NOUNS-FIRST, VERBS-FIRST AND COMPUTATIONALLY-EASIER FIRST: A PRELIMINARY DESIGN TO TEST THE ORDER OF ACQUISITION

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF INFORMATICS

OF

THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

ENES AVCU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF

MASTER OF SCIENCE

IN THE DEPARTMENT OF COGNITIVE SCIENCE

AUGUST 2014

NOUNS-FIRST, VERBS-FIRST AND COMPUTATIONALLY-EASIER FIRST: A PRELIMINARY DESIGN TO TEST THE ORDER OF ACQUISITION

Submitted by **ENES AVCU** in partial fulfillment of the requirements for the degree of **Master of Science in Cognitive Science, Middle East Technical University** by,

Prof. Dr. Nazife Baykal _____
Director, **Informatics Institute**

Prof. Dr. H. Cem Bozşahin _____
Head of Department, **Cognitive Science**

Prof. Dr. H. Cem Bozşahin _____
Supervisor, **Cognitive Science, Middle East Technical University**

Prof. Dr. Deniz Zeyrek _____
Co-advisor, **Cognitive Science, Middle East Technical University**

Examining Committee Members:

Prof. Dr. Aylin C. Küntay _____

Psychology, Koç University

Prof. Dr. H. Cem Bozşahin _____

Cognitive Science, Middle East Technical University

Prof. Dr. Deniz Zeyrek _____

Cognitive Science, Middle East Technical University

Assoc. Prof. Dr. Annette Hohenberger _____

Cognitive Science, Middle East Technical University

Dr. Ceyhan Temürcü _____

Cognitive Science, Middle East Technical University

Date:06.08.2014

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : Enes AVCU

Signature : _____

ABSTRACT

NOUNS-FIRST, VERBS-FIRST AND COMPUTATIONALLY-EASIER FIRST: A PRELIMINARY DESIGN TO TEST THE ORDER OF ACQUISITION

AVCU, Enes

M.S., Department of Cognitive Science

Supervisor: Prof. Dr. H. Cem Bozşahin

Co-advisor: Prof. Dr. Deniz Zeyrek

August 2014, 62 pages

The primary accounts for early lexical differences can be broken down into two distinct theoretical positions that either defend early noun acquisition or provide evidence that challenges this account. This work is trying to bring a computational perspective to the problem of early lexical acquisition of words. It is a preliminary investigation to see if the underlying mechanism relates to computational complexity by which short, frequent and unambiguous words are supposed to be acquired first; and long, ambiguous or infrequent words (including nouns) are predicted not to be acquired early. Our database consists a longitudinal data of three Turkish children between 8 months and 36 months. We conducted three analyses to test this; (1) in frequency analysis, we compared the type token ratios and the number of types and tokens of nouns and verbs in both child directed speech and child speech; (2) in ambiguity analysis, we examined the role of social and attentional cues on word learning; and (3) in phonological analysis, we measured the effect of word length on learning of words. Results revealed that most frequent words did not prove any noun predominance; in place of this, the usage rates of verbs was close to nouns and sometimes much more than that. Furthermore, caretakers did not have any bias to nouns or object names on the contrary there was a preference to verbs at some parts. The high rate of verbs in child speech also challenged the noun-first view. Ambiguity analysis showed that social and attentional cues in the natural language's context were important factors for word learning.

Therefore, disambiguated words in the context were the most frequent words in child speech. Phonological complexity analysis indicated that word length affected the infant's ability of word learning. Thus, short words were more advantageous when compared to long words.

Keywords: Language Acquisition, Computational Complexity, Natural Partitions Hypothesis, Early Lexical Acquisition

ÖZ

İSİM Mİ, FİİL Mİ, BİLİŞİMSEL OLARAK BASİT OLAN MI ÖNCE EDİNLİR? ;
EDİNİM SIRASINI ÇÖZÜMLEMEK İÇİN ÖNCÜ BİR ÇALIŞMA

AVCU, Enes

Yüksek Lisans, Bilişsel Bilimler Bölümü

Tez Danışmanı: Prof. Dr. Cem Bozşahin

Eş Danışman: Prof. Dr. Deniz Zeyrek

Ağustos 2014, 62 sayfa

Erken sözcüksel farklılıklar ile ilgili başlıca çalışmalar; erken isim edinimini savunanlar ve bu tutuma karşı çıkan bulgular üretenler olarak iki farklı kuramsal temele dayandırılabilir. Bu çalışma kelimelerin erken sözcüksel edinimi problemine bilişimsel bir bakış açısı getirmeyi amaçlamaktadır. Çalışma sözcüksel edinim probleminin altında yatan mekanizmanın bilişimsel karmaşıklık ile ilgili olup olmadığını sorgulayan öncü bir araştırmadır. Bu şekilde kısa, sık ve belirsiz olmayan kelimelerin önce edinileceği; uzun, muğlâk ve sık olmayan kelimelerin daha sonra edinileceği öngörülmektedir. Çalışma üç Türk bebeğin 8 aydan 36 aya kadar olan boylamsal verilerine dayanır. Bu kapsamda üç farklı analiz yapılmıştır; (1) frekans analizinde çocukların maruz kaldığı ve ürettikleri dildeki tür-türce oranı ve sayıları karşılaştırılmıştır; (2) belirsizlik analizinde, sosyal ve ilgisel ipuçlarının kelime öğrenimi üzerindeki rollerine bakılmıştır; (3) sesbilimsel karmaşıklık analizinde ise kelime uzunluğunun kelime öğrenme üzerindeki etkisi incelenmiştir. Sonuçlar dilde sık olarak kullanılan kelimelerin herhangi bir isim yoğunluğuna kanıt olamayacağını ve fill oranının isim oranına yakın olduğunu göstermektedir. Ayrıca çocukların maruz kaldıkları dilde isimlere karşı bir eğilimden çok fiillere karşı bir eğilim olduğu saptanmıştır. Belirsizlik analizi sonuçları doğal dil bağlamında bulunan sosyal ve ilgisel ipuçlarının kelime öğreniminde önemli faktörler olduğunu göstermektedir. Bu sebeple, bağlam tarafından açıklığa kavuşturulan kelimelerin çocuklar tarafından sıklıkla kullanıldığı görülmüştür. Sesbilimsel karmaşıklık analizi ise kelime uzunluğunun kelime öğrenmeyi etkilediğini ve kısa kelimelerin uzun kelimelere göre daha kolay edinildiğini göstermiştir.

Anahtar kelimeler: Dil Edinimi, Bilişimsel Karmaşıklık, Doğal Ayrışma Hipotezi, Erken Sözcüksel Edinim

To My Wife…

ACKNOWLEDGMENTS

I wish to thank, first and foremost, Prof. Dr. Cem Bozşahin for the guidance he provided me throughout the whole study. This thesis would not have been possible without his invaluable comments, constructive advices and friendly conversations. I would like to gratefully and sincerely thank my thesis co-advisor Prof. Dr. Deniz Zeyrek for having aroused my interest in the area of Language Acquisition.

I would want to thank to Prof. Dr. Aylin Küntay who shared with me her data that were crucial for the present study. Furthermore, I would like to thank Assist. Prof. Dr. F. Nihan Ketrez and Assoc. Prof. Dr. Annette Hohenberger for helping me to expand my knowledge about the procedure.

I would also want to thank to Veysel Atalay and R.Onur Karadeniz for their precious support and to my brother, Furkan Avcu for always being there when I needed technological support.

I owe my deepest gratitude to my wife, mother, father, sister and brothers who were always there, supporting and encouraging me.

Lastly, this study would not have been possible without the infants and their parents who participated Prof. Dr. Aylin Küntay's research.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

CDS	Child Directed Speech
CS	Child Speech
TTR	Type Token Ratio
NPH	Natural Partition Hypothesis
AKT-T	Aylin Küntay's Project Turkish Fragment
CLAN	Computerized Language Analysis
CHILDES	Child Language Data Exchange System
WDLEN	Word Length

CHAPTER I

INTRODUCTION

Human beings, as a unique species, have a good deal of characteristics that separate them from other creatures. At first glance, some of our abilities such as thinking, talking or writing may seem to be simple characteristics. But they do differentiate us from our closest relative- chimpanzees. Among these abilities, the concept of language has a surpassing role which is the keystone of humanity and among one of the wonders of the world. Through this unique gift we accomplish basic needs and can produce about a hundred million different utterances. But where does the knowledge of language come from? How does an infant break into language? How does the knowledge of language emerge in the early period?

According to the theory of Universal Grammar (Chomsky 1975, 1981, 1986), infants are innately endowed with a system of richly structured linguistic knowledge, which makes linguistic input accessible to them. So they know a great deal about language-much more than what appears in what they initially say. However, we cannot assume exactly what infants know about language because we cannot observe language directly, or get inside children's brains.

The development of language, as its born, demonstrates such magnificent milestones that an infant's calls and cries convey a fixed set of messages even the noises they make mean an unlimited number of different meanings. Already at birth, infants start processing the linguistic stimuli presented in their natural environment. Acoustic cues such as syllables, phonemes and words express many things to them. This means that they are biased to pay attention to speech stimuli-first with certain ways they perceive speech and then direct the acoustic stimuli onto their native language's phonological system. At four days infants display amazing capacities; for instance, they can differentiate their native language from a foreign language (Mehler et al. 1988). However, this does not mean that they start to be interested in language only around at this age. Babies a couple hours old are able to distinguish between a foreign language and the sounds of their native language. They start absorbing the language while still in the womb. According to Moon et al (2012), during the last 10 weeks of pregnancy, unborn babies listen to their mother talk and at birth can demonstrate [evidence of] what they've heard.

At 6-8 months infants start to babble, a form of linguistic production characterized by the use of a subset of possible sounds found in natural languages (Locke, 1983). Although Jakobsen (1968) considered babbling a prelinguistic phenomenon unrelated to the acquisition of language, this view is no longer valid and babbling is now considered a fundamental step in the development of language. Thanks to the babbling phenomena, infants take a chance of trying their articulatory capacities and practice the sounds of their language, leading up to the production of words. Meanwhile, at around 10-12 months, while they are still babbling, infants produce their first meaningful linguistic expressions-first

words (Huttenlocher 1974; Benedict 1979; Oviatt 1980). This is a remarkable outcome of a compelling path that begins at around 6 months, when infants pay attention to the natural language input.

Infants, when learning words, perform two independent tasks: (1) segmenting the speech stream into word [or phrase] sized units and (2) matching these units with meanings (Peters 1983; Morgan 1986; Christophe and Dupoux 1996). A question that arises here is how do infants know that words identify objects or describe actions? Actually, we don't know whether they identify words with objects yet. But they must surely be trying to figure out which words go with which meanings. One hypothesis is that this mapping can occur through a word to world mapping procedure by which infants learn the meaning of a word by observing external and contextual cues for its use (Inhelder & Piaget, 1964; Bruner, 1978; Baldwin, 1991; Markman, 1994). However, it can hardly account for how infants learn the meaning of verbs since it is difficult to figure out the meaning of a verb by just looking at the context. Therefore a sentence to world mapping procedure handles the problem by looking at the syntactic context, in which syntactic structures narrow down the possible interpretations in order to determine the meaning of a verb (Brown, 1957; Taylor & Gelman, 1989; Bloom, 1994b; Landau & Gleitman, 1985).

An important generalization of lexical acquisition studies is that regardless of the culture in which they are brought up, infants' early lexicon consists almost exclusively of nouns (Bates, Dale & Thal, 1995; Caselli et al. 1995; Gillette et al. 1999). Verbs emerge later, and for a while they remain a minority. The disadvantage of verbs behind nouns can be attributed to the above explanations about the meaning of nouns and verbs being learned in different ways. When learning the meaning of nouns, infants, to some extent, rely on word to world mapping procedure by which the word is directed to an object to which it refers. Although when it comes to verbs, the external cues do not always back up the meaning of verbs since there may be a temporal gap between the utterance of a verb and its denoting action. Furthermore, actions are temporally extended mental objects; a fact that renders the mapping procedure ambiguous. It may not be clear to the child to which part or which aspects of this action the verb may refer.

One major issue on infants' early words focuses on what categories of words infants produce, with a controversy centered on whether they display a universal, innately endowed noun bias in early lexical acquisition (Gentner, 1982; Gentner & Boroditsky, 2001; Gillette, Gleitman, Gleitman & Lederer, 1999; Macnamara, 1982). A classic and influential study which is on the distributions of nouns and verbs in early vocabulary by Gentner (1982) is the reason of cross linguistic interest in this question. Gentner (1982) studied six typologically different languages (English, Turkish, German, Kaluli, Mandarin, and Japanese) and found a noun preference. It is stated that (1) nouns are learned earlier and more easily than other parts of speech and (2) they are more frequent than verbs in the early lexicon. Later, Gentner and her colleagues gathered their claims around *Natural Partitions Hypothesis,* which claims that there is a "preexisting perceptual-conceptual distinction between concrete concepts such as persons or things and predicative concepts of activity, change of state or causal relations" (Gentner, 1982).

However, subsequent cross-linguistic research on early composition of the lexicon has shown that children's early vocabularies exhibit a large variety of word categories; in fact, even in English, nouns are not always the dominant part of speech (Bates, Bretherton & Snyder, 1988; Bloom, Tinker & Margulis, 1993). Kauschke & Hofmeister (2002) found that early words range from relational words and personal social words to onomatopoetic words. Moreover, Bornstein et al. (2004) showed that there is not any preference for nouns over verbs in the first 50 words of the children in the language they studied. Apart from these, Childres, Vaughan and Burquest (2007) found not only any preference for nouns in the early vocabulary of Ngas children (a language spoken in Nigeria) but also a significant verb preference in comprehension. In addition, studies on a variety of non-Indo-European

languages such as Mandarin (Tardif, 1996; Tardif et al., 1997), Korean (Choi & Gopnik, 1993, 1995), the Mayan languages Tzeltal (Brown, 1998) and Tzotzil (de Léon, 1999) have shown that there is an equal proportion of nouns in children's vocabulary in Korean or even a preference for verbs in Mandarin, Tzeltal, Tzotzil in the early speech of children. Therefore it can be concluded that there is no universal noun bias that favors a higher frequency of nouns over verbs in children's early speech.

