



2015-03-01

# Modeling Children's Acquisition of Grammatical Word Categories from Adult Input Using an Adaptation and Selection Algorithm

Nicole Adele Stenquist  
*Brigham Young University - Provo*

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



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

---

## BYU ScholarsArchive Citation

Stenquist, Nicole Adele, "Modeling Children's Acquisition of Grammatical Word Categories from Adult Input Using an Adaptation and Selection Algorithm" (2015). *All Theses and Dissertations*. 5830.  
<https://scholarsarchive.byu.edu/etd/5830>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu), [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

Modeling Children's Acquisition of Grammatical Word Categories from Adult Input  
Using an Adaptation and Selection Algorithm

Nicole Adele Stenquist

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

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

Department of Communication Disorders

Brigham Young University

March 2015

Copyright © 2015 Nicole Adele Stenquist

All Rights Reserved

## ABSTRACT

### Modeling Children's Acquisition of Grammatical Word Categories from Adult Input Using an Adaptation and Selection Algorithm

Nicole Adele Stenquist  
Department of Communication Disorders, BYU  
Master of Science

Children acquire the use of grammatical categories in their native language, and previous models have only been partially successful in describing this acquisition. The present study uses an adaptation selection algorithm to continue the work in addressing this acquisition. The input for the computer model is the transcribed language of parents and caregivers towards three children, whose ages ranged from 1;1 to 5;1 during the course of sampling. The output of the model consists of the input words labeled with a grammatical category. This output data was evaluated at regular intervals through its ability to correctly identify the grammatical categories in the utterances of the target child. The findings suggest that the use of this type of model is effective in categorizing words into grammatical categories.

Keywords: grammatical word categories, evolutionary programming, language acquisition

## ACKNOWLEDGEMENTS

I thank the researchers who donated child language samples to the Child Language Data Exchange System (CHILDES) archive as well as the families whose data were sampled. I thank Dr. Channell for his help in making this thesis possible. His efforts, support, feedback, and input were all invaluable to this process. With his help, this thesis turned out to be an enjoyable learning experience for me. I also thank Dr. Nissen and Dr. Tanner for their input and efforts.

I would also like to thank my family. This thesis would not have been possible without their faith, love, and support.

## TABLE OF CONTENTS

LIST OF FIGURES.....	v
DESCRIPTION OF THESIS STRUCTURE.....	vi
Introduction.....	1
Method.....	5
Participant Samples .....	5
Computer Model .....	6
Procedure.....	7
Results.....	7
Discussion.....	9
References.....	13
Appendix: Annotated Bibliography.....	16

## LIST OF FIGURES

Figure	Page
1. Results for Adam's corpus at both mutation rates.....	8
2. Results for Peter's corpus at both mutation rates.....	8
3. Results for Naomi's corpus at both mutation rates.....	9

## DESCRIPTION OF THESIS STRUCTURE

This thesis is part of a larger research project, and portions of this thesis may be published as part of articles listing the thesis author as a co-author. 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 knowledge of the grammatical word categories (GWCs) of their language (e.g., noun, verb, preposition, conjunction) and can extend this knowledge to novel words (Berko, 1958; Ratner & Menn, 2000). Researchers agree that children are not born with a robust vocabulary and functioning system for syntax, but the question of exactly how language is acquired or how any innate endowments are realized in the child has been vigorously debated. Over the past two decades, two major theoretical perspectives have become dominant: the nativist (also known as generativist or Universal Grammar) perspective and the constructivist (also called the emergentist, socio-pragmatic, functionalist, or usage-based) perspective (Ambridge & Lieven, 2011).

The nativist perspective holds that newborn children already possess some innate knowledge of syntactic categories and are only required to label each new word with an already existing category. In contrast, constructivism holds that language is acquired through cognitive or perceptual analysis of input and that children are not born with inherent grammatical categories, needing to somehow extract these categories from the language they hear.

Both nativist and constructivist perspectives have assumed that children use some type of sophisticated processing to make sense of the language input to which they are exposed, due to limited evidence. This processing is known as bootstrapping. Bootstrapping can be semantic, prosodic, or syntactic in nature. Semantic bootstrapping refers to a process by which children use semantic information to predict how a word will be syntactically placed within a sentence structure (Pinker, 1984). Prosodic bootstrapping refers to the use of prosodic cues to detect information about language structure from speech (Morgan & Demuth 1996). Syntactical bootstrapping refers to the use of distributional information to provide insight to a word's

semantics and thus its GWC, and this form of bootstrapping has been perhaps explored the most in research in studies such as those by Redington, Chater, and Finch (1998), Cartwright and Brent (1997), Mintz (2003), and Freudenthal, Pine, and Gobet (2005).

In Redington et al.'s (1998) study, data from adults (usually parents) talking to children were taken from the Child Language Data Exchange System (CHILDES; MacWhinney, 2000) database and analyzed. Words were grouped based on their distributional similarity, and these groups were evaluated based on their similarity to conventional GWCs. The results showed favorable evidence for the usefulness of distributional information in GWC acquisition.

Cartwright and Brent's (1997) study proposed the hypothesis that children create new templates for input they receive and then merge overlapping templates together. Thus, their algorithm addressed one sentence at a time and also resulted in a smaller set of categories than those of a small study done by Kiss (1973). Cartwright and Brent's study contained five experiments, and each experiment manipulated variables such as the amount of input presented to the computer, the addition of ambiguous words to the input, the use of child-directed speech, and the addition of limited semantic information to the system. Accuracy and completeness scores were shown to improve based on distributional information, reaching an average level of about 50%.

Mintz (2003) also used a distributional approach by looking at frequent frames of child-directed speech. A frequent frame was defined as two words that frequently co-occur in succession with one word intervening (e.g., *you\_\_it, the\_\_one*). The 45 most frequently occurring frames for each corpus were analyzed. The accuracy of all the scores was found to be statistically significant, but the completeness scores were found to be relatively low. The second experiment took the number of frequency of frames into account in comparison to the total

number of frames per each corpus. A limitation of this study was that the percentage of each language sample supporting the analysis and delineation of frequent frames was small.

