ACQUISITION OF FUNCTIONAL CATEGORIES

by

Hao Wang

A Dissertation Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(PSYCHOLOGY)

August 2012

Copyright 2012 Hao Wang
Dedication

This dissertation is dedicated to my beloved parents Jiayang Wang (汪嘉阳) and Yiling Liu (刘伊玲) who have always stood by me and supported me with endless love, patience and encouragement.
“Thy firmness makes my circle just,
And makes me end where I begun.”

For Jie Yang and the memories
Acknowledgements

This dissertation would not have been possible without the immense support and mentoring of my advisor Toby Mintz, who has been extremely supportive during my graduate school years at USC, both scholarly and financially. I am still grateful that he decided to take me, an electrical engineering graduate, as his student six years ago so I was able to pursue my interests in cognitive science. I found that his mentoring style suits perfectly for me. I enjoyed the freedom of exploring and working on topics that I was mostly interested in.

I am truly indebted to the members of my qualifying and dissertation committees. Frank Manis, who served as a reader of my research papers since my first year until the qualifying exam, has provided invaluable feedback on my research problems. Bosco Tjan, who has also been a long time reader for me, is greatly appreciated for his insights and comments to the computational aspect of this work. Elsi Kaiser, whose psycholinguistic classes and lab meetings were an important part of my academic life at USC and whose comments and discussions on this research project were especially helpful, deserves my sincerest gratitude. Justin Wood, who I was honored to have on my committee, was constantly pushing me to think my research in larger contexts and taught a very interesting and inspiring class on the origins of human nature.

I would like thank Eduard Hovy from the Information Sciences Institute, who provided me not only financial support for a few years but also opportunities to pursue
my interests in natural language processing. It was great pleasure to work with him and colleagues at ISI.

I owe my deepest gratitude to Chunyan Ning and Thomas HT Lee, from whom I have learned so much about linguistics, language acquisition as well as the fun of doing research when I was a programmer at the Institute of Cognitive Science of Hunan University.

The idea of studying acquisition of functional categories first came to me after going to a fascinating talk by Virginia Valian at the BUCLD conference in 2009. I would like to thank her for introducing me to this research problem and for later discussions in Paris and through emails.

Special thanks must go to researchers who have collected all the child language data and selflessly made it public available as well as Brian MacWhinney who has created and maintained the excellent CHILDES database.

Last but not least, I was fortunate enough to have many great fellow students and friends who have brought so much fun to my life in graduate school, including Zhisen Jiang, Jia Li, Susan Geffen, Yan Zhou, Jinxia Ma, Siyi Wang, Congxing Cai, Lin Ding and Joy Xing.
# Table of Contents

Dedication ii  
Acknowledgements iv  
List of Tables vii  
List of Figures viii  
List of Equations ix  
Abstract x  

Chapter 1. Introduction 1  
Chapter 2. Early development of function words 10  
Part 1. Knowledge of abstract functional categories 35  
Chapter 3. Small sample size problem 38  
Chapter 4. Measuring knowledge of determiner category with expected overlap 45  
Chapter 5. Analysis of German determiner category 53  
Part 2. Categorizing function words 62  
Chapter 6. Categorizing function words with immediate contexts 65  
Chapter 7. Distributional analysis with frequent frames 75  
Chapter 8. General discussion 84  
References 92
List of Tables

Table 1. Examples of Adam and Eve’s immediate imitative production of parents’ utterances 19
Table 2. Examples of children’s spontaneous production around 20-month old 22
Table 3. Some questions and answers between caregivers and children around 20-month old 23
Table 4. Creating increasingly larger samples from a corpus of 20 recording sessions by cumulatively adding sessions in backward (mother’s utterances only) 39
Table 5. Statistics of the eight corpora 40
Table 6. Correlation between overlap and number of utterances (mother) 42
Table 7. Numbers of utterances for samples grouped by month 47
Table 8. German definite determiner system 54
Table 9. German indefinite determiner system 54
Table 10. Statistics for the Leo corpus (German) 56
Table 11. Target function words used in this analysis 66
Table 12. Accuracy and completeness of categorization for 8, 12 and 16 clusters (mother’s speech) 71
Table 13. Purity scores for 8, 12 and 16 clusters (mother’s speech) 72
Table 14. Accuracy and completeness of categorizing function words using frequent frames (mother’s speech) 78
Table 15. Frequent frames of function words from Thomas mother’s speech 80
Table 16. Some highly accurate frequent frames for function words with substantial size (Thomas mother) 82
Table 17. Frequent frames of function words that categorize a or the (Thomas mother) 83
List of Figures

Figure 1. Determiner as the head of Determiner phrase (DP) 12
Figure 2. Order of acquisition of 14 grammatical morphemes in three children 25
Figure 3. Overlap for samples with increasingly more utterances from the mother 41
Figure 4. Overlap for samples with increasingly more utterances (Thomas mother) 43
Figure 5. Actual and expected overlap of Lily and mother 50
Figure 6. Actual and expected overlap of Naima and mother 50
Figure 7. Actual and expected overlap of Nina and mother 51
Figure 8. Actual and expected overlap of Thomas and mother 51
Figure 9. Deviation for all data points in the eight corpora analyzed (sorted by children’s deviation) 52
Figure 10. Examples of German determiners 55
Figure 11. Actual and expected overlap of Leo and mother 58
Figure 12. Deviation of actual overlap from expected overlap for Leo and mother 60
Figure 13. Preceding and following 200 words as context 67
Figure 14. Clusters result from different cut-off level 68
Figure 15. Categorization of 82 function words into 8 clusters using Thomas mother's speech 74
List of Equations

Equation 1. Overlap between determiners a and the 36
Equation 2. Overlap between all determiners 36
Equation 3. Probability of overlap for a noun occurring f times in a sample 46
Equation 4. Estimate the probability of a noun occurring with the 46
Equation 5. Expected overlap of a sample 48
Equation 6. Deviation of a sample’s actual overlap from expected overlap 48
Equation 7. Euclidean distance of two target words 67
Equation 8. Computation of accuracy 69
Equation 9. Computation of completeness 69
Equation 10. Purity for category X 70
Abstract

This dissertation investigated the knowledge and learning mechanism of functional categories (e.g., determiner and auxiliary) in young children. The first part of this dissertation demonstrated the importance of sample size when measuring linguistic productivity using speech samples, which may have contributed to discrepancies in previous studies of young children’s knowledge of the determiner category. It then proposed a novel probabilistic method for measuring linguistic productivity that takes sample size into account. With this new method, analyses of eight English and one German longitudinal corpora of children’s naturalistic speech production showed that young children between 1;11 and 3;0 were using determiners in the same way as adults do. The results are consistent with previous findings that even 14-month-olds are processing determiners as an abstract category.

The second part of this dissertation tested the hypothesis that distributional information in the input (e.g., bigrams) is informative for categorizing function words into finer categories. The analyses demonstrated that when acoustic and phonological cues are used to identify function words and their contexts from the input, linear distributional patterns are extremely accurate in categorizing function words into substantial-sized groups of words from the same grammatical category.

In summary, this dissertation contributes to several important questions in language acquisition, including the origin of knowledge of grammatical categories, the developmental course and learning mechanism of grammatical categories. Moreover, the
demonstration of the sample size problem is illuminating for future research using corpus analysis. The novel probabilistic method proposed here for measuring young children’s linguistic productivity could also be adapted to study other aspects of linguistic structures.
Human language is a complicated system and an attribute that is specific to the human species. No animals or machines have been shown to possess such a system, although they have access to the same amount of or more input and environmental information than a human child does. A pet lives in the same house of a child or a computer program with access to all the books ever produced by human does not acquire a human language. Many cognitive foundations are shared between human and other animals (such as vision and navigation skills). Nevertheless, no animal communication system demonstrated complexities found in human language, such as recursive symbolic structures. There must be something special in the human biology that enables language learning.

For machines, excessive computation power and memories are available. They also have access to huge language data sets that are definitely richer than the input of a child. Machines were able to out-perform human experts in many cognitively demanding tasks, such as chess playing and Jeopardy!, a question-answering game. Yet no machine has demonstrated comparable performance in learning and understanding of a human language.

Even for human, learning a new language later in life is not always as easy as acquiring one’s first language. One often do not achieve the same degree of language efficiency when it is learned after puberty (Newport, 1990). All of these make first language acquisition a great achievement of a child. A crucial question is what makes it
possible. It seems that there are four aspects that all contribute to language acquisition: a) input; b) time of exposure; c) learning mechanism (because languages are different, something has to be learned); d) innate knowledge (which could be a general learning mechanism or domain-specific knowledge).

Language acquisition becomes crucial when the study of language changes from a comprehensive description of linguistic structures (i.e., descriptive adequacy) to pursuing explanatory adequacy. Therefore, beyond a description of grammars, it is important to understand the initial state, growth function, input and environment, and how the interaction between them finally leads to knowledge of language.

Chomsky formulated language acquisition as a process that some biological language acquisition device (LAD) takes language input and develops into adult grammar (Chomsky, 1965). In Bayesian approaches of language acquisition, a probabilistic model that comes with certain priors encounters relevant evidence, updates its belief, and reaches a system of knowledge represented by certain posterior probabilities (Chater & Manning, 2006; Hsu et al., 2011). Although there are huge differences between the Generative and Bayesian approaches, both of them have certain knowledge or information that is pre-encoded (LAD or priors) and both take some input to reach the competence of adult grammar. It seems that both innate mental capacity and input are critical to the successful acquisition of a language.

Regarding input, an interesting question is what is in the input that supports language learning. It follows with a question that what kind of learning mechanism is needed for a child to exploit such information.
Regarding LAD, the question is what is innate that is necessary for achieving the final state. Currently there is no tool to probe directly the initial state. The classic Poverty of the Stimulus Argument was used by Chomsky and others to argue for certain innate knowledge of linguistic structures. The basic idea is that if something in the final state that is not directly available in the input and is not possible to be learned from the input it must be innate. Therefore, the tasks for language acquisition researchers were to figure out what are learned or learnable, when they are learned and if they can be learned in limited amount of time.

These are part of the so-called *Plato’s problem*, i.e., the origin of knowledge, in the language acquisition domain. A current debate is whether young children possess an abstract representation of functional categories (e.g., determiner, auxiliary and preposition) or whether the representation of functional categories is built gradually in an item-by-item fashion. Strong nativist views held that children are innately endowed with a set of grammatical categories including functional categories. They possess abstract knowledge of grammatical categories since the beginning and use that knowledge to learn their first language. The claim that young children are innately endowed with certain grammatical categories would require an explanation under evolutionary terms, i.e., how it was evolved to the current state.

On the contrary, constructivists claim that the early representation of grammatical categories is similar to memorizing chunks of phrases and lacks generalization between items from the same category. Therefore, according to constructivist views, young children do not have abstract knowledge of grammatical categories initially. It is the
burden of constructivists to explain how children transform the item-based representation to adult-like grammar.

Language cannot be studied as a whole. Aforementioned questions have to be answered for every aspect of language. Syntactic categories, including noun, verb, adjective, adverb, determiner, preposition, auxiliary etc., are the building blocks of grammar. Grammatical rules or phrase structures in linguistic theories are stated in terms of syntactic categories. What grammar manipulates are syntactic categories rather than individual words, which leads to potentially infinite grammatical sentences. In general, a word can be replaced by another word from the same syntactic category without violating the grammaticality of a sentence. The situation may be more complicated for functional categories (which is discussed later). Syntactic categories are interesting in terms of language acquisition because they must be learned before acquiring syntactic structures. Functional categories are a subject that is particularly interesting to study because they are different from lexical categories, carry less meaning and mostly serve important grammatical purposes.

**Problems in acquiring functional categories**

Because languages use different set of grammatical categories and lexicon is completely different between languages, for both nativist and constructivist views, there are still the problems of how children map words onto grammatical categories and what information in the input and environment is available for children to categorize words. In other words, for functional categories, the problems are how children assign learned word forms to the right categories and whether information in the input could support the task
of word categorization. This task is unavoidable even if children are predisposed with notions of functional categories. There must be a learning mechanism that works on the input to categorize words into proto-categories or adult-like categories, which are necessary for acquisition of syntax.

Studying the acquisition of syntactic categories and, in particular, the acquisition of functional categories has significant impacts on theories of language acquisition, linguistic theories and categorization in other cognitive domains (such as vision) as well. Syntactic categories – lexical and functional categories – are the building blocks of syntax. Some knowledge of these categories would be a prerequisite for acquiring syntax. Therefore, the time when a child possesses the knowledge of syntactic categories would be the earliest possible point in development for his/her knowledge of syntax (Valian, 1986).

Some core questions in child language acquisition are certainly relevant to the acquisition of functional categories: what the developmental course of functional categories is, what the typical age for acquiring each functional category is, whether functional categories are acquired gradually or instantaneously, whether all functional categories are acquired at the same time, what is the order of acquisition if not acquired concurrently, how large is individual variation among normal developing children and whether there is any cross-linguistic difference.

One of the much debated problems is whether children’s early syntactic representation resembles adults’ syntactic representation (i.e., the issue of continuity), which would also have consequences on linguistic theories. In the case of functional
categories, the question is whether functional categories in adults’ grammar are available to young children in early stage of development. Nativists argue that children are predisposed with notions of some abstract syntactic categories (such as noun, verb and determiner), though the contents of those categories are yet to be determined from the language experience (Valian, 1986; Valian et al., 2009). From nativists’ perspective, young children would possess the same repertoire of abstract syntactic categories as adult speakers. Constructivists argue that children build the abstract categories by learning limited-scope formulae and generalize them to adult like syntactic categories (Pine & Lieven, 1997; Pine & Martindale, 1996; Tomasello, 2000, 2003). Hence, children’s early knowledge of syntactic categories (if any) is qualitatively different from adult speakers.

Another problem concerns how children break into syntax given that syntactic categories are not directly labeled in the speech directed to children (i.e., the bootstrapping problem) (Pinker, 1984, 1987). In addition, syntactic categories are defined in terms of their distributions in sentences. The definitions of syntactic categories could become circular (e.g., nouns are defined according to their relative position to determiners etc.; defining the category determiner, in turn, would require knowledge of the noun category). There have been different proposals suggesting that distributional cues in the input (Brent & Cartwright, 1997; Maratsos & Chalkley, 1980; Mintz, 2003; Redington et al., 1998), phonological cues (Cassidy & Kelly, 1991; Kelly, 1992; Monaghan et al., 2007; Shi et al., 1998), and the relationship between semantic properties and certain lexical categories (Braine, 1976; Grimshaw, 1981; Pinker, 1984; Schlesinger, 1971) are useful for initially bootstrapping to major lexical categories. Whether any of
these cues could be applied to the acquisition of functional categories (i.e., the mechanism for acquiring functional categories) is discussed in Part 2.

Comparing to lexical categorization, categorizing function words and morphemes probably is a more challenge task for children because there may be more individual variation among members of the same functional category. For example, determiners in English should precede nouns like *the glasses*. A naïve learner would expect one could also say *a glasses*, which is of course ungrammatical because the determiner *a* can only precede singular count nouns.