Following ZettleMoyer & Collins (2005), Steedman & Hockenmaier (2007), Coltekin & Bozsahin (2007) and Bozsahin (2012), this work utilizes computational perspective to the problem of early lexical acquisition of words. It is a preliminary investigation to see if the underlying mechanism relates to computational complexity by which short, frequent and unambiguous words are expected to be acquired first; and long, ambiguous or infrequent words (including nouns) are predicted to be acquired later. This study assumes there is a computational bias in the infants' mind toward frequent, short and unambiguous strings because these aspects can be shown to ease the task computationally.

Outline of the thesis

Chapter 2 comprises a detailed literature review which, in general, consists of elaborations on noun bias in early word learning and counter evidence to it. First, noun-first view and foundations of *Natural Partition Hypothesis* will be presented by referring to both cross-linguistic and experimental work. Secondly, studies that challenge noun-first view will be detailed. Evidence from comparative cross-sectional and longitudinal studies will be presented. Lastly, issues related to the assumptions of present study, namely the computational complexity assumption on early word learning will be discussed. Chapter 3 will present details of the first analysis's method: information on participants, procedure, and results. In Chapter 4, the second analysis's method: information on participants, procedure, and results will be presented. Chapter 5 will present details of the third analysis's method and results. In Chapter 6, the findings of the analyses will be evaluated in detail and compared with respect to the existing literature and limitations will be discussed. Lastly, Chapter 7 is a conclusion of the present study in which possible further cross sectional and longitudinal ideas will be mentioned, as well.

CHAPTER II

LITERATURE REVIEW

2.1. Noun First View and Natural Partition Hypothesis

Although, the ability of language contains different components, the core component is the lexicon since the development of grammar in the acquisition of native language is largely based on the development of vocabulary (Bachmann, 1990; Bachman and Palmer, 1996). In the studies of early lexicon, with a long historical tradition, things denoted by nouns are ontologically different and far more concrete from things denoted by verbs. Diary studies by Tracy (1893) and Dewey (1894) have demonstrated that nouns enter the lexicon before verbs. William Stern (1851-1938) who conducted another diary study reported that; until three months, the input that the child is exposed, consists entirely of nouns; between one and eight months, 78% nouns and 22% verbs; and between one and eleven months, 63% nouns, 23% verbs, and 14% adjectives (Chukovsk, 1968). In line with these findings, Macnamara (1972) claimed that acquisition order is determined by the cognitive difficulty such that names for entities comes before names for states and actions.

Many more studies have been conducted from that time which basically claim (1) nouns are learned before verbs (Huttenlocher, 1974); (2) early semantic functions are referential (Nelson, 1973); (3) early nouns mostly refer to people, foods and body parts (Brown, 1973); (4) and verbs constitute just 16% of the first ten words (Greenfield & Smith, 1976). In addition, in order to rule out the possibility of children being exposed to more nouns, some manipulated studies included artificial words for nouns and verbs but children used artificial words for nouns before that of for verbs. Gentner (1982, p. 324) argued,

> "There are in the experiential flow certain highly cohesive collections of percepts that are universally conceptualized as objects, and . . . these tend to be lexicalized as nouns across languages. Children learning language have already isolated these cohesive packages—the concrete objects and individuals—from their surroundings."

According to Goldin-Meadow, Seligman and Gelman (1976), children's early lexicon is composed mostly of nouns. This claim is affirmed by many other studies in Italian (Caselli et al., 1995), Spanish (Bates et al., 1993) and French (Bassano, 2000). Furthermore, Bornstein et al. (2004) suggested that in American English, Spanish, Italian, French, Dutch, Korean and Hebrew children's early vocabularies contain more nouns than verbs. The method used in these large scaled studies is merely counting the nouns and verbs; however, experimental studies also show the relative complexity of verbs and nouns (Snedeker & Gleitman, 2004). Bassano et al. (1998) and Bassano (2000) showed that the proportion of nouns in a longitudinal study of one child was higher than the proportion of verbs.

Although evidence of the noun advantage generally comes from the languages in which the category of the verb has a disadvantaged position such as English-in the middle of a sentence-, noun dominance has also been found for languages where verbs have advantaged positions such as Chinese and Japanese-at the end of a sentence (Tardif, 1996). Imai et al. (2005) and Imai et al. (2008) revealed that while children both in noun friendly and verb friendly languages referred novel nouns to objects, they were not successful referring novel verbs to actions. Also in some verb friendly languages, children, in order to find the meaning of a new verb, use cues from surrounding nouns such as the number and animacy of nouns (Arunachalam & Waxman, 2011). If these nouns are dropped, infants may have difficulty in learning the meaning of new verbs.

Gentner (1982) put together all of the above assumptions under the title of *Natural Partition Hypothesis* which revealed that (1) a preexisting perceptual-conceptual distinction between concrete concepts (persons or things) and predicative concepts determines the distinction between nouns and verbs; and (2) nouns are conceptually more basic than other word types since the mapping between the language and world is more tangible for nouns. It is assumed that the world is naturally divided into ontological objects that are concrete and activity objects that are abstract. *Natural Partitions Hypothesis*'s universalist position is supported by findings from various languages that nouns predominate over predicate terms. Supalla & Newport (1978) claimed that noun verb distinction also exists in American Sign Language; and in pidgins and creoles (Slobin, 1975). On the contrary, *Natural Partition Hypothesis* contradicts with Whorf's (1956) Linguistic Relativity principle which states that language makes the distinction between word types and every language does it differently. However, an equally defendable position might be that the early acquisition of words belonging to the noun category stems from the way in which nouns are treated in language.

After aspects of verb learning having come to the fore, Gentner & Boroditsky (2001) and Snedeker & Gleitman (2004) proposed that the noun-verb distinction is not a distinction between the noun or verb category but between concrete concepts and abstract concepts. For example, nouns such as *shoe*, *car* and verbs such as *kiss*, *eat* can be viewed easier than nouns such as *passenger* and verbs such as *believe*, *imagine* (Hirsh-Pasek & Golinkoff, 2006). This assumption also argues that visible actions and concrete objects are ought to be acquired first. However, Gentner and colleagues still argue that since verb meaning is dependent on the linguistic system, verbs will be harder to learn; and even learned verbs may not be used in the adult sense (Gallivan, 1988; Theakston, Lieven, Pine & Rowland, 2002).

2.2. Verbs

Until recently, how infants learn the meaning of an action has lagged behind how they learn the meaning of an object. With actions, we transmit our thoughts and beliefs. From the toddlers' talking about their own actions (e.g., look Mom!) to the adult talking about the actions of others, action is in the center of life and central to the language. When talking about the relationships between the objects and people around us, we make a functional relation from a verb to an object. This functional relationship determines the way of talking and the order of words in a sentence. The argument structure of a sentence is also determined by verbs therefore they are the building blocks of grammar. Without the verbs, we cannot tell what is going on in an utterance like *John loves Mary*. From the point of syntax, a verb is an entity that takes a subject and object (sometimes both) [as its (external and internal) argument]. They are said to be actions or states of being. For example, in (1a) swimming is a verb, while in (1b), swimming is a noun. Because of its syntactic behavior the same word is in two different categories However, semantically, *swimming* in both sentences describe an action. The true difference is that: in (1a) swimming combines with an auxiliary verb "*be*", while in (1b) it combines with a determiner *the.*

(1) a. Bill is swimming
b. The swimming is awesome

From the semantic point of view, verbs encode events, states, or conditions of existence, processes or actions (Frawley, 1992 p. 141). For example, in (2a) *fear* is not only an action but also a state of being and in (2b) *afraid* is a state of being, too. Two words are both states of being but (2a) is a verb and (2b) is an adjective.

(2) a. Jack fears spiders.
b. Jack is afraid of spiders.

It can be understood from the above examples that verbs pose great difficulties for infants both from semantic and syntactic points of view. Moreover, the verb stem can take different morphological forms such as tense, number, gender and person and they can be passivized in many languages and this adds another difficulty.

2.3. Verbs as Complex Entities

> "Besides being able to be mistaken and to make mistakes verbs can change to look like themselves or to look like something else, they are, so to speak on the move and adverbs move with them and each of them find themselves not at all annoying but very often very much mistaken. That is the reason any one can like what verbs can do " (Stein, 1957 p. 212 as cited in Gentner, 1978)

Gentner (1978) mainly claims that (1) verbs are acquired more slowly than nouns; (2) verbs enter lexicon later than nouns; (3) in the first few years, the rate of vocabulary increase is higher for nouns than verbs; (4) until the age of eight, the meanings of common verbs are not fully acquired; and (5) verbs are used by adults more broadly. To Gentner and many other colleagues, nouns are dominant in children's early speech and more importantly they are acquired quicker and easier than other word types (Fenson et. al, 1994; Goldin-Meadow, Seligman, Gelman, 1976). According to Rosch (1973, 1975) the meanings of nouns are limited by nature but that of verbs depend on abstract concepts and not constraint by physical world. The category of the verb is functional; which means there must be an agent and an action, both of which relates to each other.

Verbs are difficult to learn for word learners because they pose some challenges and they relate to actions which contain components such as manner (e.g. run, speak), instrument (e.g. pen), result (e.g. close) and path (e.g. walk, crawl). According to Talmy (1985), when learning the meaning of a verb, components of it can be the dominant focus for making the connection between form and meaning. Talmy (2000) and Slobin (2001) propose that different languages highlight different components such as in English, manner is the important component of the verb (e.g. speak); but in Spanish, it is path (e.g. advance). Furthermore, nouns have more limited meanings than verbs. When a noun is uttered, most of the time, it refers to an entity in the real word therefore it has a restricted meaning. However, when a verb is uttered, it relates to some part of the action. Pickett et al. (2000) exemplifies this difference that according to an average dictionary, the noun *ball* has only two definitions, while the verb *run* has 53 definitions.

The other reason behind nouns' being simplee is that infants prefer to attach a new word to an unknown object rather than to an unknown action. Childers & Tomasello (2002), Kersten & Smith (2002) and Huttenlocher, Smiley & Charney (1983) propose that infants, in doing the work of mapping, give preference to simple actions over complex actions and self-actions over the actions of others. Lastly, nouns' labelling objects takes places simultaneously with the word's (noun) being uttered. On the contrary, verbs describe actions, they are more temporal, naturally (Langacker, 1987; Slobin, 2001). Consider, for example, a rat teasing a cat and then leaping onto a hole. Both participant objects (the rat

and the cat) are visible throughout the scene, before, during, and after the teasing occurs. In contrast to this stability of the objects, the relationship between them (teasing) is more fleeting, observable only in the moment that the teasing takes place. Therefore mapping form word to object is considerably easier than mapping from word to action. Tomasello & Kruger (1992) claims that in child directed speech nouns label permanent objects available in the context but verbs refer to actions that has taken place before. The question that arises here is how about commands or talks about intentions, future actions? Therefore, the discussion about verbs' referents should focus on non-present referents instead of non-present actions. To sum up because of many reasons mentioned above, verbs are treated as complex entities.

2.4. Verbs in Action

The literature on early lexical acquisition has neglected other word types which are more diverse than nouns. Researchers in this field study early word learning with the lens of noun learning. Since nouns propose a good foundation for studying word learning, they are more predominant in a child's first words and are learned quickly and easily when compared to the verbs which are more complex, ephemeral and abstract. However, there are some studies informing us about the importance of studying other word types and those word types are used for a variety of functions; and even in English, nouns are not always the largest group (Nelson, 1973; Gopnik, 1981; Bloom, Lifter, Hafitz, 1980; Landau & Gleitman, 1985; Bates, Bretherton & Snyder, 1988; Bloom, Tinker & Margulis, 1993).

In response to the limited research on the other form classes and nouns being earliest in early speech, researchers started to ramify investigations. Kucera & Francis (1967) showed in their one million word corpus of written language that 20% were verbs and 6% were nouns so adults use more verbs or other predicate terms than nouns. However, the communication between adults differs from speech to young children. Gopnik (1981 and 1988) presented that before nouns, non-referential expressions are used by infants and when one word stage starts non nominal expressions such as *that, no, gone, up* and *in* were heard more frequently than nouns. Kauschke & Hofmeister (2002), in a study of 32 German children from 12 months to 36 months, found that relational words (da "there"), personal social words (ja "yes", nein "no") and onomatopoeic terms (brumm "car sound") constitute the majority of early words. Nelson, Hampson and Shaw (1993) have shown that a considerable amount of early nouns are abstract nouns and perform other functions. It can be understood from these studies that nouns or names are not always the earliest categories for children.

The noun-first view also hypothesized that nouns are very easy for children to acquire and novel nouns are extended more readily than verbs. However, Waxman & Klibanoff (2000) proposed that young word learners extend novel adjectives more broadly to a variety of objects than novel nouns and for mapping novel adjectives to object properties, infants use basic level object categories. They reported that a general cognitive process in concert with word learning promotes extension of novel adjectives to object properties. Tomasello & Akhtar (1995) have shown that nouns don't have any priority in comparison with verbs. They found that when naming of a new object and new action (such as "dax") were compared, children couldn't make any difference between learning a new noun or a new verb. According to Oviatt (1980) 12 month old children cannot differentiate nouns and verbs in comprehension. Childers & Tomasello (2002) demonstrated that the difference between nouns and verbs lies behind both the difference in understanding objects and actions and the difference in connecting new words to these actions. These studies also suggest that verbs are not always complex and obscure for infants out of the starting gate of language.

In early vocabulary acquisition a great many studies have shown nouns over verb preference in a variety of languages such as in English (Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994), Italian (Caselli, Casadio & Bates, 1999), Hebrew (Dromi, 1987), Korean (Kim,

McGregor & Thompson, 2000), Mandarin (Tardif, Shatz & Naigles, 1997) and Spanish (Jackson-Maldonado, Thal, Marchman, Bates & al, 1993). However, in a comparative cross-linguistic study of three languages (English, Mandarin and Cantonese) Tardif et al. (2008) reported that most of children's first ten words were addressed to people, not objects. They distinguished nouns for objects (and animals) and nouns for people (names). They also looked at the difference between objects nouns and verbs and found that in Mandarin Chinese, children used more verbs than nouns and in Cantonese, roughly equal numbers in first ten words. Another comparative cross-linguistic study of seven languages (English, Spanish, French, Italian, Dutch, Hebrew, Korean) conducted by Bornstein and colleagues (2004) on early vocabulary development found that in the first fifty words, there isn't any preference for nouns. Childers, Vaughan and Burquest (2007) studied an Afro-Asiatic language, Ngas and found that Ngas children understand more verbs than nouns but they produce equal numbers of nouns and verbs. In a similar vein, Bassano et al. (1998) and Bassano (2000) showed that the first verbs in a longitudinal study of one child in French were seen earlier than first nouns. These studies suggest that there is not any universal bias towards nouns and there might be cross linguistic differences.

