in their learning of syntactic categories. They proposed three stages that are involved in using distributional information in classifying syntactic categories: (a) measuring the distribution of contexts within which each word occurs (b) comparing the distributions of contexts for pairs of words (c) grouping together words with similar distributions of contexts. Samples from the Child Language Data Exchange System database (CHILDES; MacWhinney, 2000) were used in a series of experiments. The computational model utilized in the study formed dendrograms, which are visual representations of the hierarchical clustering of syntactic categories. The clusters that were the most similar syntactically were placed close together, with those most different placed further apart. Accuracy and completeness were measured based on comparing the benchmark syntactic categories with the derived clusters. Accuracy and completeness were highest when the clusters of preceding and succeeding words were analyzed. Eight different experiments were carried out to measure the effectiveness of distributional information involving position of context words in relation to target word, numbers of context and target words, effectiveness based on word class, corpus size, utterance boundaries, frequency versus occurrence information, removing function words, prior knowledge of other categories, and child-directed versus adult-adult speech. Redington et al. showed that a distributional analysis is highly informative of syntactic categories and concluded that this process would be most successful for languages which have strong order constraints because it uses sequential order information.

Mintz (2003) proposed a distributional algorithm that looks at small pieces of language input as clues for the grammatical category membership of words. In an input sequence of three words, the first and third word could be considered a frame. If the frame recurred often enough in the input data, the words contained in the frame tended to be of the same grammatical category. Mintz labeled these recurring sequence patterns as frequent frames. The notion that frames provide information about distribution and thus grammatical structure to a language learner was not new. For example, Childers and Tomasello (2001) found that children more easily acquired novel verb meanings when the verbs occurred in lexical frames that occurred more frequently in the children's input. Mintz's contribution was to generalize this idea to multiple grammatical word categories. Using six corpora from the CHILDES database consisting of language input to children of age 2;6 (years;months) or younger, Mintz compared these frequent frames with previous distributional approaches using bigrams, which tried to identify the grammatical category of a word by looking at the word before or the word after the word. Mintz found that frequent frames (i.e., those occurring 45 times or more) were effective at categorizing words. These frames consisted largely of closed class items such as determiners, prepositions, auxiliary verbs, and pronouns, and the categories identified by the frames were mainly nouns, verbs, and adjectives. Mintz concluded that because the distributional information provided by the frequent frames was robust, the frames could focus a child's learning mechanism on a relatively small number of contexts that have a broad impact on how words in the input are categorized.

Freudenthal, Pine, and Gobet (2005) discussed some of the weaknesses from past research based on co-occurrence statistics and attempted to evaluate the categories derived from previous research through the use of a computational model of syntax acquisition called Mosaic. Mosaic employs the concept of chunking, whereby items with similar co-occurrence are clustered and the most likely clusters are used to help form other chunks. Freudenthal et al. showed that increased use of the chunking mechanism was able to reduce the overall error rate by increasing the accuracy of substitutions, and it was also able to prevent the substitution of similar words in incorrect contexts. Freudenthal et al. were also able to demonstrate two strengths of Mosaic: the use of realistic child-directed speech (which, of course, had been used by previous researchers as well) and the production of utterances that can then be compared to child speech. Freudenthal et al. employed a simulation program and found that the error rates were relatively low and contained fewer flaws than previous works in this area. However, the sentences generated were fairly short (averaging 3.5 morphemes in length) and the assessment of sentence acceptability was made by only two judges.

St. Clair, Monaghan, and Christiansen (2010) furthered the study done by Mintz (2003) by using both trigrams and bigrams, or flexible frames, in analyzing the accuracy of distributional cues. Like Mintz, the authors also established a random baseline, in which all of the words that were categorized in the analysis were randomly assigned across the 45 frequent frames to create a random analysis. St. Clair et al. computed accuracy and completeness measures so as to be comparable to Mintz's study. Results similar to Mintz were found in both accuracy and completeness. However, the aX and Xb flexible frames were less accurate than the aXb analysis, likely due to their reduced specificity. The aX frames were found to be more accurate than the Xb frames, indicating that high frequency preceding words were more effective at classifying target words than succeeding words. St. Clair et al. supported the idea of a frequent frame in assisting children in learning grammatical categories, and while they found that trigrams had been found to be very accurate in the grammatical classification of words, they conceded that the trigrams only covered a small part of the language a child was exposed to.

Stumper, Bannard, Lieven, and Tomasello (2011) showed that frequent frames in German do not enable the same accuracy of lexical categorization that was found for English and French from the Mintz (2003) and Chemla, Mintz, Bernal, and Christophe (2009) studies. Stumper et al.'s analysis was carried out over a longitudinal corpus of German child-directed speech to a boy called "Leo." Each frame was evaluated for how well the distributionally-defined categories corresponded to the syntactic categories. Next, the degree to which words from the same category were found in the same frame was assessed by computing the accuracy for each syntactic category. Similar to Mintz (2003), the 45 most frequent frames were selected for further analysis. Stumper et al. concluded that the frames gathered some relatively reliable evidence of categories, but considerable variability within frames was shown by the relatively low accuracy scores compared to English or French. For partial frames Stumper et al. found that a frame that involved A_x would work the best, however, the accuracy scores found for A_x frames were considerably lower than those derived from A_x_B frames. The authors concluded that this finding for German was most likely due to its free syntactical structure compared to English and French. Stumper et al.'s work shows a continued interest in the topic area of language category acquisition and its expansion of study to languages other than English, but their findings suggest that Mintz' theory of frequent frames may not offer a general indication as to how children learn syntactic categories.

Weisleder and Waxman (2011) applied the notion of frequent frames (Mintz, 2003) to the Spanish language. Weisleder and Waxman examined the distributional evidence available to young children who are acquiring Spanish and then compared it to the evidence available to children acquiring English. Weisleder and Waxman also considered the clarity of frequent frames for identifying the grammatical categories of noun, verb, and adjective in Spanish. Differing from Mintz's work, these authors also considered phrase-final sequences, or "end-frames." Weisleder and Waxman selected six parent-child corpora from the CHILDES database and analyzed the input when children were at ages 2;6 or younger; three corpora were in English and three in Spanish. The English corpora were among those previously examined by Mintz (2003) to ensure that execution of the frequent frames was comparative to Mintz's work, and thus able to be compared to Spanish. Weisleder and Waxman began by gathering the frames, defined as two linguistic elements with one word intervening. After identifying the intervening words, framed words were then assigned to a grammatical category by a native speaker of each language. Weisleder and Waxman found that the accuracy for English was higher than that for Spanish for the frame-based and frame-type categories. As one of the only studies performed in both English and Spanish, Weisleder and Waxman's results suggest the need for further investigation of the process of learning the grammatical categories of Spanish words.

Each of the studies above has presented theories or models that have attempted to explain and model algorithms as to how children acquire grammatical categories. However, none of these models even made an attempt to classify every word in the input data into its most likely grammatical word category. In addition, the study of Spanish language grammatical category acquisition has received minimal attention.

An Algorithm Using Adaptation and Selection

An adaptation and selection algorithm is derived from the idea of evolutionary computing. Evolutionary computing involves applying evolutionary biology principles such as 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). Evolutionary programming in particular involves a population of solutions that randomly mutates to create offspring from the parent solutions. The most-fit offspring from the parent solutions are then chosen to become the parents of the next generation (Fogel, 2006). In view of the fact that the system never receives feedback as to what aspects of the chosen solutions are correct, the incorrect parts are as likely to mutate in the next generation as are the correct elements. Nonetheless, over many selections, reproductions, and mutations, an offspring evolves that is an acceptable solution to the question of interest.

Evolutionary algorithms have been successfully applied to several disciplines. For example, evolutionary algorithms have been used to develop seemingly trivial programs, such as computers that learn how to play checkers, to more significant advances such as systems that aid in early breast cancer detection by providing better interpretation of the radiographic features of mammograms (Fogel, 2002; Fogel, Watson, Boughton, & Porto, 1998). In another discipline, Siegler (1996) applied evolutionary algorithms to help explain the development of human cognitive strategies.

The present study applies evolutionary principles using an adaptation and selection algorithm to the area of the acquisition of the grammatical categories in language. Due to the varied, confusing, and often complicated input that young children receive, the application of this algorithm might better model the processing required by a child during early language learning and production. Previous studies of the application of an adaptation/selection model to the problem of grammatical word category acquisition (e.g., Cluff, 2014) have focused only on its application to the English language. The present study expands its application to the Spanish language.

Method

The language sample corpora used in this research study had been collected previously by other researchers for various purposes. In the current study, the task was to correctly format and grammatically tag the Spanish language sample corpora, to run the corpora through the evolutionary algorithm modeling program, and to tabulate and present the findings.

Participants

Five corpora from the CHILDES Spanish database (MacWhinney, 2000) were used for input for the adaptation-selection computational algorithm. Two of these corpora are the same samples used by Weisleder and Waxman (2011), while three are additional corpora taken from the database to increase the pool of subjects. The amounts of background information available differ among the participants.

Koki. Koki's first language was Spanish: Mexican Spanish from her father and other influences and Argentine Spanish from her mother. Koki was the first child of a middle-class professional couple. Both of her parents were linguists. The data collected includes thirteen 30 to 45 minute audio recordings of Koki interacting with her parents in her home in Michoacan, Mexico. The earliest recording was made when she was 1;7 and the last one when she was 2;11. Language samples were recorded during "play sessions" or daily routines (lunch, bath, and so forth). The number of child-directed utterances in the corpus is 4, 231 with a total of 14,778 word tokens representing 1,493 word types (Montes, 1987).

Emilio. Emilio was a Spanish-speaking boy who was audio recorded (with some gaps) from 0;11 to 4;8. Emilio was born May 20, 1980. The number of child-directed utterances in the corpus is 9,607 with a total of 30, 971 words representing 2,578 word types (Vila, 1990).

Irene. Irene was a Spanish monolingual speaker from a Northern region of Spain (Asturias). Her parents were also both monolingual. The samples recorded for Irene took place in intervals of two weeks or monthly from 0;11 to 3;2. She was born August 23, 1997. The number of child-directed utterances in the corpus is 16, 960 with a total of 88,417 words representing 4,909 word types (Ojea & Llinas-Grau, 2000).

Yasmin. Yasmin was born December 13, 1999 in Barcelona, Spain. Her mother is trilingual (Catalan, Spanish, English) and her father was also trilingual (English, Urdu, Spanish). Her mother speaks both Spanish and English to her and her father speaks to her mostly in English but also in Spanish. Yasmin was also exposed to Catalan when she was taken to a day-care center. The recordings take place from the age 1;10 to 2;9. The number of child-directed utterances in the corpus is 5,330 with a total of 20,732 words representing 1,598 word types (Ojea & Llinas-Grau, 2000).

Mendía. Mendía was a monolingual middle class girl living in Madrid, Spain whose native tongue was European Spanish. She had not had contact with any other languages. Her mother reported that Mendía acquired language development milestones and psychomotor skills within standard percentiles. The samples were taken between June 2006 and January 2007 in the context of playing with objects and reading illustrated books for children. At this time Mendía was 1;8 to 2;3 years old. The number of child-directed utterances in the corpus is 20,828 with a total of 71,766 words representing 3,342 word types (Nieva, 2013).

The Modeling Program

The program begins by opening a text file of already transcribed, grammatically-tagged (coded) utterances directed to 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 the evaluation process. An output file is opened which records the results.

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 then made of all 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 computer program. A list of the most-likely tags for each word in the corpus is also made so that the level of agreement on word types can be calculated. A population of 500 dictionaries is created such that every word in each dictionary has a grammatical tag entry randomly assigned.

The adaptation/selection process is cycled through for 5,000 generations. The odd-numbered utterances in the input file are designated to be the basis for the evaluation of each of the 500 dictionaries. The program tags the odd utterances of each dictionary during each cycle. The program examines the odd-numbered input file utterances of the 500 dictionaries to determine which dictionary has the highest accuracy in tagging the odd-numbered utterances. After 5,000 cycles the most accurately tagged dictionary is determined and the tally of the number of correct tags is increased by one. The program then calculates the percent of words in the odd-numbered utterances of that dictionary that were tagged correctly. The odd-numbered utterances serve as the basis for the dictionary with the highest accuracy. In order to quantify the generalization accuracy of the most accurate dictionary of the tagged odd-numbered utterances, this dictionary and its tags are applied to the even-numbered utterances in the input file. This demonstrates the accuracy of which the highest-scoring dictionary tags the word tokens and word types. Both of these token and type accuracy levels are then written to the screen and to the output file.