The role of input in the acquisition of functional categories, such as the relationship between frequency, stress and syllable length of functional elements and the order of acquisition, is also worth a comprehensive examination.

**Hypotheses**

In this dissertation, the following hypotheses are tested to contribute to the current debate on abstract knowledge of functional categories in young children and to understand the potential mechanisms of the acquisition of functional categories.

1. Young children who just start to produce combinatorial speech (between two and three years old) possess knowledge of abstract functional categories.

2. Together with acoustic cues, distributional or statistical information in child-directed speech is informative for categorizing function words into separate functional categories.
Method

Past studies use a variety of methods to investigate the acquisition of functional categories, including naturalistic longitudinal observations, diary reports, behavioral experiments measuring both comprehension and production, and corpus analysis. Many of the studies looked at children’s spontaneous production data. This dissertation tested the above hypotheses with analyses of naturalistic longitudinal corpora of children’s and mothers’ speech.

Scope

The super-ordinate set of functional categories includes a number of similar yet distinctive categories. Each of them deserves detailed examinations. In addition, the acquisition of functional categories concerns a number of populations, such first language learners, second language learners, bilinguals and multi-linguals, sign language learners and children with language disorders. Moreover, a comprehensive study of the acquisition of functional categories would also require cross-linguistic evidence from typologically different languages. It is not possible to examine every aspect of the acquisition of functional categories in this dissertation.

I will limit my analyses to the acquisition of functional categories in normal developing children acquiring English or German. Since determiners are the most frequent words in English and many other languages, they are among the earliest acquired functional categories, and the determiner category among functional categories in English was extensively debated, this dissertation focuses on the acquisition of English and German determiner categories. In the second part of this dissertation, prepositions,
pronouns and auxiliaries were analyzed along with determiners. The development of other functional elements is not analyzed. However, the overall findings in this dissertation should be generalizable to the acquisition of other functional categories and should be applicable to many other languages.

**Contributions**

This work concerns with several important questions in language development, such as what the stages or milestones of the development of function words and morphemes are, when children start to show knowledge of abstract functional categories, and whether young children’s knowledge of abstract functional categories resembles that of adults. Some descriptive work has been done on the development of function words and morphemes (e.g., Brown, 1973). Recent development in corpora and methodologies provides a good opportunity to test conclusions from previous studies and to supply evidence for the current debate between nativism (under the UG framework and generative grammar) and constructivism on the nature of young children’s early representation of functional categories. Previous corpus studies (Pine & Lieven, 1997; Pine & Martindale, 1996; Valian, et al., 2009) suffered from the sample size problem as demonstrated in Chapter 3 of this dissertation. Hence, their conclusions may not be based on robust evidence. Chapter 4 proposed a new probabilistic method to solve this problem thus provides stronger evidence for the argument. This work also concerns with the effect of input. For example, input may have differential effect on the development.
Chapter 2. Early development of function words

This chapter first reviews evidence that supports the distinction between lexical and functional categories. Mean length of utterance (MLU), a widely used measure of early syntactic development is then briefly described. It follows with an overview of classic longitudinal studies on the development of functional elements and more recent experimental studies on function word use. Finally, the chapter presents state-of-the-art views on the nature of early representation and the acquisition of functional categories.

**Distinction between lexical and functional categories**

All the languages studied so far have a set of syntactic categories, which may be different from one language to another. The set of syntactic categories usually can be divided into two subsets of categories: lexical categories and functional categories.

Lexical categories contain open-class words from the vocabulary (such as nouns, verbs and adjectives) with each class has a moderate to large number of words. Lexical categories often have new words added and old words dropped out in a relatively short period. Functional categories contain closed-class words and morphemes (such as determiners, auxiliaries, complementizers, various kinds of particles and inflectional morphemes for number, tense, gender and case) with each class has a small number of words and/or morphemes. The items in functional categories often are relatively stable across time. Languages usually utilize a different subset of functional categories. They may be realized differently as words (e.g., the determiner category in English) or grammatical morphemes (e.g., the plural morpheme –s in English).
In traditional grammar, open- and closed-class words were also referred to as content and function words, respectively. As one can tell from their names, content words often carry meanings like objects, actions and properties, while function words usually carry less meaning but mostly serve as anchor points in sentence structures. Comparing to content words, function words in general possess some distinctive acoustic, phonological and distributional properties, such as shorter duration, fewer numbers of syllables, carrying weak or no stress, high frequency, occurring more often at phrase boundaries etc.

The distinction between lexical and functional items has a long history in the study of languages. In structural linguistics, researchers were making the distinction between content and function words. They were formalized as lexical and functional categories in the generative framework that became popular in the last fifty years. Since the dawn of generative linguistics, functional categories have played a critical role in theories of syntax as syntactic heads under the X-bar scheme (Abney, 1987; Chomsky, 1986; Cinque, 2006). For instance, determiner (D) was treated as the head of DP (Determiner Phrase) that includes a determinant and a noun phrase (see Figure 1 for an example). Furthermore, using a cross-linguistic comparative method, Cinque (2006) has proposed a universally fixed hierarchy of functional heads in the IP (Inflectional Phrase) domain, in which a language can adopt or manifest a part of the hierarchy of functional heads. Treating functional categories as heads of phrases could explain why function words often occur at phrase boundaries. At last, some theorists even suggest that functional heads are the only item subject to parametric variations (Fukui, 1988).
Regardless the theoretical analyses of linguistic structures, functional categories are crucial components of a grammar.

![Figure 1. Determiner as the head of Determiner phrase (DP)
Adapted from (Abney, 1987)](image)

Besides theoretical motivations of the distinction between lexical and functional categories, support of the distinction also came from language acquisition, language processing and neurobiological and aphasia studies.

Many different kinds of cues in child-directed speech have been shown to be able to distinguish function words from content words in a number of languages (Cutler, 1993; Gervain et al., 2008; Hochmann et al., 2010; Nespor & Vogel, 1986; Shi, 1995; Shi, et al., 1998; Shi et al., 1999). For example, Gervain (2008) demonstrated that 28-month-old Japanese and Italian infants were sensitive to the differential frequency distribution of content words and function words by testing the infants’ preference to word order in an artificial language. In another recent study, 17-month-old infants have been shown to be able to differentiate content words and function words solely based on word frequency (Hochmann, et al., 2010).

Shi (1995) investigated the question that whether lexical and functional items in English infant-directed speech can be divided into two groups according to their acoustic,
phonological and distributional properties like syllable duration, syllable structure, relative syllable amplitude, vowel quality and position in utterances. Since functional items in English tend to have short syllable duration, different syllable structure, relatively low amplitude, centralized vowels and occur more often in utterance-initial position, collectively these cues are very effective in classifying items into a lexical group and a functional group. However, the two groups have significant overlap on all the dimensions so none of the cues individually could predict group membership. Shi et al. (1998) asked the same question in another two typologically very different languages - Mandarin Chinese, a morphologically isolating language and Turkish, a morphologically agglutinative language. It is often assumed that across languages functional items are universally minimal comparing to lexical items in terms of production and perception, which may be guided by certain evolutionary motivations. For instance, functional items in many languages are realized as inflectional morphemes or clitics. The realization of minimality may differ in languages. For example, vowels in English functional items tend to be reduced or centralized; vowels in Turkish grammatical morphemes are harmonized (changed) to the same vowel in the stem that the suffix is attached to; functional items in Mandarin Chinese tend to have neutral tone instead of lexical tone. Similar to the results for English, Shi and colleagues found that the set of cues was sufficient for a Kohonen-style self-organizing network to classify infant-directed speech accurately into lexical and functional groups for both Mandarin Chinese and Turkish. Moreover, behavioral experiments with newborn infants using the high amplitude sucking procedure have
revealed their incredible ability in discriminating content words and function words based on perceptual cues (Shi, et al., 1999).

Regarding the developmental timeline, content words and function words (as well as lexical and functional categories) follow two different paths. The first words produced by children are usually nouns and verbs for most languages. Some functional items (e.g., the present progressive –ing, plural –s and prepositions in English) do not appear in production until 2 to 2;6 (although in languages like Turkish grammatical morphemes occur much earlier than other languages), while the production of first recognizable word is around age one. After a slow start, children’s vocabulary grows rapidly. They may reach a repertoire of 500-600 words by age of two, while their acquisition of functional words and morphemes follow a relatively fixed schedule and the rate of the inventory growth is in no way comparable to lexical words, although the functional items have much higher frequency than lexical items in the input (Brown, 1973; Clark, 1993).

Differences between content and function words were also observed in the neural substrates or processes underlying the processing of content and function words using functional Magnetic Resonance Imaging (fMRI) in normal adults (Diaz & McCarthy, 2009) or using Event-Related Potential (ERP) in normal adults, patients with Broca’s aphasia and congenitally deaf adults (Neville et al., 1992; ter Keurs et al., 1999). In an experiment attempted to elicit ERPs in normal adults when sentences were presented word by word, Neville and colleagues (1992) found that the two word types elicited qualitatively different ERPs that are compatible with the different roles of the two word types in language processing and that cannot be explained by the differences in word
frequency and word length. The findings were also supported by conducting the same experiment with congenitally deaf subjects who have mastered English vocabulary but not the grammar. They found the same ERP signature in both deaf and normal hearing subjects for open class words (semantic processing) but markedly different ERP signatures for closed class words (grammatical processing).

A series of studies on the processing of open- vs. closed- class words in Broca’s aphasia patients (particularly their difficulties in the processing of closed class words) have revealed significant insights into the differential processing/organization of the two types of words in the lexicon (Berndt & Caramazza, 1980; Biassou et al., 1997; Blount, 1969; Boyle & Gerken, 1997; Fenson et al., 1993; Friederici & Kilborn, 1989; Friederici et al., 1991; Fukui, 1986; Kelly, 1992; Pulvermüller, 1995; Rosenberg et al., 1985; Rydin, 1971). Although there are different views on what underlies the difficulty – the closed-class words themselves or deficits in syntactic processing, there is no doubt that the two word types are processed differentially and they play distinctive roles in language processing.

Summarizing all the evidence supporting the distinction between lexical and function words, it is justified to study the development of function words and the acquisition of functional categories separately from content words and lexical categories.

**MLU as a measure of syntactic development**

Since early twentieth century, mean length of utterance (MLU) has become a standard measure of children’s gross language development. Many studies that are discussed in this paper used MLU as an indicator of morphosyntactic development. It is
necessary to introduce the term here. Mean length of utterance is computed by dividing the total number of words in utterances that are longer than one word by the total number of utterances in a sample of the child’s speech. This was referred to as mean length of utterance in words (MLUw). After its initial introduction, a few number of variations of MLU have been proposed, such as MLU in syllables that measures utterance length by counting the number of syllables in an utterance and MLU in morphemes (MLUm) that measures utterance length by counting the number of morphemes in a utterance (Brown, 1973). MLUm is computed by dividing the total number of morphemes by the total number of utterances in a speech sample. MLUm, firstly used in Brown (1973), is now a primary measure of early morphological and syntactic development in the field. It has been shown to be well correlated with the early development of children’s morphological and syntactic knowledge in English (Brown, 1973; de Villiers & de Villiers, 1973). In their cross-sectional study of the acquisition of grammatical morphemes in 21 children, de Villiers and de Villiers (1973) found that “[MLUm] is a far better predictor … than is chronological age, which adds very little to the predictiveness of [MLUm] alone”. In Brown’s (1973) classical study of the development of 14 grammatical morphemes, the development was divided in to five stages according to the children’s MLUm with Stage I has an MLUm of 1.75, Stage II 2.25, Stage III 2.75, Stage IV 3.50, Stage V 4.00. Brown suggested that when MLUm is over 4.0 it is not an accurate measure of linguistic knowledge anymore. Each stage represents different achievements by children, some of which are discussed in the next sections.
Early development of function words

Since the middle of the 20th century, researchers have been systematically studying lexical development in young children using different paradigms, such as longitudinal observations of a small number of children and cross-sectional studies of a large group of children. For example, Brown (1973) summarized his observations and analyses of lexical development of three English-learning children in the first four years. As another example, Fenson et al. (1993) conducted a study with 1789 children to measure the rate of lexical development. The course of early lexical development from the first word to the second birthday has been well plotted out for both comprehension and production (Barrett, 1995; Benedict, 1979; Fenson, et al., 1993; Nelson, 1973), although some small variations exist between different studies. There is usually a several-month delay between the comprehension and production vocabularies. Children normally produce their first word between six to 12 months of age. They acquire an average of ten words by 13 months of age, 50 words by 17 months, and 310 words by 24 months (Fenson, et al., 1993). The vocabulary during early lexical development does not always grow at the same rate. The acquisition of the first few dozens of word is usually slow. After the children have acquired about 20-40 words, many of them experience a vocabulary spurt that the growth of vocabulary becomes much faster so they reach a vocabulary up to 500-600 words by the second birthday (Bloom, 1973; Dromi, 1987; Halliday, 1975; Nelson, 1973). Note that a huge variation was also observed for the time of the first word, the size of the vocabulary at different ages, the growth rate of vocabulary and whether there is a word spurt (Benedict, 1979; Fenson, et al., 1993;
Goldfield & Reznick, 1990; Nelson, 1973). For example, Goldfield and Reznick found that five of the 18 children they studied did not show a word spurt and instead, their vocabularies grow steadily.

The development of function words, including grammatical morphemes, has received special attention from researchers. Some phenomena have been consistently observed and are typical in the development of functional elements, such as initial omission of function words and grammatical morphemes by young children, similar order of acquisition across individuals and high correct rate when children start producing particular functional elements. These phenomena are crucial to the study of development of function words and to uncover how children acquire functional categories. Therefore, relevant studies were reviewed in detail in the following two sections.

**Initial omission of functional elements**

Young children’s early speech often lacks functional elements such as function words (e.g., determiners in English) and inflectional morphemes (e.g., past tense -d and plural -s morphemes in English). Early speech was characterized as ‘telegraphic speech’ (Brown & Fraser, 1963) since people usually only include content words that are important for conveying the meaning and leave out function words and punctuations in telegraphs (or text messages in the modern times) where there is a constraint on the number of words. The receiver is still able to decode and understand the highly condensed message. This phenomenon has been observed and documented in many studies of early language development (Brown, 1973; Brown & Bellugi, 1964; Gerken et al., 1990; Kemp et al., 2005; Radford, 1990).
Brown and Bellugi (1964) described three processes in the acquisition of syntax in two children - Adam and Eve. The process that is most relevant to the current discussion is the first process, imitation and reduction, in which children’s imitation of mothers’ speech often omits the functional elements (see Table 1 for examples). The two children’s speech of interaction with mothers were tape-recorded every second week for two hours in a total of 38 weeks. They were 27-month old and 18-month old at the beginning of the recording. Their MLUm increased from 1.84 to 3.55 and from 1.40 to 3.27, respectively.