A great many studies in early lexical acquisition used the MacArthur-Bates Communicative Development Inventory (CDI) which is a vocabulary checklist containing two parts; comprehension and production. CDI is completed usually by mothers or caretakers and judged to be biased towards nouns such that the presentation of word categories is aimed to steer caretakers in favor of nouns. Moreover, in CDI English version there are 249 nouns and 57 verbs and the first 13 subgroups of words presented in the checklist are nouns (Stoll et al., 2011). Because of these reasons using checklists in early lexical acquisition such as CDI creates a noun bias and it would be a good idea to look distribution of word types in naturalistic data.

Researchers have also looked at children's use of nouns and verbs in some particular contexts. Gelman & Tardiff (1998) reported that children between 12-18 months use more nouns than verbs. However, children around 36 months use more verbs in the context of playing with toys and use more nouns in the context of reading books (Gelman & Xu, 1999). Although Gentner (1982) studied many typologically different languages while formalizing *Natural Partition Hypothesis*, the acquisition order and frequency of use are not universal cognitive parameters. Gopnik and Choi (1990) have shown that Korean speaking children's first words include far more verbs than nouns. Moreover, in a comparative study of Korean and English Kim et al. (2000) found that English speaking children use more nouns and Korean speaking children use more verbs; and this difference is attributed to the input children are exposed to. They concluded that the early lexicon is shaped by general socio cultural, cognitive and language specific factors. Another phenomenon that needs to be taken into consideration is that child language may include some concepts instead of verbs but that may not actually be verbs in adult language. For example, a child may say *that* when s/he want an object (it is not required for him to say *give that*) (Choi and Gopnik 1993, 1995, Tomasello 1992). More than that children's early data in two Mayan languages (Tzeltal and Tzotzil) showed that children prefer verbs and these findings judge universal noun bias (Brown, 1998; de Leon, 1999).

2.5. Computational Complexity Assumption on Early Lexical Acquisition

Camaioni and Longobardi (2001) claimed that there is a verb preference in Italian after they studied 15 Italian infants' child directed speech. However, by contrast, Tardif et al. (1997) found that children of Italian showed no difference between nouns and verbs. In a similar vein, Choi and Gopnik (1995) found that in English child directed speech the distribution of nouns and verbs were equal whereas in Korean, there were more verbs. However, English children produced more nouns and Korean children showed no difference. Therefore frequency in the language of mothers alone cannot account for the distribution of nouns and verbs. Apart from the frequency distribution in the input, the acquisition of nouns and verbs

must depend on some additional factors. Computationalism suggests two more factors in addition to (1) frequency: (2) how the context disambiguates words, and (3) phonological complexity of words.

According to computationalism, a problem's being easy or difficult can be determined by some aspects such as time complexity, non-determinism and algorithmic space. Whether they are implemented in a computer or not, computationalist models are hypotheses which connect representations to solutions to eliminate alternatives in the hypothesis space. Reinterpretation of the results above might suggest that early noun acquisition is merely a conceptual bias towards names, objects and their first appearance in child language. Therefore a computationalist perspective was proposed by Zettlemoyer and Collins (2005) for machine learning and adopted for language acquisition by Steedman & Hockenmaier (2007) and Çöltekin & Bozsahin (2007).

The computationalist acquisition perspective sees early lexical acquisition as a continuing problem space composing of words, phrases, sentences, utterances and even propositions. From the linguistic side, when the child comes across a phonological form (the mother utters the word "cat") and its referent in real world (the child sees the cat), the aim of the child is to learn the syntactic category of that word (in this case, cat is a noun). The more the child is exposed to a word, the more she learns the correct combinations of that word with other words. Therefore, the child starts to learn frequent structures in which the target word occurs. For example, she would learn the word "cat" is mostly followed by auxiliaries or preceded by determiners. In this situation, the child proceeds when she does the correct matching of phonological form with the referent.

From the computationalist perspective, the child starts with an empty lexicon, when she is exposed to a word she generates a hypothesis that helps her understand that word in child directed speech. After that the child updates her lexicon according to this hypothesis, and restarts with the new lexicon. In the end all the strings the child is exposed to would be in the lexicon and the child learned the syntactic category of the word (Gold 1967 as cited in Bozşahin, 2012). The mapping procedure can be thought of as an algorithm which shows computational bias towards some aspects. We know that if some strings are contiguous the algorithm can easily generate hypotheses on the basis of encountering. Thus the child in the end can manage to learn the correct word type through experience. Therefore frequency is the first aspect that eases the task computationally. Bozşahin (2012) proposed that only contiguous substrings are allowed to bear types, therefore to carry a meaning, and short strings are considered more feasible, because the algorithms must consider all such possible pairs. Moreover, when the greatness of hypothesis space is thought, it can be understood that only short strings are feasible in learning a model. Therefore word length is the second aspect of computational bias. Since many words in natural language are ambiguous, a computer program analyzing the speech will need to decide which reading is intended, in order to be able to come up with the appropriate meaning. The ambiguous context, discourse, sentence or word causes a different particular reading. Therefore, a computer program will come up with ridiculous readings and for longer sentences, hundreds or even thousands of readings arise. That is why ambiguity is the third aspect of computational bias.

When natural language data is reconsidered, it can be seen that both Gentner's (1982) and Dromi's (1987) data suggest that long words seem to have telegraphic character and be rhythmic repetitions, they are simple and clear. Bozşahin (2012) argued that Tzeltal verbs in Brown's (1998) data are quite argument-specific therefore less ambiguous than opaque verbs. For example, eating tortillas, eating beans, and eating in general are different words. Being argument specific for early verbs is not the only criteria; some of them are not argument-specific, but they are the most frequent verbs in the language. Moreover, all of the verbs that Tzeltal children produced have a maximum of four phonemes. This shows that short words were acquired easier than longer words. All in all, computationalist aspects were seen to ease language acquisition task for infants.

In line with Karl Popper's (1963) *falsificationism* principle, which stressed that what is unfalsifiable is classified as unscientific, the computationalist models are also falsifiable. Bozşahin (2012) made some predicitons about this topic and stressed that if the phonological complexity (word length), ambiguity and frequency do not have an affect over children's early word learning, then the computationalist assumptions will be wrong. To exemplify this, Bozşahin (2012) asserted that if some short verbs are not learned early even when they are frequent and unambiguous; if some infrequently used long nouns can be learned early and some infrequent but short nouns can be learned early, then the computationalist model will fail to explain such results.

2.6. Methodological Issues

2.6.1. Synopsis of the Present Methodology

The present study aimed to assess the role of computational complexity assumption on early lexical acquisition. Contrary to noun first view, frequent, unambiguous and short words are expected to be acquired first.

The data of the present study is the product of a longitudinal, cross linguistic project funded by Eurocores, ESF, conducted partially by Prof. Aylin Küntay. It includes 3 Turkish children's data (annotated in CLAN format) starting from 8 months of age to 36 months.

Computational complexity assumption was aimed to be assessed by three different analyses; frequency, ambiguity and phonological complexity analyses. Frequency analysis, which is conducted on three infants' data, has two parts; child directed speech analysis and child speech analysis. Ambiguity analysis is conducted on one infant's child directed speech which is comprised of 18 videos. Phonological complexity analysis is again conducted on one infant's both child directed speech and child speech. Since one of the aims of the study was to obtain a general idea on how frequency, ambiguity and phonological complexity are related with early word learning, we preferred to make use of various measures for each analysis. In the frequency analysis part, early word acquisition was assessed throughout type token ratio, noun to verb ratio and arithmetic geometric mean ratio. In the ambiguity analysis part, how the context of natural language disambiguates or obscures the words that infants are exposed was assessed throughout social and attentional cues. Lastly, in the phonological complexity analysis, word length was used for assessing if a word is short or long.

2.7. Hypotheses

Frequency. *As opposed to a universal* noun bias on early lexical acquisition, we expect that frequent words will be acquired first no matter what their category is.

Ambiguity. We expect that words which are disambiguated by the context will be acquired early independently from the category of word.

Phonological Complexity. *We expect that* short words will be acquired earlier than longer words, independent of the lexical category of the word.

CHAPTER III

FREQUENCY ANALYSIS

Psycholinguistic theories of language acquisition state that language use is the source of all linguistic units. Thus, frequency is a key factor in language acquisition in that rules of language are extracted from distributional characteristics of the language.

De Villiers (1985), Naigles and Hoff-Ginsberg (1998) and Theakston et al. (2004) have all found that the acquisition order of some verbs were matched with the frequency of use in child directed speech. Moreover, they showed that the constructions of adults in which verbs were used were correlated with that of children's. Rowland and Pine (2000) showed that the infant they studied produced correctly inverted wh-auxiliary sequences that were highly frequent in the language of caretakers and for the lower frequency sequences he made errors. Frequency effect is not limited to the early stages of language development. Diessel and Tomasello (2001) that early matrix clauses in utterances with finite complements were also the most frequent in the input. Weird word order experiments (Matthews et al., 2005) and grammaticality judgments (Theakston, 2004; Ambridge et al., 2008a,b) also show the effects of frequency on children's linguistics representations.

According to some studies, infants' use of word categories can easily be observed in the child directed speech. For example, if a language allows noun drops, children will also produce fewer nouns; if there are fewer noun drops in child directed speech, children will produce more nouns (Stoll et al. 2011). The other example comes from Tardif et al. (1997) who studied six Italian, six English and ten Mandarin children starting 1; 10 to 2; 0. They closely looked at the distribution of nouns and verbs in these three languages and found that Mandarin children used more verbs than nouns like their caretakers did. As for the English and Italian, children showed an equal distribution of nouns and verbs. Therefore they match their input.

In this part, it is assumed that contrary to the views of Gentner (1982), early lexical acquisition relates to computational complexity by which frequent words are acquired first. It is claimed that *Natural Partition Hypothesis*'s noun dominance in child speech is not universal. In order to achieve this goal we have made two different analysis; the first one basically focuses on the language that is spoken to infant, namely every word that the child is exposed to; the second one mainly depends on the language that is spoken by the infant, namely every word that is uttered by the infant.

3.1. Methods

Since we used the data of Aylin Küntay's project as detailed below and did not do the data collection, we will just report their properties in some parts.

3.1.1. Description of AKT-T[1]: Participants

The frequency analysis is based on a sample of longitudinal corpus of three Turkish infants.[2] These infants (one girl and two boys) were recorded who were age 0; 8 at their first recordings. The children live in the city and child caregiver interactions were video-recorded at the homes of infants (Burcu, Ekin and Can), for one hour every two weeks. Burcu's parents both had 12 years of education; Can's parents both had college degrees; and Ekin's parents both had doctoral degrees. Table 3.1 provides the basic characteristics of the datasets.

Table 3.1. Characteristics of the Databases

	Burcu	Can	Ekin
Number of sesssions	44	26	28
Start Age	00;08;02	00;07;28	00;08;01
End Age	03;00;03	02;10;07	01;09;28
Total number of types in CDS*	14,230	12,189	12,809
Total number of tokens in CDS	89,300	70,951	86,920
Total number of types in CS*	3,606	3,907	1234
Total number of tokens in CS	17,676	19,057	5762

Note: CDS= Child Directed Speech, CS= Child Speech

3.1.2. Description of AKT-T: Procedure

The infants were video recorded for about two hours per month over a period of 28 months. The recordings took place within two weeks of that month, distributed over two sessions in that month. However, a few of the video recordings were excluded because of poor video quality and some technical problems such as audio missing.

In order to obtain linguistic data the infant must be alert and interacting with mother and recorder and that was the only criteria about recordings. The recordings were carried out with a video camera and an external microphone, which was placed close to the area where the infants were playing. All of the recordings took place inside the house. A Turkish research assistant recorded the infants in their natural environment. The context was always the same and no influence was imposed on the context. There were sometimes a number of adults present during the recording, either interacting with the child or talking to each other. Situations included mostly free play, roaming around, having a snack and those kind of natural things. Sometimes the recorder interacted with the infants (again part of the natural environment of the children). In a few cases, the interaction took place to induce children to talk, but mostly the recorder did not actively take part in the interactions filmed but rather took care of the technical arrangements. A main characteristic of our recordings was that usually two people were around; the mother and the recorder and this mirrored the typical daily life of the target infants. However, the amount of interaction with adults varied from recording to recording. Afterwards, trained native speakers transcribed and morphologically

[1] Aylin Küntay's Project Turkish Fragment

2 This sample corpus is part of a project which was carried out Prof. Aylin Küntay and includes eight infants. The data were collected within a large scale cross-linguistic project funded by Eurocores, ESF, conducted partially in collaboration with Dr. Sophie Kern at Laboratoire Dynamique du Langage at University of Lyon II. This is the first study in Turkish child language that attempts to document changes from babbling to first words, and early grammar and social-pragmatics from 8 months of age on to 36 months.

glossed the spoken language from the videotapes using the CHAT transcription format provided by the CHILDES project (MacWhinney, 2000).[3]

3.1.2.1. Our Analysis of AKT-T Data: Procedure

In order to analyze child directed speech and child speech separately, by means of CLAN (Computerized Language Analysis) program, we divided each of age of recording into two parts: the language that the infants are exposed to and the language that the infants produced. The separation of data was followed by counting the noun and verb types and tokens. We included the raw counts of nouns and verbs used per adults and per target children for each recording cycle in Appendix. In order to evaluate the reliability of our counting of types or tokens of word types, a second coder produced independent annotations for two representative chat formats. Since coders were free to assign multiple parts of speech to each word, we assumed that words for which multiple parts of speech were indicated for a particular word contained multiple opportunities for agreement. A statistical measure of to what extent our implementation of coding or measurement system works is Cohen's kappa.[4] It ranges from –1.0 to 1.0, where values near to one mean better reliability, values near zero suggest that agreement is attributable to chance, and values less than zero signify that agreement is even less than that which could be attributed to chance. Figure 3.1 shows the results of the inter-rater reliability analysis which is kappa = 0.829 with $p < 0.01$. This measure of agreement, which is statistically significant, is convincing.