After all 500 dictionaries are evaluated in one cycle, the one with the highest accuracy is used as the starting point for populating the next generation of dictionaries. Each word entry in the offspring dictionary has a slight chance (1 in 800, 1 in 1200, 1 in 1600, 1 in 2400) chance of having its grammatical tag entry replaced with a randomly chosen tag, regardless of if that particular tag entry was correct in the parent dictionary or not. This rate of change reflects the mutation rate. Through this process a population of new offspring dictionaries is created, evaluated, and the highest-scoring dictionary becomes the basis for future dictionaries. After the 5000 evolutionary cycles are completed, the final data regarding token and type accuracy 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 computational model for 5,000 generations of adaptation and selection. The grammatical coding used the set of categories used by Harmon (2012). The adaptation-selection modeling program was designed to tag the words from the language sample corpus input with its most likely grammatical category. Generally, tagging each word in a corpus with the most likely grammatical tag for the word results in about 92% of the words being tagged correctly in English (Charniak, 1993). This level was assessed for the five Spanish corpora and found to average 91.82%. This level thus provided a context for evaluating accuracy.

Results

The model calculated the effects of mutation rate on the accuracy of two different variables: token and type percentages. The token percentages are the percent of accuracy with which the model tagged the words in the even utterances in each corpus after the highest-accuracy dictionary was selected for each generation. The type percentage is the percent of how many of the different words in each corpus were tagged correctly. The following table and graphs will show increased token and type accuracy for each of the corpora across 5,000 cycles for the four different mutation rates.

Table 1 presents the word-token performance levels of the model for the last 200 of the 5,000 cycles for each corpus at each of the four mutation rates. It can be seen in this table that the 1/1200 and 1/1600 resulted in the same accuracy at these levels for all of the children. Generally, 1/2400 shows the highest percent accuracy. When the program allowed fewer mutations per generation, it progressed to more accurately tag the words in the sample. Table 2 presents the word-type performance levels of the model at the same cycles for each of the mutation rates. However, these data do not show a strong correlation between mutation rate and percent type-accuracy for any of the children.

Table 1

Mean Token Accuracy for Generations 4800-5000 at Each Mutation Rate for Each Child's Corpus

Corpus	1/800	1/1200	1/1600	1/2400
Koki	83.15	82.96	82.96	83.08
Emilio	85.48	85.59	85.59	85.57
Irene	85.03	85.32	85.32	85.44
Yasmin	85.21	85.27	85.27	85.48
Mendía	87.33	87.38	87.38	87.44

Table 2

Mean Type Accuracy for Generations 4800-5000 at Each Mutation Rate for Each Child's Corpus

Corpus	1/800	1/1200	1/1600	1/2400
Koki	71.03	70.53	70.66	70.61
Emilio	68.21	68.61	68.44	68.70
Irene	64.79	67.68	68.54	68.91
Yasmin	71.16	71.14	71.58	72.07
Mendía	73.55	74.21	73.98	74.10

Figures 1 and 2 present the token and type accuracy mean percentages for all five corpora. Figures 3-12 present the token and type data for each child across all mutation rates. It can be seen in these figures that the token accuracy increased rapidly within the first 800 generations for each child's corpus. This increase began to taper off through the rest of the cycles until it reached and maintained the highest percent accuracy at about the 3400th generation. Token accuracy increased most rapidly with the highest mutation rate (1/800), followed sequentially by each successively lower mutation rate. Type accuracy results generally showed a more gradual incline in accuracy across generations. Figure 10 shows that the type accuracy of the Yasmin corpus increased most quickly, reaching peak accuracy at the 1600th generation, while the other children's corpora reached peak accuracy between 3000 and 4000 cycles. The 1/2400 mutation rate produced the slowest incline in accuracy but resulted in maintaining the highest type accuracy for each child's corpus. The 1/1600 mutation rate had a faster incline than others and maintained its trajectory for improvement of its accuracy level, usually reaching second highest in type accuracy by the last generation.

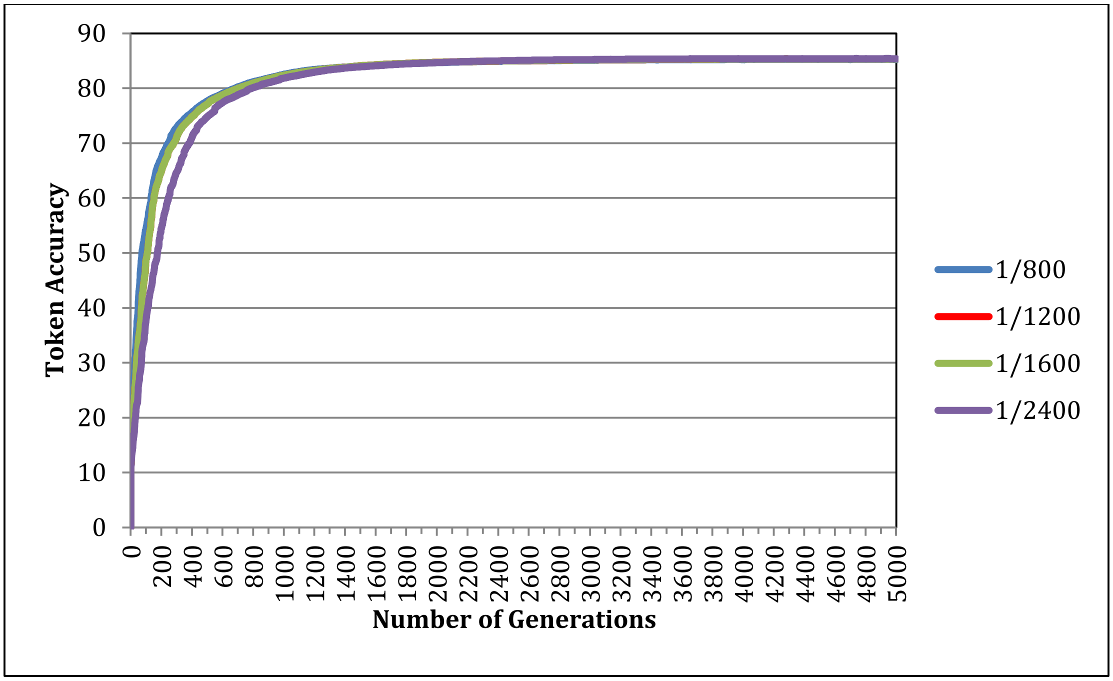

Figure 1. Token accuracy percentages for all corpora and mutation rates.

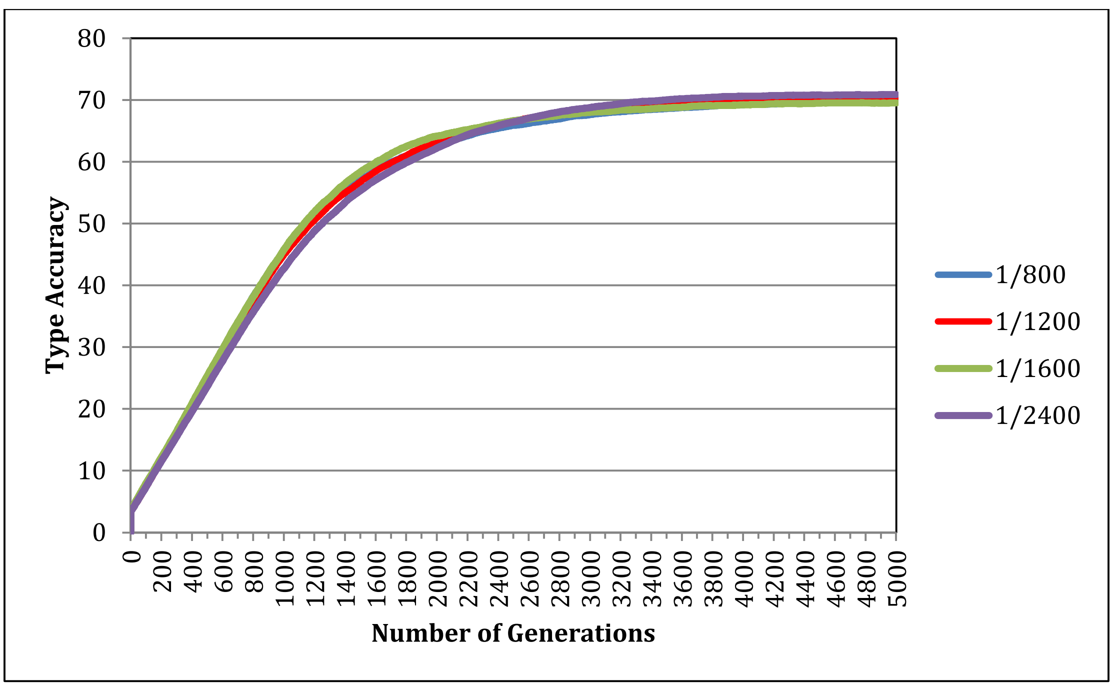

Figure 2. Type accuracy percentages for all corpora and mutation rates.

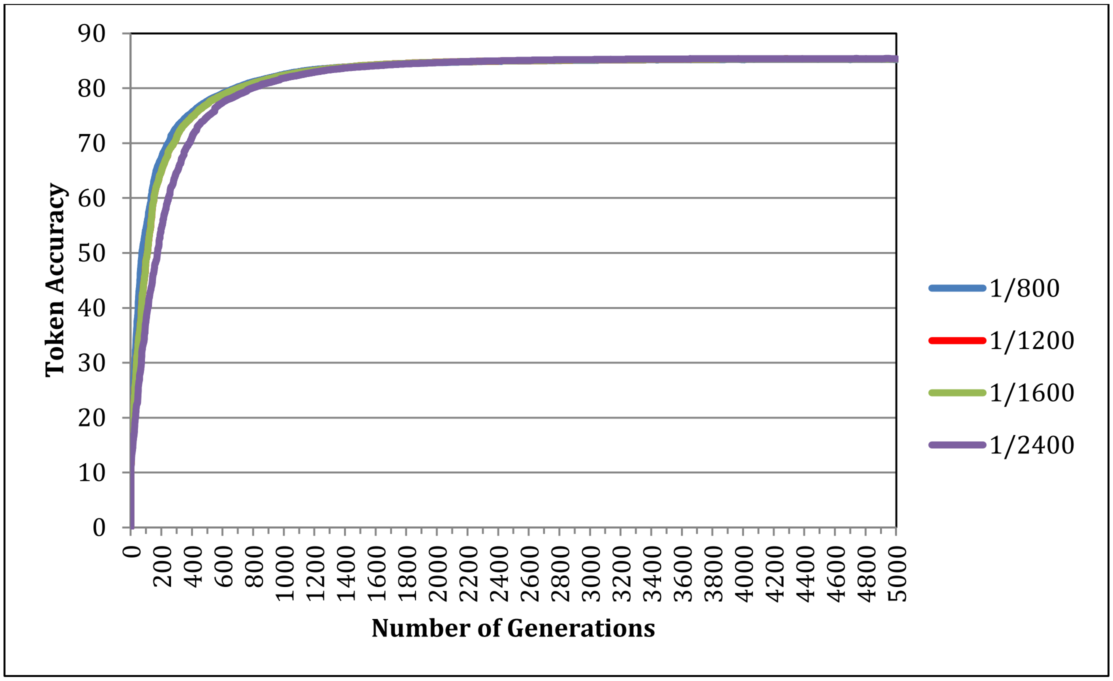

Figure 3. Token accuracy results for Koki's corpus at all mutation rates.