**Table 1. Examples of Adam and Eve’s immediate imitative production of parents’ utterances**

<table>
<thead>
<tr>
<th>Parent</th>
<th>Child</th>
</tr>
</thead>
<tbody>
<tr>
<td>There goes one</td>
<td>There go one</td>
</tr>
<tr>
<td>Daddy’s brief case</td>
<td>Daddy brief case</td>
</tr>
<tr>
<td>Fraser will be unhappy</td>
<td>Fraser unhappy</td>
</tr>
<tr>
<td>He’s going out</td>
<td>He go out</td>
</tr>
<tr>
<td>That’s an old time train</td>
<td>Old time train</td>
</tr>
<tr>
<td>It’s not the same dog as Pepper</td>
<td>Dog Pepper</td>
</tr>
<tr>
<td>No, you can’t write on Mr. Cromer’s Shoe</td>
<td>Write Cromer shoe</td>
</tr>
</tbody>
</table>

*Adapted from Brown & Bellugi (1964) Table 2*

Several interesting phenomena have been noticed in the two children’s imitative production. Firstly, it is easy to see that the children’s imitation was not exactly the same as the model utterances. Apparently, some elements were missing in the children’s utterances, such as the auxiliary will, the copula *be*, the determiners *the* and *an*, the negation words *not* and *No*, the inflectional morpheme for present progressive *-ing,*
genitive ‘s and 3rd person singular -es. The omissions appear to be very systematic. The children were not just randomly skipping words. It is evident that most of the content words that are useful to convey means were reserved but function words and morphemes were largely omitted.

Secondly, Brown and Bellugi observed that words in the children’s imitative production were perfectly ordered as in the model utterances. The children did keep the word order of the words when they repeat, although many elements in the model utterances were missing. Retaining the order is not a natural consequence of omission, they argued. The children could have produced those words in potentially other ways, such as backwards or in random order, if they were simply holding the words in the working memory. Therefore, it was suggested that the children were actively processing the utterances in the input even if they had very limited knowledge about the structure of the target language. Since word order is an important aspect of the English grammar that the two children were acquiring, it would be very interesting to know whether the early speech of children acquiring free word order languages (e.g., Turkish) resembles the same pattern. Slobin & Bever (1982) did an imitation test with 30 Turkish-learning children with an average age of 3;9. Each of the 18 stimulus sentences was composed of two nouns and one verb in different orders (e.g., NNV, NVN or VNN). The test found that majority of the sentences (73%) was imitated correctly and the children tend to reorder sentences with non-canonical word order (e.g., NVN and VNN) to canonical ordering (NNV). The finding that children were sensitive to canonical word order even in free word-order language is in line with Brown and Bellugi’s conclusion.
Lastly, the maximum length of children’s utterances during that period was about four morphemes. Children omitted words so their production is usually between two to four morphemes. In other words, some words were omitted in order to keep the length constraint. The underlying mechanism that is responsible for the length constraint and the initial omission is still uncertain. One can explain that the length constraint originates from certain domain-general cognitive processing or working memory limitations.

Nativists would argue that children’s grammar consists of only a few simple grammatical rules (e.g., no rule for functional elements) so they could not produce longer utterances. Their speech production directly reflects their limited syntactic knowledge.

Synthesizing evidence from many naturalistic studies of early language development, Radford (1990) analyzed omissions of functional elements in young children’s speech production to argue for lack of functional categories in children’s earliest English grammar (around 20 months). Similar to Brown and Bellugi’s (1964) findings, he found that young children mostly use nominals without a determiner when a determiner is required in adult grammar. His examples of children’s naturalistic and imitative production drawn from various corpora well demonstrate the initial omission phenomenon. Some examples of children’s spontaneous production that apparently all miss some functional elements are shown in Table 2.
Table 2. Examples of children’s spontaneous production around 20-month old


Open door. Want ball. Want car (Stephen 19)

Blanket gone. Stick gone. Finger there (Bethan 20)


Adapted from Radford (1990) Chapter 4 (1)

Children’s responses to what-questions are even more revealing. The question-answer pairs in Table 3 clearly show that children at this point of development did not include determiners, such as a, in their responses when it is actually required in the adult grammar.

Researchers have been trying to understand why those functional elements are often omitted in early speech. Brown (1973) evaluated three possibilities: 1) content words with semantic meaning are easy to learn; 2) content words carry more information; 3) content words often have heavier stress which is easier to notice by children. Gerken, Landau and Remez (1990) examined children’s early sensitivity to functional elements in perception and production. Boyle and Gerken (1997) and Kemp (2005) investigated the problem that whether children are more likely to omit articles from sentences containing novel nouns and verbs.
Table 3. Some questions and answers between caregivers and children around 20-month old

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>What's this? Telephone</td>
<td>Paula 18</td>
</tr>
<tr>
<td>What's this? Spoon</td>
<td>Paula 18</td>
</tr>
<tr>
<td>What's this? Ballon</td>
<td>Dewi 18</td>
</tr>
<tr>
<td>What's that one? Check</td>
<td>Stephen 19</td>
</tr>
<tr>
<td>What's that? Tree</td>
<td>Elen 19</td>
</tr>
<tr>
<td>What's daddy built for you? Castle</td>
<td>Helen 20</td>
</tr>
<tr>
<td>What's the lady got? Mouse</td>
<td>Hayley 20</td>
</tr>
<tr>
<td>What are you doing with the brush? Butterfly</td>
<td>Lucy 20</td>
</tr>
<tr>
<td>What do you want? Want cup</td>
<td>Daniel 23</td>
</tr>
<tr>
<td>What's that? Good book</td>
<td>Leigh 24</td>
</tr>
<tr>
<td>What's that one? Naughty cow</td>
<td>Leigh 24</td>
</tr>
<tr>
<td>What have you got on there? Frog</td>
<td>Anne 24</td>
</tr>
</tbody>
</table>

Adapted from Radford (1990) Chapter 4 (2)

Brown’s (1973) characterization of Stage I as telegraphic speech mainly based on his English data. He also sought support from studies of early language development in about half a dozen other languages- Finnish (Bowerman, 1973), Swedish (Rydin, 1971), Hebrew (Bar-Adon, 1971), Mandarin Chinese (Chao, 1951), Russian (Slobin, 1966), Korean and Luo (Blount, 1969).

Order of acquisition

Function words and grammatical morphemes are acquired gradually after its first occurrence. Functional elements are not acquired all at the same time. Rather, they are
acquired at different points of grammatical development. In general, the rate is slower comparing to lexical acquisition. Some functional categories are acquired very late (later than Stage V).

Brown (1973) systematically studied the acquisition of 14 grammatical morphemes using longitudinal data from three children-Adam, Eve and Sarah. He was able to find some common patterns of development of the grammatical morphemes in the three children. As I mentioned earlier, the development was broke into five stages according to the MLUm. Function words and inflectional morphemes start to appear in Brown’s stage II (MLUm 2-2.5, average 2.25), e.g., prepositions like in and on, copular am, is and are, plural and possessive inflections on nouns, and progressive, past and third-person singular inflections on verbs. The development of grammatical morphemes extends a long period, sometimes even over Stage V (Figure 2).

One disadvantage of naturalistic studies is that they are sometimes difficult to distinguish between children’s imitative and creative use of words. Well-designed behavioral experiments often provide more control on the phenomenon that one is trying to probe. A body of experimental research has been done on young children’s comprehension and production of functional elements.
Figure 2. Order of acquisition of 14 grammatical morphemes in three children
Reproduced from Brown (1973) P. 271
Nature of early representation of functional categories

Last section illustrated the developmental timeline of functional elements. Young children gradually acquire classes of function words or morphemes from their second birthday. Before that, functional elements are often omitted in early child speech. Several critical questions remain: 1) what is the nature of the early representation of functional categories? 2) when do young children acquire and possess knowledge of adult-like abstract functional categories? 3) how do they acquire that knowledge? This section reviews recent research that is relevant to the first two questions. The next section focuses on the third question.

Kemp et al. (2005) tested elicitation and priming of determiners in English-learning children age from two to four years. In the first experiment, two novel toys were used to pull for determiners the or a from the child. One of the toys was first introduced with either a or the. It was then presented singly or with another three identical toys. Two scenarios were designed to pull for determiner-noun combinations the X and a X, respectively. They found that few two-year-olds but almost all 3- and 4-year-olds produced the novel noun with a different determiner from the initial introduction of the toy. This was taken as evidence that children’s initial knowledge of determiners is limited rather than being fully abstract from the beginning. However, they admitted that “cognitive demands of memorizing, recalling, and producing the novel nouns may have caused the younger children, especially, to revert to omitting determiners or to failing to use the whole range of determiners with which they were familiar.” The results from some experimental studies of determiners with even younger children suggest that what
Kemp et al. has observed in two-year-olds was very likely due to performance limitations rather than lack of competence or knowledge of determiners.

Children produce almost no determiners in every language around 14 months. However, they may already have some knowledge about the function words and their contexts in the target language. Behavioral experiments with German- and French-learning infants have demonstrated that infants as young as 14 months old are able to store and represent specific function words (Höhle et al., 2004; Shi & Melançon, 2010). Mintz (Höhle, et al., 2004; Mintz, 2006; Shi & Melançon, 2010) has shown that 12 months English-learning infants are able to categorize novel nouns using frequent frames, a joint co-occurrence context of immediate preceding and following words. Children age from one to three years old respond better to speech with correct function words (Gerken & McIntosh, 1993; Petretic & Tweney, 1977).

Using a head-turn preference method, Höhle et al. (2004) familiarized German-learning 12-month-olds and 15-month-olds with two novel words (monosyllabic pseudowords) in noun context (the novel word follows the indefinite article *ein ‘a’*) or in verb context (the novel word follows the personal pronoun *sie ‘she’*). The children were immediately tested after the familiarization phase with short passages in which a novel word was consistently used in noun contexts or verb contexts. It was found that 15-month-olds who were familiarized in the noun context showed a differential listening time to the ‘noun passages’ and ‘verb passages’. There was no significant difference for 15-month-olds familiarized in the verb context and both groups of 13-month-olds. The results suggest that 15-month-olds but not 13-month-olds were able to classify the novel
words to the noun class based on the co-occurrence statistics between the indefinite article and its following words. This and many other studies together provide robust evidence that one-year-olds are sensitive to word distribution and transitional probability in the input (Höhle, et al., 2004; Mintz, 2006; Saffran et al., 1996; Shi & Melançon, 2010).

Shi and Melançon (2010) tested young children’s knowledge of determiners and whether they can generalize between different determiners. Two novel words were combined with three French determiners-des (‘some’), ton (‘your’) and le (‘the’) to create Determiner + Noun sequences and combined with three French pronouns-je (‘I’), il (‘he’) and tu (‘you’) to create Pronoun + Noun sequences. French-learning 14-month-olds were familiarized with either the Determiner + Noun or Pronoun + Noun sequences. The two groups of children were both tested with the same set of items that contain both determiner and noun sequences. Shi and Melançon found that children familiarized with Determiner + Noun sequences showed significantly longer looking time for sequences contain a pronoun. Since the determiners in the test trials have never appeared in the familiarization phase, the difference in looking time can only be explained by the children’s prior knowledge of relationship between the determiners in the familiarization and test phases. Therefore, the result indicates that children as young as 14 months old are treating some determiners as a group so they could generalize and transfer knowledge of co-occurrence statistics between determiners. It further suggests that there is a primitive, if not completely abstract, determiner category in the grammar of young children. This would predict that young children with knowledge of such a determiner
category would use determiners as adults do. The result is also in line with a corpus study of young children’s initial determiner use conducted by Valian more than twenty years ago.

Valian (1986) examined speech samples of six English-learning children aged from 2 to 2;5 (MLU 2.93 to 4.14) for evidence of knowledge of six syntactic categories: Determiner, Adjective, Noun, Noun Phrase (NP), Preposition and Prepositional Phrase (PP). Distributional-based criteria were used to assess the children’s abstract knowledge of the six syntactic categories. She found evidence for all categories, except the borderline performance of Adjectives and PPs for the lowest MLU child. With the result, she argued that children must possess abstract knowledge of syntactic categories very early on and argued against semantic hypotheses on the origin of syntactic knowledge.

Two of the six syntactic categories she investigated, determiner and preposition, are relevant here. I will first describe Valian’s criteria for these two categories and then go through children’s performance according to the criteria developed. The method she used were based on mainly Brown’s distributional analysis of child corpora (Brown & Bellugi, 1964), Bloom’s (1970) rich interpretation and Chomsky’s (1975) limits of taxonomic analysis.

Words in children’s utterances were assigned syntactic categories as if they were in adult speech according to linguistic and social contexts. Procedure for testing category assignment is: 1) the preceding and following expressions; 2) substitutability test in children’s use; 3) multiple appearances test. A category should appear in all possible
locations; 4) subcategories (restricting different words to different subclasses, e.g., a to
singular nouns).

The Criteria used are that determiners must precede adjectives or nouns but not
precede other determiners and cannot occur alone. For prepositions, they take NP objects
like verbs but do not take tense or verbal complement. Prepositions like determiners can
precede nouns and they can also precede determiners but determiners cannot.

A small set of determiner types can account for most of determiner tokens (the
and a: 72%; the, a, my, this/that: 87%). All three criteria were error free: no ordering
error (two determiners in a row) and no determiners alone were found.

In, on and of can account for 61% of all preposition tokens. Children did not
inflect prepositions for tense or use a verbal complement after a preposition, as they
would for verbs. The two lowest MLU children sometimes did not use a preposition
when required (e.g., used go instead, ‘go that’).

Therefore, it supports the argument that children between two and two and a half
(MLU 3-4) have abstract knowledge of these categories. In the discussion, she also tried
to rule out the semantic basis of such syntactic knowledge.

Ihns & Leonard (1988) studied determiners and NP in Adam’s spontaneous
speech (Brown, 1973). They looked at the use of determiners with nouns according to
Valian’s (1986) criteria. Only one exception was found. They also found gradually
increasing usage of determiners in obligatory contexts with Adam’s age. Therefore, it
supports Valian’s finding that children use determiners correctly since early on and there
is no semantic limited scope, which in turn supports theories claim continuity in development.

Pine & Martindale (1996) aimed to replicate Valian’s (1986) results of determiners using corpora of seven children (age range from 1;11 to 2;7). They argued that Valian’s criteria are too generous and could be passed by children possess a relatively small amount of limited scope knowledge. They then proposed a new method for assessing young children’s knowledge of the determiner category by examining overlap in contexts in which different determiners occurred. For example, the overlap of determiners *a* and *the* is computed as the number of nouns used with both *a* and *the* divided by the number nouns used with either *a* or *the* (see page 35 for more discussions on overlap). The presence of an abstract determiner category was argued to enable acquired knowledge of one determiner to be available to other category members. Children with the category were expected to use multiple determiners with a noun and thus show similar overlap as adults. Due to the small sample size in Pine & Martindale (1996), Pine & Lieven (1997) analyzed speech samples from 11 children and Valian et al. (2009) analyzed speech samples from 21 children to mitigate the small sample size problem. Pine & Martindale (1996) and Pine & Lieven (1997) found that young children’s determiner use did not support the existence of grammatical knowledge of a determiner category, while the children in Valian et al. (2009) used determiners just like adults. Those analyses have come to conflicting results as to whether children’s overlap compared to their mothers’ indicated a determiner category. Simply analyzing larger or more samples does not seem to increase the reliability of measuring determiner
productivity using overlap. In addition, the findings in Pine & Martindale (1996) and Pine & Lieven (1997) appeared to be incompatible with the results from recent experimental studies that even 14-month-olds have some knowledge of a determiner category.