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Measure of Agreement	Kappa	,829	,020	39,749	,000
N of Valid Cases		1364			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Figure 3.1. Cohen's kappa Analysis

In our count we excluded proper names; and like other research on the issue, we identify nouns and verbs based on adult grammar. In the counting of types, we used two different parameters. In the first one (lexical type analysis); for just infant I (Burcu), [5] when measuring type ratios, a form like *geldi* "(he or she) came" or *geldiler* "(they) came" was counted as one, irrespective of how often the different forms actually occurred. Each noun and verb form of paradigm was not counted as an individual type. Therefore, *geldi* "(he or she) came" or *geldiler* "(they) came" were counted as a same type because they are just morphological forms of one lexeme. In the second one (word form analysis), when measuring type ratios, a form like *verdi* "(he or she) gave" was counted as one, irrespective of how often the form actually occurred. Each noun and verb form of paradigm was counted as an individual type. Therefore, *verdi* "(he or she) gave" was counted as a different type than *verdim* "I gave". When measuring tokens, we counted the actual number of occurrence

[3] The recordings were transcribed and coded in the Clan format. The transcripts represent all of the adult–child spontaneous and guided conversation during the course of visits. All words are in lower case; only proper names are capitalized. Uncertain transcriptions are enclosed in parentheses; standard equivalents of child or colloquial forms are given in square brackets. Child utterances are separated into morphemes by hyphens.

[4] A measure of reliability in an n-alternative decision that corrects for chance guessing of frequent options

[5] Our main aim in the analysis of just one infant is to discuss the necessity of counting morphologically different types as one.

of each specific form. However, we excluded incorrect forms such as forms in which one morpheme was lacking because some forms were not clearly recognizable as word types.

After we extracted all the nouns and verbs types or tokens that occurred during one individual recording cycle, we used three different measures[6] in order to measure lexical diversity or the distribution of word categories in child directed speech and child speech. The first mechanism is type-token ratio which is the proportion of word types (nouns or verbs) over the word tokens (nouns or verbs). It was measured separately for each target child, at each age of recording both for child directed speech and child speech. The second one is noun to verb ratio which is the proportion of nouns divided by the sum of nouns and verbs in types or tokens. It was measured only for Infant I (Burcu) both for child directed speech and child speech. The last mechanism is number of types-tokens analysis which is the number of types-tokens of nouns and verbs. It was measured only for Infant I both for child directed speech and child speech. It represents the pure number of different word forms for each word category.

3.1.3. Tools

3.1.3.1 Type Token Ratio

Our analysis is mostly based on Type-Token Ratio (TTR) which is a measure of vocabulary variation within a written text or a person's speech. Type-Token Ratio is the relationship between the number of tokens and the number of types. Lexical variety within a text can be shown by the type-token ratio. TTR basically measures the range of vocabulary for a typical speech sample. It is a useful index of lexical diversity. If the TTR is large, the repetition of words is smaller. It has been used in a wide variety of studies ranging from individual differences studies (Lieven, 1978) to longitudinal case studies (Fletcher, 1985). According to Miller (1981) and Fletcher (1985), a TTR of 0.5 can be used as a baseline and ıf the TTR falls above or below 0.5 it can be concluded that the lexical diversity is not normal. TTR analysis is exemplified with the following example.

There is a text below from Alice's Adventures in Wonderland by Lewis Carroll,

> *"Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, 'and what is the use of a book,' thought Alice 'without pictures or conversation?"*

Here we have a total of 57 words therefore the number of tokens is 57. As can be seen from the passage, some words are repeated such as "pictures" and "conversations" and also there are three occurrences of "of". When we look at the frequency of occurrence of the words, it can be seen that there are 41 different words, types. Therefore, the TTR would be 0.719 which shows a lexically varied text. We can say that a high TTR indicates a large amount of lexical variation and a low TTR indicates relatively little lexical variation. Carroll (1964) suggested another measure for lexical diversity;

> "A measure of vocabulary diversity that is approximately independent of sample size is the number of different words divided by the square root of twice the number of words in the sample." (Carroll, 1964 p. 54 as cited in Richards, 2009)

While TTR is negatively correlated with sample size, Carroll's measure is positively correlated with the number of tokens.

[6] These mechanisms will be detailed under the Tools part.

3.1.3.2 Noun to Verb Ratio

Noun to verb ratio is the proportion of nouns divided by the sum of the proportion of verbs and proportion of nouns. When its value is greater than 0.50, it indicates a noun bias and when its value is lower than 0.50, it indicates a verb bias. The output is a bounded scale between 0 (verbs only) and 1 (nouns only). Noun to verb ratio is performed for types and tokens, separetely. It is exemplified again by the above text. In the text above, there are 10 nouns (proper names are excluded) and 7 verbs so the noun to verb ratio is 0.588 which shows a slight noun bias.

3.1.3.3 Number of Types-Tokens

Word type-token analysis shows the number of types and tokens that each word category has. Its aim is to visualize the development of word types-tokens throughout the milestones of language development.

3.2. Results

In order to analyze the distributions of nouns and verbs, we divided our corpus of each child into two parts; child directed speech which includes the words child exposes and child speech which includes the words uttered by the infant.

3.2.1. Child Directed Speech Analysis

3.2.1.1 Type Token Ratio Analysis

3.2.1.1.1 Lexical Type-Token Ratio Analysis

As we noted above, we conducted a different lexical type analysis in order to include the difference of inflectional suffixes that a verb or noun takes in Turkish. It is a kind of lexical analysis since a unit of lexical meaning exists regardless of the number of inflectional endings it may have. According to Göksel &Kerslake (2005), in Turkish nouns can be inflected for number, person and case but verbs are inflected for voice, negation, tense, aspect, modality and person. Therefore making a lexical type distinction would make a reasonable difference when counting the types in a Turkish corpus. Table 3.2 gives an impression of the ten most frequent noun and verb types found in the three infant's child directed speech.

Table 3.2. The Ten Most Frequent Verb and Noun Types in Child Directed Speech

Verbs	*Nouns*
bak- (intr.) "look"	*kız- "daughter"*
gel- (intr.) "come"	*oğul- "son"*
al- (trans.) "take"	*abla- "sister"*
de- (trans.) "say"	*baba- "father"*
ol- (intr.) "become"	*anne- "mother"*
ver- (trans.) "give"	*bebek- "baby"*
yap- (trans.) "do"	*oyuncak- "toy"*
dur- (intr.) "stop"	*top- "ball"*
koy- (trans.) "put"	*park- "park"*
otur- (intr.) "sit"	*süt- "milk"*

The most frequent nouns in the three infants' child directed speech are mostly kinship terms. There may be a universal bias towards kinship terms in child adult spontaneous speech. Due to the fact that we have not included proper names into our analysis, proper names are not

included in the table either. Most frequent verb types reflect some kind of variety and there isn't any noticeable difference

Table 3.3. Number of Noun & Verb Types-Tokens per session across Infant I's CDS

Age	Noun Types	Noun Tokens	Verb Types	Verb Tokens
00;08;02	73	336	53	334
01;00;03	40	247	43	471
01;00;17	23	97	31	138
01;01;15	56	495	60	396
01;02;05	40	219	38	238
01;02;23	46	210	52	389
01;03;09	58	358	49	357
01;03;26	55	293	41	277
01;04;10	81	460	54	456
01;05;11	63	321	54	288
01;06;00	67	305	45	386
01;06;17	38	174	33	198
01;07;00	27	169	36	217
01;07;14	30	121	26	188
01;08;01	31	149	36	244
01;08;16	30	131	28	208
01;10;00	32	172	48	370
01;10;14	38	182	37	175
01;10;28	29	170	31	250
01;11;27	56	200	34	217
02;00;12	27	103	24	119
02;00;26	47	241	39	261
02;01;17	34	123	33	211
02;02;01	38	182	36	291
02;02;14	46	193	39	261
02;02;29	22	87	34	216
02;03;12	34	146	32	361
02;03;28	82	358	53	404
02;04;09	59	429	68	622
02;04;23	86	445	59	426
02;05;05	71	391	65	630
02;05;21	67	366	59	612
02;06;05	49	237	56	526
02;06;20	64	283	52	556
02;07;01	76	345	48	422
02;07;17	91	403	55	580
02;08;02	60	320	44	355
02;08;28	66	329	58	561
02;09;24	76	341	48	411
02;10;08	46	187	47	285
02;10;22	49	260	50	479
02;11;04	46	209	44	276
02;11;20	50	307	59	502
03;00;03	53	300	42	457

Table 3.3 shows the relative number of noun and verb types and tokens in Infant I (Burcu)'s data. It can be observed that noun types mostly outnumbered verb types. However, noun tokens were lagged behind by verb tokens. Therefore Infant I heard many more different kinds of nouns but more verb tokens during the course of development.

In figure 3.2, it can be observed that TTR in child directed speech represent the lexical variety of the context. In noun-first view, it seems that nouns dominate the conversation between the child and caregivers. And to others, despite the fact that the child heard not a variety of verbs, she did heard the verbs many more times than nouns.

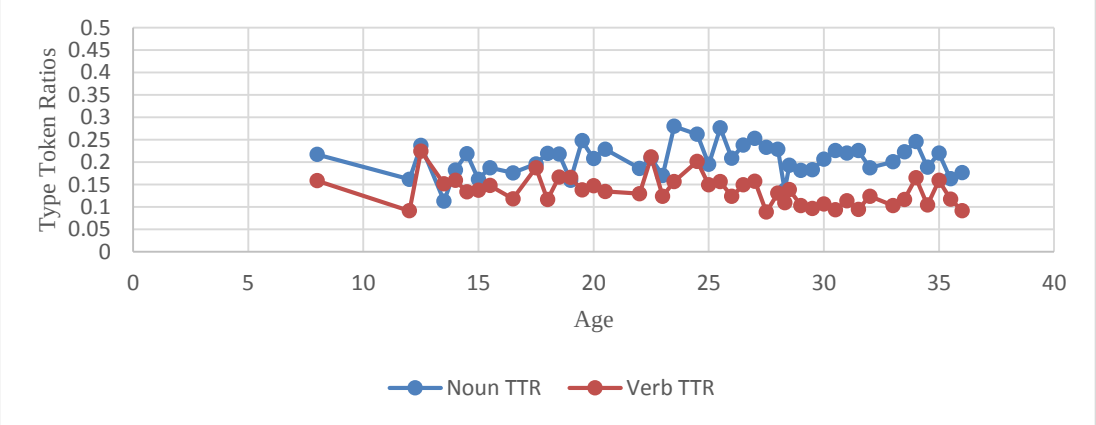


Figure 3.2. Lexical Type Analysis: Nouns' and Verbs' Type Token Ratios for Infant I

At first glance, lexical type analysis supports noun first view in that verb TTR is lags behind. However, it may not be a reliable analysis due to the some reasons such as assuming infants know the roles of inflections in language.[7]

3.2.1.1.2 Word Form Type-Token Ratio Analysis

Haspelmath (2002) defines a word form as a text in terms of whether it is separated by spaces in writing or not. Word forms can be pronounced and used in writing and every word form belongs to one lexeme e.g. the word form *lived* belongs to the lexeme *live*. In word form type analysis, we counted all the morphological forms of a lexeme as different words. Figure 3.3 shows the development of nouns and verbs based on TTR. Taken on their own, the distribution of nouns in child directed speech does not reflect any bias towards nouns. Adults use the same amount of nouns and verbs. The natural speech surrounding the infant includes even more verbs in some parts. The data for Infant I and II suggest an intensive use of verb category between 15 and 20 months. After that time, the verb type token values were decreased and subject to random fluctuations.

[7] Details will be presented at discussion part.