Freudenthal et al. (2005) suggested that the only way to judge the effectiveness and the functionality of generated grammatical categories is through producing a model that generates sentences using these categories. Freudenthal et al. used a computational model called MOSAIC, which uses a chunking mechanism to sort words into grammatical categories, to test this idea. Freudenthal et al. suggested that this method of generating sentences was effective in inferring the goodness of derived GWCs, but their software model yielded utterances with an average length of only 3.5 morphemes. These obtained small sentence lengths illustrate the difficulty present in generating and testing GWCs without comparing these categories to those in human-coded transcripts.

The studies described above have used syntactic bootstrapping in some way to address the question of GWC acquisition. Recently a new way of addressing this question without the use of bootstrapping has been explored using an adaptation and selection algorithm. In this type of algorithm, GWCs start off as being randomly paired with words, and given but limited exposure to data, increasingly better sets of these links evolve to yield an adequate level of correctness (Channell, Nissen, & Tanner, 2014). Preliminary results suggest that such a model is capable of modeling the acquisition of GWCs based on exposure to child-directed utterances, but this capability has only been tested by noting the level of agreement with conventional GWCs in the same set of adult utterances used to train the model.

Young (2014) examined a model that implemented aspects of evolutionary biology, such as variation, adaptive change, self-regulation, and inheritance. Young applied this model to six English language corpora. The computer model used created dictionaries based on the words in

each corpus and matched those words with a single grammatical tag. These dictionaries evolved over 5,000 generations, and four different mutation rates were used in creating offspring dictionaries. Young found that the accuracy achieved by the model in correctly matching words with tags approached 90%.

Judd (2014) applied an evolutionary model to the corpora of language addressed to five Spanish-speaking children, whose ages ranged from 0;11 to 4;8 (years; months). As in Young (2014) the model evolved dictionaries that linked words to their most likely grammatical tags and was run for 5000 cycles; four different rates of mutation of offspring dictionaries were also assessed. The accuracy for coding the words in the corpora peaked at about 85%.

Cluff (2014) used an adaptation and selection model to assign words to their grammatical categories using language addressed to five children who ranged in ages ranging between 1;1 and 5;1. In Cluff's model, each evolving dictionary could have up to three tags per word; however, the average number of tags per word for each corpus, which ranged from 1.12 to 1.16, was a target which along with tag accuracy helped to determine the fitness of a dictionary and thus control the evolutionary process. The accuracy for coding the ten corpora used averaged 92.74% after 4000 evolutionary cycles.

The results obtained by Young (2014), Judd (2014), and Cluff (2014) suggest the feasibility of an adaptation and selection model of children's acquisition of the GWCs of words. However, these studies assessed the accuracy of that model only using alternate utterances from the same corpora as were used for training the model. This decision may have led to overestimation of the accuracy of the model. An evaluation using utterances not drawn from the training corpus would lead to a more general assessment of accuracy. The present study evaluates a computer model based on adaptation and selection principles by examining the

accuracy with which a sample of child's utterances may be GWC coded by the model when it is trained using a corpus of the caregiver input to that same child.

## **Method**

### **Participant Samples**

Samples from the CHILDES database (MacWhinney, 2000) were used as input for the computational model. Utterances included spoken language. The language data available for each child was divided into two corpora: one had the utterances spoken by adults (usually a parent) in the context of or directed at the child, including questions, statements, and commands. The other corpora contained the child's utterances. For each of the children, the last 500 non-repeated multi word utterances were used as the child or test utterances. More background detail was available for some children and interactional contexts than for others.

**Adam.** (Brown, 1973). Adam came from a middle-class, African-American family that spoke Standard American English. His parents were well educated and were employed as a minister and an elementary school teacher. His language samples were taken in his home with his parents and few other adults. They started when he was age 2;3 and ended when he was 4;10. The total number of samples collected was 55. In these samples 19,301 of the utterances spoken by adults in his presence were used (Brown, 1973).

**Naomi.** (Sachs, 1983) Naomi was born June 8, 1968. A total of 93 language samples were collected. These samples were collected in her home, during parent-child interaction, when she was age 1;1 to 5;1. This sample consisted of 12,034 child-directed utterances.

**Peter.** (Bloom, Hood, & Lightbown, 1974; Bloom, Lightbown & Hood, 1975) Peter was the oldest child of his family and was born on December 27, 1969. He belonged to a middle- to upper-middle class family that lived in a university community in New York City. His samples

were taken largely from interactions with multiple adults during play in his home. From his 20 language samples, there were 20,827 child-directed utterances spoken by adults. Peter was age 1;9 during the first sample and 3;2 in the last sample.

The adult corpora, used for training the model, and the child corpora, used for testing the model, were formatted for automated GWC coding, coded, and this coding was then corrected by a graduate student in speech-language pathology.

### **Computer Model**

The computer modeling program used in the present study was called `ns_ev` (Channell, 2015). The `ns_ev` program evolves a set of GWCs for English words, given exposure to English language input, using an adaptation and selection model. The set of words from the corpus, each paired with one GWC, forms one dictionary. At the program's start, a set of 500 dictionaries is randomly populated such that each word is given a randomly chosen GWC. The program is then run for 4,000 evolutionary cycles. In each cycle, each of the dictionaries codes a section (5000 utterances) of the training corpus. These codes are compared to the automated tagging (with manual revisions) of the same utterances, and the total number of agreements is saved as the score for each dictionary. The 20 highest scoring dictionaries are identified and allowed to reproduce. In the present model's context, reproduction means that a child dictionary is created identically in GWC definitions to the parent except that, with a pre-set level of mutation, an alternate GWC definition is randomly drawn for a word without any information as to whether the word's previous GWC definition was correct or not. These child dictionaries compete by coding utterances in the training corpus, the best dictionaries of that generation are allowed to reproduce, and this process is repeated 4000 times.