Because of the roles functional categories play in current syntactic theories and the recent rebuttal on the nature of representation of the determiner category—whether young children possess knowledge of an abstract determiner category or their early production is guided by limited scope formulae, it is important to reconcile the conflicting findings from different studies. Part I of this dissertation examined the validity of the overlap measure and devised a new overlap-based probabilistic method to quantify determiner productivity that has overcame a fatal problem of the original overlap measure.

**Learning mechanism**

Even if young children are predisposed with notions of abstract functional categories, they still have to assign the word forms in the target language to those categories because word forms and members of a category differ between languages and have to be learned from the input. In other words, a child has to map words in the target language to the right categories. This section discusses potential mechanisms for this word categorization process.

There have been a number of approaches to word categorization. One of them proposed that prosodic cues could be used to differentiate nouns and verbs into different categories. Semantic bootstrapping argues that the relation between syntactic categories
and semantic features (e.g., object and action) is useful to establish some major lexical categories. Based on the distributional information in children’s input, different word co-occurrence patterns (e.g., bigrams) have also been suggested to be very effective in categorizing words into major lexical categories.

A number of cues are informative about the categories of content words, such as phonological properties of words (Cassidy & Kelly, 1991; Kelly, 1992; Monaghan, et al., 2007; Shi, et al., 1998), the semantic commonalities across words that belong to the same syntactic category (Braine, 1976; Grimshaw, 1981; Pinker, 1984; Schlesinger, 1971), and certain word co-occurrence patterns (Cartwright & Brent, 1997; Mintz, 2003; Mintz et al., 2002; Redington, et al., 1998). However, it is reasonable to question the informativeness of those cues in categorizing function words.

Function words have different prosodic properties, such as weak or no stress, which is shared by words in different functional categories. Shi (1998) have found that phonological and acoustic information could be used to make the distinction of lexical and functional categories. However, it may not be enough to make finer distinctions within functional categories.

In addition, most function words carry few or no meaning. The semantic relationship between syntactic categories and semantic classes (e.g., object and action) does not seem to be helpful when categorizing function words.

Distributional information in the input would be the last resort because the categories of function words are also defined in terms of the contexts they occur. The distributional environments that have been studied so far (e.g., bigrams and frequent
frames) are good at categorizing content words (e.g., nouns and verbs) in general. Part 2 of this dissertation investigates whether current distributional environments are enough to categorize function words into proper categories when working with acoustic cues in the input.
Part 1. Knowledge of abstract functional categories

An important aspect of having abstract knowledge of grammatical categories is the understanding that members within the same category share certain properties and can be used to replace each other without violating the grammaticality of the sentence. This is often called the substitution test. Linguists use this principle to determine word categories when analyzing a new language. Psycholinguists use this principle to test children’s knowledge of grammatical categories. Without doubt, this principle also applies to functional categories, which are a subset of grammatical categories. For example, in the sentence “The cup is on the table”, table can be substitute by any noun and the can be substitute by other determiners like a or my. The result sentence would be perfectly grammatical.

In language acquisition, it is possible that young children may simply memorize small chunks of phrases (limited scope formula as in Pine & Martindale (1996) and Pine & Lieven (1997)) at the beginning. They have not made any generalization for the members of a category so they would not use one member of a category in a different context – another noun for determiner category. For example, children without abstract knowledge of a determiner category would not produce “the apple” and “an orange” even if he/she has already learned “an apple” and “the orange”. At some point of development, children must have achieved the state of an adult grammar when they have obtained abstract knowledge of functional categories. After this point of development, they will be able to use one member of a category to replace another member of the same category.
Sometimes, they would even over-generalize when they have not learned the differential constraints of members of the same category. A classic example, though not in category acquisition, is the U-shaped developmental pattern of past tense grammar in English-learning Children.

If a child have abstract knowledge of functional categories, it is expected to observe that he or she will use both determiners *a* and *the* before nouns to a degree that resembles adults’ usage. *Overlap of a and the* is computed by dividing the number of noun types that were used with both *a* and *the* by the number of noun types that were used with either *a* or *the* (Equation 1). The computation of overlap can be easily generalized to all determiners with Equation 2.

\[
Overlap_{a, the} = \frac{\text{# nouns occurring with both } a \text{ and } the}{\text{# nouns occurring with either } a \text{ or } the}
\]

*Equation 1. Overlap between determiners a and the*

\[
Overlap_{all, det} = \frac{\text{# nouns occurring with any two determiners}}{\text{# nouns occurring with any determiner}}
\]

*Equation 2. Overlap between all determiners*

The *overlap test*, proposed in Pine & Martindale (1996), compares children’s overlap to adults’ overlap (usually their mothers) to determine the time point when children have possessed the knowledge of abstract categories (Pine & Lieven, 1997; Pine & Martindale, 1996; Valian, et al., 2009). However, there are indications that these studies suffered from a problem of small sample size. Clearly, enough child and adult utterances are needed to compute the overlap accurately. Chapter 3 explores this problem.
by using utterances accumulated over sessions to increase sample size and obtain an estimate of the range of real overlap. After establishing that even a very large corpus is not enough, Chapter 4 proposes a new baseline—expected overlap of a sample that is independent from sample size—instead of the mother’s overlap. Expected overlap is computed using probabilistic method based on the nouns in a specific sample thus it would avoid the small sample size problem as shown in Chapter 3. By comparing expected to actual overlap for both children and mothers, it tries to show whether young children use determiners as their mothers do. Chapter 5 compares expected and actual overlap of a German child because the determiner system in German is more complex than the English determiner system. This analysis hopes to show cross-linguistic validity of the method proposed in Chapter 4. Analysis in Chapter 3 was conducted on mothers’ utterances only, Chapter 4 and Chapter 5 on utterances of both children and mothers.
Chapter 3. Small sample size problem

There are several indications that overlap is positively correlated with size of the sample that is used to compute the overlap. Intuitively, a bigger sample contains more noun tokens thus offers more opportunity of overlap. The problem is that when using small samples mothers’ overlap could be under-estimated, which means that mothers’ overlap could potentially be higher if a bigger sample is available. However, in previous studies, mothers’ overlap served as the baseline when assessing children’s overlap. When a bigger sample is analyzed, it is likely that mothers’ overlap will be higher and children’s overlap could be higher or the same. In summary, mothers’ overlap from small samples does not seem to be a good baseline. This may invalidate certain claims from previous studies on children’s knowledge of an abstract determiner category.

This analysis tests the hypothesis that overlap is positively correlated with sample size by computing overlap on increasingly larger speech samples of the mothers. To build larger samples, a mother’s utterances from several recording sessions are combined together to form larger samples for each individual mother. For the reason that a mother may adjust the complexity of her speech according to the language development of her child, the samples were built cumulatively from the last recording session to the first recording session. Table 4 illustrates how 20 samples with increasing sizes were built from a corpus with ten recording sessions. The first sample contains all utterances from the last recording session. The second sample contains utterances from the last two recording sessions. The last and largest sample contains utterances from all recording
sessions. When overlap has been computed for all the samples, the relationship between
overlap and sample size can then be determined.

**Table 4. Creating increasingly larger samples from a corpus of 20 recording sessions by
cumulatively adding sessions in backward (mother’s utterances only)**

<table>
<thead>
<tr>
<th>Analyzed sample</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>…</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recording</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sessions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>included</td>
<td>20</td>
<td>19 - 20</td>
<td>18 - 20</td>
<td>17 - 20</td>
<td>…</td>
<td>1 - 20</td>
</tr>
</tbody>
</table>

**Data**

The data includes speech of eight mothers in naturalistic settings from corpora in
the CHILDES database (MacWhinney, 2000): Eve (Brown, 1973), Nina (Suppes, 1974),
Anne (Theakston et al., 2001), Aran (Theakston, et al., 2001), Lara (Rowland & Fletcher,
2006), Lily (Demuth et al., 2006), Naima (Demuth, et al., 2006), Thomas (Lieven et al.,
2009). Only utterances from the first recording session to the last recording session when
the child is younger than three years old and the child’s MLU is less than four are
analyzed because this is the period when children start to combine words together to
produce multi-word utterances and begin to show some knowledge of grammatical
categories. Table 5 lists for each child the age at the time of the first and last recordings,
the child’s MLU in morpheme for the first and last sessions and number of utterance for
both child and mother.
Table 5. Statistics of the eight corpora

<table>
<thead>
<tr>
<th>Child</th>
<th>Files</th>
<th>Begin age</th>
<th>End age</th>
<th>Begin MLU</th>
<th>End MLU</th>
<th>Child utt</th>
<th>Child utt/m</th>
<th>Mother utt</th>
<th>Mother utt/m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anne</td>
<td>1-34b</td>
<td>1;10.07</td>
<td>2;9.10</td>
<td>1.55</td>
<td>3.28</td>
<td>19868</td>
<td>1806</td>
<td>36220</td>
<td>3293</td>
</tr>
<tr>
<td>Aran</td>
<td>1-34b</td>
<td>1;11.12</td>
<td>2;10.28</td>
<td>1.41</td>
<td>3.55</td>
<td>17111</td>
<td>1426</td>
<td>34487</td>
<td>2874</td>
</tr>
<tr>
<td>Eve</td>
<td>1-20</td>
<td>1;6</td>
<td>2;3</td>
<td>1.56</td>
<td>2.77</td>
<td>10859</td>
<td>1086</td>
<td>10239</td>
<td>1024</td>
</tr>
<tr>
<td>Lara</td>
<td>2-11-25</td>
<td>1;09.13</td>
<td>2;11.25</td>
<td>1.73</td>
<td>3.71</td>
<td>32232</td>
<td>2149</td>
<td>53551</td>
<td>3570</td>
</tr>
<tr>
<td>Lily</td>
<td>1-65</td>
<td>1;1.02</td>
<td>2;11.27</td>
<td>1.19</td>
<td>3.06</td>
<td>31102</td>
<td>1352</td>
<td>57393</td>
<td>2495</td>
</tr>
<tr>
<td>Naima</td>
<td>1-46</td>
<td>0;11.28</td>
<td>2;1.17</td>
<td>1.43</td>
<td>3.51</td>
<td>22240</td>
<td>1711</td>
<td>33763</td>
<td>2597</td>
</tr>
<tr>
<td>Nina</td>
<td>1-38</td>
<td>1;11.16</td>
<td>2;10.28</td>
<td>2.07</td>
<td>3.89</td>
<td>21147</td>
<td>1762</td>
<td>22356</td>
<td>1863</td>
</tr>
<tr>
<td>Thomas</td>
<td>2-00-12 2-11-28</td>
<td>2;0.12</td>
<td>2;11.28</td>
<td>1.48</td>
<td>2.59</td>
<td>119077</td>
<td>9923</td>
<td>196160</td>
<td>16347</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td>1.55</td>
<td>3.30</td>
<td>34205</td>
<td>2652</td>
<td>55521</td>
<td>4258</td>
</tr>
</tbody>
</table>

* utt = utterances, utt/m = utterances per month

Procedure

For each of the eight corpora, increasingly larger samples were created by successively adding the mother’s utterances from recording sessions. To avoid effect of mothers adjusting speech to the development of their children, utterances were added in reverse chronological order, from the latest to the earliest session (see Table 4). A computer program then automatically searched for noun phrases with a determiner in every sample. Finally, mother’s overlap between a and the was computed on each sample using Equation 1.

Utterances in the corpora were annotated with morphological information (e.g., plural) and part of speech using MOR and POST programs in CLAN (MacWhinney, 2000). The search of noun phrases with a determiner starts from a determiner in the
morphological annotation of an utterance and looks to the right for a noun until reaching a verb, comma or end of utterance.

**Results**

![Graph showing overlap values](image)

**Figure 3. Overlap for samples with increasingly more utterances from the mother**

For each corpus, this analysis produced a series of overlap values from samples of different sizes. The numbers of utterances in the samples are increasing because the samples were created by successively adding more utterances. Figure 3 shows the overlap values for seven of the eight mothers (Thomas mother’s overlap is shown in Figure 4). There is a clear trend that overlap becomes higher when more utterances are analyzed. The overlaps are all quite small for the first samples, 0.10 on average. When the samples
are larger than 15,000 utterances, all of the overlaps become higher than 0.2. The
growing speed of overlap varies by corpus, the overlap of Nina’s mother reaches 0.4
around 20,000 utterances and the overlap of Lara’s mother is only 0.26 even if about
53,000 utterances are analyzed. Overlap and number of utterances in a sample were
strongly positively correlated, mean r=.89 (ranging from 0.84 to 0.93) for the eight
corpora (Table 6).

Table 6. Correlation between overlap and number of utterances (mother)

<table>
<thead>
<tr>
<th>Child</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anne</td>
<td>0.88</td>
</tr>
<tr>
<td>Aran</td>
<td>0.93</td>
</tr>
<tr>
<td>Eve</td>
<td>0.93</td>
</tr>
<tr>
<td>Lara</td>
<td>0.87</td>
</tr>
<tr>
<td>Lily</td>
<td>0.91</td>
</tr>
<tr>
<td>Naima</td>
<td>0.86</td>
</tr>
<tr>
<td>Nina</td>
<td>0.93</td>
</tr>
<tr>
<td>Thomas</td>
<td>0.84</td>
</tr>
<tr>
<td>Mean</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Seven of the corpora have sizes between 10,000 and 50,000 utterances from the
mothers. The Thomas corpus is a dense corpus (five hours of recording per week) that
has about 200,000 utterances from the mother, which is four times more than the largest
of the other seven corpora. It provides a good opportunity to investigate “how big is
enough” – whether overlap becomes asymptotic when the sample reaches certain size.
Figure 4 shows the overlap for the samples created from Thomas mother’s utterances.
The overlap reaches 0.35 when the sample has 50,000 utterances. However, it keeps increasing when three times more utterances are analyzed. Although it seems the overlap becomes stable for the last samples, it is very likely that the overlap is just growing slower and will be higher when more utterances are available.