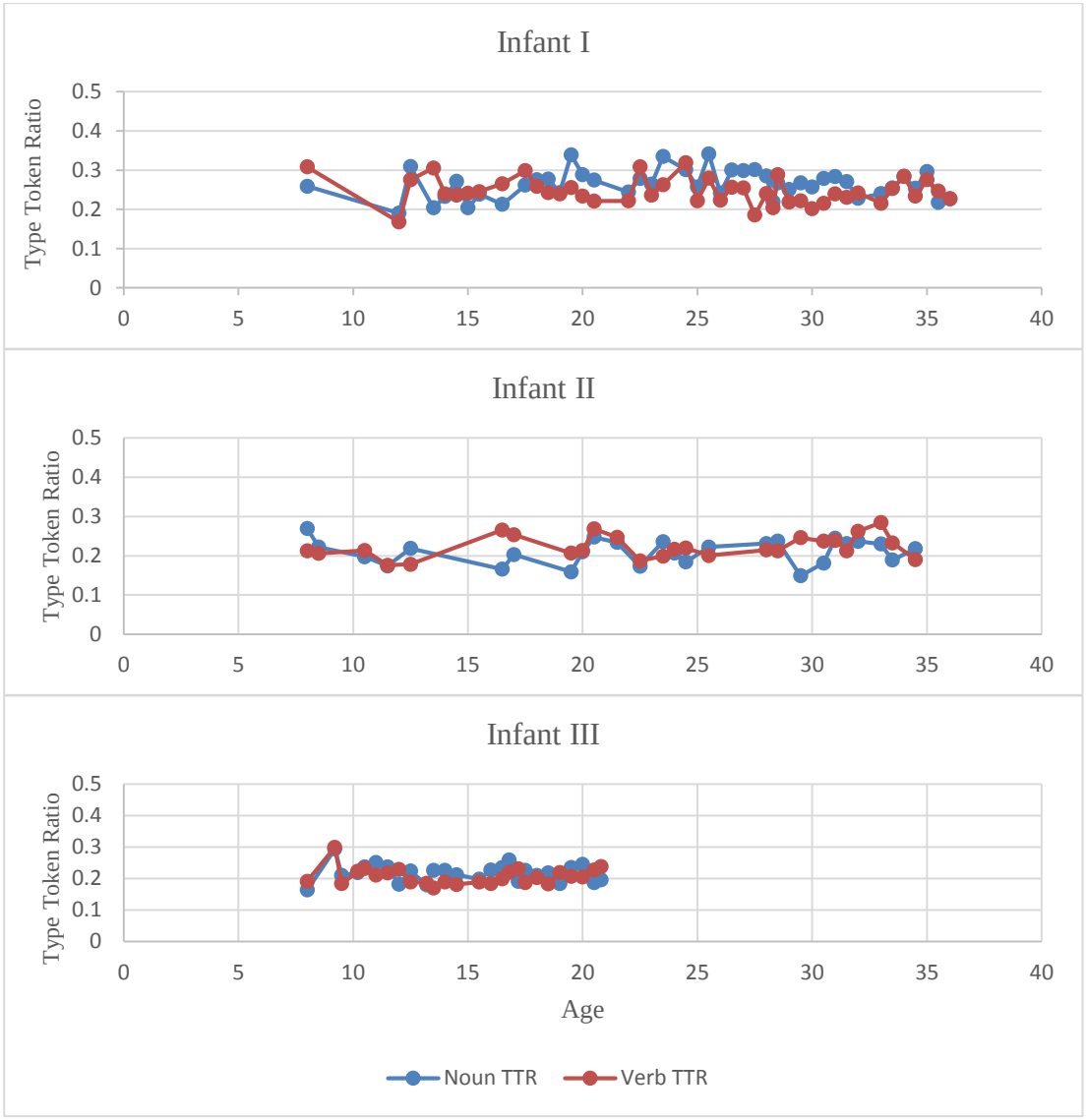


Figure 3.3. Word Form Type Token Ratios of Nouns and Verbs in Child Directed Speech

When both figure 3.2 and figure 3.3 are examined, it can be said that there is a slight preponderance of TTR of nouns over verbs in the type analysis but not in the word form type analysis. This means that verbs usually have more word forms (because of their rich inflection) and therefore catch up with the nouns.

As explained earlier, our hypothesis is that the language spoken to infants does not present any universal noun bias or any kind of noun's predominance over verbs. To test this, we examined the cumulative sum of noun and verb types in child directed speech. Figure 3.4 shows the results, there is a clear predominance of verb types over noun types throughout the milestones of language development. As it is explained earlier, since Turkish verbs are inflected much more, the more verb types are included in child directed speech. This may be a reason behind the verb type dominance. However, in order to go beyond this reason we conducted a lexical type analysis for Infant I which considers verbs' high rate inflections in Turkish. Figure 3.5 presents a cumulative sum of our lexical type analysis of Infant I.

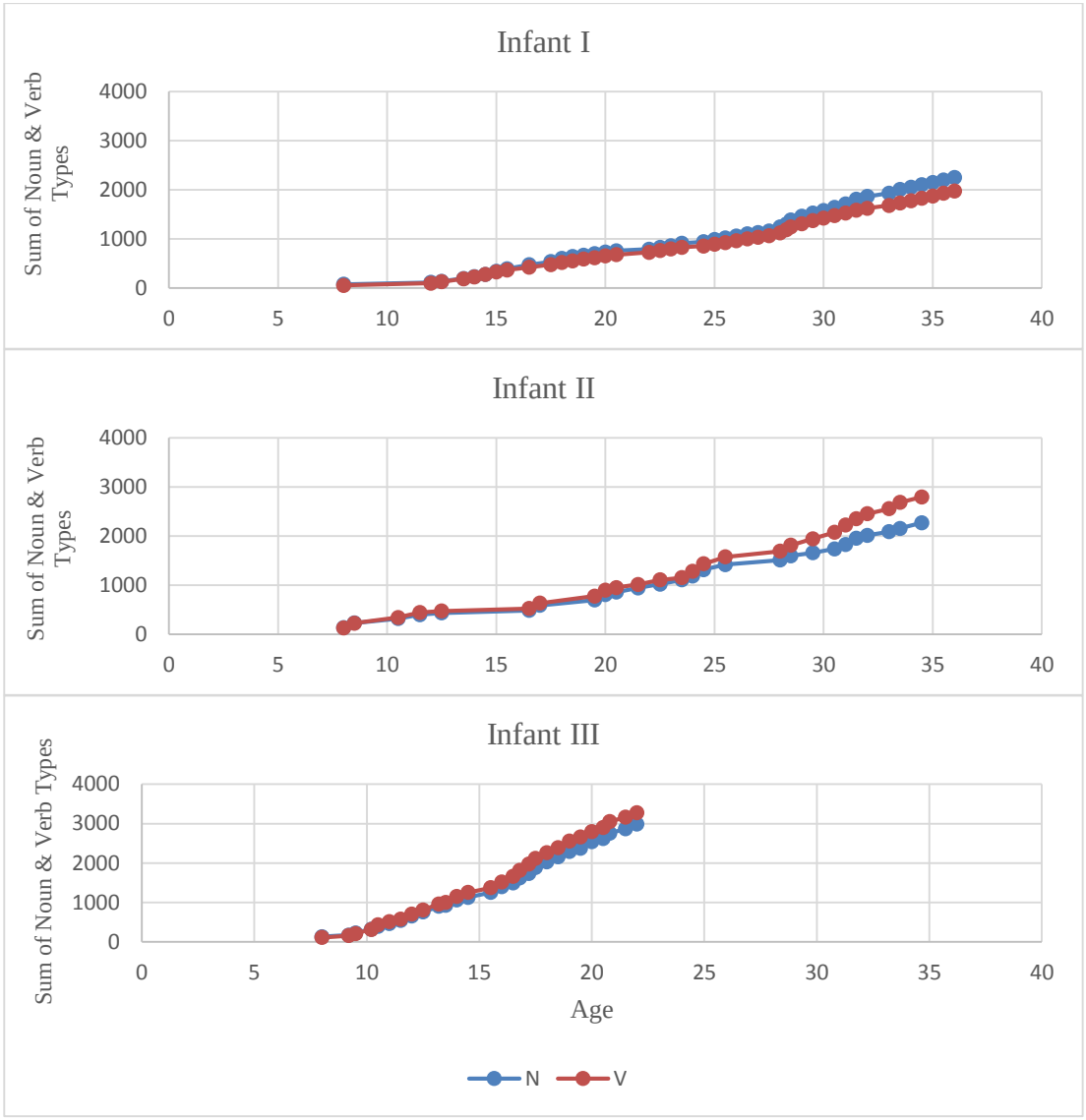


Figure 3.4. Cumulative Sum of Noun and Verb Types in Child Directed Speech

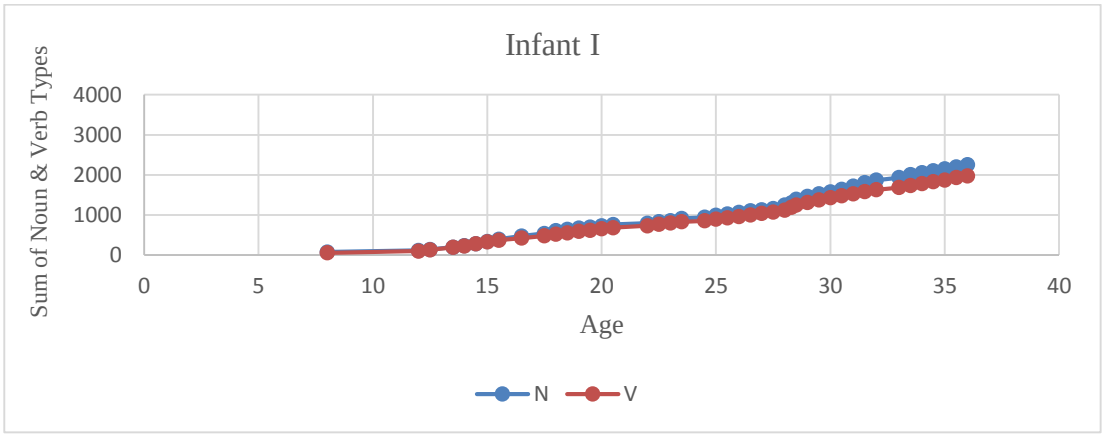


Figure 3.5. Cumulative Sum of Lexical Word Types in Child Directed Speech

When lexical word type analysis is taken into consideration, noun types outnumbered verb types after 20 months. However, as we said earlier lexical type analysis is somewhat problematic.

3.2.1.2 Noun to Verb Ratio Analysis

Besides TTR, we measured the distribution of nouns and verbs by the means of noun to verb ratio. Noun to verb ratio computes the proportion of nouns over the total of nouns and verbs. If it is zero, there are only verbs; and if it is one, there are only nouns. Noun to verb ratio is measured for both types and tokens; and only for Infant I at each age of recording. Figure 3.6 shows the noun to verb ratio for both types and tokens. Since noun to verb ratio is a bounded scale between 1 and 0, a higher noun to verb ratio means more nouns; a lower noun to verb ratio means more verbs. As it can be observed from the Figure 3.6, noun to verb ratio in Infant I's child directed speech fluctuates between 0.6 and 0.3 which means no noun predominance both for types and tokens.

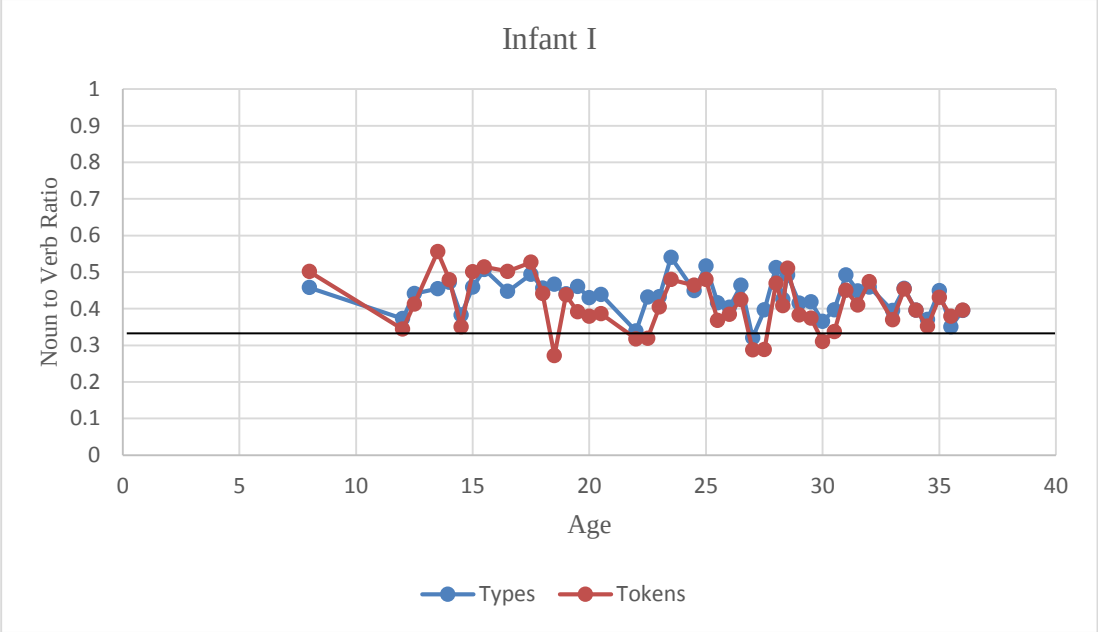


Figure 3.6. Noun to Verb Ratio in Child Directed Speech

3.2.1.3 Number of Types-Tokens Analysis

Number of types-tokens analysis's results, which can be seen at figure 3.7, show that noun types were outnumbered by verb types. This result was expected because of the high number of inflections that a verb takes in Turkish. When it comes to tokens, the development of tokens were parallel with number of types. Verb tokens were much greater then noun tokens especially in the later samples starting from 28 month.

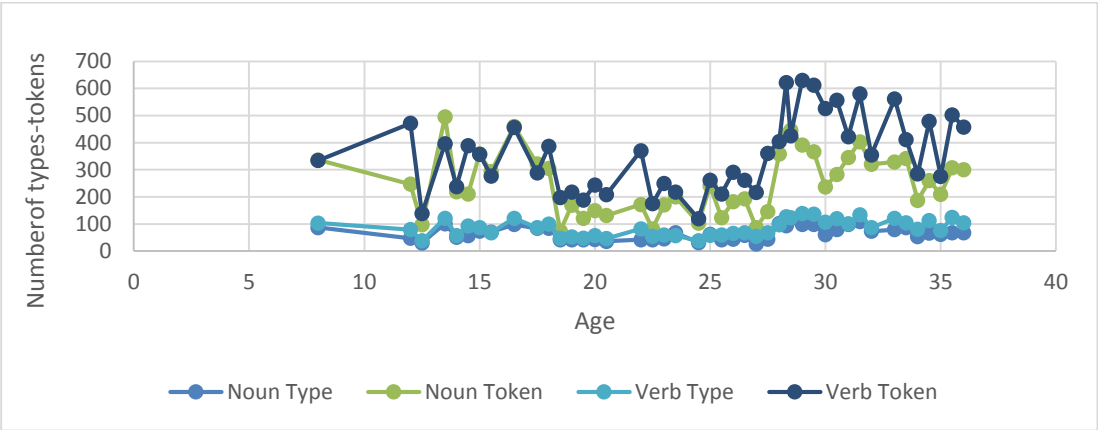


Figure 3.7. Development of noun and verb types-tokens in Child Directed Speech (Infant I)

3.2.2. Child Speech Analysis

3.2.2.1 Type Token Ratio Analysis

3.2.2.1.1 Lexical Type-Token Ratio Analysis

In order to take into account include Turkish's many more verb inflections than noun inflections, we conducted a lexical type analysis for child speech- the language produced by the infants, as well. Table 3.4 presents the ten most frequent noun and verb types found in three infant's speech.