At 500-cycle intervals, the best-scoring dictionary codes the target child's corpus. This tagging is compared to the automated tagging of that corpus, and the percent of accuracy is written to a file. However, no feedback from this test is returned to the evolving set of dictionaries.

### **Procedures**

Corpora were formatted for automated analysis by removing extraneous punctuation, coding, and notations, and by dividing them so that there was only one independent clause per line. Utterances were in lower case characters except proper nouns and the pronoun *I*. Words in the corpora were then software coded as to GWC, and the correctness of these GWCs was manually checked and corrected if necessary. The adult training corpora together with the child test corpora were then processed using the computer model. At 500-cycle increments, the computer model reported the accuracy of agreement between the best dictionary's coding of the child's corpus and the manual coding of the same corpus.

### **Results**

Growth curves for each child's corpus at mutation rates 1/1600 and 1/2400 are found in Figures 1-3. As shown by the figures, a rapid increase in accuracy was seen in generations 1-1500, with a significant increase until generation 3000, and a gradual increase until generation 12,000 for all three participants studied. The percent accuracy for both Adam and Peter both leveled out at slightly above 80, while the percent accuracy for Naomi leveled out at just less than 80. All three participants had higher percentages with a mutation rate of 1/2400, though the difference made by mutation rate was quite small.

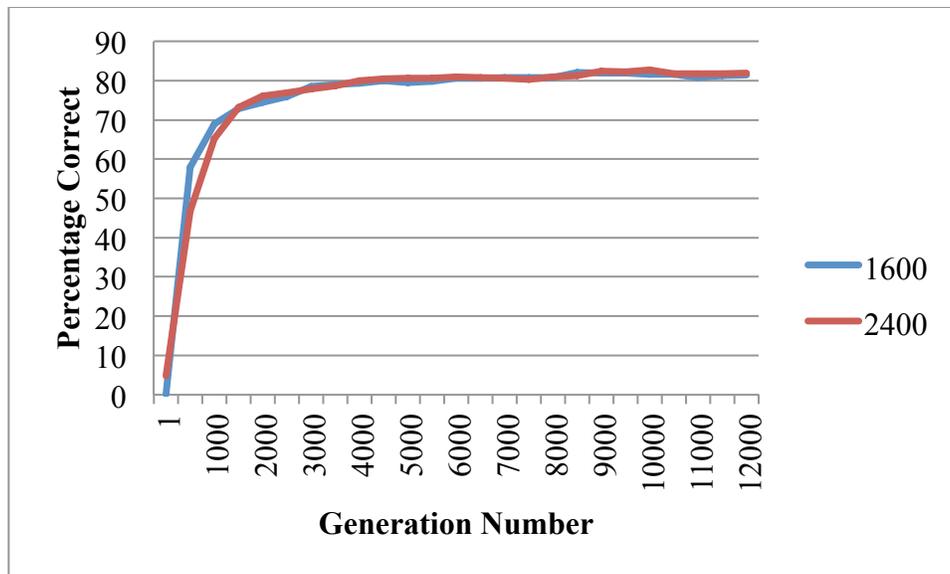


Figure 1. Results for Adam's corpus at both mutation rates.

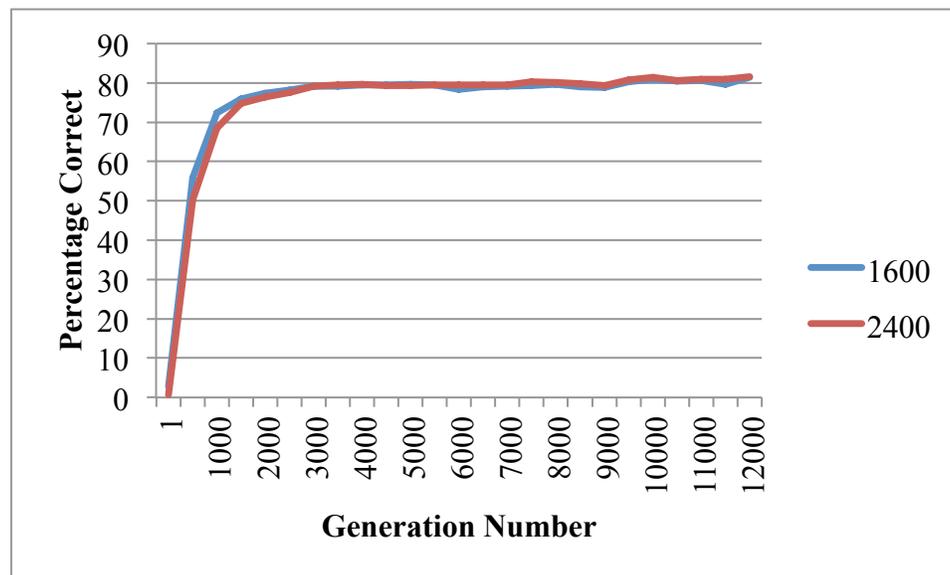


Figure 2. Results for Peter's corpus at both mutation rates.