**Figure 4. Overlap for samples with increasingly more utterances (Thomas mother)**

If samples used in previous studies (between 400 and 800 utterances) were large enough for an accurate estimation of overlap, significant increase in overlap for larger samples is not expected. Apparently current analysis has shown that overlap increases as more utterances analyzed, which suggests that previous studies under-estimated mothers’ and maybe children’s overlap by using small samples. Mother’s overlap could reach 0.3 to 0.4 when a large sample is analyzed while Valian et al. (2009) reported 0.11 as average overlap of 21 mothers. Children’s overlap could be higher or the same when more utterances are analyzed. It is problematic to compare a child’s overlap to the mother’s overlap using a small sample that under-estimates at least the mother’s overlap. Simply
using larger samples to compute overlap does not seem to resolve the problem because most of the corpora that have been collected are relatively small and more importantly we do not know how big a sample is enough for computing overlap accurately.

Instead of directly using adult’s overlap as a baseline for measuring a child’s knowledge of abstract categories, Chapter 4 proposed a new measure that compares a sample’s overlap (*actual overlap*) to the *expected overlap* for that sample computed using a probabilistic method based on the nouns used in the sample.
Chapter 4. Measuring knowledge of determiner category with expected overlap

Chapter 3 has shown that overlap is positively correlated with sample size thus large sample of utterances is required to compute a reliable overlap between determiners. Since there are not many dense corpora available and overlap does not become asymptotic even with very large samples, a new method is needed to avoid the requirement of large samples. Intuitively, the overlap of a sample depends on several factors: 1) some nouns are more probable to show overlap than others (e.g., book vs. sun); 2) the more noun tokens the higher probability of overlap. This analysis proposes a probabilistic method that computes expected overlap of a sample based on the nouns in that sample. When considering only determiners a/an and the, the probability of overlap for a noun is the total probability mass (1) minus the probability of the noun only occurring with a/an or the. Let \( p \) be the probability of the noun occurring with the and 1 - \( p \) the probability of the noun occurring with a/an. If a noun occurs \( f \) times in a sample, its probability of only occurring with the is \( p^f \) and its probability of only occurring with a/an is \( (1 - p)^f \). Equation 3 computes the probability of overlap for a noun occurring \( f \) times in a sample. Each noun occurrence is considered as a Bernoulli trial that the outcome is either a/an or the. Most nouns do not occur with all the determiners with equal opportunities. The probability of a noun occurring with a/an and the is unknown but can be estimated from a large corpus, such as all adults’ speech from corpora in the CHILDES database or a particular mother’s speech. An estimation of the probability is
how many times a noun occurs with *the* comparing to how many times the noun occurs with both *a/an* and *the* in the large corpus (Equation 4).

\[
p(\text{overlap}) = 1 - p^f - (1 - p)^f
\]

*Equation 3. Probability of overlap for a noun occurring *f* times in a sample*

\[
p = \frac{\# \text{nouns occurring with } \text{the}}{\# \text{nouns occurring with either } \text{the or a/an}}
\]

*Equation 4. Estimate the probability of a noun occurring with the*

With sample-specific expected overlap (i.e., \(p(\text{overlap})\) in Equation 3), this analysis tests the hypothesis that young children’s actual overlap deviates from expected overlap to the same degree as their mothers. The extent of deviation from expected overlap would be similar for children and mothers if the children were using determiners in the same way as their mothers. On the other hand, larger deviation of children’s actual overlap from expected overlap would indicate that they were using determiners in a more idiosyncratic way.

**Data**

The data analyzed includes the utterances of children and mothers from the same eight corpora used in Chapter 3. For each corpus, utterances from recording sessions of the same month of the child’s age is combined together to form a single sample. Only utterances with at least a determiner were selected for analysis. The number of utterances per month for each child and mother is shown in Table 7.
Table 7. Numbers of utterances for samples grouped by month

<table>
<thead>
<tr>
<th>Subject</th>
<th>Utterances/month</th>
<th>Total utterances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anne</td>
<td>138</td>
<td>1651</td>
</tr>
<tr>
<td>Aran</td>
<td>195</td>
<td>2334</td>
</tr>
<tr>
<td>Eve</td>
<td>133</td>
<td>1326</td>
</tr>
<tr>
<td>Lara</td>
<td>187</td>
<td>2808</td>
</tr>
<tr>
<td>Lily</td>
<td>151</td>
<td>3016</td>
</tr>
<tr>
<td>Naima</td>
<td>254</td>
<td>5841</td>
</tr>
<tr>
<td>Nina</td>
<td>492</td>
<td>4924</td>
</tr>
<tr>
<td>Thomas</td>
<td>747</td>
<td>8960</td>
</tr>
<tr>
<td>Mother</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anne</td>
<td>610</td>
<td>7318</td>
</tr>
<tr>
<td>Aran</td>
<td>805</td>
<td>9660</td>
</tr>
<tr>
<td>Eve</td>
<td>214</td>
<td>2135</td>
</tr>
<tr>
<td>Lara</td>
<td>583</td>
<td>8750</td>
</tr>
<tr>
<td>Lily</td>
<td>663</td>
<td>15245</td>
</tr>
<tr>
<td>Naima</td>
<td>634</td>
<td>15852</td>
</tr>
<tr>
<td>Nina</td>
<td>868</td>
<td>8684</td>
</tr>
<tr>
<td>Thomas</td>
<td>5061</td>
<td>60736</td>
</tr>
</tbody>
</table>

**Procedure**

The probability of a noun occurs with *the* \((p)\) was estimated from the entire set of mother’s speech from a corpus. \(p\) was estimated separately for each dyad (or corpus). Hence, the \(p\) for the same noun may differ for two dyads. A computer program searches for noun phrases with a determiner in all utterances as described in Chapter 3. The search was conducted on mother’s speech only. It then counts the number of times *a/an* occurred with a noun and the number of times *the* occurred with that noun. Finally, the probability
of a noun occurring with *the* was calculated using Equation 4. When computing expected overlap for a child and the mother as described below, the same *p* that is estimated from the mother’s speech was used. This ensures that the mother’s speech instead of the child’s own speech is treated as the baseline of the measurement.

To compute the expected overlap of a sample, a computer program searches in the sample for nouns phrases and records the frequency distribution of nouns (i.e., how many times a noun occurs). With the probability of a noun occurs with *the* and the number of occurrences of a noun, the expected overlap of a noun in this sample is computed using Equation 3. Expected overlap for a sample is the summation of the expected overlap of all nouns in this sample rounded half up to the nearest integer (Equation 5). The actual overlap of a sample was computed using Equation 1. Finally, Equation 6 computes the deviation of a sample’s actual overlap from expected overlap.

\[
\text{Expected Overlap} = \left[ 0.5 + \sum_{\text{All nouns}} p(\text{overlap}) \right]
\]

*Equation 5. Expected overlap of a sample*

\[
\text{Deviation} = \frac{\text{Expected Overlap} - \text{Actual Overlap}}{\text{Expected Overlap}}
\]

*Equation 6. Deviation of a sample’s actual overlap from expected overlap*

**Results**

Figure 5 to Figure 8 show the actual and expected overlap for four of the eight children and mothers. In general, children produced fewer noun phrases than the mothers did thus their expected overlap are smaller than the mothers. When comparing the actual
overlap to expected overlap, children’s deviations from expected overlap are quite small for most of the data points in the children’s second year.

Figure 9 shows the deviation from expected overlap for all data points in the eight corpora analyzed. A paired t-test was conducted on the children and mothers’ deviations. There was no significant effect for deviation, t(150) = 1.98, p = 0.96. The means of the two groups (0.214 and 0.212) were very close but the variance of the children’s deviations (0.093) is about six times larger than the variance of the mothers’ deviations (0.015). The deviation of some data points that is close to the extreme values -1 or 1 is likely due to the way deviation is calculated (e.g., an actual overlap of 0 and an expected overlap of 1 would result in a deviation of 1).

This analysis demonstrates that children’s actual overlap deviates from expected overlap to the same degree as their mothers. The results suggest that English-learning young children use determiners in the same way as adults do.
Figure 5. Actual and expected overlap of Lily and mother

Figure 6. Actual and expected overlap of Naima and mother
Figure 7. Actual and expected overlap of Nina and mother

Figure 8. Actual and expected overlap of Thomas and mother
Figure 9. Deviation for all data points in the eight corpora analyzed (sorted by children’s deviation)
Chapter 5. Analysis of German determiner category

Chapter 4 demonstrated that the deviation of actual overlap from expected overlap is a better measure than directly comparing children’s and mothers’ actual overlap. The eight English-learning children in the last analysis were using English determiners in the same way as adults do since two years old. Questions arise on the usefulness of this method to other languages and whether young children learning languages other than English show similar developmental pattern of determiner category. German is interesting here because its articles and determiner system are much more complex than the English determiner system. It encodes not only definiteness but also gender, number and case. Such an intricate system may delay children’s age of acquisition simply because more features need to be learned. On the other hand, a complex system provides more data and more variety in the input that could actually facilitate learning. It is an empirical question whether German-learning children acquire the determiner category in the same period as English-learning children.

Unlike the and a/an in English, German uses six definite determiners and six indefinite determiners that mark number, gender and case. Table 8 and Table 9 illustrate the definite and indefinite determiner systems in German (Gross, 2004). Gender is a feature that the English determiner system lacks. Figure 10 gives some examples of German noun phrases with a determiner (Wiltschko, 2009). The structure of German noun phrases is also more complicated than that of English noun phrases. Nevertheless, they have the same basic structure – determiner adjectives noun. Determiners are
obligatory in most cases except for uncountable nouns. Adjectives are optional. Therefore, German is a good test case for the method developed using English in Chapter 4 because it has a similar but more complex determiner system comparing to English. This chapter tests the hypotheses that expected overlap is a good predictor of actual overlap in German and a German-learning child uses determiners in different context to the same extent as adults.

Table 8. German definite determiner system

<table>
<thead>
<tr>
<th>Gender</th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominative</td>
<td>der</td>
<td>das</td>
</tr>
<tr>
<td>Accusative</td>
<td>den</td>
<td>das</td>
</tr>
<tr>
<td>Dative</td>
<td>dem</td>
<td>dem</td>
</tr>
<tr>
<td>Genitive</td>
<td>des</td>
<td>des</td>
</tr>
</tbody>
</table>

Table 9. German indefinite determiner system

<table>
<thead>
<tr>
<th>Gender</th>
<th>Singular</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominative</td>
<td>ein</td>
<td>ein</td>
</tr>
<tr>
<td>Accusative</td>
<td>einen</td>
<td>ein</td>
</tr>
<tr>
<td>Dative</td>
<td>einem</td>
<td>einem</td>
</tr>
<tr>
<td>Genitive</td>
<td>eines</td>
<td>eines</td>
</tr>
<tr>
<td>Gender</td>
<td>der Mann</td>
<td>die Frau</td>
</tr>
<tr>
<td>--------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>D.MASC man</td>
<td>D.FEM woman</td>
<td>D.NEUT child</td>
</tr>
<tr>
<td>‘the man’</td>
<td>‘the woman’</td>
<td>‘the child’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number</th>
<th>der Mann</th>
<th>die Männ-er</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.MASC.SG man</td>
<td>D.PL man-PL</td>
<td></td>
</tr>
<tr>
<td>‘the man’</td>
<td>‘the men’</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>die Frau</th>
<th>die Frau-en</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.FEM.SG woman</td>
<td>D.PL woman-PL</td>
<td></td>
</tr>
<tr>
<td>‘the woman’</td>
<td>‘the women’</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>das Kind</th>
<th>die Kind-er</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.NEUT.SG child</td>
<td>D.PL child-PL</td>
<td></td>
</tr>
<tr>
<td>‘the child’</td>
<td>‘the children’</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case</th>
<th>der Mann hat den Apfel des Schülers gegessen</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.MASC.SG.NOM man AUX D.MASC.SG.ACC apple</td>
<td>D.MASC.SG.GEN student eat.PART</td>
</tr>
<tr>
<td>‘The man has eaten the student's apple.’</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>der Mann hat dem Schüler geholfen</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.MASC.SG.NOM man AUX D.MASC.SG.DAT student help.PART</td>
<td></td>
</tr>
<tr>
<td>‘The man has helped the student.’</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 10. Examples of German determiners*
*Adapted from (Wiltschko, 2009)*)
Data

The corpus is spontaneous speech production of a monolingual High-German-learning child Leo (Behrens, 2006) from 1;11.13, the onset of multiword speech, up to 4;11. The recordings captured mainly his interaction with his parents and the research assistant doing the recordings. Between 2;0.00 and 2;11.29, the corpus is very dense and consists of five one-hour sessions each week supplemented by daily diary notes. The data analyzed in this experiment is Leo’s and his mother’s utterances between his age 1;11.12 and 2;11.29 (see Table 10). Utterances from recording sessions of the same month of the child’s age is combined together to form a single sample. The number of child’s utterances in this corpus is substantial and comparable to that of Thomas. The number of mother’s utterances is similar to the child’s, but much less than that of Thomas’ mother.

Table 10. Statistics for the Leo corpus (German)

<table>
<thead>
<tr>
<th>Child</th>
<th>Files</th>
<th>Begin date</th>
<th>End date</th>
<th>Begin MLU</th>
<th>End MLU</th>
<th>Child utt</th>
<th>Child utt/m</th>
<th>Mother utt</th>
<th>Mother utt/m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leo</td>
<td>le011112 le021129</td>
<td>1;11.12</td>
<td>2;11.29</td>
<td>1.12</td>
<td>3.42</td>
<td>115857</td>
<td>9655</td>
<td>99899</td>
<td>8325</td>
</tr>
</tbody>
</table>

* utt = utterances, utt/m = utterances per month

Procedure

The procedure is similar to Chapter 4 with some minor differences. This analysis included the six definite and six indefinite articles in German as listed in Table 8 and Table 9. The indefinite inflectional markings for plurals are excluded. German determiners also mark gender and case besides number and definiteness. Since determiners marking different genders and cases may not be used in the same context,
there would be no or little overlap between these determiners. Thus, the current analysis
tests the overlap between the two groups of determiners that mark distinct definiteness as
the analysis of English determiners in the last chapter. All definite articles are treated as
one group and all indefinite articles are treated as another group.

Because this corpus is not annotated with morphological information and
grammatical categories the cooccur command from the CLAN program (MacWhinney,
2000) was used to extract all the words that follow one of the twelve determiners from
both the child’s and the mother’s utterances. Therefore, in this analysis, the contexts that
a determiner occurs with are not only nouns but also include many adjectives and adverbs.

The probability that a word follows a definite determiner, $p$, was estimated from
the entire set of mother’s speech. A computer program counts the number of times a word
follows an indefinite determiner and the number of times that word follows a definite
determiner. Finally, the probability of the word following a definite determiner is
calculated using a variation of Equation 4.

To compute the expected overlap of a sample, a computer program records the
frequency distribution of words that follow a determiner (i.e., how many times a word
follows a determiner) in the sample. With the probability of a word following a definite
determiner ($p$) and the number of occurrences of a noun, the expected overlap of a word
in this sample was computed using Equation 3. Expected overlap for a sample is the
summation of the expected overlap of all words that follow a determiner in this sample
rounded half up to the nearest integer (Equation 5). The actual overlap of a sample was
computed using a variation of Equation 1. Finally, Equation 6 computes the deviation of a sample’s actual overlap from expected overlap.