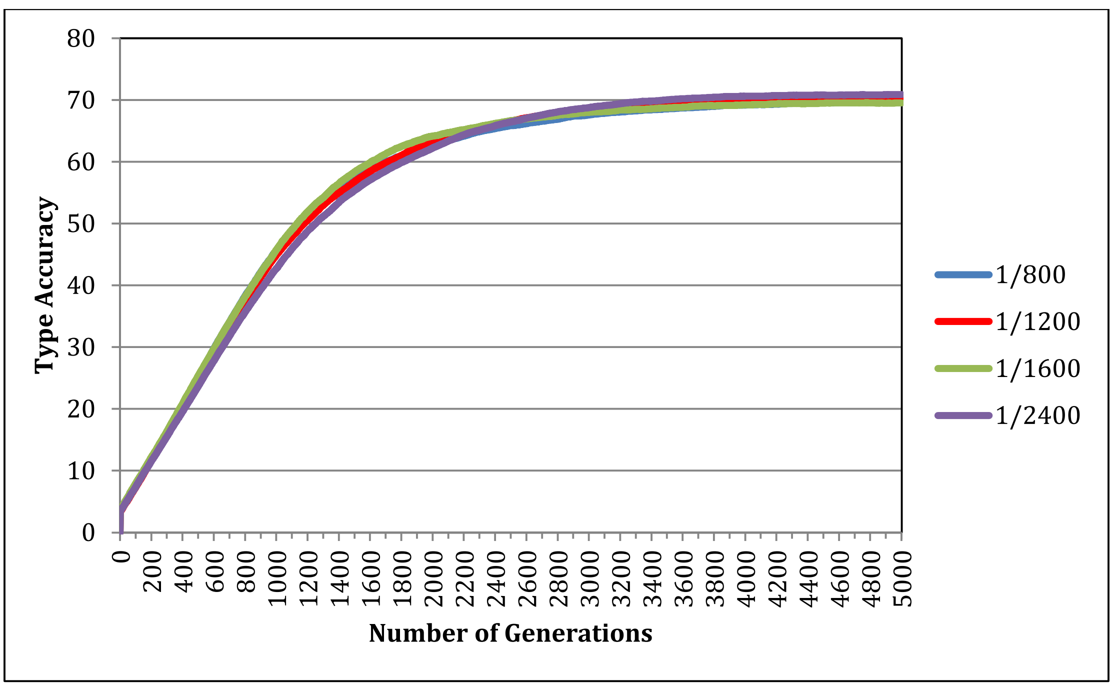

Figure 4. Type accuracy results for Koki's corpus at all mutation rates.

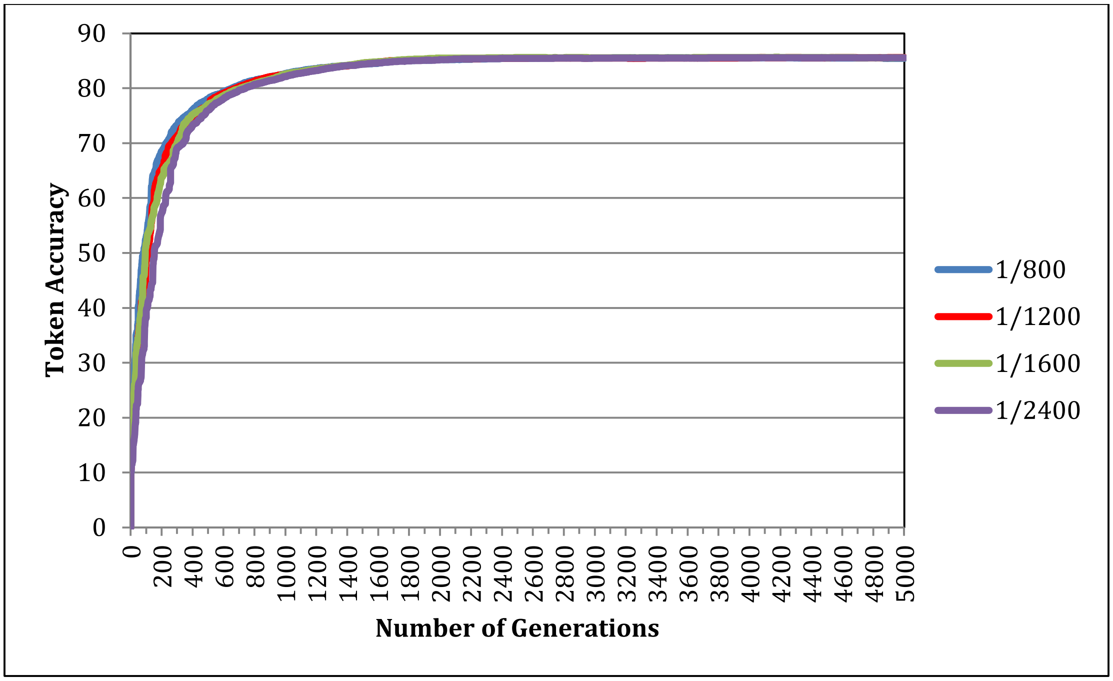

Figure 5. Token accuracy results for Emilio's corpus at all mutation rates.

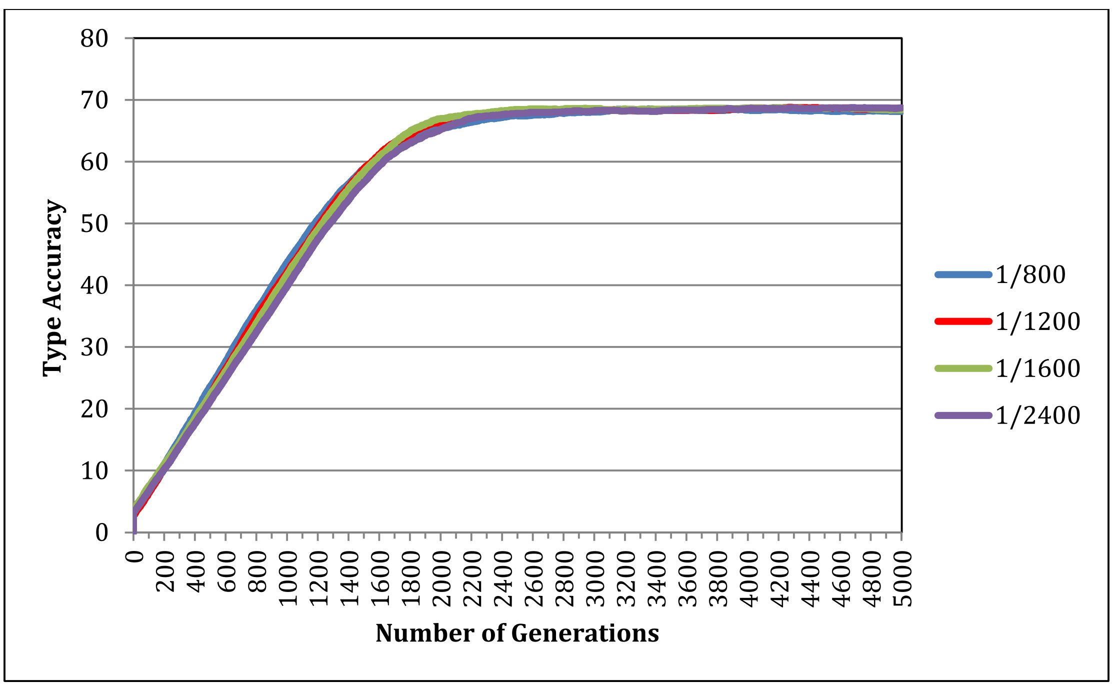

Figure 6. Type accuracy results for Emilio's corpus at all mutation rates.

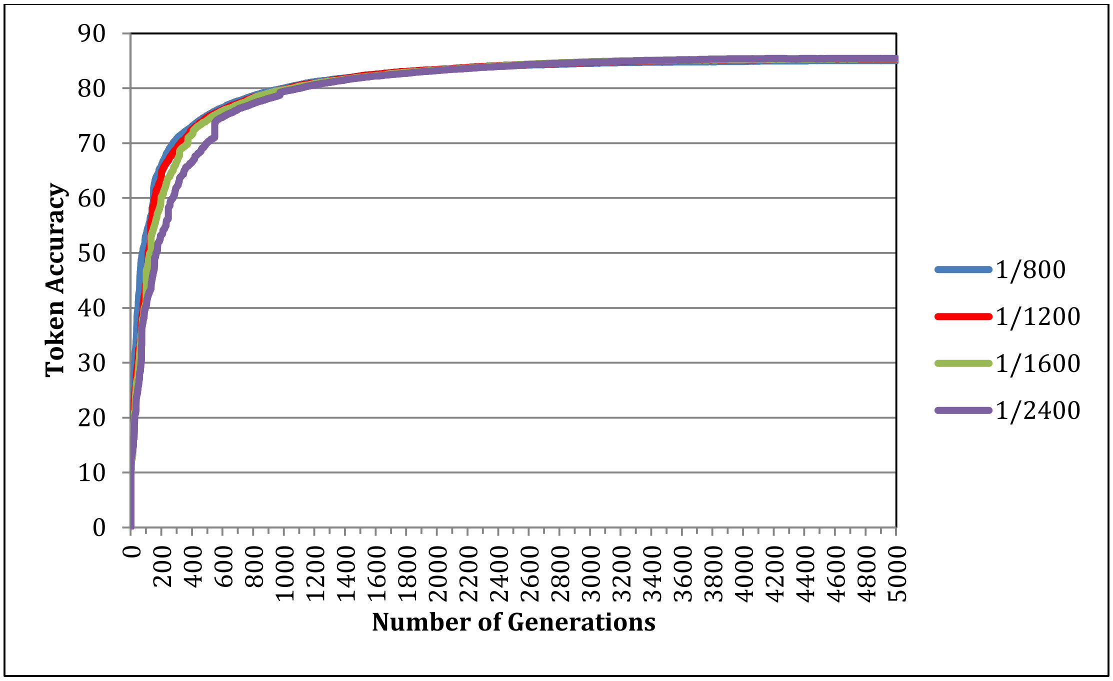

Figure 7. Token accuracy results for Irene's corpus at all mutation rates.

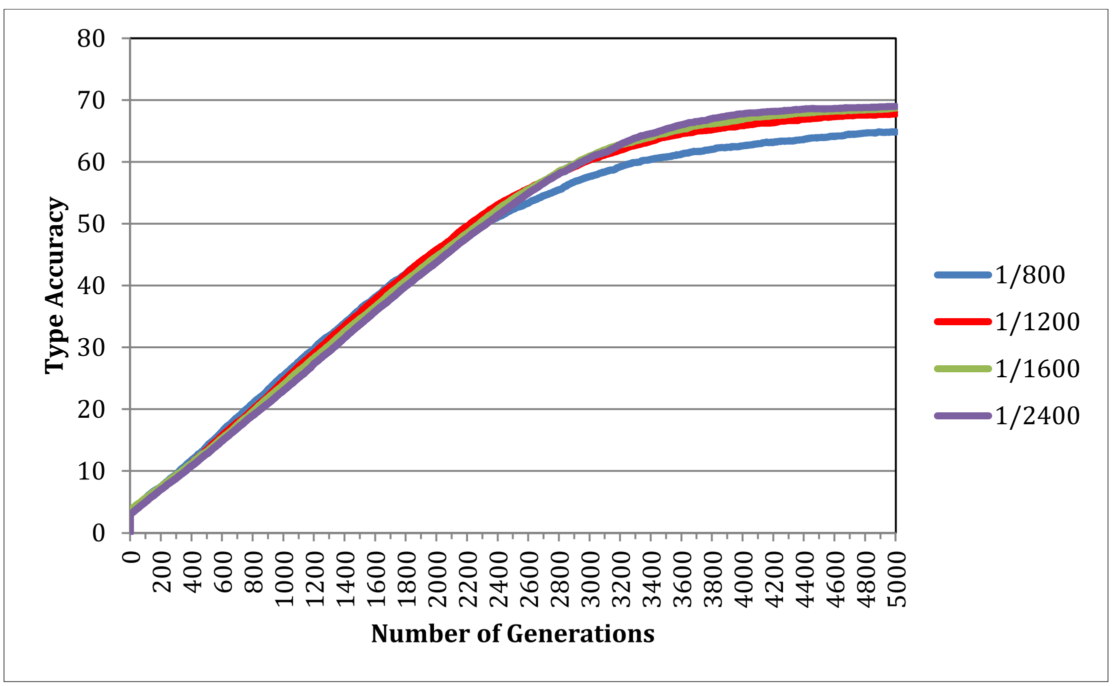

Figure 8. Type accuracy results for Irene's corpus at all mutation rates.