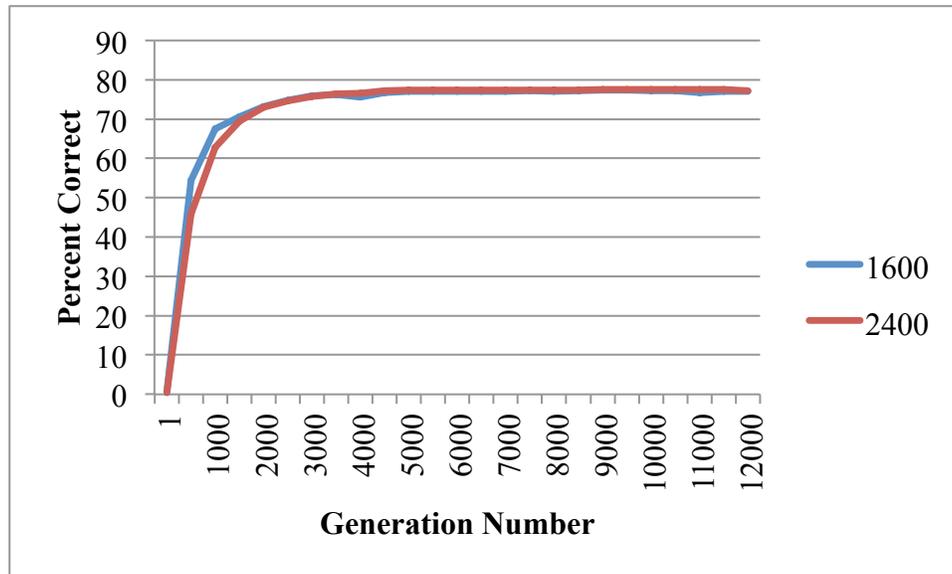


Figure 3. Results for Naomi's corpus at both mutation rates.

### Discussion

In the present study, the model implementing an adaptation and selection algorithm was successful in achieving high levels of accuracy in assigning words to their grammatical categories when trained on the adult utterances from a corpus and tested on the utterances produced by the child in the same corpus. The program's accuracy levels increased greatly over the first 1,500 generations, increased steadily through generation 3,000, and leveled off from then until generation 12,000.

A slight difference between the accuracy levels of the different children was found, with Adam and Peter's results being slightly higher than those of Naomi. Adam and Peter's accuracy levels both slightly exceeded 80% while Naomi's accuracy levels approached 80%, but never reached that level. The reason for this disparity is not immediately apparent. Future studies could be conducted on larger numbers of participants in order to determine a normative level of accuracy under the same circumstances. Additionally, a slight difference between the mutation rates was found with 1/2400 achieving higher levels of accuracy across participants than 1/1600.

Whether 1/2400 is the optimal mutation rate will also have to be explored through further research.

This study achieved higher levels of accuracy than a similar study just completed by Chatterton (2015). In Chatterton's study, five corpora of adult utterances were used as a training set for an adaptation and selection model, and 30 child samples from Channell and Johnson (1999) were used to evaluate the accuracy of the GWCs assigned. Chatterton found that the average accuracy level for the 30 participants was 77% and only one child attained 80%, while in the current study two out of three of the participants had achieved accuracy levels of greater than 80%. Future studies may well use a larger number of participants than the present study, but one possible reason for this difference in accuracy levels is probably the influence of the language style of the adult on the child that hears his or her speech.

Three other recent studies that investigated the question of grammatical category acquisition using an adaptation selection algorithm were done by Cluff (2014), Judd (2014), and Young (2014). These studies, like the present study, also found that a lower mutation rate achieved a higher level of accuracy. These studies achieved levels accuracy levels between 89-94%, notably higher those obtained in the current study. These higher accuracy levels were likely due to two things. First, the current computer program evaluates only a window of the corpus at a time during training, while these two past studies used the odd numbered utterances of the entire corpus for training and the even numbered utterances for testing. Accuracy levels of these past studies were also tested using utterances of the same speaker upon which the program had been trained. Cluff's work achieved the highest accuracy levels, likely due to the allowance of multiple GWC tags per word, a feature not present in this study or those done by Judd or Young.

Three other studies that investigated the same question of GWC acquisition to which the present results can be compared were those of Mintz (2003), Redington et al. (1998), and Cartwright and Brent (1997). Each of these studies used a different distributional approach to addressing the question of GWC acquisition. Mintz did this through the use of frequent frames, Redington et al. created and investigated distributional dendrograms, and Cartwright and Brent looked at distributional templates. One major contribution of the present study not utilized in past studies is the direct evaluation of parent language using language produced by that particular parent's child. Additionally, the present study examines a much larger amount of language in many cases than these past studies. For example, Cartwright and Brent's templates examined one sentence at a time instead of analyzing larger chunks of an entire corpus. When compared to Mintz's work, the present study was able to also attain high levels of accuracy, but much higher levels of completion, as frequent frames were found to be unsuccessful in accounting for large parts the corpora studied. In general, the present study used a very different approach from these previous studies and was still able to attain high levels of accuracy. The use of parent utterances for training and child utterances from the same corpus for testing was also possibly a more representative model of child language learning than had been used in these studies.

One limitation of the present study is allowing the evolving dictionaries to have only one possible category per word. Allowing multiple categories per word has been shown previously (Cluff, 2015) to be effective in attaining higher levels of accuracy, but it is unclear if holding the average number of allowed GWCs per word to the corpus actual average as done by Cluff represents adequate control as an evaluating measure. A second limitation was that the current study only used language data from three corpora; a larger number of participants would allow more confidence in the findings. Third, the effect of various possible settings in the

computational model is still unexplored. This would include setting such as the mutation rate, the chunk size, and the dictionary size that would lead to optimal levels of operation of the model. Finally, the samples used in the present study represented only about 100 hours of adult speech toward children, which is a relatively small amount of language when considering the amount to which children are exposed over the course of their development. Using corpora containing larger amounts of child-directed speech might lead to higher accuracy levels.

Nevertheless, the results of the present study suggest that further research in the area of adaptation and selection algorithms applied to the question of GWC acquisition are warranted. This study presents a promising approach to the question of grammatical acquisition using a new parent-to-child adaptation and selection model. The findings suggest that the development and testing of adaptation and selection models of language acquisition might offer useful insight into the development and impairment of language.