**Results**

Figure 11 shows the numbers of expected and actual word types that overlap between definite and indefinite determiners. In the first four months of the data (1;11 to 2;3), Leo produced fewer word types than his mother. Thus, both Leo’s actual and expected overlap was very low in those four months. After that period, his word production is similar to his mother in quantity. The expected overlap for Leo and his mother was somewhere around 150 to 250.

![Figure 11. Actual and expected overlap of Leo and mother

The pattern of actual and expected overlap for Leo and his mother is very similar to those of the English-speaking children and mothers in the last chapter. When comparing the mother’s expected and actual overlap, they are quite close for most of the
months, which suggests that expected overlap is a rather accurate estimation of actual overlap. For Leo, his actual overlap follows very closely to his expected overlap except at 2;2. It appears that his deviation of actual overlap from expected overlap is no larger than his mother’s deviation.

Figure 12 shows the deviation of actual overlap from expected overlap for both Leo and his mother. A paired t-test was conducted on Leo’s and his mother’s deviations. There was no significant effect for deviation, $t(12) = 2.18, p = 0.47$. Leo’s mean deviation (0.199) was slightly higher than his mother’s mean deviation (0.145). The variance of the child’s deviations (0.099) is about ten times larger than the variance of the mother’s deviations (0.010). The deviation for Leo at 1;11 is rather high comparing to the others due to the way deviation is calculated - he produced zero overlapped word type and the expected overlap is one word type (as shown in Figure 11) thus the deviation is 1.

This result does not seem to suggest that Leo has any delay in using different determiners in the same contexts. He used determiners in the same way as his mother since 2;3 when he started to produce significant amount of determiners. It is not clear whether the higher deviations from 2;0 to 2;2 were due to Leo’s development of syntactic knowledge or fewer words produced in those three months. Analysis of data from more children would be helpful to answer this question.
This analysis shows that the German-learning children achieved abstract knowledge of functional categories at the same time period as English children, which suggests some universal developmental pattern that is favored by nativist views. The analyses in the last chapter and this chapter examined children’s production and provide valuable data on the time point when children start to possess abstract knowledge of function categories.

**Discussion**

The analyses showed that young children at the first half of their second year of life already possess abstract knowledge of some functional categories like determiner. Even before the second birthday, they already process function words/morphemes as abstract categories (Shi & Melançon, 2010). When they start producing combinatorial speech, they are able to quickly generalize nouns to different determiners. These evidence
strongly suggests that they are actively using abstract knowledge in language processing and production.

Their early sensitivity to function words/morphemes and abstract knowledge of functional categories indicate that they can use functional items to categorize nouns and verbs.

The results also bear on the question that whether children come to the world with abstract grammatical categories or they develop those categories from limited-scope formulae. According to nativist views, early emergence of functional categories is taken to be evidence that these categories are innately endowed and a part of UG. Children’s task is to determine which categories exist in their native languages and assign words to the pre-existing categories. The current study is not able to decisively conclude on the existence of innate grammatical categories, but it is in favor of nativists’ argument. If the categories are not innately available to young children, there must be a powerful categorization mechanism that can form adult-like categories in limited time with very limited input and without negative evidence (the Poverty-of-Stimulus Argument).

Regardless which theoretical view one takes, young children have to map words to categories or group words from the same category together. This is the word categorization problem, which is explored in the second part of this dissertation.
Part 2. Categorizing function words

Word categorization is necessary to map words onto categories for language acquisition. There have been many research on categorizing content words (e.g., nouns and verbs). Currently there are three major theoretical approaches to the categorization problem: semantic bootstrapping, phonological bootstrapping and distributional bootstrapping. This dissertation focuses on the distributional approach. Several distributional cues (including bigrams, frequent frames and preceding function words) were shown to be informative for categorizing nouns, verbs and some adjectives in typologically different languages. Although the above-mentioned distributional patterns were able to capture a few groups of function words or morphemes, most of the groups or frames were dominated by lexical items. No research to date has specifically studied the categorization of functional items using distributional information.

Previous research has shown that there are some acoustic, phonological and prosodic cues that distinguish function words from content words (Shi, et al., 1998). However, function words belong to a number of different functional categories (such as determiner, auxiliary, conjunction, preposition and postposition). Those cues may not be able to make finer distinctions between functional categories. Semantics are not so helpful here because 1) function words usually carry less meaning than content words. It would be difficult to form functional categories based on their meanings; 2) children may not have learned the meaning of function words when they start to categorize words because some researchers claim that the reason for their initial omission in production is
that no meaning is attached to the form and they only use it when they have learned both form and meaning (Naigles, 2002; Shi & Melançon, 2010). Similar to content words, function words display great regularity in their co-occurrence patterns and distributions in general. For example, they often occur in phrase boundaries; they often occur at the beginning or end of an utterance; they often followed or proceeded by a variety of content words from the same category (such as a determiner+Noun, pronoun+Verb, prep+NP in English). The hypothesis is that distributional information in the input is a very informative cue for categorizing function words if they could be differentiated from lexical items using acoustic cues. Existing distributional contexts like bigrams and frequent frames are not very effective in capturing function words. Often function words serve as contexts but not targets. One of the reasons for this is that function words are often highly frequent so they are selected to be the contexts based on the definition of frequent frames and other distributional cues. Given that newborn infants can distinguish function words to content words and 14-month-olds can represent the forms of function words, it looks like they must have already established the distinction of content vs. function words when they start to categorize words from one and a half years old (Shi & Melançon, 2010; Shi, et al., 1999). It is plausible that they may use this information as a constraint to word categorization, i.e., they develop functional categories only involving function words, which make the tasks different and relatively easier. Two analyses were carried out to test whether the distributional information in the input plus the constraints of only including function words as targets leads to accurate categorization. Chapter 6 uses immediate one-word contexts similar to Mintz (2002). Chapter 7 uses frequent
frames (Mintz, 2003). The outcomes are both measured with accuracy (which is explained in Chapter 6).
Chapter 6. Categorizing function words with immediate contexts

Mintz (2002) has shown that immediate preceding and following words of the word to be categorized (the target word) performed reasonably well in categorizing words into groups of words that are from the same category. The most frequent 200 words in a corpus were used as target words. There are a few groups that contains primarily function words, such as he her here his it Linda my Nina Nina’s she that the there this your and I’m he’s she’s they’re you’re. The majority of the target words are content words. This analysis uses the same context - immediate preceding and following one word. However, all the target words were the most frequent function words. The analysis in this chapter tests the hypothesis that distributional information in the input, specifically linear contexts is informative for categorizing function words into category-like groups.

Data

Mothers’ speech from the same eight English corpora in Chapter 3 was analyzed. 82 high frequency English function words from the determiner, auxiliary, preposition and pronoun categories were selected as target words (see Table 11).
Table 11. Target function words used in this analysis

<table>
<thead>
<tr>
<th>Determiner</th>
<th>a another any four no one other some that the these this those three two what</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronoun</td>
<td>he her him his i me my she their us we you your</td>
</tr>
<tr>
<td>Preposition</td>
<td>above among at below between by during for from in inside near of on over since to under until within</td>
</tr>
<tr>
<td>Auxiliary</td>
<td>am are be been can could did didn't do does doesn't don't had hadn't has hasn't have haven't is isn't may might must shall should was wasn't were weren't will won't wouldn't would</td>
</tr>
</tbody>
</table>

Procedure

The number of preceding and following words of each target function word was recorded in two context lists, respectively. Only word types that are in the most frequent 200 words in the mother’s speech were used as the context. Hence, the two context lists have 200 or less word types. The two context lists were concatenated together to form a single vector (see Figure 13). The distance between any two function words was calculated using a number of distance measures, such as Euclidean (Equation 7) and Canberra (Lance & Williams, 1967) distances. The Canberra distance measure significantly outperformed the others in both accuracy and purity (see next section Quantitative evaluation for definitions of accuracy and purity). The numbers reported in this analysis were all computed using the Canberra distance. Note that the general method is similar to Mintz et al. (2002) but the target words were manually selected.
function words and the distance measure and the hierarchical clustering algorithm were different.

\[ \theta = \cos^{-1}\left(\frac{\text{context}_i \cdot \text{context}_j}{|\text{context}_i||\text{context}_j|}\right) \]

**Equation 7. Euclidean distance of two target words**
*Adapted from Mintz et al. (2002)*

The distance between every pair of function words was computed and resulted in a symmetrical distance matrix. A hierarchical clustering algorithm *ward* was run on the distance matrix to group similar function words into clusters. The grouping of function words differs at different heights of the hierarchy. A certain height is manually selected to evaluate the grouping (such as four or two clusters in Figure 14). Height is selected so
that the groups are not too small; they contain at least a number of members; and include as few as possible function words from other categories.

In this analysis, the hierarchy was cut off at three different levels that resulted in 8, 12 or 16 clusters. The cut-off levels were chosen to make sure that each cluster has a relatively large number of target words while the clusters contain primarily target words from one of the categories.

![Diagram of clusters resulting from different cut-off levels]

*Figure 14. Clusters result from different cut-off level*

*Adapted from Mintz (2002)*
**Quantitative evaluation**

The categorization outcome is evaluated with standard measures in the literature - *accuracy* and *completeness* (Cartwright & Brent, 1997; Mintz, 2003; Redington, et al., 1998), which are similar to the metrics *precision* and *recall* in signal processing literature. To compute these measures, categorized words were labeled with their actual grammatical categories. For each cluster, the categories of every pair of target words were compared to each other. A hit was recorded when two items were from the same grammatical category (e.g., when two determiners were in the same cluster). A false alarm was recorded when two items were from different grammatical categories (e.g., when a determiner and an auxiliary were in the same cluster). Misses were computed by comparing all possible pairs of target words regardless of clusters. A miss was recorded when the members of a pair belong to the same grammatical category but were not clustered together (e.g., when two determiners were in two different clusters). Accuracy and completeness were then computed using Equation 8 and Equation 9.

\[
\text{Accuracy} = \frac{\text{hits}}{\text{hits} + \text{false alarms}}
\]

*Equation 8. Computation of accuracy*

\[
\text{Completeness} = \frac{\text{hits}}{\text{hits} + \text{misses}}
\]

*Equation 9. Computation of completeness*

Accuracy is penalized when words belong to different grammatical categories are grouped together. Accuracy ranges from zero to one, where a score of one indicates that
each distributional category contains items from only one grammatical category.

Completeness also ranges from zero to one, where a score of one indicates that words of the same grammatical category were classified together. Completeness is penalized when words belong to the same grammatical category appear in different distributional categories. For the problem of bootstrapping, forming categories that are linguistically accurate is arguably more important than forming comprehensive categories. Thus, accuracy is a more relevant measure of categorization success for current analyses.

The clustering was also assessed with purity that was devised in Mintz (2002) so the results of this analysis can be compared to Mintz’s (2002) study on clustering the most frequent 200 words. Purity of a category is calculated with Equation 10 where $X$ is the total number of words for category $X$, $C$ is the number of clusters, $x_i$ is the number of words in the $i$th cluster that are from $X$, $w_i$ is number of words in the $i$th cluster. The purity of a hierarchy is the mean purity values of all categories.

$$Purity = \frac{\sum_{i=1}^{C} \left( \frac{x_i^2}{w_i} \right)}{X}$$

*Equation 10. Purity for category $X$*

**Results**

Table 12 shows the accuracies and completeness at the three cut-off levels. Overall, the mean accuracies of the eight corpora ($M=.59, .74, .76$) were reasonably good. Apparently, more clusters improved the accuracy but yielded lower completeness. The accuracies and completeness were similar for the eight corpora. The Thomas corpus has
somewhat higher accuracies and completeness, which is probably due to the significantly larger size of the corpus.

*Table 12. Accuracy and completeness of categorization for 8, 12 and 16 clusters (mother’s speech)*

<table>
<thead>
<tr>
<th>Subject</th>
<th>8 clusters</th>
<th></th>
<th>12 clusters</th>
<th></th>
<th>16 clusters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acc*</td>
<td>Comp*</td>
<td>Acc</td>
<td>Comp</td>
<td>Acc</td>
<td>Comp</td>
</tr>
<tr>
<td>Anne</td>
<td>0.55</td>
<td>0.27</td>
<td>0.79</td>
<td>0.23</td>
<td>0.83</td>
<td>0.18</td>
</tr>
<tr>
<td>Aran</td>
<td>0.62</td>
<td>0.30</td>
<td>0.69</td>
<td>0.21</td>
<td>0.80</td>
<td>0.16</td>
</tr>
<tr>
<td>Eve</td>
<td>0.64</td>
<td>0.33</td>
<td>0.78</td>
<td>0.26</td>
<td>0.81</td>
<td>0.18</td>
</tr>
<tr>
<td>Lara</td>
<td>0.52</td>
<td>0.27</td>
<td>0.71</td>
<td>0.23</td>
<td>0.72</td>
<td>0.16</td>
</tr>
<tr>
<td>Lily</td>
<td>0.58</td>
<td>0.26</td>
<td>0.71</td>
<td>0.23</td>
<td>0.72</td>
<td>0.16</td>
</tr>
<tr>
<td>Naima</td>
<td>0.61</td>
<td>0.28</td>
<td>0.71</td>
<td>0.25</td>
<td>0.69</td>
<td>0.15</td>
</tr>
<tr>
<td>Nina</td>
<td>0.52</td>
<td>0.30</td>
<td>0.70</td>
<td>0.23</td>
<td>0.71</td>
<td>0.14</td>
</tr>
<tr>
<td>Thomas</td>
<td>0.70</td>
<td>0.35</td>
<td>0.79</td>
<td>0.26</td>
<td>0.82</td>
<td>0.19</td>
</tr>
<tr>
<td>Mean</td>
<td>0.59</td>
<td>0.29</td>
<td>0.74</td>
<td>0.24</td>
<td>0.76</td>
<td>0.17</td>
</tr>
</tbody>
</table>

* Acc = Accuracy, Comp = Completeness

Purity values, listed in Table 13, were also comparable to those obtained in Mintz (2002). In Mintz (2002), more target words (200 words) were categorized but the hierarchy was also cut into more clusters (between 20-30 clusters in total and yielding approximately 9–10 substantial clusters). In the current analysis, purity values show the same pattern as accuracies – the more clusters the higher purity. There was also little variation among the corpora. The Thomas corpus again resulted in the best purity values at all three cut-off levels.
Table 13. Purity scores for 8, 12 and 16 clusters (mother’s speech)

<table>
<thead>
<tr>
<th>Subject</th>
<th>8 clusters</th>
<th>12 clusters</th>
<th>16 clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anne</td>
<td>0.60</td>
<td>0.80</td>
<td>0.87</td>
</tr>
<tr>
<td>Aran</td>
<td>0.62</td>
<td>0.71</td>
<td>0.80</td>
</tr>
<tr>
<td>Eve</td>
<td>0.68</td>
<td>0.78</td>
<td>0.80</td>
</tr>
<tr>
<td>Lara</td>
<td>0.58</td>
<td>0.71</td>
<td>0.74</td>
</tr>
<tr>
<td>Lily</td>
<td>0.61</td>
<td>0.76</td>
<td>0.78</td>
</tr>
<tr>
<td>Naima</td>
<td>0.62</td>
<td>0.76</td>
<td>0.78</td>
</tr>
<tr>
<td>Nina</td>
<td>0.63</td>
<td>0.72</td>
<td>0.76</td>
</tr>
<tr>
<td>Thomas</td>
<td>0.74</td>
<td>0.82</td>
<td>0.83</td>
</tr>
<tr>
<td>Mean</td>
<td>0.63</td>
<td>0.76</td>
<td>0.80</td>
</tr>
</tbody>
</table>

The accuracies and purity values are not the same across categories. Two of the four categories, pronoun and auxiliary, are better clustered together than the other two categories, determiner and preposition.