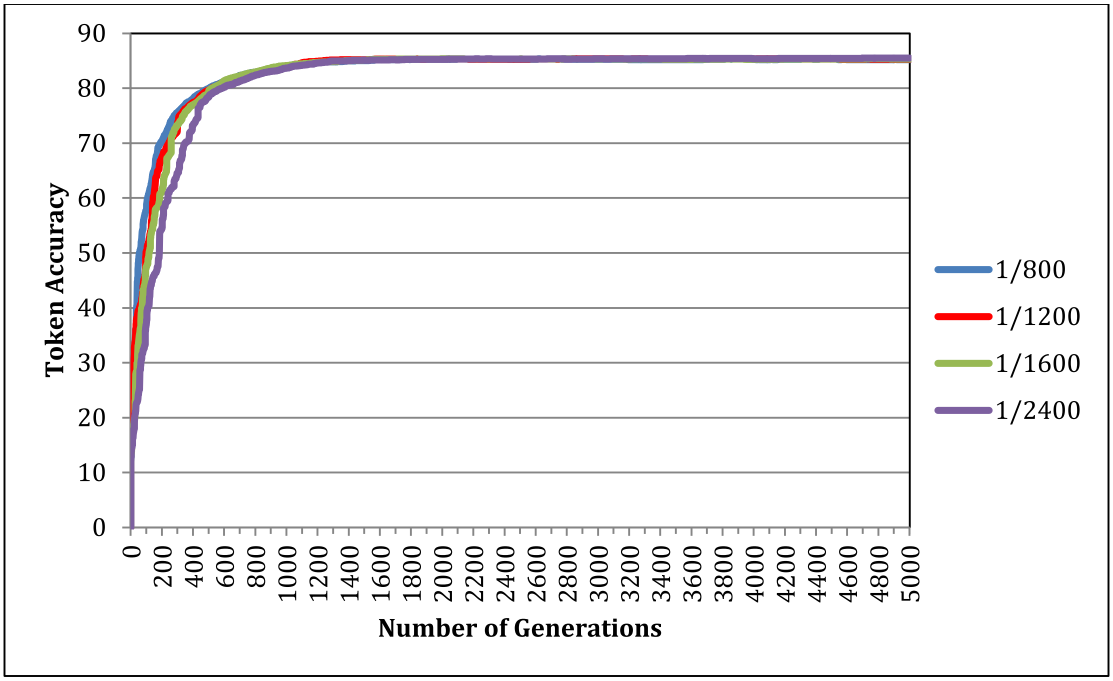

Figure 9. Token accuracy results for Yasmin's corpus at all mutation rates.

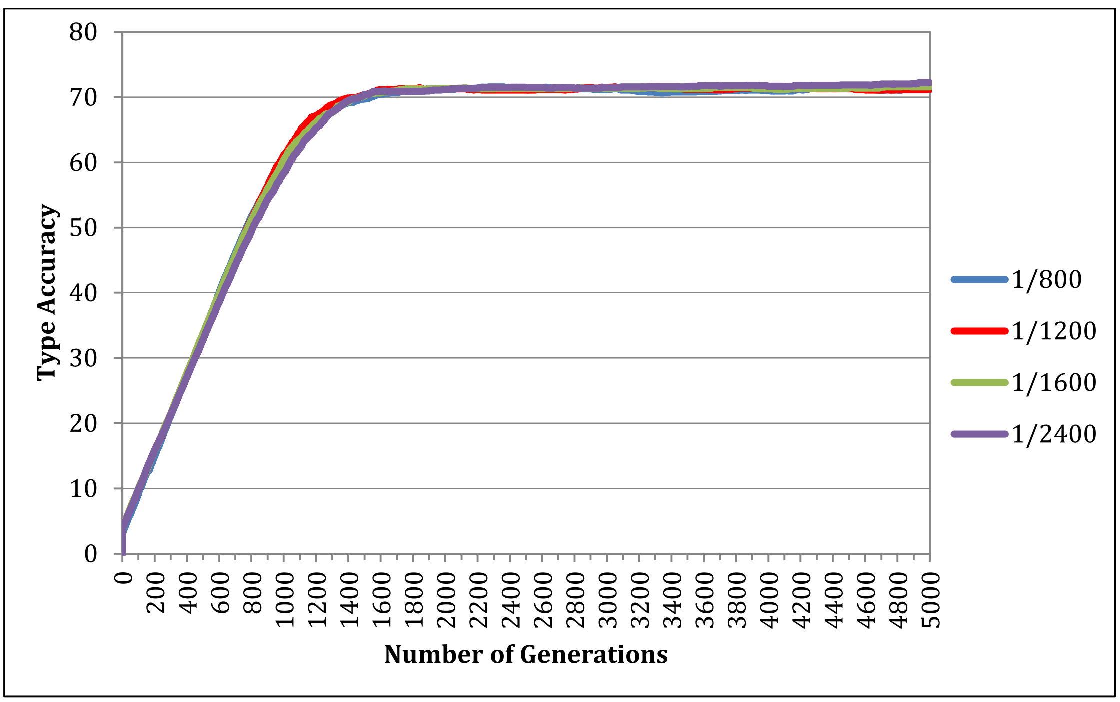

Figure 10. Type accuracy results for Yasmin's corpus at all mutation rates.

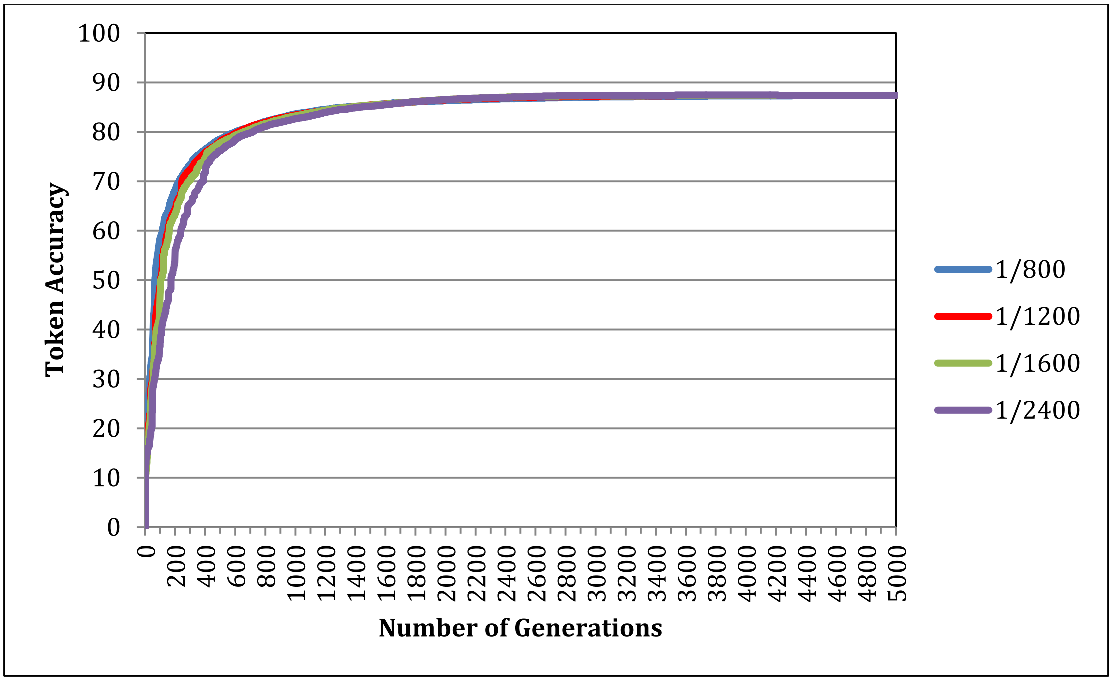

Figure 11. Token accuracy results for Mendía's corpus at all mutation rates.

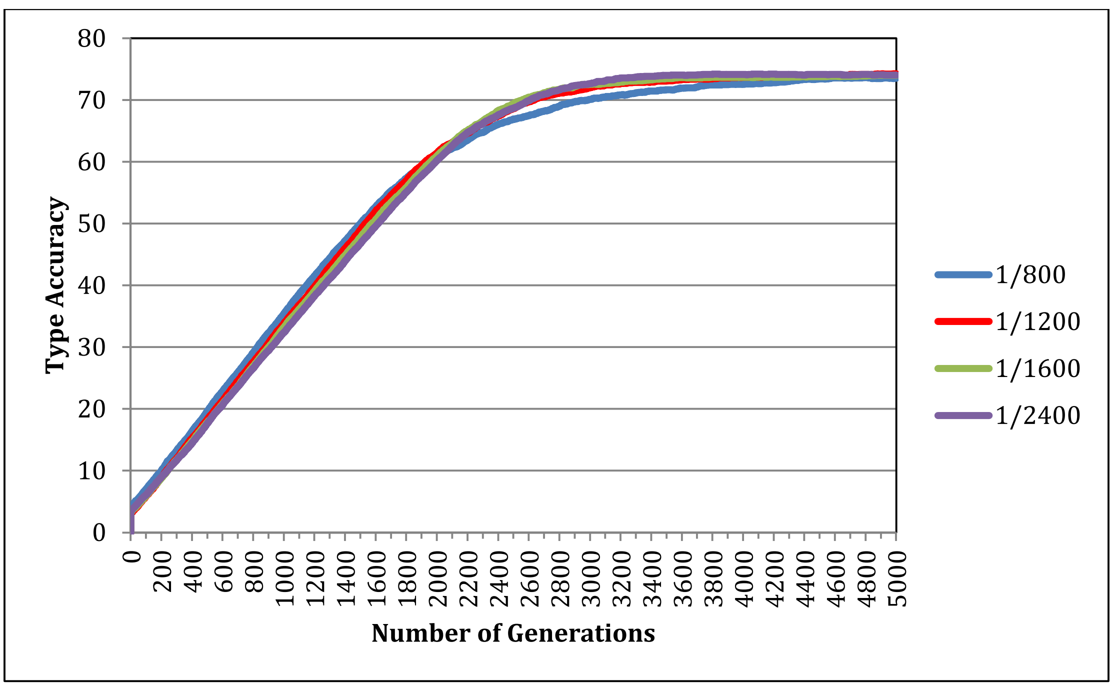

Figure 12. Type accuracy results for Mendía's corpus at all mutation rates.

Discussion

This study examined the accuracy of an adaptation-selection computational algorithm which modeled the learning of the grammatical categories of words from Spanish language corpora. The accuracy of learning the grammatical categories of words was examined both for word tokens (all the words used in the corpus) and word types (the different words used in the corpus). For word tokens, the algorithm led to a rapid increase of accuracy during the first 800 generations which was followed by a more gradual increase until reaching a plateau at about 2,400 generations. Generally, lower mutation rates yielded a higher accuracy in tagging word tokens. Each mutation rate, however, produced a fairly high accuracy of word tokens, usually reaching about 85%. For word types, the model produced a slower incline in accuracy across generations, with the peak accuracy being somewhat lower than the level of word token accuracy, usually around 72%. Word type accuracy was also best when using the lowest rate of mutation.

The findings of the present study offer an interesting contrast to those of previous studies which had investigated the ability of young children to learn the grammatical categories of words. The study most comparable to the current work was recently completed by Cluff (2014). Cluff's study utilized a similar model of adaptation and selection to analyze the learning of the grammatical categories of words in five English corpora. Similar to the present study, Cluff found that as the mutation rate decreased, the overall accuracy of the program increased. Also as in the current study, the accuracy of coding word tokens increased rapidly during the first 500 generations and then showed a slower but steady increase until the 4000^{th} generation, which was the last generation in her study. Cluff's study obtained slightly higher accuracy results across generations than were found in in the present study. However, the reason for this difference is as yet unclear as several key variables differed between the studies, with the main difference being Cluff's use of English corpora in contrast to the present study's use of Spanish corpora. The present study's version of the program implementing the adaptation-selection model also differed from Cluff's study in terms of the number of tag options allowed per word and the number of grammatical tags in the set used to code the corpora, and thus direct comparison with the present study is not possible.

However, like the study of Cluff (2014), the present study used an adaptation-selection model of grammatical category acquisition. This model is a profoundly different approach to modeling the learning of the grammatical categories of words. Cartwright and Brent (1997) had proposed a distributional strategy that allows children to group words with similar syntax into more distinct categories. Redington et al. (1998) had used dendrograms to model another distributional theory of grammar acquisition. Mintz's (2003) work using frequent frames influenced Weisleder and Waxman's (2011) study that examined the distributional evidence available to young Spanish speakers and compared it to the evidence available to children acquiring English. However, this frequent frames approach was found to be successful in Spanish only in certain categories such as nouns and verbs. The current study looked at a model's categorization ability across all grammatical categories, not just a small number of categories. While each of these previous studies differed in their approaches, each has contributed to further understanding of grammatical category acquisition. The current study has provided a new lens to view the question of one aspect of language acquisition in young children.

The current work is not without its limitations. One limitation of the program version used in this study was that, similar to other studies, the program was set to use only the most-likely tag per word. However, many words are grammatically ambiguous and function as more than one grammatical category, such as the English word *farm* that can be either a noun or a verb. The program version used by Cluff (2014) implementing the adaptation and selection model used one, two, or three possible tags per word, thus increasing the program's probability to match a correctly tagged word token. By using only a single tag per word, the highest accuracy that this program could achieve was 91.82%, and this model performed quite well relative to this maximum possible level, reaching about 85% accuracy at each mutation rate. Future studies with Spanish samples could increase the number of tags allowed per word to examine the effect that that program capability would have on the accuracy of the evolutionary model.