Table 3.4. The Ten Most Frequent Verb and Noun Types among Child Speech

Verbs	*Nouns*
bak- (intr.) "look"	*anne- "mother"*
al- (trans.) "take"	*abla- "sister"*
gel- (intr.) "come"	*baba- "father"*
git- (intr.) "go"	*mama-"food"*
aç- (trans) "open"	*bebek- "baby"*
ver- (trans.) "give"	*el- "hand"*
yap- (trans.) "do"	*süt- "milk"*
otur- (intr.) "sit"	*araba- "car"*
bit- (intr.) "finish"	*ayak- "foot"*
ol- (intr.) "become"	*gol- "goal"*

The important thing is that the most frequent words in child directed speech are more or less the most frequent words in child speech. Likewise, in child directed speech, most of the nouns are family names. According to TTR analysis of Infant I's speech which can be observed from Figure 3.8, until the thirtieth month there is not any categorical difference between word types. However, after the twenty fifth month, the noun category starts to elude from the verbs. But it is not true for noun tokens which lag behind verbs. Even our lexical type analysis challenges Natural Partition Hypothesis in that infants first words are not nouns or names and the noun category is not the earliest.

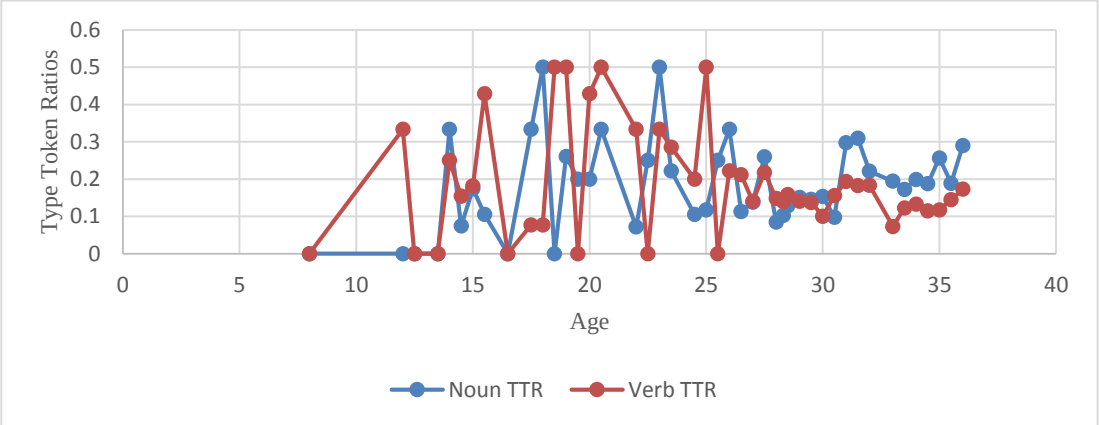


Figure 3.8. Lexical Type Analysis: Nouns and Verbs' Type Token Ratios in Child Speech (Infant I)

3.2.2.1.2 Word Form Type-Token Ratio Analysis

The word form TTR analysis differs from lexical type analysis in that the first one assumes all the morphological forms of a lexeme as different words. For example, *gitti* "s/he went" and *gider* "s/he goes" are counted as different words. The reason behind this assumption is

that infants up to two years maynot discriminate the word stem and inflection fully and in order to plot the data from a different perspective, word form TTR was conducted. Figure 3.9 shows the development of nouns and verbs based on TTR throughout three infant's speech.

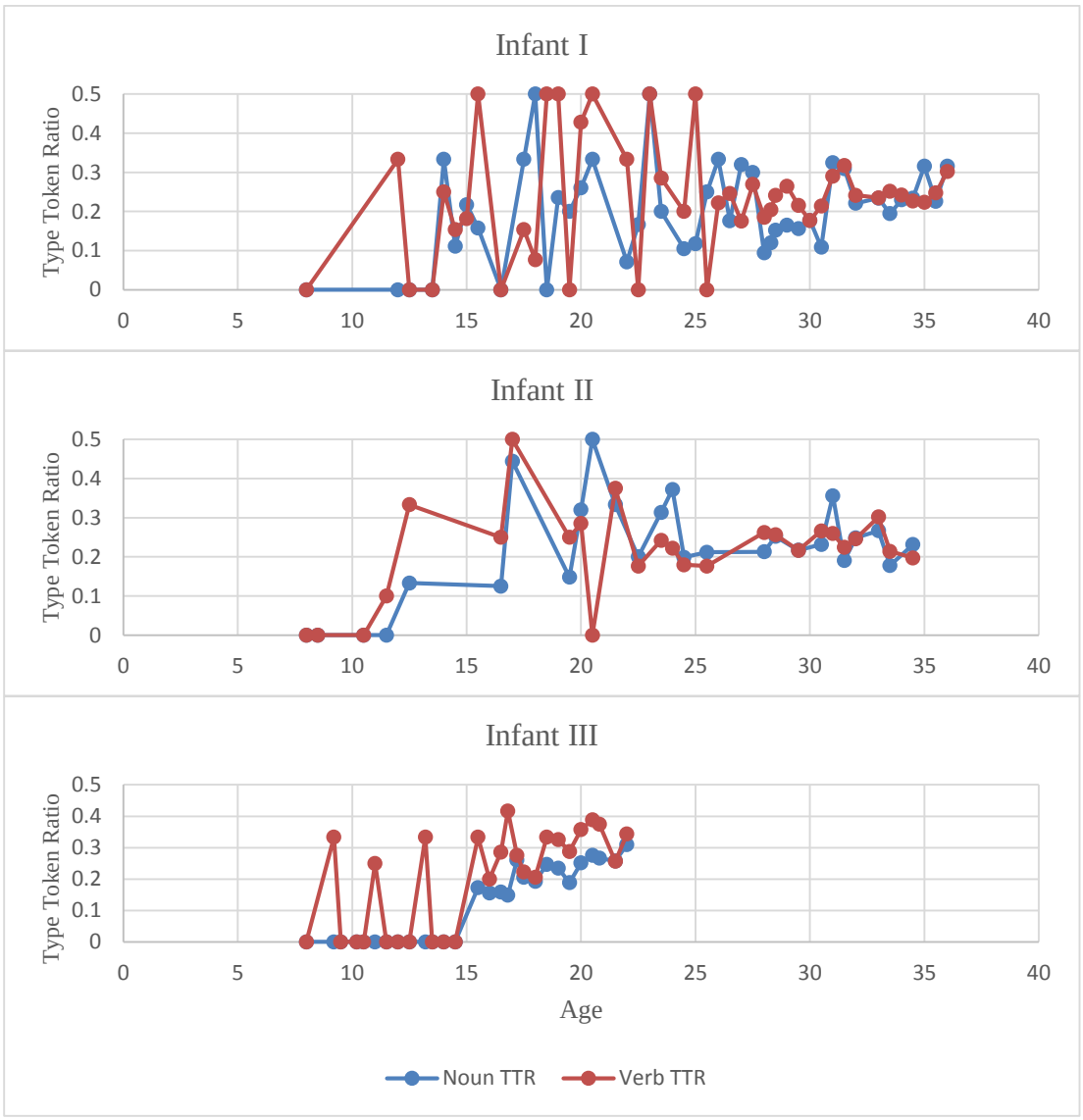


Figure 3.9. Type Token Ratios of Nouns and Verbs in Child Speech

It can be observed that the data for Infant I and II suggest that neither of the two word categories in child speech have an edge over the other. However, in Infant III's speech the distribution of verbs is more dominant than that of nouns. The data above definitely shows that early words are not nouns or object names. As explained earlier, the language spoken to infants does not present any universal noun bias or any kind of noun predominance. This can be easily spotted in the cumulative sum of noun and verb types in child speech in which three infant's data are pooled (see, Figure 3.10)

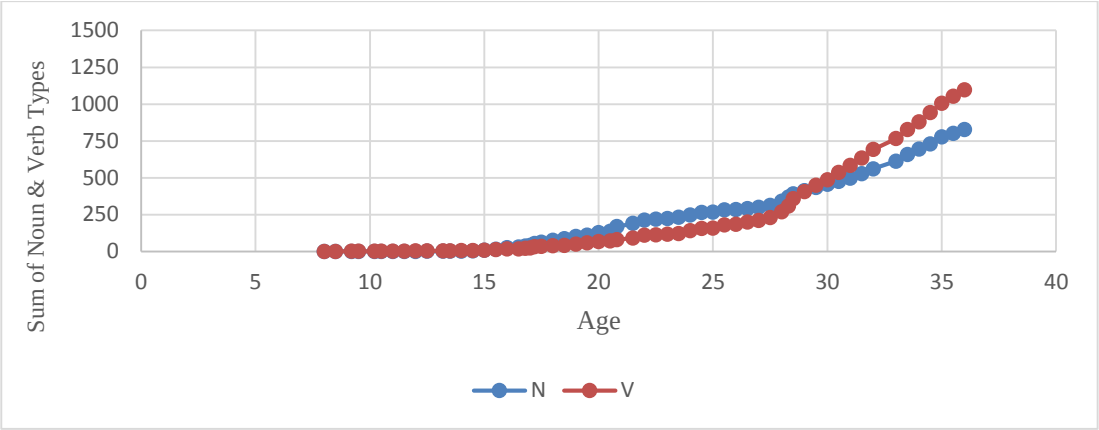


Figure 3.10. Cumulative Sum of Noun and Verb Types in Child Speech (Infants I, II and III pooled)

From Figure 3.10, it can be understood that the process of verb production is stable until the fifteenth month, as in the process of nouns. After the fifteenth month, noun production outnumbers verb production until the twenty eighth month. However, with the start of the burst of speech after the twenty eighth month, the verb production exceeds the nouns strikingly. As a result of several analyses conducted so far and especially in the cumulative graphs, when the vocabulary spurt starts around 19-21 months, the curve starts increasing. The growth rates become more impressive at 28 months, however, the start of that non-linear increase is earlier. It can be seen through the sigmoid shape and also the steepness is more visible at later times.

3.2.2.2 Noun to Verb Ratio Analysis

In order to handle the data from various points, we measured the distribution of nouns and verbs by the means of noun to verb ratio which is the proportion of nouns over the total of nouns and verbs.[8] It is measured for both types and tokens; and only for Infant I at each age of recording. As it can be observed from the Figure 3.11, noun to verb ratio in Infant I's speech fluctuates between 1 and 0 until two years which means no noun predominance both for types and tokens. Furthermore, after 26 months noun to verb ratio standardized and sailed between 0.6 and 0.2. This quite striking settling of the two lines after 26 months can be attributed to the vocabulary spurt. Before the vocabulary spurt (about 25 months), it looks quite erratic, then the lines become steadily around the mean of 0.4 and show less fluctuation. Therefore, it is concluded that these values do not support noun dominance on their own.

[8] Details about noun to verb ratio can be found in Stoll et all. (2011)

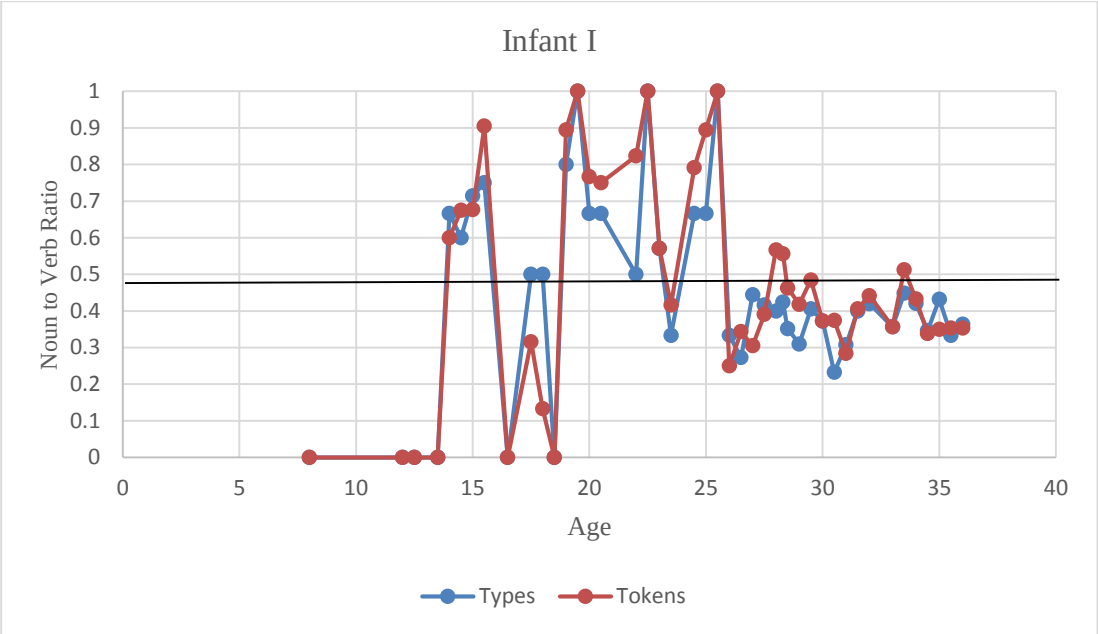


Figure 3.11. Noun to Verb Ratio in Child Speech

3.2.2.3 Number of Types-Tokens Analysis

The development of type-token numbers in child directed speech were the same as child directed speech. The number of noun types lagged behind verb types and the number of noun tokens was also outnumbered by verb types. Figure 3.12 also showed that until the twenty eighth month, the production is limited; whereas after this time, infant I started to use more verbs.