If immediate preceding and following one-word contexts were able to categorize the target words accurately, function words from the same categories would be expected to be clustered together. There might be several clusters for one category but each cluster should be dominated by the members from that category. This is exactly the case when manually examining the clusters. Figure 15 displays the eight clusters that were generated using the speech of Thomas’ mother. Five of the eight clusters contain only members from one category: three pure clusters of auxiliaries and two pure clusters of pronouns.
Another cluster was dominated by prepositions with a few auxiliaries. The rest two clusters were a mix of determiner, pronouns and prepositions.

In summary, this analysis showed that function words are clustered into groups that correspond to categories very well. The limited contexts of adjacent words can be very informative for inductive analysis of categories of function words.
Figure 15. Categorization of 82 function words into 8 clusters using Thomas mother's speech

* det = determiner, pp = preposition, pro = pronoun, aux = auxiliary
Chapter 7. Distributional analysis with frequent frames

Researchers have examined a variety of distributional contexts ranging from unigrams, bigrams to whole sentences. They have also analyzed different kinds of computational mechanisms for deriving categories from these contexts (Cartwright & Brent, 1997; Mintz, 2003; Mintz, et al., 2002; Redington, et al., 1998). One distributional environment, frequent frames, has been shown to be particularly informative for categorizing words in English (Mintz, 2003). This chapter investigated using frequent frames as distributional contexts for categorizing function words.

A frame is defined as two jointly occurring words with one word intervening. The frames occurring most often in a corpus are selected as frequent frames. For example, you__it is a frequent frame in many English corpora of child-directed speech. In this frame, you and it jointly serve as contexts, and words occurring between you and it are treated as a single frame-based category. Mintz (2003) analyzed the frequent frames in speech directed to six children under the age of 2;6. Overall, his analyses showed that frequent frames are a robust cue to word categories in English. For example, Mintz reported that in one corpus, the you__it frame contains 433 tokens comprised of 93 unique words (types), all of which are verbs. He concluded that frequent frames could provide information for bootstrapping into the syntactic categories of a language as an initial basis for categorization.

In this analysis, frequent frames involving only function words were used to test whether they can accurately categorize function words.
Data

The speech of the eight mothers as in Chapter 3 was analyzed here.

Procedure

The frequent frame analysis as described in Mintz (2003, Experiment 1) was carried out on utterances from each of the eight mothers. The method was then modified to include only frames that involve a function word as the target. To conduct the frequent frame analysis, each utterance was segmented into frames. All frames with a function word as the target were tallied through the entire corpus and ranked by their frequency. The most frequent 45 frames were then selected as frequent frames for function words. Each frequent frame defines a category that consists of all the word tokens that occurred within the frame in the corpus. To evaluate the success of the categorization, category of each target function word was labeled.

To further evaluate the success of the frame-based categorization, for each corpus analyzed there is a ‘control categorization’ that randomly selects two frame-based categories and exchanges a randomly selected token from one category with a randomly selected token from the other, and repeats this process for one million trials (see (Stumper & Lieven, 2009) for a similar control procedure). This differed from the procedure used in Mintz (2003), in which control categorization was based on analyzing a scrambled corpus. The current method was simpler to implement, but may yield more accurate ‘control categories,’ since there is more homogeneity in the linguistic categories pre-selected by frequent frames than in the entire corpus. This control method thus provides a stringent comparison for testing the informativeness of frequent frames.
Quantitative evaluation

The categorization outcome is evaluated with accuracy and completeness as in Chapter 6. The difference is that in this analysis accuracy and completeness were assessed on frame-based categories rather than clusters.

Results

Table 14 lists the accuracy and completeness for each of the eight corpora as well as the two measures from the control condition where categorized words were randomly switched. The Accuracy of the experimental group is significantly higher than the control group, t(7) = 2.36, p < 4.04E-09. The completeness of the experimental group is also significantly higher than the control group, t(7) = 2.36, p < 7.8E-07. The accuracies are slightly higher than that in the first experiment of Mintz (2003). The completeness is comparable to that in the first experiment of Mintz (2003). It is interesting to observe that all eight corpora reached similar accuracy while their sizes are quite different. The Thomas corpus has about 18 times more utterances than the Eve corpus while the accuracies and completeness for these two corpora are almost the same.
Table 14. Accuracy and completeness of categorizing function words using frequent frames

(*mother’s speech*)

<table>
<thead>
<tr>
<th>Subject</th>
<th># of tokens categorized</th>
<th># of frame tokens</th>
<th>Experiment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Accuracy</td>
<td>Completeness</td>
</tr>
<tr>
<td>Anne</td>
<td>6186</td>
<td>37808</td>
<td>1.00</td>
<td>0.11</td>
</tr>
<tr>
<td>Aran</td>
<td>9824</td>
<td>54680</td>
<td>1.00</td>
<td>0.15</td>
</tr>
<tr>
<td>Eve</td>
<td>1656</td>
<td>11152</td>
<td>0.99</td>
<td>0.12</td>
</tr>
<tr>
<td>Lara</td>
<td>6603</td>
<td>48739</td>
<td>0.99</td>
<td>0.14</td>
</tr>
<tr>
<td>Lily</td>
<td>6362</td>
<td>83364</td>
<td>0.98</td>
<td>0.10</td>
</tr>
<tr>
<td>Naima</td>
<td>3639</td>
<td>52772</td>
<td>0.97</td>
<td>0.08</td>
</tr>
<tr>
<td>Nina</td>
<td>6440</td>
<td>38113</td>
<td>1.00</td>
<td>0.11</td>
</tr>
<tr>
<td>Thomas</td>
<td>26227</td>
<td>328415</td>
<td>0.99</td>
<td>0.11</td>
</tr>
<tr>
<td>Mean</td>
<td>8367</td>
<td>81880</td>
<td>0.99</td>
<td>0.11</td>
</tr>
</tbody>
</table>

*Acc = Accuracy, Comp = Completeness*

One may argue that the high accuracy was due to that very few word types in each frame-based category. By examining the target words of the frequent frames, this could be ruled out (see Table 15 for a list of the frequent frames for the Thomas corpus). Only a few frequent frames were dominated by one or two word types. For example, for the Thomas corpus, four of the 45 frames (*do__want, on__floor, going__have, going__put*) contain only one word type; 14 of the 45 frames included less than three word types. However, the rest of the frames included significant numbers of word types and tokens while maintain a high accuracy, such as *what__you, I__think, you__a, it__the*. Table 16
listed some highly accurate frequent frames for function words with substantial size from Thomas mother’s speech.

Many of the frequent frames in Mintz (2003) used function words like *a, the* and *to* as framing context. It appears to be a similar case in this analysis, although the target words here were all function words. An examination of the contents of the frequent frames for function words revealed that determiners *a* and *the* were captured by some frames but they were rarely captured by the same frame (Table 17). The frequent frames in this analysis did not take utterance boundaries as framing units. Since determiners *a* and *the* often appear at utterance beginning, it is likely that frequent frames with utterance boundaries can take advantage of this information thus improve the categorization of *a* and *the*.

This analysis shows that frequent frames are very effective in categorizing function words. Most of the frequent frames contain mainly words from the same functional category. In addition, words in many of the frequent frames demonstrated significant variety (a large number of word types in a frame). Certain post-analysis processing may be necessary to construct prototype functional categories from those frame-based categories.
Table 15. Frequent frames of function words from Thomas mother’s speech

<table>
<thead>
<tr>
<th>Frame</th>
<th>Token</th>
<th>Type</th>
<th>Primary category*</th>
<th>% primary category</th>
<th>Category (Token)</th>
</tr>
</thead>
<tbody>
<tr>
<td>what__you</td>
<td>2291</td>
<td>13</td>
<td>aux</td>
<td>100%</td>
<td>aux (2291)</td>
</tr>
<tr>
<td>there__are</td>
<td>1865</td>
<td>3</td>
<td>pro</td>
<td>99%</td>
<td>pro (1864) aux (1)</td>
</tr>
<tr>
<td>would__like</td>
<td>1822</td>
<td>7</td>
<td>pro</td>
<td>99%</td>
<td>pro (1820) aux (1) pp (1)</td>
</tr>
<tr>
<td>are__going</td>
<td>1183</td>
<td>2</td>
<td>pro</td>
<td>100%</td>
<td>pro (1183)</td>
</tr>
<tr>
<td>can__see</td>
<td>1170</td>
<td>4</td>
<td>pro</td>
<td>100%</td>
<td>pro (1170)</td>
</tr>
<tr>
<td>I__think</td>
<td>975</td>
<td>10</td>
<td>aux</td>
<td>99%</td>
<td>pro (3) aux (972)</td>
</tr>
<tr>
<td>do__think</td>
<td>965</td>
<td>2</td>
<td>pro</td>
<td>100%</td>
<td>pro (965)</td>
</tr>
<tr>
<td>on__floor</td>
<td>815</td>
<td>1</td>
<td>det</td>
<td>100%</td>
<td>det (815)</td>
</tr>
<tr>
<td>back__the</td>
<td>678</td>
<td>8</td>
<td>pp</td>
<td>100%</td>
<td>pp (678)</td>
</tr>
<tr>
<td>I__know</td>
<td>659</td>
<td>5</td>
<td>aux</td>
<td>99%</td>
<td>pro (1) aux (658)</td>
</tr>
<tr>
<td>out__the</td>
<td>641</td>
<td>11</td>
<td>pp</td>
<td>99%</td>
<td>det (2) pp (639)</td>
</tr>
<tr>
<td>you__a</td>
<td>616</td>
<td>21</td>
<td>aux</td>
<td>91%</td>
<td>det (2) aux (565) pp (49)</td>
</tr>
<tr>
<td>do__remember</td>
<td>606</td>
<td>2</td>
<td>pro</td>
<td>99%</td>
<td>det (1) pro (605)</td>
</tr>
<tr>
<td>it__the</td>
<td>592</td>
<td>20</td>
<td>pp</td>
<td>83%</td>
<td>det (3) aux (97) pp (492)</td>
</tr>
<tr>
<td>have__look</td>
<td>554</td>
<td>3</td>
<td>det</td>
<td>97%</td>
<td>det (540) pp (14)</td>
</tr>
<tr>
<td>go__the</td>
<td>520</td>
<td>12</td>
<td>pp</td>
<td>99%</td>
<td>aux (2) pp (518)</td>
</tr>
<tr>
<td>is__what</td>
<td>511</td>
<td>2</td>
<td>det</td>
<td>100%</td>
<td>det (511)</td>
</tr>
<tr>
<td>what__we</td>
<td>463</td>
<td>11</td>
<td>aux</td>
<td>100%</td>
<td>aux (463)</td>
</tr>
<tr>
<td>it__a</td>
<td>462</td>
<td>13</td>
<td>aux</td>
<td>82%</td>
<td>det (1) aux (382) pp (79)</td>
</tr>
<tr>
<td>to__a</td>
<td>453</td>
<td>8</td>
<td>aux</td>
<td>99%</td>
<td>pro (2) aux (449) pp (2)</td>
</tr>
<tr>
<td>do__want</td>
<td>446</td>
<td>1</td>
<td>pro</td>
<td>100%</td>
<td>pro (446)</td>
</tr>
<tr>
<td>have__got</td>
<td>418</td>
<td>4</td>
<td>pro</td>
<td>99%</td>
<td>pro (417) pp (1)</td>
</tr>
<tr>
<td>you__have</td>
<td>408</td>
<td>17</td>
<td>aux</td>
<td>98%</td>
<td>pro (2) aux (402) pp (4)</td>
</tr>
<tr>
<td>look__the</td>
<td>405</td>
<td>11</td>
<td>pp</td>
<td>99%</td>
<td>det (2) pp (403)</td>
</tr>
</tbody>
</table>

* det = determiner, pp = preposition, pro = pronoun, aux = auxiliary
<table>
<thead>
<tr>
<th>Frame</th>
<th>Token</th>
<th>Type</th>
<th>Primary category*</th>
<th>% primary category</th>
<th>Category (Token)</th>
</tr>
</thead>
<tbody>
<tr>
<td>why__you</td>
<td>394</td>
<td>9</td>
<td>aux</td>
<td>100%</td>
<td>aux (394)</td>
</tr>
<tr>
<td>going__have</td>
<td>351</td>
<td>1</td>
<td>pp</td>
<td>100%</td>
<td>pp (351)</td>
</tr>
<tr>
<td>did__say</td>
<td>339</td>
<td>5</td>
<td>pro</td>
<td>100%</td>
<td>pro (339)</td>
</tr>
<tr>
<td>one__the</td>
<td>335</td>
<td>10</td>
<td>pp</td>
<td>97%</td>
<td>det (2) aux (7) pp (326)</td>
</tr>
<tr>
<td>do__know</td>
<td>334</td>
<td>2</td>
<td>pro</td>
<td>100%</td>
<td>pro (334)</td>
</tr>
<tr>
<td>you__see</td>
<td>333</td>
<td>11</td>
<td>aux</td>
<td>97%</td>
<td>pro (3) aux (325) pp (5)</td>
</tr>
<tr>
<td>do__say</td>
<td>333</td>
<td>3</td>
<td>pro</td>
<td>100%</td>
<td>pro (333)</td>
</tr>
<tr>
<td>shall__put</td>
<td>332</td>
<td>2</td>
<td>pro</td>
<td>100%</td>
<td>pro (332)</td>
</tr>
<tr>
<td>then__can</td>
<td>329</td>
<td>6</td>
<td>pro</td>
<td>99%</td>
<td>det (1) pro (328)</td>
</tr>
<tr>
<td>there__a</td>
<td>326</td>
<td>12</td>
<td>aux</td>
<td>97%</td>
<td>aux (319) pp (7)</td>
</tr>
<tr>
<td>what's__matter</td>
<td>317</td>
<td>2</td>
<td>det</td>
<td>100%</td>
<td>det (317)</td>
</tr>
<tr>
<td>one__three</td>
<td>312</td>
<td>2</td>
<td>det</td>
<td>99%</td>
<td>det (310) aux (2)</td>
</tr>
<tr>
<td>going__put</td>
<td>309</td>
<td>1</td>
<td>pp</td>
<td>100%</td>
<td>pp (309)</td>
</tr>
<tr>
<td>it__be</td>
<td>308</td>
<td>9</td>
<td>aux</td>
<td>99%</td>
<td>aux (305) pp (3)</td>
</tr>
<tr>
<td>when__were</td>
<td>302</td>
<td>2</td>
<td>pro</td>
<td>100%</td>
<td>pro (302)</td>
</tr>
<tr>
<td>is__the</td>
<td>298</td>
<td>13</td>
<td>det</td>
<td>71%</td>
<td>det (213) pro (1) pp (84)</td>
</tr>
<tr>
<td>are_saying</td>
<td>296</td>
<td>2</td>
<td>pro</td>
<td>100%</td>
<td>pro (296)</td>
</tr>
<tr>
<td>oh__you</td>
<td>294</td>
<td>11</td>
<td>aux</td>
<td>96%</td>
<td>det (9) pro (1) aux (284)</td>
</tr>
<tr>
<td>I__see</td>
<td>294</td>
<td>7</td>
<td>aux</td>
<td>100%</td>
<td>aux (294)</td>
</tr>
<tr>
<td>you__want</td>
<td>287</td>
<td>5</td>
<td>aux</td>
<td>100%</td>
<td>aux (287)</td>
</tr>
<tr>
<td>look_that</td>
<td>286</td>
<td>4</td>
<td>pp</td>
<td>100%</td>
<td>pp (286)</td>
</tr>
</tbody>
</table>