This study was also limited by the number of available Spanish corpora, using five corpora from MacWhinney's (2000) CHILDES database. As previous studies in this area had done, data were extracted by the model from longitudinal corpora collected over periods of two or more years, and thus the number of contributed, useable corpora was quite small. Increasing the number of analyzed corpora would increase the generalizability of studies in this area. Should future longitudinal studies of language acquisition in Spanish become available for use, future studies might include corpora that evaluated the adaptation-selection model on factors such as the different dialects of Spanish. Also, because the corpora were obtained from online source, many variables could not be manipulated or controlled in the current study. Such variables include the development of each child, the level of education of the parent or caregiver, and the setting of obtaining each corpus. Each of these variables has a likely influence on the type of data input that was used in the study. Perhaps obtaining samples that were similar in child development level, caregiver education level, and language sample setting would allow future researchers to view the effect of each of these variables on the model's grammatical category acquisition. In addition, each of the five samples was collected and transcribed by different researchers for different purposes, and thus the care given to transcribing the caregiver's utterances varied, and each corpus required some reformatting before being grammatically coded. In spite of these limitations, however, it is encouraging to see that this model was able to function so well given the difference in language corpus collection conditions and in the heterogeneity of the regional dialects of the children's conversational partners whose language was sampled.

As this was a new approach to grammatical category acquisition, this was the first time an evolutionary model was applied to Spanish language samples. The best settings for the program in regards to the number of dictionaries evolved, the mutation rate, and the number of generations to study have yet to be determined. Future work in this area could inspect which settings would be the most beneficial for the highest growth in the model's accuracy by modifying each of these settings. This program also used the same set of grammatical word categories used by Harmon (2012), and the effect that the set and number of grammatical categories has on the accuracy of this and similar programs has yet to be investigated.

Based on the evidence from this preliminary study, it is still unclear if children might use some sort of adaptation-selection algorithm to acquire grammatical categories in Spanish. Siegler (1996) promoted the idea of applied evolutionary algorithms to explain the development of human cognitive strategies. However, available data in this area are limited and thus any conclusions would be speculative and premature. Further research is warranted in this area and would be advantageous to furthering our understanding of the nature and mechanisms involved in human language development.

Overall, though, the present study has made a contribution to understanding possible mechanisms of learning the grammatical categories of words. The adaptation selection model used was able to reach respectable levels of accuracy across five rather varied Spanish language corpora. Changes in accuracy associated with progress through the generations of adaptation and selection as well as changes associated with altering the rate of mutation were similar across corpora. The findings of the present study are promising and support the need for further work with evolutionary models of grammatical category acquisition. Subsequent studies might seek to apply this adaptation-selection model to other aspects of language development and to other aspects of human cognitive development.

References

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

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

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

Charniak, E. (1993). *Statistical language learning.* Cambridge, MA: MIT Press.

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

Childers, J. B., & Tomasello, M. (2001). The role of pronouns in young children's acquisition of the English transitive construction. *Developmental Psychology, 37,* 739-48.

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.

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

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

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

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

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

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

Harmon, T. G. (2012). *Accuracy of automated grammatical tagging of narrative language samples from Spanish-speaking children.* (Unpublished master's thesis). Brigham Young University, Provo UT.

Harris, Z. (1951). *Methods in structural linguistics.* Chicago, IL: University of Chicago Press.

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

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. doi:10.1016/S0010-2077(03)00140-9

Mintz, T. H. (2006). Frequent frames: Simple co-occurrence constructions and their links to linguistic structure. In E. V. Clark & B. F. Kelley (Eds.) *Constructions in acquisition*. Stanford, CA: CSLI.

Montes, R. (1987). *Secuencias de clarificación en conversaciones con niños* (Morphe 3- 4): Universidad Autónoma de Puebla, Puebla, Mexico.

Montes, R. G. (1992). *Achieving understanding: Repair mechanisms in mother–child conversations.* (Unpublished doctoral dissertation). Georgetown University, Washington, DC.

Nieva, S. (2013). *Nieva Longitudinal CHILDES Corpus. URL: http://childes.talkbank.org/data/Romance/Spanish/Nieva.zip, ISBN 1-59642-433-8.*

Ojea & Llinas-Grau. (2000). *El desarrollo de las categorías gramaticales: Análisis contrastivo de la adquisición lingüística temprana del inglés, castellano y catalán.* Universitat Autònoma de Barcelona, Barcelona, Spain.

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.

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.

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

Vila, I. (1990) Adquisición y desarrollo del lenguaje. Barcelona: Graó.

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.

APPENDIX: ANNOTATED BIBLIOGRAPHY

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

This chapter outlines the major theoretical approaches to language learning. The two main approaches discussed are the nativist and constructivist theories. The nativist approach assumes that linguistic knowledge is innate rather than learned, while constructivists assume that language is not innate and children must acquire aspects of linguistic knowledge such as grammatical categories. The chapter focuses on how children learning languages governed by syntax (word-order) actually learn the rules governing the word order of their language. They give three issues that include the debate of the opposing theories of language acquisition: semantic bootstrapping, prosodic bootstrapping, and distributional analysis. The authors discuss Pinker's theory of semantic bootstrapping and identify some inherent violations to language rules. This has led some researchers to discredit this theory. The authors also discuss prosodic bootstrapping as a possible approach of language acquisition. The chapter reviews distributional approaches such as frequent frames, chunking, merging templates, as well as possible acquisition through phonological cues. By presenting each of these topics from the viewpoint of each acquisition theory, the reader is able to see that while research supports many theories, the question of child language acquisition remains unanswered.

Arias-trejo, N., & Alva, E. (2013). Early Spanish grammatical gender bootstrapping; Learning nouns through adjectives. *Developmental Psychology*, *49*, 1308-1314.

Purpose: Arias-trejo and Alva investigate recent research that found that children can benefit from the morphophonological cues marking gender and number to infer a familiar referent. Learning grammatical gender in Spanish is predictable largely from morphophonological information: Nouns ending in *a* tend to be feminine and those ending in *o* tend to be masculine. This study looks to see if toddlers learning Spanish are able to use gender information from the final vowel (*a* or *o*) of familiar adjectives and transfer that information to associate novel labels to novel objects (nouns).

Method: They tested a group of thirty-seven 30-month-old toddlers. This age group was chosen based on their ability to distinguish feminine and masculine determiners to guess a target noun. Also, around the age of 3 years toddlers begin to consistently use gender cues in their spontaneous speech. Each toddler received both audio and visual stimuli. For the audio stimuli the infants heard two novel nouns: masculine=*pileco*, and feminine=*betusa*. The visual stimuli consisted of two computer images unfamiliar to toddlers: a fire hydrant and a diabolo, with eight variants of each (color and size). The experiment consisted of 16 trials divided into two blocks; each block was composed of two parts: training and test. All trials lasted only 5,000 ms. An experimenter assessed the trials offline to view how long the child looked at the object vs. at a distracter.

Conclusions: A two-way repeated analysis of variance was performed with the children's' prenaming vs. postnaming of the objects and block (1 vs. 2); these were the within-subject variables. The analysis revealed a significant correlation between naming and block. The overall results showed that the 30-month year olds were able to associate two novel words with two novel objects after being exposed to morphophonological gender cues. However, the children did not show this ability until they had been exposed to a second training block. Regardless, the work gave evidence that children have an early ability to learn novel word-object associations. These conclusions coincide with previous research done with toddlers' sensitivity to gender cues.

Relevance to the current work: This study shows that word learning is a flexible process and that inflectional bootstrapping strategies pair with word learning constraints, syntactic computations, and social cues to establish novel word-object associations. The outcome of this work also suggests the large role that child-directed speech plays in language learning.

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

Purpose: This paper was an extension of Mintz' work done in 2003. Stumper et al. show that frequent frames in Dutch (Erkelens 2009) and German do not enable the same accuracy of lexical categorization that was found for English (Mintz 2003) and French (Chemla 2009). They also explore the relationship between the accuracy of frames and their potential usefulness.

Method: Stumper et al.'s analysis is carried out over a longitudinal corpus of German child-directed speech to a boy called "Leo". Similar to Mintz' work in 2003, the 45 most frequent frames were selected for further analysis. Each frame was evaluated for how well the distributionally defined categories corresponded to the syntactic categories. Next, the degree to which words from the same category were found in the same frame was assessed by computing the accuracy for each syntactic category. Following Mintz' work, two different methods of categorization were used: Standard and Expanded labeling. Token and type accuracy for all frames and categories was computed out of each of these trials and the original and random accuracy scores were compared. The authors utilized Fisher's Omnibus test to combine the p-values into a single measure of overall significance.

Results: All accuracy scores for frame accuracy were higher than random. Two conclusions followed the study's results: the frames do gather some relatively reliable evidence of categories, but there is considerable variability within frames as shown by the relatively low accuracy scores compared to English or French. The scores found are similar to the Dutch accuracy scores. For category accuracy they found that words from the same grammatical category tended to occur in many different frames and the distribution of the same category across frames seemed arbitrary. As with English, French, and Dutch, one single frame does not offer a reliable cue to one category. For partial frames they found that a frame that involved A_x would work the best, however, the accuracy for A_x frames were considerably lower than those derived from A_x_B frames.

Conclusions: Stumper et al. conclude that the reason that frequent frames do not work as well in German is most likely because of its free syntactical structure compared to English or French. German has a morphologically more complex determiner system; determiners in German are marked for case, number, and gender. When comparing the accuracy of frames to their usefulness they found that even the high token-based accuracy have limited value because of their placement in "fact set" phrases with little or no variability in the slot position. The authors discuss that a frame showing both high accuracy and high diversity may be a guide to learn the category of an intervening word. Authors admit that the corpus used was quite small and only a snapshot of the linguistic environment to which the child is actually exposed.

Relevance to the current work: Stumper's work shows a continued interest in the topic area of lexical category acquisition and the expansion of this work to languages other than English. Because of their findings, we know that Mintz' theory of frequent frames is not a perfect indication of how children learn syntactical categories and further work is needed in this area.

Bedore, L. M. (2001). Assessing morphosyntax in Spanish-speaking children. *Seminars in Speech and Language, 22*, 65.

Purpose: This article outlines the nature and acquisition of the Spanish morphosyntactic system in children. Clinicians should have an understanding of this system to best help children in therapy. This study analyzes spontaneous language samples and structured assessment tasks from and completed by Spanish-speaking children in order to understand morpho-syntactic performance.

Method: The article gives an overview of the Spanish language: articles match nouns in number in gender, pronouns are marked for number and gender, adjectives must agree in number and gender of the noun that they modify, etc. The tables in the article point the reader to see how differences in meaning have been encoded morphosyntactically in the children's language samples. The assessment tasks in the study varied in the dimensions of naturalness of the task, and the level of linguistic demand.

Conclusions: Comparatively to English, Spanish has a relatively free syntactic structure, but with the basic word order being "Subject-verb-object" (SVO). The article discusses the ages for children's typical first productions of grammatical forms. What is not clear, however, is if at these times that the children are able to fully contrast these forms. In regards to Spanish-speaking children with language impairment, literature shows that they seem to use less complex syntax compared to their peers and recommend further work in morphosyntax in Spanish.

Relevance to current work: While this article does not discuss computational models for grammar acquisition, this work gives insight to the way that morphosyntax is acquired in the Spanish language. The current study contributes to the knowledge and understanding of language learning in Spanish children.

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

Purpose: In their study, Cartwright and Brent propose a strategy that allows children to group words with similar syntax into more distinct categories. The distributional strategy they propose is based partly on the idea of sentential minimal pairs. The strategy uses categorization cues from other domains such as semantics and phonology. They seek to prove this strategy through series of computer simulations. A central goal for them is also to uncover the problem that syntactic categories are defined by their environments, yet the environments are defined by the permitted sequences of categories. By using their computational models they hope to discover the distributional component of this theory.