## 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.
- Bloom, L., Hood, L., & Lightbown, P. (1974). Imitation in language development: If, when and why. *Cognitive Psychology*, 6, 380-420.
- Bloom, L., Lightbown, P., & Hood, L. (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development*, 40 (Serial No. 160).
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Cartwright, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition*, 63, 120-170.
- Channell, R. W. (2015). ns\_ev. [Computer software]. Provo, UT: Brigham Young University.
- Channell, R., Nissen, S., & Tanner, K. (2014, November). *Modeling children's acquisition of grammatical word categories using an adaptation and selection algorithm*. Paper presented at the meeting of the American Speech-Language Hearing Association, Orlando, FL.
- Chatterton, M. (2015). *External validity of grammatical word category classification using an adaptation and selection model* (Unpublished master's thesis). Brigham Young University, Provo, UT.
- Cluff, S. Z. (2014). *A model of grammatical category acquisition using adaptation and selection*. (Master's thesis, Brigham Young University). Retrieved from <http://scholarsarchive.byu.edu/etd/4086>

- Freudenthal, D., Pine, J., & Gobet, F. (2005). On the resolution of ambiguities in the extraction of syntactic categories through chunking. *Cognitive Systems Research*, 6, 17-25.  
doi:10.1016/j.cogsys.2004.09.003
- Judd, C. L. (2014). *A model of grammatical category acquisition the Spanish language using adaptation and selection*. (Master's thesis, Brigham Young University). Retrieved from <http://scholarsarchive.byu.edu/etd/4196>.
- 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
- Morgan, J. L., & Demuth, K. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Erlbaum.
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- 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.
- Sachs, J. (1983). Talking about the there and then: The emergence of displaced reference in parent-child discourse. In K. E. Nelson (Ed.), *Children's language*, Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Retrieved from <http://scholarsarchive.byu.edu/etd/4148>

### Annotated Bibliography

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

The introduction of this book summarizes and defines the major theoretical approaches to child language acquisition. These categories include the more innate-centered approaches of nativist, generativist, and universal grammar, and the more learned-centered approaches of constructivist, emergentist, socio-pragmatic, functionalist, and usage-based. Chapter 6 of this work discusses how children learn the syntax of word-order languages. This chapter looks at the research and problems with generativist and constructivist theories such as semantic bootstrapping, prosodic bootstrapping, distributional analysis, frequent frames, semantic assimilation, distributional learning, and learning syntactical categories with the help of phonological cues. Ambridge and Lieven pointed out the limitations in these various approaches. The limitations that they point out show that there are still many challenges in understanding how children acquire language, which are challenges that this study strives to continue exploring.

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.

This study expanded the work of Mintz (2003) to samples in the German language. They reference the work of Erkelens (2009), which showed low accuracy results when analyzing the Dutch language, and wanted to see if similar results would be yielded in German. The researchers analyzed child directed speech collected during 58 one-hour recordings directed towards a boy they call "Leo" while he was between the ages of 2;0 and 2;2. This is the largest sample of child directed speech that exists in German. For each frame found, all of the words intervening were stored together in a group and the 45 most frequent frames were selected for further analysis (modeled after Mintz's work). In order to measure accuracy, a "hit" was recorded when two words from the same grammatical category ended up in the same frame, and a "miss" was recorded when two words from the same category ended up in different frames. In order to measure completeness, a "hit" was recorded when two words from the same category ended up in the same frame and a "miss" was recorded when two words from the same category ended up in different frames. The accuracy was found to be higher when using the frequent frames than was measured at random, but was still relatively low, showing that the frames do provide some information about categories, but there is still considerable variability within frames. These results were similar to those found for the Dutch language. It is thought that these results are due to German having less strict syntax than English and French. This study shows that the question of grammatical category acquisition has yet to be fully answered.

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

Cartwright and Brent show that though many computational models have been used in past studies, there has been one (Kiss, 1973) that has used a computational model to suggest a theory of syntactic category acquisition. Kiss used a hierarchical cluster analysis to group

together words with similar distributional patterns. The differences of Cartwright and Brent's study from Kiss's work include that their system is more efficient in that it produces a set of discrete categories of words. It also differs in that it is less time consuming because it works on one sentence at a time, forgetting the previous sentences, while Kiss's hierarchical cluster analysis processes the all of the input at once. The author's proposed that children create templates (or generalized minimal pairs) of the language they hear based on certain rules of preference. They merge overlapping templates together until they have reached optimal templates. They conducted five experiments to test this proposal. The first experiment tested if this strategy was effective in learning an artificial language, and the second added ambiguous words to the input. The other experiments used child-directed speech in English, then added longer input files, and lastly tested the effect of adding semantic information to the system. Accuracy and completeness scores showed that distributional input was effective for merging groups, and these scores were higher when semantic information was added to the distributional information. This work shows that using computational models to study theories of syntactic category acquisition is a valid approach, and suggests further hypotheses for how children learn language.

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

This proposal discusses successful results with a previous version of the computational model that uses evolutionary programming. The idea of children's use and understanding of grammatical categories and previous theories of possible methods of acquisition are discussed. The paper then presents an entirely different approach to the question through the use of evolutionary programming. It mentions that this type of computing uses principles of evolution such as variation, adaptive change, self-regulation, and inheritance. This type of model uses populations of solutions and randomly mutates to create offspring, which compete to become the parents of future generations. Over many cycles, the program creates an appropriate solution.

The adult utterances from the six corpora used in Mintz's 2003 study were used for this study as well. Three thousand total evolutionary cycles were run. During each cycle, 100 dictionaries compete. Each dictionary contained all of the words from the corpus randomly assigned to a grammatical category. From the corpora, odd numbered utterances were examined to see if the dictionary contained the correct tag for each word, and these levels of accuracy are used for the fitness criteria to determine which dictionary would "reproduce". The even numbered utterances in the input file were used to determine the generalization accuracy of the best dictionary. The results indicated that the average level of accuracy for all six corpora was only 2.27% less than the 92.92% average maximum accuracy attainable with one grammatical category tag per word.