* det = determiner, pp = preposition, pro = pronoun, aux = auxiliary
Table 16. Some highly accurate frequent frames for function words with substantial size

*(Thomas mother)*

<table>
<thead>
<tr>
<th>Frame</th>
<th>Word categorized (Token)</th>
</tr>
</thead>
<tbody>
<tr>
<td>you__have</td>
<td>do (816) are (600) did (289) would (191) have (183) can (143) were (51) don't (7) will (5) should (2) could (2) had (1) might (1)</td>
</tr>
<tr>
<td>would__like</td>
<td>you (1811) he (5) she (2) we (1) I (1) be (1) to (1)</td>
</tr>
<tr>
<td>I__think</td>
<td>don't (933) didn't (25) do (6) I (3) did (2) should (2) shall (1) might (1) could (1) can (1)</td>
</tr>
<tr>
<td>back__the</td>
<td>in (281) of (217) to (78) on (72) from (17) inside (6) at (6) over (1)</td>
</tr>
<tr>
<td>out__the</td>
<td>of (538) in (54) on (14) for (13) at (11) to (5) over (2) from (1) during (1) that (1) what (1)</td>
</tr>
<tr>
<td>go__the</td>
<td>to (316) in (108) on (62) over (10) under (6) at (5) inside (4) of (4) near (2) by (1) is (1) do (1)</td>
</tr>
<tr>
<td>what__we</td>
<td>have (128) do (102) did (82) shall (52) can (24) would (8) were (8) could (4) should (3) didn't (1)</td>
</tr>
<tr>
<td>to__a</td>
<td>have (273) be (154) do (20) me (2) on (1) has (1) is (1) in (1)</td>
</tr>
<tr>
<td>you__have</td>
<td>can (201) don't (40) must (35) could (31) might (24) didn't (24) would (13) should (12) do (6) won't (5) will (4) shall (4) to (4) you (2) are (1) have (1) did (1)</td>
</tr>
<tr>
<td>look__the</td>
<td>at (327) in (40) on (16) for (6) inside (6) under (3) over (2) what (2) above (1) near (1) by (1)</td>
</tr>
<tr>
<td>why__you</td>
<td>are (171) don't (108) do (49) did (28) have (25) were (6) would (3) didn't (2) won't (2)</td>
</tr>
<tr>
<td>one__the</td>
<td>of (211) in (35) on (34) for (25) at (14) from (6) is (4) has (3) that (2) to (1)</td>
</tr>
<tr>
<td>you__see</td>
<td>can (252) could (21) didn't (19) don't (12) won't (9) might (7) to (5) I (3) did (2) will (2) should (1)</td>
</tr>
<tr>
<td>there__a</td>
<td>was (160) is (72) are (38) isn't (24) were (16) wasn't (5) in (4) for (2) been (2) have (1) of (1) should (1)</td>
</tr>
</tbody>
</table>
Table 17. Frequent frames of function words that categorize a or the (Thomas mother)

<table>
<thead>
<tr>
<th>Frame</th>
<th>a or the categorized (Token)</th>
</tr>
</thead>
<tbody>
<tr>
<td>on_floor</td>
<td>the (815)</td>
</tr>
<tr>
<td>have_look</td>
<td>a (529)</td>
</tr>
<tr>
<td>it_a</td>
<td>a (1)</td>
</tr>
<tr>
<td>what's_matter</td>
<td>the (315)</td>
</tr>
<tr>
<td>is_the</td>
<td>the (4)</td>
</tr>
</tbody>
</table>
Chapter 8. General discussion

This chapter discusses the implications of the results from this study.

**Determiner overlap**

Chapter 3 demonstrated that determiner overlap is positively correlated with sample size. It implicates that adult overlap was under-estimated by using small samples in the previous studies. If a larger sample of a mother’s speech were available, the observed overlap would be higher. While overlap of adults was used as the baseline for measuring young children’s determiner productivity, the results obtained are likely to be unreliable because of the correlation between overlap and sample size. The analysis in Chapter 3 also showed that overlap keeps increasing even the sample is composed from a very dense corpus. This suggests that it would be better not to rely on direct comparison between adults’ overlap and children’s overlap since it is impossible to obtain a corpus of every possible noun phrase one knows.

Chapter 4 proposed a better measure to overcome this sample size problem – deviation of actual overlap from expected overlap. Expected overlap is computed from the nouns in a sample and the probability of a noun occurring with a particular determiner. Therefore, expected overlap is a sample-specific measure and is independent of sample size. It is solely determined by the frequency distribution of nouns in a sample. The results in Chapter 4 showed that the deviation of an adult’s actual overlap from expected overlap is quite small. Actual and expected overlap was highly correlated for all eight mothers.
Theories have different predictions on the developmental pattern of the determiner category. Nativist theories claim that children have knowledge of abstract categories since very early on. Thus, they would predict that children should use determiners in the same way as adults when they start to produce multi-word utterances. On the contrary, constructivists would predict that children first use determiners in memorized chunk of phrases (i.e., limited-scope formula) and gradually progress to an abstract determiner category. Hence, they would predict large deviation from expected overlap for young children around two years old and that the deviation becomes smaller over time and finally reaches adult level. The analysis of eight English-learning children found that the children’s deviations from expected overlap were quite small and not significantly different from their mothers’ deviations from 1;11 to 3;0. No systematic change in children’s deviations was observed. The results are in favor of nativist views on the acquisition of the determiner category. It is also in line with findings from experimental work that tested whether infants can generalize from one article to another (Höhle, et al., 2004; Shi & Melançon, 2010).

To test the cross-linguistic validity of the proposed method, Chapter 5 investigated the developmental pattern of the German determiner category because German has a more complex determiner system than English. Nativists would predict a similar developmental pattern of the determiner category between German-learning and English-learning children. Constructivists would predict a longer or at least a similar learning curve (decrease of deviation). The result of one German child showed the same pattern of the English children, deviations were not significantly different from the adults’
deviations. This result verified the cross-linguistic validity of the method. It is also in favor of nativist views on knowledge of the determiner category.

Yang (2011a, 2011b) proposed a similar method for computing sample-specific expected overlap while the method in this dissertation was developed. Yang’s method also uses the nouns in a sample to compute expected overlap. He obtained a rather accurate expected overlap for several corpora. However, the probability distribution of nouns occurring with determiners ($p$) was always 1/3 for every noun. In addition, the expected overlap was calculated for an entire corpus rather than over time. The current analysis therefore provided a more accurate computation of expected overlap and a more fine-grained view of the developmental pattern of determiners. The independent creation of two similar methods at the same time in itself affirms the move from comparing actual overlap to using expected overlap as the baseline.

Determiners, such as *a/an* and *the* for English, were selected for analysis here because they are the most frequent words in English and German. Nevertheless, the methods used in this dissertation should be applicable to the full range of determiners. It could also be applied to other functional categories and many other languages. The developmental pattern of the determiner category would be indicative to other functional categories. There may be some performance-related differences between determiner and other functional categories. For example, the size of input for other functional categories would be much smaller than determiner according to Zipf’s law (Zipf, 1949). It would not be surprising that children’s acquisition of those functional categories is later than
determiner. However, according to nativist views, it is expected that they would use those function words in the same way as adults do once they start producing them.

Results obtained in the first part of this dissertation could be explained by claiming innately endowed functional categories (i.e., nativist views on acquisition of functional categories) or some kind of powerful learning mechanism. The time point of knowledge of an abstract functional category is moved even earlier to the third year. The origin of the knowledge, which could be due to innate proto-categories or just some mental capacity that enables fast learning and word categorization, is still unclear.

If children are not predisposed with such knowledge, a crucial question is how young children acquire an abstract functional category in limited amount of time-first two years. Dewar & Xu’s (2010) demonstration of 9-month-olds’ capability of inductive learning through overhypothesis formation provides an interesting alternative to this question. In their experiments, 9-month-old infants were able to form second-order generalization about categories even the evidence was limited to a few objects in some categories. With such a powerful inductive learning mechanism, infants should be able to quickly acquire abstract functional categories inductively with limited input. However, it remains to be tested whether infants use such learning mechanism in the domain of acquisition of grammatical categories.

**Distributional analysis**

Chapter 6 showed that immediate one-word contexts were reasonably good in grouping common English function words into clusters of function words from the same categories. Chapter 7 tested categorizing the same set of English function words using a
different distributional context – frequent frames. Frequent frames for function words were extremely accurate in categorization while the completeness was comparable to the vanilla frequent frames. Many of the frequent frames for function words captured a variety of word types from the same grammatical categories. The Thomas corpus has about 18 times more utterances than the Eve corpus while the accuracies and completeness for these two corpora were almost the same. This suggests that small input data is enough to achieve accurate categorization of function words, which could explain why young children use determiners in the same way as adults when they just start to combine words together and produce multi-word utterances.

Since even very young children can use phonological and acoustic information to distinguish lexical and functional items (Shi, et al., 1998), they may be able to attend specifically to the contexts of function words in the input when they are learning their native languages. With that assumption, distributional information in the input, which is discoverable using immediate one-word contexts or frequent frames, would be a good source for an initial coarse grouping of function words. Function words from the same grammatical category would be grouped together. However, function words from a given grammatical category may be distributed in several groups. Some kind of joining algorithm can be useful to combine those groups into larger proto-categories. For example, Mintz (2003) suggested a method for grouping frame-based categories according to the overlap of word types in the groups. Two frames that have many identical word types from a category (e.g., preposition) will be joined to form a larger category.
The cross-linguistic validity of frequent frames has been tested with two typologically very different languages – German and Turkish – at both word and morpheme levels (Wang et al., 2010). Frequent morpheme frames perform quite well in those languages with rich morphologies. The distributional context analyzed in Chapter 7, frequent frames for function words, should be applicable to those rich-morphology languages as well. When the frequent frames for function words operate at the morpheme level, it is expected that they would capture both grammatical stems and morphemes.

An implication of the results from the two analyses is that even if semantic information is not available to children, acoustic and distributional cues in the input would be enough to bootstrap children from scratch to some kind of prototypical categories. At last, this is favorable in terms of simplicity of theories because the bootstrapping of both lexical and functional categories can be accounted for under the distributional approach.

**Importance of sample size for corpus analysis**

There has been growing awareness in the field that sample size is critical for reliably measuring children’s linguistic productivity using corpus analysis (Baayen, 2001; Rowland & Fletcher, 2006; Yang, 2011a, 2011b). For general descriptive statistics of a sample of text, Baayen (2001) has shown that most of the summary statistics, such as sample mean frequency and parameters of the zeta (Zipf) and lognormal distributions systematically depend on sample size. In the context of language acquisition, Rowland et al. (2008) summarized the problem as follow
“... the accuracy with which any one sample assesses productivity is affected by sample size, by the frequency statistics of the language, and by the vocabulary size of the child. Importantly, even collecting much bigger samples will not overcome these problems. There will still be an impact of sample size and frequency statistics on measures of productivity, no matter how many utterances are collected. In addition, children’s limited vocabulary knowledge will still affect the range and variability of the syntactic structures they produce. In order to attribute limited productivity to children reliably it is important to control for the effect of sample size and vocabulary, while taking into account the frequency statistics of the language.”

Following the same reasoning, the first analysis in this dissertation clearly demonstrated the sample size problem. The second and third analyses circumvented the sample size problem by comparing actual overlap to expected overlap of the same sample, instead of a direct comparison between an adult’s overlap and a child’s overlap.

**Conclusion**

This study demonstrated that young children around 24 months old possess abstract knowledge of functional categories. Functional categories are available in children’s grammar since very early on. Function words are late in production is not because of lack of knowledge of functional categories but probably due to performance-related factors. Functional categories are used by young children in the same way as adults do. The results are in favor of nativist views or a strong learning mechanism. With the knowledge of functional categories, children can start acquire syntactic constituents.
like NP and VP (because functional categories are the heads of the projections) and more complex syntactic structures.

Distributional information in the input was able to accurately categorize function words with the help of acoustic cues. Therefore, it could be a primary source for initial categorization of function words and bootstrapping of functional categories.

The sample size problem demonstrated here and the proposed method for measuring linguistic productivity in children are valuable for corpus studies in future language acquisition research. The method may be applied to study children’s productivity of other grammatical categories and linguistic structures.
References


English in Secondary Schools (pp. 20-37). New South Wales, Australia: English
Teachers' Association.
Infants' Speech Processing: The Role of Determiners in the Syntactic Categorization of
Lexical Elements. Infancy, 5(3), 341-353. doi: 10.1207/s15327078inf0503_5
Hsu, A. S., Chater, N., & Vitányi, P. M. B. (2011). The probabilistic analysis of language
acquisition: Theoretical, computational, and experimental analysis. Cognition, 120(3),
data. Journal of Child Language.
592-609.
10.1515/cogl.2009.022
Lawrence Erlbaum Associates.
ontogenesis and representation of syntactic categories. In K. E. Nelson (Ed.), Children's
Cognition, 90(1), 91-117.
coherence hypothesis: Cross-linguistic evidence in language acquisition. Cognitive
Naigles, L. R. (2002). Form is easy, meaning is hard: resolving a paradox in early child language.
Cognition, 86(2), 157-199. doi: 10.1016/s0010-0277(02)00177-4
Research in Child Development, 38(1-2, Serial No. 149).
Foris.
Subsystems with Different Sensitive Periods. Cerebral Cortex, 2(3), 244-258. doi:
10.1093/cercor/2.3.244