Method: The authors summarize theories of category acquisition. These include theories on the use of distributional analysis, semantics, phonology, as well as previous computational models such as hierarchical cluster analysis (HCA). Cartwright and Brent discuss Kiss's work (1973) which used HCA to group together words whose distributional patterns were similar. Their approach differs slightly from Kiss's approach. The two main differences are that their strategies results in a set of discrete categories of words, whereas HCA results in a large number of categories. Also, Cartwright and Brent's learning strategy results in a set of discrete categories of words that is incremental; it operates on only one sentence at a time, forgetting previous sentences. They propose a theory in which children use a particular set of preference in order to decide which arrangement of words (groups) to merge. They look at the contexts in which the words occur using sentence templates--or sequence of group tags. A series of five experiments using the computer simulation model was conducted to investigate its effectiveness in categorizing words. The templates utilize preferences that children use in deciding which groups to merge as well as children's preferences of arranging words into groups.

Results: Computer simulations showed that this strategy is indeed effective at categorizing words. They found this true for both artificial-language samples as well as natural child directed speech samples. They found that the strategy performs best when it is given semantic information about concrete nouns.

Conclusions: Cartwright and Brent show for the first time that categorical inference can be done incrementally, one sentence at a time. They show that it is possible to define a form of distributional analysis using these increments, as well as how distributional and semantic information could be combined in category acquisition. The authors state that it appears possible to connect their theory to others in literature about category acquisition.

Relevance to the current work: The current study also utilizes computer programs using an algorithm to look at the role of distributional analysis in grammatical category acquisition. Cartwright and Brent agreed that their experiments would need to be expanded in order to gain more conclusive data. The current work also looks to expand knowledge on this topic and uses an evolutionary algorithm to do so.

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.

Purpose: Chang, Lieven, and Tomasello employ several different algorithms of syntactic category learning to test distributional information from child and adult utterances. This study tests these algorithms across twelve different languages. The authors compare these algorithms to evaluate, measure, and to predict word order in order to try to find an algorithm that replicates knowledge used to create the corpora.

Method: The authors use an evaluation measure known as Word Order Prediction Accuracy (WOPA) to evaluate twelve different languages (Cantonese, Croatian, English, Estonian, French, German, Hebrew, Hungarian, Japanese, Sesotho, Tamil, and Welsh). Given a candidate set of words made from the utterance they wanted to predict, the system attempted to predict the order of words. The WOPA score is then the number of correctly predicted utterances (two words or more). They compare this WOPA score to six syntax acquisition algorithms whose basis lie in computational linguistics and child language. The six algorithms are the Lexstat Learner, Prevword Learner, Freqframe Learner, Token/Type learner, and Type/Token learner.

Results: The category and statistics in the Prevword and Freqframe learners were found to be useful in characterizing the order in child speech, but their tendency to discover more broad categories such as nouns and verbs made it hard to order words of the same category relative to another. The Token/Type learner was able to yield more specific categories. The Type/Token learner on the other hand performed better with frames with a higher lexical diversity. The Lexstat learner and Type/Token learner of these were best able to account for more complex utterances. However, in the end, none of the learners were able to learn standard linguistic syntactic categories (i.e nouns, verbs, adjectives, determiners).

Conclusions: The authors determined that they are able to better characterize the order of words in child and adult speech when using more specific categories. The authors concluded that possibly using combinations of broad and specific categories would allow the algorithms to work better and suggest the need for future studies.

Relevance to the current work: Chang, Lieven, and Tomasello illustrate the growing interest in using computational methods to study child syntax acquisition. The authors' review of these methods show what these different algorithms show about how speech is syntactically formulated in each of these different languages.

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

Purpose: Chang, Lieven, and Tomasello address the fact that computational syntax acquisition systems typically involve theories and or language-specific assumptions that do not allow the system to transfer over to other languages. In order to look past these theories and assumptions, a "bag of words" incremental generation task (BIG) along with an automatic sentence prediction accuracy (SPA) evaluation measure were developed. Utilizing BIG and SPA measures the authors seek to compare theories of syntax acquisition in multiple languages.

Method: The article presents a method for evaluating how well a program called the Adjacency Prominence learner as well as more simple learners are able to learn syntactic categories from the input they have been given. The authors use the SPA measure along with the BIG task in order to evaluate algorithms that have been previously put forth in computational linguistic and developmental psycholinguistics. They use child-adult interactions that have been collected from speakers of twelve typologically-different languages (Cantonese, Croatian, English, Estonian, French, German, Hebrew, Hungarian, Japanese, Sesotho, Tamil, and Welsh). In the first run through *n*-gram-based learners are evaluated with the BIG-SPA measures. Then the psycholinguistic learner, or Adjacency-Prominence learner, was compared to the simpler learners. The simplest learners were a Bigram (two adjacent words) and a Trigram (three adjacent words) learner. Also, several other *n*-gram learners were tested, as was as a Chance learner whose performance percentage for an utterance was 100/*n!*. The learners only differed by their Choice function, or probability of producing a specific word from the bag of words at each point in the sentence. The two main parts of the BIG-SPA included collecting statistics on the input and then predicting the test utterances.

Results: The Chance learner was statistically lower than both the Bigram learner and the Trigram learner. This suggested that the *n*-gram statistics in these learners were useful for predicting word order within the BIG task. The combined Bigram+Trigram learner showed an improvement over both the Bigram and Trigram learner. This suggests that trigram statistic improved the prediction accuracy over the bigram. The BIG-SPA sought to allow comparisons of learners from different domains. The authors examined a corpus-based learner based on an "incremental connectionist model" of sentence production and syntax acquisition called the Dual-path model. This model accounted for a large range of syntactic occurrences. The two paths were sequencing and meaning pathways: sequencing incorporates a network that learned statistical relationships over sequences and the meaning pathway had a representation of a message to be produced that was dependent on the sequencing pathway. After running each of the tests, the authors found that the Adjacency-only learner was better than Bigram learner, the Adjacency-only, and the Prominence-only learner.

Conclusions: These results suggest that the adjacency and prominence statistics express different parts of the problem of word order prediction. Overall the authors conclude that the BIG-SPA test can be used to compare syntactic learners in typologically-different languages. It can also be used to look for biases in particular algorithms which can aid in the search for a more correct syntax acquisition algorithm. The authors suggest making syntactic learning theories more explicit while testing them with the BIG-SPA task in order to help researchers discover how humans learn syntax.

Relevance to the current work: Research has shown that syntactic constraints differ across languages. As the authors discussed is difficult to adapt a particular theory of syntactic categories or constraints to languages that differ typologically. This study shows the need for further research in understanding how young children are able to learn language (particularly grammatical categories) from the language input given them.

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

Purpose: One goal of this article was to begin to test the validity of frequent frames cross-linguistically. The authors extend Mintz's 2003 study of frequent frames to the French language. The French language contains a function word system that allows patterns that are potentially detrimental in a frame-based analysis procedure. They show that the importance of having target words framed by the context words is crucial for the mechanism to be efficient.

Method: The authors perform three different experiments in this study. Experiment 1 consisted of we testing the validity of frequent frames in French. This language was chosen because its many features that could be problematic in the frame-based procedure. They obtained a French corpus from the CHILDES database and only used the mother's child-directed speech. An analysis was run by "Cordial Analyseur" to code each instance of a word with its syntactic category. In Experiment 2 the authors compared frames in French and English with other types of contexts that are outwardly quite similar to frames in terms of their basic content and structure: [A B x] and [x A B]. This was done to characterize the "core computational principles" that make frequent frames good environments for categorization. Experiment 3 included the investigation of a recursive application of the frame-based procedure using both French and English corpora. First they performed an initial analysis to derive frame-based categories. Afterwards they then reanalyzed the corpus defining the frames based on the categories of words which were obtained from the initial analysis. These looked not just at the words surrounding the target word, but the category of the surrounding words.

Results: The authors were surprised by their findings of Experiment 1; the notion of frequent frames holds true for French. Because of French's more varied and ambiguous system of function words they had supposed it to yield quite different results than the English corpora. Experiment 2 tested discontinuity of environments: front texts and back texts [A B x] and [x A B] respectively. The front and back context elements combined did not yield effective categorization compared to frequent frames. Experiment 3's investigation of recursivity found that recursivity was not a robust feature of syntactic category learning.

Conclusions: The authors conclude that their analysis provides strong support for frames as a basis for the acquisition of grammatical categories in young children. Due to the findings of Experiment 1, Chemla et al. conclude that what they thought to be the potentially problematic characteristics of French did not actually appear to be problematic in the frequent-frames approach. The results of Experiment 2 suggested that co-occurring context elements (front/back) must frame a target word. Experiment 3 resulted in something they found somewhat counterintuitive: the recursive application of the frame-based procedure resulted in relatively poor categorization. They state that this finding suggests that computation based on specific items (actual words as opposed to categories) is a core principle in categorizing words. Discontinuity and item-specificity appear to be crucial features, while recursivity does not.

Relevance to the current work: Past investigations have demonstrated that lexical co-occurrence patterns found in child-directed speech could provide source of information for children in correctly categorizing nouns and verbs, as well as some other form-class categories. This study looked at how the frequent frame suggested by Mintz in 2003 applied when cross-linguistically examined in the French language. Results were surprisingly similar to corpora in English. The current study looks at an evolutionary computer model with Spanish corpora.

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.

Purpose: This paper argues that past research on syntactic categories in regards to grammatical acquisition is flawed. Freudenthal et al. discuss these flaws from past research in co-occurrence statistics and attempt to evaluate the categories that were derived from previous research through the use of Mosaic, a computational model of syntax acquisition.

Method: Studies by Redington and Finch, Pinker, and Mintz compared human-labeled framed words to other human-labeled words. Thus they didn't see the problem that these categories would generate faulty utterances. Also, similar verb forms such as plain present tense and the infinitive verb form lead to faulty utterance labeling. Because of these flaws, these authors utilize a new version of Mosaic which employs the concept of chunking. Mosaic is used to simulate phenomena in child speech. The current model is a simple discrimination net that is headed by a "root node," or a new candidate for a syntactic category. For nodes to be added, it is required that whatever follows the word to be encoded in the input must have already been encoded in the model. In order to produce utterances, Mosaic must output all the utterances that it has encoded. Then when two nodes occur in the same contexts more than 20% of the time, they (and the words they link to) are seen as mutually substitutable, and thus sentences other than those seen in the data can be generated. This is called a generative link. The simulations were run using two English corpora of child-directed speech. The simulations were run using varying levels of chunking.

Results: The Mosaic model was trained on the corpus, and then used to generate 500 utterances having an MLU of 3.5. Two human raters scored the generated utterances in terms of syntactic errors; semantic mistakes were acceptable (*Cut them with the knife* vs. *Cut them with the puzzle*). However, other errors such as word class, subject-verb agreement, missing verb arguments, wrong verb particle, wrong determiner, were not acceptable.

Conclusions: The authors concluded that using this model clear word class orders do occur, but not at very high rates. Thus Pinker's concerns from his work in 1987 are not significant. This research showed that the chunking mechanism was able to reduce the overall error rates. It was also able to prevent the substitution of similar words in incorrect contexts. This study was able to demonstrate two strengths of MOSAIC: the use of realistic child-directed speech and the production of utterances that can then be compared to child speech.

Relevance to the current work: This paper outlines some of the flaws that have been done in past research in child language acquisition. Many of these flaws are outcomes of human-labeled words. This study employs a computer-simulated program and found that the error rates were relatively low and contained thus contained an inferior amount of flaws than previous works in this area.

Guijarro-Fuentes, P., Rothman, J., Demuth, K., Patrolia, M., Song, J., & Masapollo, M. (2012). The development of articles in children's early Spanish: Prosodic interactions between lexical and grammatical form. *First Language*, *32*, 17-37.

Purpose: This study examined longitudinal data collected from two 1 to 2 year-old Spanish speaking children from Spain (Irene and Emilio). The purpose of the study was to evaluate the first spoken articles in the Spanish speakers compared to studies done in English and French. Preliminary studies suggest that first articles in Spanish are found in larger prosodic structures compared to English and French speakers.

Method: To examine the prosodic development of articles and nouns, they collected all of the nouns and nominal adjectives (*el grande* which is *the big one*) where an article is required. Each of the nouns was coded for the target number of syllables, the target word stress pattern, and which syllables were omitted. Each article was then coded for gender, number, definiteness, target number of syllables and the number of syllables produced. The total number of target article contexts was 1166 for Irene and 267 for Emilio.

Results: Irene produced 827 articles which was 70% of the contexts of where an article was required. Emilio, who was slower to develop, produced 160 articles. This constituted 60% of obligatory article contexts. There was a significant increase of the use of articles when they began to use multiword utterances. They found that both of the children truncated their multisyllabic (three or four) words. Irene tended to truncate more nouns, while Emilio only truncated once in the context of an article. Interestingly, before Emilio began producing articles there was a much higher occurrence of truncation. Compared to English and French speakers, the Spanish-speaking children did not show any difference in the use of articles with one-and two-syllable words. They also found that as the number of syllables in the noun increased, the use of articles decreased.