These results indicate that the model is a promising approach to categorizing words into grammatical categories, and a promising previous step to the present study.

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

Cluff studied a computational model that used an adaptation selection theory in order to address the question of grammatical category acquisition. The same five corpora from the CHILDES database that were used in Mintz's (2003) study were used for this study as well. The program initially created 100 dictionaries containing all of the words from the corpus randomly assigned to a tag that represented a grammatical category. At the end of each generation, the program compared these random tags to the correct tags that were assigned by tagging software, and those dictionaries that received the best scores (the greatest number of correct grammatical categories assigned), was able to "reproduce", or create 100 other dictionaries with a certain chance of mutation. This study by Cluff examined the effect of allowing multiple tags per word, and the effect of different mutation rates on accuracy levels. The most successful mutation rates were found to be those that were lower (1/800, 1/1200, 1/1800), with the least successful being 1/400. Also, allowing multiple tags per word resulted in better accuracy. The findings in this study indicate that this model warrants further study. The present study adds the process of using the child-directed speech of a particular child to that child's speech.

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

This chapter discusses how Darwin's definition of natural selection can be seen as an algorithm. Dennett argues that this idea of natural selection was an algorithm based on the definition of an algorithm. The definition provided for an algorithm is that it is a process that consistently yields a certain type of result when it is run. He says that an algorithm must consist of substrate neutrality, underlying mindlessness, and guaranteed results. Substrate neutrality refers to the idea that the process itself is made of a logical structure, and the outcome is not dependent upon the properties of the materials involved. Underlying mindlessness refers to the fact that each step in the process is relatively simple, and the concept of guaranteed results indicates that an algorithm always accomplishes its goal. Dennett suggests that Darwin's theory matches this definition of an algorithm. In the present study, Darwin's idea of natural theory is used as an algorithm within a computational model and applied to the question of how children learn the use of grammatical categories.

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

The introduction to this book describes that Blondie24 is a computer program that beats humans at the game of checkers. It is based on evolutionary programming, which models human cognition or intelligence through its ability to find the best solution to a problem (or the highest point on a function). It does this using a decision-making system that chooses one stimulus over another with a certain defined goal that provides criterion for success. Fogel compares this search to trying to find the highest peak in the Rocky Mountains while hiking in dense fog so that the hiker can only see what is around him or her. Just like the Rocky Mountains, functions can have multiple peaks, many much shorter than others. In this way, once a peak is found, the

hiker does not know if the highest peak has been found, or simply a shorter localized peak. Evolutionary processes help avoid this problem through the process of random variation in which most offspring (in the case of the present study, dictionaries) are very similar to their parents, but a certain amount experience mutation, giving their significantly different compositions from their parents. These mutations help to explore alternative areas on a function (or in the Rocky Mountains). Also, survival of the fittest helps to speed up the process. In the mountain climbing analogy, this concept would be represented by using an army of mountain climbers, each carrying a walkie-talkie to be in constant communication with all others. As soon as one mountain climber found a slope towards a higher peak than the others, all other mountain climbers would be pulled into this area (in the present study, more poorly performing dictionaries will “die off” or not have the chance to reproduce).

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

The authors of this study argue that previous studies on syntactic category acquisition had been inherently flawed. They point out that it may be that the co-occurrence statistics (or the quantification of the likelihood that one word is surrounded by other words) are more accurate over longer units. They also state that the only way to know if a set of generated grammatically categories are accurate is to have a model that generates sentences. They suggest that chunking should be used to measure the form of utterances. They use a computational model called MOSAIC to demonstrate syntactic category acquisition based on these hypotheses. This model incorporates a chunking mechanism in which frequent phrases are treated as one unit, which would result in chunked single words no longer being substituted except when substituted as part of a chunk. They found that this method of chunking did decrease the number of syntactic errors by MOSAIC. The sentences formed from the generated categories were an average of 3 words long. This study pertains to the current study because it shows that the use of computational models has been helpful in learning about syntactic category acquisition.

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

For this experiment, children ages four through seven were asked to provide the correct morphological change to nonsense words. Based on the grammatical category of the word presented, they were asked to apply the correctly for of plural, verb tenses, possessive, derivations, and compounds to these words. The results of this study showed that children from these ages have an understanding of basic morphological rules, and how to apply them based on grammatical categories. This study showed that children do not solely memorize the language that is presented to them; they are able to also categorize and use grammatical categories.

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

Hills looked at the difference between child-directed speech and adult-directed speech and investigated the differences between these two that could lead to improved language acquisition. This study examined multiple aspects of language input, such as associative structure

(the chance of a word occurring with its free associates), contextual diversity, word repetitions, and frequency. This article summarizes the word from longitudinal studies of six language corpora. Four of these were from child-directed input directed toward children ages 1;0 to 5;0 and two of the corpora included adult-directed input. This study found that the child-directed language was more associative, included more repetition, and greater frequency of individual word use, and relied more on usage consistency. This was found to be especially true for words learned by age 1;0. Additionally, the properties more frequently found in child-directed speech better predicted word acquisition than those in adult-directed speech. These findings indicate that child directed speech is structured in ways such that it facilitates children's language acquisition. These findings also support the current study's use of corpora containing child-directed speech.

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

This study examined the question of whether or not children can detect the sound patterns of familiar words in the context of a sentence. The authors built this study based on past research, which had shown that infants are attuned to the sound structure and prosody of their native language halfway through their first year. The question was investigated based on head turning patterns amidst sentences containing words to which the infants had previously attuned, indicating familiarity. The study was split into two experiments. The first examined infants age 7 ½ months and the second examined infants age 6 months. The findings from this study indicated that 7 ½ month old infants do have the capacity to identify sound patterns of familiar words in the context of a sentence. This ability is an important foundation to learning grammatical categories, as assigning a word to a grammatical category must be preceded by recognizing the word within the context of a sentence.