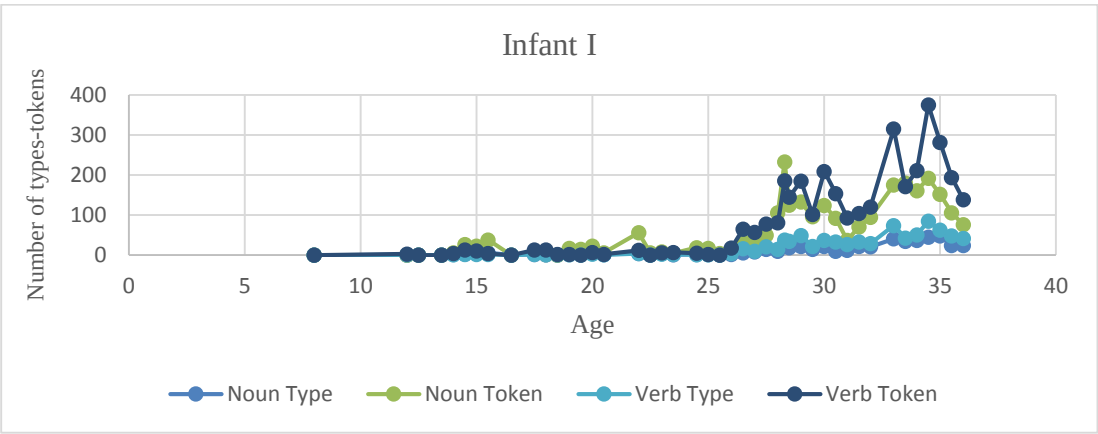


Figure 3.12. Development of noun and verb types-tokens in Child Speech

3.3. Discussion

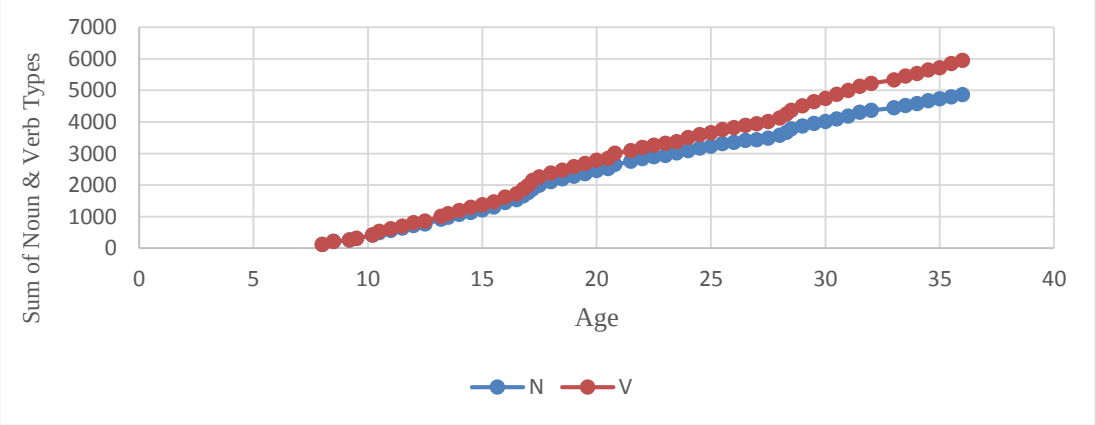
3.3.1 Child Directed Speech

The difference of inflectional suffixes that a verb or noun takes in Turkish is the main reason for conducting lexical TTR analysis. In this analysis, a lexical unit was counted as one type. The ten most frequent nouns are mostly kinship terms even the most frequent five words. Instead a universal noun bias, as Kauschke and Hoffmeister (2002) stated, there can be a general tendency for early nouns towards personal social words and kinship terms. The second outcome of lexical TTR analysis is that since Turkish verbs are inflected much more than nouns, noun types outnumbered verb types whereas noun tokens stayed behind verb

tokens. This also shows that infant I heard many different nouns but many more forms of the same verbs.

Although lexical type analysis, at first glance, seems to back up noun first view, there are some deficits to it. According to Ekmekçi (1982), verbal inflections can be observed at early stages of language development in Turkish since semantic concepts are generally expressed by means of inflections. However, at initial stages of language development, the inflections do not reflect the child's conceptual development. For example, the creative use of the plural form can be seen at 1; 9 and causative and passive voice markers at 2; 0. These findings show that at early stages Turkish infants may not know the place of inflections in language. However, both in Ekmekçi (1982) and in this study, the criteria is just using or using productively as in the many language acquisition studies. Due to the fact that lexical type analysis indirectly assumes infants know the role of inflections from the beginning, it may not be reliable. For this reason we conducted word form TTR analysis in which all morphological forms of one lexeme were counted as different words. The word form TTR analysis results show that caretakers use more verbs but mostly nouns' and verbs' TTR are close to each other. One of the main findings of this analysis is that between the fifteenth month and twenty fourth month, caretakers use many more verbs than nouns. The reason of this is that after one year when children start to walk and act independently, caretakers use more action words such as *stop, give, sit and do not touch.*

The cumulative sum of word types also shows that caretakers' language does not seem to present any noun bias or noun predominance. Moreover, Infant II and III's caretakers used more verbs than nouns. This finding, of course, does not support any verb bias in Turkish children's child directed speech nor do we claim that. However, Turkish-speaking children's vocabularies contain a much higher proportion of verbs. The reason of this cannot be the "verb bias" in Turkish-speaking children's mind or "noun bias" in English-speaking children's mind. The fact that verbs are learned so early and in such a great number in Turkish, or nouns in English, must be explained in terms of computational complexity which claims frequent word types may be acquired first. Figure 3.13 shows the cumulative noun and verb change across the three infants' child directed speech.[9] We can easily observe that the proportion is initially identical, however, after 18 months starts to diverge with verbs dominating nouns. Moreover, one can clearly find that what the infant is exposed to is directly related with what the infant produced. Our three subjects are exposed to more verbs than nouns after 25 months. Another thing that needs to be mentioned is when the children get older, contrary to the common belief that the distribution of nouns and verbs are stable in child directed speech, the proportion of verbs in the context is increased. This pattern represents the conventional use of their language.



[9] Since Ekin's (Infant III) data stops around 23 months, the divergence after 27-28 months excludes it.

Figure 3.13. Cumulative N/V Change of three Infants

In order to measure early lexical acquisition with various scales and reach a general conclusion without depending on just one scale, we used noun to verb ratio and raw numbers of nouns and verbs types-tokens. In noun to verb ratio scale, as it can be understood from its name, higher noun to verb ratio means more nouns. Infant I's child directed speech shows that noun to verb ratio shuttled between 0.6 and 0.3. Mean rate for types in caretaker's speech is 0.43 and 0.41 for tokens. Therefore values below 0.5 show that no noun predominance was observed. Moreover, the number of word types-tokens in infant I's child directed speech showed that verb types and tokens were much greater than that of nouns. There might be two reasons behind this finding; (1) the first one was the high rate of verb inflections compared to noun inflections in Turkish but this was valid only for types and (2) the second one was the frequency effect.

In conclusion, according to our scales and contrary to noun first view, Turkish child directed speech does not represent any universal noun bias.

3.3.2 Child Speech

Our main aim in making the same analyses for child speech is to compare the results with the language that is spoken to infants. Firstly, when the ten most frequent word categories in child speech are observed, one can see that frequent words in child directed speech are also frequent in child speech. Most of the nouns are kinship terms and body parts. The main finding of lexical TTR analysis is that until 30 months there is not any categorical difference between nouns and verbs and usages are very limited. After 30 months, noun category dominates the conversation between children and caretakers. However, first words are neither nouns nor names.

Secondly, the outcome of the word form TTR analysis shows that the production of nouns or verbs is limited; there are more nouns at one recording of age and more verbs at another. Nevertheless, after 25 months when the infants start to speak more adult like, the usages of noun and verb categories start to increase like an integrated spiral. The question of whether infants produced more nouns or verbs in the end of their developmental process can be easily spotted in the cumulative sum of noun and verb types. According to cumulative sum graph, until the sixteenth month, the distribution of noun and verb categories is constant. The children produced both nouns and verbs. However, after the sixteenth month, nouns outnumbered verbs. The reason of this is that there may be a general tendency towards verb dropping during these months in Turkish infants. Since the sixteenth month indicates the production of more than one words which naturally increases nouns; and because the lexicon also contains more nouns than verbs, it, thus, can be filled up more easily with nouns. After the twenty eighth month, verb production exceeds noun production strikingly. The reason of this can be the subcategorization frames of the verbs' having been established more firmly. Thus, nouns may be dropped again because it is clear that there are these noun positions (argument positions) which may or may not be overtly filled which is an interaction with the syntactic development. After the child is the master of her language, she represents the frequencies of the language. Lastly, noun to verb ratio and number of types-tokens results were paralleled with TTR results in that Turkish children do not show any tendency to nouns.

CHAPTER IV

AMBIGUITY ANALYSIS

In this part, it is assumed that contrary to the views of Gentner (1982), early lexical acquisition relates to computational complexity by which unambiguous words are acquired early no matter they are; verbs or nouns.

Making reference in language use is the central language function which is a prerequisite to all else. For an infant who is in the acquisition process of a language guessing the meaning of a novel noun, the best option is that word refers to an object in the context (Pinker, 1989; Yu & Smith, 2007; Siskind, 1996). Furthermore, Childers & Tomasello (2002) proposes that infants prefer to attach a new word to an unknown object rather than to its unknown action. It has been argued that the child at some point has a "naming insight", i.e., it suddenly understands that "all things have names". This enlightenment may go hand in hand with the vocabulary burst. However, one must be cautious here because the direction of causality is not clear. It may be that this insight itself is triggered by the higher number of words (especially nouns) being acquired rather than the insight alone triggering the increase of nouns (Elman, 1996). However, according to Gleitman (1990), this generalization may not be true for verbs. Moreover, it is possible for infants just ignore the information in such highly ambiguous learning context and wait for a context in which the referents of heard words are more certain (Brent & Siskind, 2001). Kachergis, Yu and Shiffrin (2013) proposes,

> "As words and their intended referents are observed in different situations over time, learners can apprehend the correct word-object or word-action mappings. Because the presence of known high frequency pairs reduces ambiguity, highly ambiguous situations containing some familiar referents become feasible learning opportunities." (pg. 200-201).

Therefore behind the computational word learning mechanism, there lies two assumptions: (1) words are relevant to the context, and (2) infants remember the co-occurrence of multiple words and objects in a scene.

Despite many challenges facing human infants such as uncertainty and ambiguity in the language environment, they learn words quite quickly. Since regularities in the co-occurrences of word types and referents (objects or actions) have the most important role in acquiring the words. The role of sharing attention through social cues to joint attention is emphasized by many theoretical works in early word learning (St. Augustine, 397/1963; Bloom, 2002; Clark 2003) and empirical data supports the view that infants use signals like the eye-gaze of speakers to infer what the speaker is talking about (Baldwin, 1993; M. Carpenter, Nagell, & Tomasello,1998; Hollich, Hirsh-Pasek, & Golinkof, 2000).

The core problem for early lexical acquisition is famously presented in Quine (1960). According to Quine's (1960) thought experiment an anthropologist who observes a speaker saying ''gavagai'' while pointing in the general direction of a field. The intended referent might be a rabbit, grass, the field, or rabbit ears but it is indeterminate from this experience. The solution to this indeterminacy problem requires that the learning system be somehow constrained and these constraints are social, attentional, linguistic and representational cues.

Table 4.1. Brown (1998) Tzeltal Data: Mik and Xan's first verbs

Mik (1;5 – 2;0)	Xan (1;3 – 2;2)
ba "go/allgone"	we' "eat tortillas"
la' "come!"	chu' "suckle breast"
we' "eat tortillas"	ay "exist, be located"
ak' "give"	boj "cut with machete"
tzak "take, grasp in hand"	k'ux "eat beans, crunchy things"
: : : :	: : : :
lo "eat fruit, soft things"	lo "eat fruit, soft things"

Brown (1998)'s data suggests that early acquisition of verbs seems to be possible in Tzeltal language. The interesting point about the verbs in Tzeltal language is that they are more argument specific than opaque verbs (Bozşahin, 2012). Therefore, verbs become less ambiguous. For example, the difference between *lo* "eat fruit" and *we'* "eat tortillas" summarizes the point. Eating tortillas, eating fruit and eating in general are different words in Tzeltal. However, we are not suggesting a verb first alternative against Gentner (1982)'s noun first view, neither is Brown (1998). We just suggest that disambiguated words are acquired early. All in all, in order to answer the question of how infants use the context to acquire the words, we have made two different analyses in child directed speech; the first one basically focuses on the utterances which contain nouns; the second one mainly depends on the utterances which contain verbs.

4.1. Methods

Since we used the data of Aylin Küntay's project and did not do the data collection, we just report their properties at some parts.

4.1.1. Description of AKT-T: Participants

Since this study is a preliminary one, the ambiguity analysis is based on a sample of longitudinal corpus of just one Turkish infant.[10] This infant (one girl) was recorded who was aged 0; 8 at her first recording. The child lived in the city and the child caregiver interactions were video-recorded at the home of the infant (Burcu), for one hour every two weeks. The child's parents both had 12 years of education. Table 4.1 provides the basic characteristics of the dataset.

[10] This sample corpus is again part of the project which was carried out by Prof. Aylin Küntay and details of the project were given at Chapter 3

Table 4.2. Characteristics of the Database

	Burcu
Number of video sessions	18
Start Age	00;08;02
End Age	01;11;27
Total time	17h 51m 46s
Total number of utterances that contain nouns	470
Total number of utterances that contain verbs	922
Total number of utterances analyzed	1392

4.1.2. Our Analysis of AKT-T Data: Procedure

In the ambiguity analysis, we examined closely the period when the infant uttered her first words so our video recordings start at 8 months to 24 months. The procedure about video recordings and transcriptions were the same as in the frequency analysis part.

Infants are not continually in communication with caregivers as in many Western cultures and they are exposed to the interactions not addressed to them. For example, an infant who is playing with his toys in a room may hear the interactions of his mother and father or interactions of his brother and sister or even interactions of two people in television. This example reflects the typical case in most studies of children's language development. According to Lieven (1994), the main question is whether infants treat all the surrounding speech as a language learning environment, or whether they prefer focusing more on the speech addressed to them. Although both kinds of speech are always simultaneously present, we decided to narrow down the hypothesis space and include just the utterances which were directly spoken to child. Therefore we excluded the utterances which were between the caregiver (mother) and recorder or between the mother and father.

Following Frank, Tenenbaum & Fernald (2013) we developed two video coding charts for nouns and verbs which can be seen from Table 4.3 and 4.4 each have its feature sets in them. For every verb or noun, there are possible (social and attentional) cues in the context and the features below determine relevant ambiguity position (clearly disambiguated, somewhere in between and clearly ambiguous).

Social Cues for Nouns

a. Mom Eyes: If the caregiver (mother or recorder in our data) uses eye gaze (look at the referred object) when s/he utters a noun, we put a cross in the box.
b. Mom Hands: If the caregiver uses his or her hands and touches the object in question, we put a cross in the box.
c. Mom Point: If the caregiver uses his or her hands and point the object in question, we put a cross in the box.