Conclusions: Because of the increase in the use of articles in multiword utterances the authors concluded that previous to this point Irene and Emilio had acquired some of the syntactic and semantic knowledge that would be needed to produce articles where required. Also, the fact that article production is so closely related to MLU suggests that it could be linked to syntactic knowledge development. Because of the higher amount of mutlisyllabic words in Spanish, it is more likely that they will attempt three- and four-syllabic words before their English or French-speaking peers.

Relevance to the current work: Both of the children in the study were also involved in the current work (Irene and Emilio). They found different results in each child; Irene expanded prosodic structure first, while Emilio permitted article production even with the use of truncated nouns. This article provides an addition to the literature on the understanding of how prosody influences the acquisitions of grammatical forms.

Gutierrez-Clellen, V. F., Restrepo, M. A., Bedore, L. M., Peña, E. D., & Anderson, R. T. (2000). Language sample analysis in Spanish-speaking children: Methodological considerations. *Language, Speech, and Hearing Services in Schools, 31*, 88-98.

Purpose: This article discuss how spontaneous language samples can assist in assessing children who are speakers of both Spanish and English and shows how these procedures can apply to research and clinical aims. The authors use Developmental Assessment of Spanish Grammar (DASG) and other methodological considerations to view measures of Spanish grammar with diagnostic potential.

Method: The authors discuss several factors that influence the morphosyntactic skills of Spanish-speaking children. Available language analysis procedures are compared and recommendations are offered for appropriate assessment of Spanish-speaking children. A reduction of inflectional morphology is seen in Spanish speakers acquiring English among other errors, but the errors found really are quite variable. The authors state which points to take into consideration when selecting measures of Spanish grammar. One should always keep in mind the most current developmental data regarding productivity of morphemes. The DASG was developed based on children from Mexican and Puerto Rican children living in Chicago. It describes morphosyntactic skills in six categories: use of indefinite pronouns and noun modifiers, personal pronouns, primary verbs, secondary verbs, conjunctions, and interrogative words. The categories each receive a weighted score in DASG according to developmental complexity. MLU and MLTU (mean length of terminable unit) of utterance are other measures. Sociolinguistic factors should also be taken into consideration. By investigating and describing these factors and assessments, the authors seek to give the reader further understanding about how to use them clinically. These principles are also encouraged to be applied by clinicians with bilingual Spanish-English children as well as those with language impairment.

Relevance to the current work: Studies regarding Spanish-speaking children are limited. The authors express the serious need that exists for developmental data to be done for Spanish-speaking children to help diagnose language learning and impairment in these individuals.

Long, S. H., & Channell, R. W. (2001). Accuracy of four language analysis procedures performed automatically. *American Journal of Speech-Language Pathology*, *10*, 180-188. doi:10.1044/1058-0360(2001/017)

Purpose: Long and Channell seek to find the accuracy of four language analyses that were performed automatically. This accuracy is compared to human coding accuracy.

Method: The authors assessed 69 language samples of children aged 2;6 to 7;10 years. The samples formed a diverse group of participants in order to simulate the range of challenges that would be present in a true clinical language analysis. The authors wished to see how the software would cope with varying demands, just as clinicians would need to do. Each child language sample consisted of between 132 and 232 complete utterances, with MLU that ranged from 1.02 to 10.56. The analyses were completed with Computerized Profiling after being run through a coding program called

GramCats. Each language analysis was performed under two conditions: all coding and tabulation done by CP, and in Condition 2 the codes generated by CP were reviewed by two human judges. The total accuracy was calculated comparing both of the conditions. The four analyses assessed MLU, LARSP, Developmental Sentence Scoring (DSS), and Index of Productive Syntax (IPSyn).

Results: The simplest measure, MLU, was the most accurately calculated. The IPSyn and DSS scores were also highly correlated between Condition 1 and Condition 2. For LARSP, the coded structures were reasonably accurate at the word, phrases, and clause levels, but was not very accurate when coding subordinate clauses.

Conclusions: As found in child language research, interrater agreement in this study showed acceptable (LARSP) to good (IPSyn, DSS) to excellent (MLU) reliability when compared to corrected analyses (.85, .90, .95 respectively). The authors found that the language samples analyzed by the CP program were equivalent to human coding. The authors recommend the use of this software to aid, not replace, the task of analyzing language data.

Relevance to the current work: The current study utilizes computer software to grammatically code Spanish language samples. Using computer software such as discussed in the article has facilitated work in language data analysis.

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

Purpose: Mintz introduces the concept of frequent frames. The study seeks to determine the usefulness of information given from frequent frames in grammatical categorization. The study advances research done on distributional patterns in grammar. Arguments in the reliability of distributional patterning led to further investigation. The potential problems did not significantly undermine how informative the distributional patterns were.

Method: Mintz selected six corpora from the CHILDES database. The author states that he chose to look at individuals rather at larger samples because analyzing individual input allows evaluation of the informativeness of patterns in the input to the individuals. This then, is ultimately the database from which children learn individually. Childers and Tomasello in their 2001 study found that children more easily acquire novel verb meanings when the verbs occur in lexical frames that occur more frequently in the children’s input. This and other studies provide evidence that these frames are psycholinguistically relevant for language learners. In this article Mintz compares these frames with previous distributional approaches using bigrams. In order to analyze the frames an exhaustive tally was made of all the frames. Then a subset of these frames was selected as the set of frequent frames. A frame-based category was then created. The number of times each word occurred within a frame was also recorded. Accuracy scores were computed for each frame based-category by looking at all possible pairs of word tokens within the category. The pairs were classified as a Hit or a False Alarm. Accuracy was defined as the proportion of Hits to the number of Hits plus False Alarms. In order to do this the tokens were labeled with their true grammatical category. The study also measured completeness, or the degree to which the analysis grouped in the same distributional category words that belong to the same grammatical category.

Mintz performed a second experiment in order to examine the categorization outcome when a frame selection method was used that is sensitive to the frequency of frames relative to the total number of frames in a corpus. In Experiment 2 the threshold used provided a type of normalization of the method that was used in Experiment 1. In Experiment 2 the frequent frames which consisted of only one or two word types were removed from the set of frequent frames. This was done in order to guard against minimal frame-based categories contributing to high accuracy.

Results: In Experiment 1 the mean token accuracy and completeness for Standard and Expanded Labeling were both significantly higher than baseline. The major finding of the frequent frames was that they are extremely effective at categorizing words. This is especially impressive as one considers the restricted distributional contexts of only using the 45 most frequent frames. The frames consisted largely of closed class items such as determiners, prepositions, auxiliary verbs, and pronouns. Because the distributional information was robust, the frequent frames can focus a child on a relatively small number of contexts that can have a broad impact on how words in the input are categorized. Mintz states a limitation in his experiment: the set of frequent frames was selected by the same absolute threshold for all corpora. He suggests analyzing the corpora using a frequency threshold for each corpus that is based on a relativized frequency criterion. As with Experiment 1, Experiment 2 results found the mean token accuracy and completeness for Standard and Expanded Labeling both significantly higher than baseline.

Conclusions: The study found considerable consistency across the corpora. The author suggests that two frame-based categories can be unified if they surpass a threshold of lexical overlap. Mintz also suggests that future studies will be needed in order to determine whether children actually make use of frame-like information.

Relevance to Current Work: This work showed frequent frames to be extremely effective and efficient sources of information for categorizing words in children's input. The outcomes of this study provided the basis for the study done by Waxman and Weisleder in analyzing how children learn grammatical categories. However, we know that while found to be effective for categorization, frequent frames have limitations in their ability to show how children learn grammatical categories.

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

Purpose: Pinker outlines the fact that language acquisition is most fundamentally viewed as the input and output of language. The chapter discusses the fundamental problem of getting a child started in forming the correct types of rules for language acquisition. He discusses three solutions that do not work, another solution from a past paper of his (1984), and then sketches out an alternative class of language acquisition. Most importantly, Pinker discusses how all suggested models contribute to the debate on semantic bootstrapping.

Method: Pinker reviews the three methods of bootstrapping: correlational, prosodic, and syntactic. According to the correlational hypothesis a child uses distributional information (such as serial position, position relative to other words, inflections, and other semantic ideas within sentences) to learn grammatical categories.

Pinker notes that this hypothesis cannot be true; this is too large of a generalization for a child to make. Pinker notes that any theory of language acquisition must account for the effects of constraints on languages and how the child can acquire full competence in language. Prosodic bootstrapping is based on the belief that a child is able to record the intonation contour, stress pattern, relative timing of elements, and pauses in input sentences in order to infer the phrase structure of a sentence. However, we know that phrase structure is only one determine of prosody, but not the final word. While we know prosody is important in language bootstrapping, this hypothesis does not provide sufficient information. Supporters of syntactic bootstrapping argue that possible grammars are subject to many innate constraints so that even a small distributional analysis is sufficient to yield correct categorization of language. Citing work from Lasnik (1987), Pinker argues that Lasnik's evidence for this theory begins with too many presuppositions on what a child might actually know about the language. Pinker then reviews the semantic bootstrapping hypothesis which was first assumed by Wexler and Culicover in 1980. This hypothesis requires four background assumptions. Pinker notes that it is important to understand that the semantic bootstrapping hypothesis claims that certain semantic elements are sufficient conditions for the use of syntactic in motherese of early child speech, not necessary conditions. Then Pinker outlines ten distinct problems with the semantic bootstrapping theory.

Conclusions: Pinker discusses what these problems are able to tell us about a child's bootstrapping mechanism. He concludes that these problems point to what he calls a "smoking gun" assumption as the real problem. I other words--the assumption being that grammatical development is driven by a distinguished set of cues that are perceivable in the input and are also uniquely diagnostic syntactic rules. These rules then serve as the premises of deductive generalizations that then yield grammar. He proposes that a model of language acquisition should have the following properties: 1. A child's inferences do not need to flow from a specially-designated subset of correlated phenomena, but rather can flow from any member of the set that is exemplified in the input. 2. Partially diagnostic input cues should be usable to make guesses about the language. 3. Rule hypotheses, as well as distributional analyses, should be tentative and easily revised. 4. A decision that is favored by the input cues and that leads to the greatest degree of successful further learning is retained.

Relevance to the current work: Pinker gives an overview of many of the theories of language acquisition. He ensures that the reader understands that these theories and methods of data analysis are incomplete. The current study seeks to form a more complete picture of language acquisition in children. This study especially incorporates the idea that decisions that are favored should be retained. This correlates well with the idea of adaptation and selection featured in this work.

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

Purpose: Redington, Chater, and Finch seek to demonstrate that distributional information does in fact aid children in their learning of syntactic categories. Through discussion of information gathered by a distributional analysis of a corpus of speech from the CHILDES project, they show that a distributional analysis is highly informative of syntactic categories. They also argue that this information can be drawn by some psychologically probable mechanisms.

Method: The authors discuss the problem of learning words' syntactic categories and consider with the reader possible sources that could assist in solving this problem. They outline four main sources of information in linguistic input that could be useful in learning syntactic categories. They include using distributional information, semantic bootstrapping (relating the linguistic input to the situation in which it occurs), phonological cues, and prosodic information. They also discuss the innate knowledge of language learning and its effects on the current issue. The authors acknowledge the arguments against the usefulness of distributional analysis and refute these arguments given primarily by Pinker in 1984. They outline previous work accomplished in favor of distributional learning methods that are potentially psychologically relevant. These include distributional analyses in linguistics, within the neural network, and in statistics.