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

This study was one of the first to use a computational model to study the acquisition of grammatical categories. This model examined distributional information using neurologically plausible mechanism in order to classify words. This study used only a relatively small sample of utterances, but its output contained grammatically appropriate classes, which exhibited certain aspects characteristic of word class systems of children. Through this study, we see that computational models have had successes in helping us answer the question of how children learning grammatical categories.

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

MacNamara discusses the nature of language learning, specifically the acquisition of semantics, syntax, and phonology. He suggests that when learning language and analyzing input, children first comprehend meaning before applying syntax and phonology. He suggests that children first have an idea before applying syntax to communicate this idea. For example, in the sentence *The boy struck the girl* vs. *The girl struck the boy*, the child must first understand what is happening before applying syntax to the sentences. He also discusses the fact that grammatical

forms that are connected with meaning are developed before syntactic forms that have less to do with meaning. In his conclusions, he states that, though much of what he discussed suggested a sequential nature in language learning, he acknowledges that acquiring each aspect of language is somewhat interdependent on the other aspects.

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

In this study, Mintz looked at distributional analysis and its role in learning the grammatical categories of an artificial language. Forty undergraduate students were the participants for the study. These participants listened to sentences from this new artificially constructed language for six minutes. After this listening segment, sentences were played for them and they were asked to identify whether or not they heard that sentence and also to rate how confident they were in their answer. The listening segment contained frequent frames with multiple words intervening. The idea of the study was that if the participants used distributional information to medial words into a category, they would be more likely to incorrectly recognize words containing the same frame with a different intervening word. The results of this study indicated that participants did use distributional analysis through the lexical co-occurrence properties of medial words in certain frames. These medial words did lead to the formation of a category that was recognized. The short amount of time to which the participants were exposed to the language (6 minutes), indicates that distributional analysis provides information that the brain can use readily. Some limitations of the study included the simplicity of the training sentences, representing a language much less complex than real languages, and the fact that adults were studied instead of children. This study provides further information by way of grammatical category acquisition, but is further evidence that more research is needed in this area.

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

In this article, Mintz suggests and tests a way that children may learn grammatical categories through distributional patterns that they hear in speech. The distributional patterns that he addresses are called frequent frames. Frequent frames exist where there is a pair of words that occur together frequently with one word in between them. The words that occur between this pair of words would be categorized together into the same grammatical category. Mintz performed two experiments in order to study these frequent frames. In the first experiment, 6 corpora from the CHILDES database were analyzed. The only sections of each corpus that were analyzed were those in which the child was 2;6 or younger. In this first experiment, Mintz only analyzed the 45 most frequently occurring frames. This experiment yielded very well on measures of accuracy compared to the baseline data. Completeness scores were also significantly higher than those for the baseline, but were still quite low. A limitation of this study was the fact that it only analyzed the 45 most frequently occurring frames. Mintz therefore performed a second experiment that analyzed the frequency of frames relative to the total number of frames per corpus. Another limitation of this study was that it only labeled each word with one grammatical category. This resulted in lower scores of completion. Though this study did have

limitations, it showed significant evidence that children could learn grammatical categories from distributional information.

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

Some information contained in this study that is pertinent to the present study is its description of syntactic bootstrapping. It explains that the words that surround a verb provide information about the verb's meaning (there exist correlations between verb syntax and semantics). The example provided by the study was using the word "see". This verb can appear in front of directional prepositions, as in the sentence *I can see all the way to the mountains*. This syntactic context provides the listener with the information that this verb involves motion. This word can also be followed by sentence complements as in the sentence *I don't see what you're getting at*. This syntactic context provides information that "see" can also have a more cognitive meaning. This study also explains that syntactic bootstrapping hypothesizes that the more syntactic frames in which a child hears a verb, the easier the verb will be to learn because each frame provides additional information about meaning.

The purpose of this study was to investigate the effects of a mother's input on a child's order of acquisition of verbs. Speech from 57 mothers directed to their Stage I children (in terms of Roger Brown's stages) was examined. These children were just beginning to combine words at the time of the researchers first visit. Then 10 weeks later, speech from these children was recorded for investigation. These interactions were examined for the production of 25 verbs, which were chosen because they vary in both semantic and syntactic properties.

The first major finding of this study was that the more frequently a verb appeared in input, the more often and in more diverse contexts it was used later by the child. Also, the more often a verb appeared in utterance-final position, the fewer different syntactical frames it appeared in the child speech corpus. The authors predicted that this is due to the lack of a complete frame from which to draw syntactic information. They also found that the diversity of syntactic environments in which verbs appeared in input was a significant and position predictor of the frequency and diversity of use in the child corpus. This finding supports the multiple frames tenant of syntactic bootstrapping. This study showed that language input affects child language acquisition, and also that children do gain information from the syntactic information provided by child-directed speech.

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

This study investigates a distributional approach to learning grammatical categories. These authors first present previous linguistic and neurological research. Neural research shows that language within the same linguistic category is topographically similar. This study analyzed language samples from the CHILDES database. From this database, words were grouped together based on their distribution within utterances. These words were then organized into dendrograms, which grouped words with similar syntactic distributions more tightly. These groups were tested in a series of experiments based on their accuracy and completeness. These

groups were found to be strongly related to grammatical categories, thus this study presents strong evidence for distributional approaches to syntactic category acquisition.

Rauh, G. (2010). *Syntactic categories: Their identification and description in linguistic theories*. New York, NY: Oxford University Press.

This book discusses the nature and use of parts of speech and syntactic categories. Rauh points out that humans use categorization as a way to simplify the massive amount of sensory information with which they are constantly presented. Categorization is used in science in particular, and for this to be effective, each scientific category must have clearly defined boundaries and purpose. The boundaries for particular parts of speech, however, are not as clearly defined as often thought. Though parts of speech and syntactic categories carry some similarities, they are not identical. Syntactic categories occur when two words can be found in the same location within a sentence, or when two words can be replaced for each other in the same position within a sentence. Exactly how syntactic categories are mapped and defined is different based on linguistic perspective. This book outlines how each theoretical perspective defines the syntactic categories. For the purposes of the present study, it is assumed that syntactic categories refer to categories of words and also that words within the same category are mutually substitutable. This work also gives strong evidence that the use of these categories is valid.

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

In order to learn a language, children must distinguish between linguistic elements in the speech they hear. There are multiple cues, which help children achieve this task. There are cues that are inherent in the language as well as cues that caregivers use when talking to children. An example of a cue inherent in language is frequently occurring grammatical morphemes such as the determiner “the” being placed before noun phrases. An example of a caregiver cue would be the use of shorter sentences. Many studies have shown that infants are sensitive to these cues. This study was made up of two experiments. For both of these experiments, 60 children between the ages of 2;0 and 2;2 participated. A robot instructed children about nouns and then told the children to point to the noun on a page. In the first experiment, the role of prosody and morphological cues was examined, and an interaction between these two was investigated. In order to examine prosody, pauses were inserted at natural and unnatural boundaries and in order to examine grammaticality, an auxiliary that was either grammatical, not grammatical, or a nonsense syllable were used. This experiment found that children performed better on sentences with natural prosody, and a lack of interaction between prosody and grammaticality was found. These results indicate that children use both of these types of cues during sentence comprehension. In the second experiment, the relation between grammatical cues provided by frequently occurring grammatical morphemes and caregiver cues (short utterance length and placement of key were in utterance final position) was examined. The findings from this experiment indicated that children use both of these types of cues and that the presence of one type of cue does not appear to decrease the role of another (the presence of short utterance or key words at the end of an utterance did not decrease the child’s use of frequently occurring grammatical morphemes). This study shows that children pay attention to many types of cues

during language acquisition, but does not completely answer the question of how children learn language.

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

St. Clair, Monaghan, and Christiansen (2010) compared frequent frames to flexible frames, using a computational model to perform the study. Multiple experiments showed the benefits of combining input information from bigrams (e.g., aX, Xb) and trigrams (e.g., aXb) into flexible frames (e.g., aX +Xb), which overcomes the weaknesses of strictly analyzing bigrams or trigrams independently. Accuracy increased, and a larger amount of the amount of the language sample was analyzed, suggesting that allowing a less rigid distributional form may provide more information during input for children learning language and acquiring grammar.

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

This study examined the use of frequent frames developed by Mintz to predict syntactic categories. This study states that these frequent frames seem to yield accurate results in English, as well as in French (Chemla, Mintz, Vernal, & Christophe, 2009). A previous study by Erkelens in 2009 found that this approach was not useful in grammatical category acquisition in Dutch. The study done by Stumper et al. also shows a lack of usefulness in categorization of grammatical categories using frequent frames when applied to the German language. Additionally, they found that there was a negative correlation between the accuracy of frames and the lexical diversity of the middle position of the frame. This finding is likely due to the less strict syntactical structure of German when compared to English and French. This study suggests that children likely employ more than just this method of frequent frames when categorizing grammatical categories.

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

This chapter reviews empirical findings regarding how structure-dependent regularities (such as language) as represented in the human brain at a behavioral, cognitive, and neural level. They discuss that language uses a finite number of symbols or signs and a finite number of rules in order to produce infinite expressions. This is done by generating and interpreting hierarchical structures. The higher levels of this hierarchy include sentences, lower level constituents as phrases and lower level entities as word classes such as determiners, adjectives, nouns etc. These pieces create a structure-dependent recursive process that can be reiterated infinitely. An important piece upon which this structure is based is on what the author calls lexical categories. These categories include information about the meaning, phonology, and syntactic properties of a word. The human brain stores words into these categories in order to use them within the different levels of the language system (phonological, syntactic, morphosyntactic, and semantic).

The authors suggest that a possible explanation for why there are still many questions about language acquisition is that the levels of language organization are only viewed at a surface level, when these relations are actually established on a deeper level. Children are not instructed regarding the properties of human language they must learn to communicate, but most of them develop the capacity to use and understand language within the first few years of life. The authors suggest that some aspects of human language developed based on the human brain's affinity to learn these properties. This work also discusses that ERPs and FMRI's have been used to show that infants have the ability to differentiate phonemic and prosodic contours, an ability that likely helps them to segment speech into individual words, which children can do at no later than 7 months of age. In order to develop both expressive and receptive language, children must learn to assign words to grammatical categories, and use these categories to generate phrases and clauses (and sentences). The authors state that the way children do this is still largely unknown. They cite Cartwright and Brent (1997) in stating that grammatical categories may be assigned to words using co-occurrence regularities, and that phonological, prosodic, and morphological cues in addition to the distribution of function and content words all may contribute to a way to accomplish this categorization. These authors also mention studies that have found importance in the left inferior frontal cortex of the brain and 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. doi:10.1017/S0305000909990067

This study expands and replicates Mintz's study by looking at frequent frames in both Spanish and English. It also looks at not only the types of frames that Mintz studies, but also end-frames. An end-frame occurs at the end of the sentence and includes the end of the sentence as a boundary for the frame. The results of this study showed that accuracy scores were statistically significantly higher compared to baseline for both English and Spanish, as well as for both mid-frames and end-frames. When comparing accuracy levels across different groups, they found that accuracy was higher for English than it was for Spanish due to homophony among function words and noun-drop that exist in the Spanish language. Accuracy was also higher for mid-frames than for end-frames. The accuracy for verb-frames was found to be higher than it was for noun-frames, and higher for noun-frames than for adjective frames. This means that the clarity of distributional information that is available in frequent frames is different for different languages, and that within languages it is different for different grammatical categories. Overall, this study supports the idea that distributional information such as frequent frames provides information about grammatical categories to children across two different languages.