Results: Redington, Chater, and Finch propose three stages using distributional information in classifying syntactic categories. 1. Measuring the distribution of contexts within which each word occurs. 2. Comparing the distributions of contexts for pairs of words. 3. Grouping together words with similar distributions of contexts. The experiments performed used transcribed speech from the CHILDES database. They included a benchmark classification of each target word, which is the syntactic category within which it most commonly occurs. They evaluated the accuracy and completeness of the scores. Information-theoretic scoring was also utilized, which produces only one measure that reflects both accuracy and completeness. In Experiment 1 they found that the preceding context appears to be much more useful than the succeeding context. However, the best results were found by combining the two contexts. In Experiment 2 of varying the number of target and context words, they discovered distributional analysis works well even when there is a small vocabulary of target and context words so long as the majority of the target words are content rather than function words. In Experiment 3 they found that distributional information is more useful for content words rather than for function words. In Experiment 4 they varied the corpus size in order to ask how much input is required for the method to be effective. They report that given more input, it seems probable that a slight increase in performance could be expected. In Experiment 5 the authors looked at the distributions across utterance boundaries because they tend to contain a high amount of noise. By looking only at context items that were in the same utterance as the target it reduced the quantity of noise. However, they found that besides improving classification, marking utterance boundaries provided little extra benefit. Experiment 6 looked at the difference between frequency and occurrence. They discovered that co-occurrence can be used to constrain words' syntactic categories when frequency information is not available; however, the distributional method works better when frequency information is included. Experiment 7 looked at removing function words. They found that removing function words has a large impact of the amount of information that is provided by the distributional method, but it still provides a relatively great amount of useful information. Experiment 8 led the authors to conclude that frequency information about lexical items in each class could be important in exploiting distributional information. Experiment 9 discussed the concept of motherese and found that the distributional mechanism is not dependent on the use of motherese because of the evidence that shows that children who are not recipients of motherese do not acquire language any less quickly.

Conclusions: Following the eight experiments the authors concluded that distributional information is a powerful cue in learning syntactic categories. However, they do not disregard the fact that innate knowledge is also critical.

Relevance to the current work: The authors of this work highlighted the fact that distributional analysis is both informative and useful to children learning syntactic categories. They conclude that this process would be most successful for languages which have strong order constraints because it uses sequential order information. In the current study we look at how grammatical categories are acquired in the Spanish language using an evolutionary algorithm. Spanish has a different syntactical structure than English. They guess some success for languages in which word order is relatively free. Using processes of adaptation and selection we will discover how true their guess proves to be.

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

Purpose: Young children must locate and distinguish linguistic units in the speech they hear in order to learn a language. This study looks at whether 60 children (ages 2;0-2;2) used grammatical and caregiver cues in their sentence comprehension. The study looks to see how these different types of cues interacted and how they affected children language comprehension.

Method: Numerous studies have found that infants are sensitive to prosodic cues in speech as well as highly frequent grammatical morphemes. Other cues not related to grammar such as placement of key word in utterance-final position and shorter utterances could possibly aid the child in linguistic differentiation. This study was composed of two main experiments; Experiment 1 examined distribution and prosodic cues and their role in comprehension, and Experiment 2 examined the role of distributional and positional cues. In Experiment 1 children listened to sentences that asked them to point to a picture representing a target noun. Each sentence was created with a pause in one of three locations in order to test the role of prosody. Each sentence also included words that were either grammatical or not for that context. Experiment 2 examined whether utterance position and length continue were important at age 2;0 to 2;2. It also examined if what the authors found in Experiment 1 held true if the target words were paced in utterance-final position of short sentences. Experiment 2 used the same comprehension task as in Experiment 1.

Results: The analysis for Experiment 1 revealed main effects for both prosody and grammaticality. Experiment 2 found a significant main effect of position; children showed better comprehension in utterance-final targets than in utterance-internal targets. Grammaticality was also a main effect in Experiment 2.

Conclusions: Experiment 1 showed that prosodic and grammaticality cues aid children in their early sentence comprehension. Experiment 2 found a main effect for utterance position and length which supports the data results found by Fernald & McRoberts that caregiver cues continue to play an important role in sentence comprehension when children are two years of age. The authors concluded that both language internal and caregiver cues are important in early sentence comprehension.

Relevance to the current work: This article supports that a caregiver's child-directed speech plays a key role in a child's ability to learn and comprehend language. This study also showed the importance of grammatical sentences in language comprehension.

Shapiro, K., & Caramazza, A. (2003). The representation of grammatical categories in the brain. *Trends in Cognitive Sciences, 7*, 201-206.

Purpose: Shapiro and Caramazza seek to answer two fundamental questions about language: "How is knowledge about different grammatical categories represented in the brain, and what components of the language production system make use of it?" The authors use evidence from neuropsychology, electrophysiology and neuroimaging to illustrate that information about a word's grammatical category may be represented in the brain. The study focuses on nouns and verbs.

Method: To this point, experiments on the retrieval of grammatical category information in the brain have not been conclusive, and conflicting at best. Electrophysiological measures have suggested that nouns and verbs are processed by distinct neural generators. When neuroimaging has been put in place, little evidence has been found to support an anatomical distinction. In fact, these neuroimaging results have led some to believe that words are stored in distributed networks. Event-related potentials (ERPs), neuroimaging (PET and fMRI) and transcranial magnetic stimulation (TMS) are three techniques that have been used to test hypotheses about the role of brain regions in processing words of different grammatical categories. The parts of the brain that are discussed for possible involvement in grammatical category processing are the left frontal cortex, the left hemisphere, the parietal cortex, prefrontal cortex, and the temporal lobe. Basically, because of conflicting results of past studies, no real conclusive data has been brought forth. In this article the authors review these different studies in attempt to make a more solid conclusion. The authors found that generally nouns and verbs are represented in the brain, but they are represented separate from their meaning at the levels of word form and morphological computation.

Conclusions: Considering each of the presented studies, the authors advocated that researchers in this topic area should define the level of representation in the brain at which grammatical category is being investigated. When this is known, tasks should be chosen which are sensitive to computations at that level of representation.

Relevance to the current work: As the authors discuss, having an understanding of how words are represented and retrieved in the brain is crucial to any future neurobiological theory of language. Expansive research has been done to see how language is learned, but to date research still lacks a conclusive answer. The current study uses a theory based on evolutionary principles to attempt to contribute a solution to this unanswered question.

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.

Purpose: The authors seek to discover if bigrams or trigrams are more useful as a distributional cue for learning grammatical categories. They hypothesize that children construct more accurate "flexible frames" (called trigrams) spontaneously from lower-order distributional patterns (called bigrams) that have broader coverage and are thus less accurate. They base their study on Mintz's 2003 study on frequent frames.

Method: The authors selected the same six corpora of child-directed speech from the CHILDES corpus used by Mintz in 2003. The 45 most frequent frames within each corpus were then selected with all the words that occurred within them. These frames were labeled "aXb", where a__b refers to the "non-adjacent co-occurrence" frame, and X refers to the set of words that occur in this context between "a" and "b". To analyze the fixed frames the most frequent a_b "non-adjacent co-occurrences" were selected. The words (X) that intervened between the a and the b word were then grouped together. They then analyzed both the preceding bigram (aX) and the succeeding bigram (Xb). For each bigram analysis they selected the 45 most frequent words. Both token (every word occurrence was counted) and type (only distinct words were counted) analyses were performed. The authors also established a random baseline, in which all of the words that were categorized in the analysis were randomly assigned across the 45 frequent frames to create a random analysis. The authors computed accuracy and completeness measures so as to be comparable to Mintz's 2003 study. To determine accuracy, the number of hits (when two words occurring in the same frame were of the same grammatical category) was divided by the number of hits plus the number of false alarms (when two words occurring in the same frame were from a different grammatical category) (accuracy = hits/(hits + false alarms)). Accuracy gave an overall measure of how successful the distributional cues were at grouping words of the same grammatical category together. Completeness measured how well the distributional cues grouped all words from one grammatical category together in the same distributional cue grouping. Completeness was the number of hits divided by the number of hits plus the number of misses (when two words of the same category occurred in different frames) (completeness = hits/(hits + misses)).

Results: The authors found results similar to Mintz in both accuracy and completeness. In summary, the aX and Xb frames were less accurate than the aXb analysis, likely due to their reduced specificity. However, the aX frames were found to be more accurate than the Xb frames, indicating that high frequency preceding words were more effective at classifying target words than succeeding words.

Conclusions: St. Clair, Monaghan, and Christiansen concluded that their study provides support for a new view of the distributional information that a child uses in order to determine grammatical categories. They conclude that their replicated study supported Mintz's 2003 findings.

Relevance to the current work: These trigrams are what Mintz used in his 2003 study on frequent frames. Bigrams on the other hand, have a more broad coverage of the input but have a lower accuracy in grammatical classification. The authors argued that there are several reasons to suspect that fixed trigram frames are not likely to be initial cues that are utilized by children in their learning of grammatical categories. Their study helped show that the idea of a frequent frame is useful in assisting children in learning grammatical categories. Trigrams have been found to be very accurate in the grammatical classification of words, but only cover a small part of a child's language exposure.

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

Purpose: The authors' purpose of this chapter is to present past studies' finding on how language is represented in the human brain at a behavioral, cognitive, and neural level. Empirical research has found that grammar, syntax in particular, is based on structure-dependent, recursive hierarchical computation. The authors explain what this hierarchical computation is: higher-level structures (sentence-level clauses) are composed of a combination of lower-level entities (determiners, nouns, adjectives). The recursive algorithm then is able to generate an infinite amount of sequences. Tettamonti and Perani state that it is due to the hierarchical nature of languages that humans are able to infer the intended conceptual meaning of a phrase from just a linear sequence of speech. While language is organized by multiple levels of structure-dependent hierarchical levels, most levels are only viewed a surface level of language production and perception. Most hierarchical relations are established at a deeper, more abstract level. They conclude that this structure is why there are so many remaining questions about human language acquisition. ERPs and FMRIs have shown infants' ability to differentiate phonemic and prosodic contours. This ability has been assumed to help determine segmental transitions across words. Neurophysiological studies on grammatical category acquisition have shown correlations between structures in regards to timing and the anatomy and processes that occur in the brain in infancy. The authors cite research done on grammatical category learning (Cartwright & Brent, etc.) and the fact that such research lacks conclusive evidence on how children learn to assign words to grammatical categories. The authors share what has been determined with studies done with children with SLI that has emphasized the role of the anterior left hemisphere in syntactic processing.

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.

Purpose: The aim of this study was to advance the evidence found by Mintz in 2003 regarding the notion of frequent frames assisting in young children learning distinct grammatical categories (noun, verb, adjective). Mintz's work supported the hypothesis that distributional information in the input could support the discovery of distinct grammatical forms. Weisleder and Waxman look to advance this notion across two languages: English and Spanish.

Method: Weisleder and Waxman first examine the distributional evidence available to young children who are acquiring Spanish and then compare it to the evidence available to those acquiring English. They also consider the clarity of frequent frames for identifying the grammatical categories of noun, verb, and adjective. Differing from Mintz's work in 2003, Weisleder and Waxman also consider phrase-final sequences, or "end-frames."

Weisleder and Waxman selected six parent-child corpora from the CHILDES database and analyzed the input when children were aged 2;6 or younger. Three of these corpora were in English, three in Spanish. The English corpora were those that were previously examined by Mintz in 2003. This was helpful to ensure that their execution of the frequent frames was comparative to Mintz's work, and thus able to be compared in Spanish. They began by gathering the frames, defined as two linguistic elements with one word intervening. They consider both "mid-frames" (A__B) and "end-frames" (A__). Each adult utterance was segmented into three element frames. They did not include frames that crossed an utterance boundary.

Next, they selected the forty-five most frequent mid-frames and forty-five most frequent "end-frames." They then identified the intervening words (both types and tokens) The intervening words were then assigned to a grammatical category by a native speaker of each language. The assignments were then checked by a Spanish-English bilingual.Accuracy, or consistency of the frame-based categories was computed by comparing every pair of words that occurred within any frame. They were labeled as either a "Hit" or a "False Alarm." The authors focused on Accuracy for token frequencies in particular. They also computed the Completeness of the frame-based categories, but only the results of accuracy were reported in their study. A Monte Carlo method was used to obtain a baseline categorization measure for computing accuracy scores for random word categories.

Results: For each corpus they compared the Accuracy score for the frame-based categories to the corresponding baseline measure. In order to view systematic differences in accuracy between the two languages they found the accuracy score for the frame-based categories and frame-type for each language. They used a three-way ANOVA and Tukey's HSD. They found that the accuracy for English was higher than Spanish. They also found that a main effect for frame-type indicated that accuracy was higher for mid-frames than end frames. Verb-frames revealed higher accuracy than noun-frames and noun-frames had higher accuracy than adjective frames. Each difference that they found was statistically reliable.

Conclusions: The results of this study support the previous evidence found by Mintz. It also extends his work by examining input to children acquiring Spanish as compared to English, considering mid-frames as well as end-frames, and considering the accuracy of the distributional evidence available for each of the grammatical categories within the frequent frames.

Relevance to the current work: They found that frequent frames did contain robust cues for grammatical categories for nouns and verbs. Weaker cues were found for adjectives. This pattern was more pronounced in Spanish than English. This is one of the only studies performed both in English and in Spanish.