Attentional Cues for Nouns

a. Kid Eyes: This feature checks whether the infant is looking at the object when the name of the object is uttered.
b. Kid Hands: This feature checks whether the infant uses its hands and touches the object when the name of the object is uttered.

Social Cues for Verbs

a. Mom Non-Ostensive Gestures: This feature checks whether the caregiver (mother or recorder) uses gestures other than ostensive gestures such as mimicing.

Table 4.3. Sample Video Coding Chart for Nouns

					Social Cues			Attentional Cues		Ambiguity Position		
Video	Utt.N.	Utterance	Objects Presented	Objects Referred	Mom eyes	Mom hands	Mom point	Kid eyes	Kid hands	Clearly Disambiguated	Somewhere Between	Clearly Ambiguous
B1	1	iremin dişleri gitti	dişleri	diş	x	x						x
B1	2	birazda kucağımıza alalımmı	kucağımıza	kucak	x		x					x
B1	3	ne diyorsun kız	kız	kız	x	x						x
B1	4	saçların ne kadar az	saçların	saç	x	x		x			x	
B1	5	uykun mu geldi?	uykun	uyku								x

Table 4.4. Sample Video Coding Chart for Verbs

						Social Cues				Attentional Cues		Ambiguity Position		
Video	Utt.N.	Utterance	Action Presented	Action Referred	Non-ostensive Gestures	Ostensive Gestures	First Utter Then Action	Object Presented	Physical Acting	Perlocutionary Act	Kid eyes	Clearly Disambiguated	Somewhere Between	Clearly Ambiguous
B1	1	gitme	uyarmak	gitmek	x									x
B1	2	gel	gelmek	gelmek		x							x	
B1	3	burcu bak	bak	bakmak		x				x	x	x		
B1	4	yürü hadi	yürü	yürümek		x	x		x	x		x		
B1	5	kucakla hadi	kucaklamak	kucaklamak	x	x					x	x		

b. Ostensive Gestures: This feature checks whether the caregiver points to the object when s/he utters it.
c. First Utter Then Action: This feature checks whether the caregiver acts according to the verb she just uttered (But these two actions must be consecutive).
d. Objects Presented: This feature checks whether there is an object when a transitive verb is uttered.
e. Physical Acting: This feature checks whether the caregiver acts according to the verb she uttered.

Attentional Cues for Verbs

a. Perlocutionary Act: If the caregiver utters a verb that has an effect on the infant and if the infant takes action, then this feature is fulfilled.
b. Kid Eyes: If the infant is looking to the object when the name of the object is uttered, then this feature is fulfilled.

After the videos are coded according to the features above, the coders assign a position (clearly disambiguated, somewhere between, and clearly ambiguous) to the noun or verb in question according to the context. For example, when the mother utters the word "*abla"* and looks the girl in the context, points her and the child looks the girl; then this noun is said to be clearly disambiguated in the context. The videos were also coded off-line by a native Turkish speaker and in order to obtain the reliability of our coding of videos, a second coder produced independent annotations for two representative videos (one for the nouns and one for the verbs). Since coders were free to assign multiple features to each word, we assumed that words for which multiple features were indicated for a particular word contained multiple opportunities for agreement.

A statistical measure of how well our implementation of coding or measurement system works is Cohen's kappa,[11] which ranges from –1.0 to 1.0, where large numbers mean better reliability, values near zero suggest that agreement is attributable to chance, and values less than zero signify that agreement is even less than that which could be attributed to chance. Figure 4.1 shows the results of the inter-rater reliability analysis for nouns' ambiguity analysis which is kappa = 0.693 with $p < 0.01$. This measure of agreement, which is statistically significant, is convincing.

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Measure of Agreement	Kappa	,693	,077	7,177	,000
N of Valid Cases		57			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Figure 4.1. Cohen's kappa Analysis for Nouns

Figure 4.2 shows the results for verbs' ambiguity analysis which is kappa = 0.754 with $p < 0.01$. This measure of agreement is also statistically significant and convincing.

[11] A measure of reliability in an n-alternative decision that corrects for chance guessing of frequent options

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Measure of Agreement	Kappa	,754	,068	8,284	,000
N of Valid Cases		65			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Figure 4.2. Cohen's kappa Analysis for Verbs

4.2. Results

The main aim of the ambiguity analysis is to show early lexical acquisition relates to computational complexity by which unambiguous words are expected to be acquired earlier than ambiguous words. In order to measure how the context of natural language disambiguates or obscures the words that infants are exposed to, we developed video coding charts for nouns and verbs. With these video coding charts, we assume that infants use social and attentional cues for mapping the word to an object or to an action.

In the nouns' ambiguity resolution part, according to our analysis, of all the 470 utterances, there are 218 nouns which are clearly disambiguated in the context, 178 clearly ambiguous nouns and 74 nouns which are somewhere between these two. In the verb ambiguity resolution part, of all the 922 utterances, there are 449 verbs which are clearly disambiguated in the context, 332 clearly ambiguous verbs and 142 verbs which are somewhere between these two. In this analysis, there are clearly disambiguated 449 verbs but it does not mean that there are 449 verbs since many of them are the same verbs. (There are 128 different verbs and 93 different nouns which are used in different contexts).

For our ambiguity analysis to be reliable, the words that are produced by the infant as soon as she starts speaking must be the words that are clearly disambiguated by the context. Table 4.5 shows the ten most frequent words of Infant I and their frequency. It can be observed that the most frequent word is a verb and there are only three nouns which are kinship terms (the reason behind this can be the universal bias to kinship terms as explained earlier). Among the most frequent ten words of Infant I, there are seven verbs; three of which are morphological forms. Thus, there are four lexical types.

Table 4.5. Ten Most Frequent Words and their Frequencies in Infant I's Speech

Words	Frequency
bak- (intr.) "look"	469
anne- "mom"	405
abla- "sister"	378
al- "take"	200
baba- "dad"	141
almış- "has taken"	100
aç- "open"	96
aldı- "took"	65
geldi- "came"	63
açayım- "I open"	62

Table 4.6 shows the results of ambiguity analysis across the ten most frequent words in Infant I's speech. There are 78 utterances which contain *bak* "look" and of all these 78 utterances, in 69 utterances the context is very clearly disambiguated. In 5 utterances the

context is somewhere between and in 4 utterances the context is ambiguous. Therefore we claim the data is consistent with our claim that the disambiguation of *bak* by social and attentional cues is one of the reasons of its being the most frequent word. Although these most frequent forms are also mostly disambiguous forms, the reason why they are disambiguous must be considered. Thus, "bak" should indeed be unambiguous because it is usually accompanied by the noun towards which the child is supposed to look, otherwise the verb would make no sense. Likewise for the other verbs one could argue that they are obligatorily accompanied by their internal and/or external argument (object, subject). Here verbs have a natural advantage over nouns of being disambiguous. Moreover, it can be easily observed that despite the fact that the noun *baba* "father" is among the most frequent words, it is not clearly disambiguated by the context. The exceptional reason of this is that it might be serving as a proper name and in none of the video recordings the father can be seen; and since we coded the video recordings we have not coded baba as clearly disambiguated.

Table 4.6. Ambiguity Analysis of Infant I's Most Frequent Ten Words

Words	Clearly Disambiguated	Somewhere Between	Clearly Ambiguous	Total
bak- (intr.) "look"	69	5	4	78
anne- "mom"	19	3	5	27
abla- "sister"	30	3	6	39
al- "take"	33	5	3	41
baba- "dad"	0	6	29	35
almış- "has taken"	33	5	3	41
aç- "open"	9	1	0	10
aldı- "took"	33	5	3	41
geldi- "came"	64	10	32	106
açayım- "I open"	9	1	0	10

Furthermore, the word *yağmur* "rain" is not included in table 4.6 because it is not a frequent word nor is it a clearly disambiguated word. Other words can be seen from Table 4.6.

Table 4.7 shows the most frequent nouns' ambiguity analysis and Table 4.8 shows the most frequent verbs' ambiguity analysis. It can be observed from both tables infant's most of the frequent words (which means they are acquired) are disambiguated by the context. For the most frequent nouns, the most important feature is *kid eyes* for disambiguation. In other words, for disambiguation of nouns *kid eyes* is the most indicative feature. What follows *kid eyes* are the *mom eyes* and *mom point*, respectively. As for the verbs, the most important feature is again kid eyes which shows that no matter if it is a noun or verb while infants are acquiring the words, they look to the objects or reflections of the actions. The second and third important feature in the disambiguation of verbs are the non-ostensive gestures and perlocutionary act, respectively.

Table 4.7. Ambiguity Analysis of Infant I's Most Frequent Ten Nouns

Nouns	Clearly Disambiguated	Somewhere Between	Clearly Ambiguous	Total
anne- "mom"	19	3	5	27
abla- "sister"	30	3	6	39
baba- "dad"	0	6	29	35
bebek- "baby"	19	0	4	23
oyuncak- "toy"	15	1	2	18
çikolata- "chocolate"	4	0	1	5
el- "hand"	9	2	0	11
annem- "my mom"	19	3	5	27
çiş- "pee"	5	1	0	6
çanta- "bag"	18	3	6	27

Table 4.8. Ambiguity Analysis of Infant I's Most Frequent Ten Verbs

Verbs	Clearly Disambiguated	Somewhere Between	Clearly Ambiguous	Total
bak- (intr.) "look"	69	5	4	78
al- "take"	33	5	3	41
almış- "has taken"	33	5	3	41
aç- "open"	9	1	0	10
aldı- "took"	33	5	3	41
geldi- "came"	64	10	32	106
açayım- "I open"	9	1	0	10
ver- "give"	21	5	3	29
bitti-"finished"	6	2	8	16
yap- "do"	18	8	21	47

4.3. Discussion

The goal of this study was to measure the role of computational complexity by which unambiguous words are expected to be acquired first. To address this question we introduced the contributions of various sources of non-linguistic information children use to acquire words. We used a longitudinal corpus of videos of child directed speech starting at 8 months to 24 months. We annotated the videos with information about the social and attentional cues for the acquisition of noun and verb categories. We claimed that the words that the child should acquire first should be the ones that the mother has used in an unambiguous way first such that the child would acquire it easily. What one would want to see is that the most unambiguous words of the caregiver are the ones that are acquired first by the child. We founded that the words disambiguated by the context are the words that Infant I acquired no matter what their category is.

In both the nouns and verbs' ambiguity analysis, the *kid eyes* social cue had the most impact on infants' acquisition of words. The reason of this may be the caregivers' laboring, following in and talking about objects or actions that infants are interested in (Baldwin, 1991). Another possible explanation is that general continuity of conversational topics over

time determines children's view point (Frank, Johnson & Demuth, 2012). Frank et all. (2012) show that in natural speech the topic of the previous utterance is most likely the topic of the current one. Therefore children's attention has been drawn and word categories can easily be disambiguated.

CHAPTER V

PHONOLOGICAL COMPLEXITY ANALYSIS

In this part, our computational complexity assumption on early lexical acquisition claims that phonologically short words are acquired first no matter if it is a verb or a noun. By means of phonological complexity, computationalist assumption refers to word length.

The effects of phonological form on early word learning facilitate or slow down the acquisition. McMurray (2007) points that phonological complexity contributes to word difficulty. It is not straightforward to measure the impact of phonological complexity during early word learning since the basis of an infant's lexico-phonological representation is not yet well understood. However, computational complexity considers word length as a proxy for phonological complexity and hence word difficulty.

Computational mechanism asserts that infants have a computational bias towards short strings because this aspect can be shown to ease the task of early lexical acquisition. From the psychologist view any theory of immediate memory accounts for word length which says short words are faster learned than long words. Storkel & Rogers (2000) points out that variables which affect the formation of mental lexicon constrain lexical acquisition. Phonological variables such as prosody, word length and syllable structure affect an infant's ability to extract and represent the phonological form. For example, shorter words are predicted to be easier to acquire due to the facilitation at the sub-lexical level.

5.1. Methods

5.1.1. Description of AKT-T: Participants

The phonological complexity analysis is based on a sample of longitudinal corpus of one Turkish infant.[12] This infant (one girl) was recorded who was aged 0; 8 at her first recording. The child lives in the city and child caregiver interactions was video-recorded at the home of infant (Burcu), for one hour every two weeks. The child's parents both had 12 years of education. Table 5.1 provides the basic characteristics of the dataset which includes number of recordings, start and end age, total number of types-tokens in child directed speech and child speech and mean length of words in child directed speech and child speech.

[12] This sample corpus is again part of the project which was carried out by Prof. Aylin Küntay and details of the project were given at Chapter 3

Table 5.1. Characteristics of the Database

	Burcu
Number of recordings	44
Start Age	00;08;02
End Age	03;00;03
Total number of types in Child Directed Speech	14,230
Total number of tokens in Child Directed Speech	89,300
Total number of types in Child Speech	3,606
Total number of tokens in Child Speech	17,676
Mean length of words in Child Directed Speech	5,034
Mean length of words in Child Speech	3,817

5.1.2. Our Analysis of AKT-T Data: Procedure

In the phonological complexity analysis, we looked at both child directed speech and child speech closely the period when the infant uttered her first words. The procedure about video recordings and transcriptions was the same as in the frequency analysis part. After the transcriptions were coded in chat format, we divided the data into child directed speech and child speech with CLAN (Computerized Language Analysis) program (MacWhinney, 2000).

This was followed by typing commands like *wdlen* (word length) to obtain the length of words in both child directed speech and child speech. Our basic criteria for word's being long or short is the letters that the word has. For example, *bak* "look" has three letters and *çikolata* "chocolate" has eight letters. Therefore the first word is a short verb and the second word is a long noun. We have made a rather general two fold distinction (short and long) in word length since choosing a three fold distinction (short, medium and long) can cause uncertainty between short and medium and medium and long words. Therefore, words which had *2-6* letters were evaluated as short words and words which had *7-10* letters were evaluated as long words. In word length analysis, any noun or verb distinction was not made rather they were treated as words.

5.2. Results

The main aim of the phonological complexity analysis is to show early lexical acquisition relates computational complexity by which short words are expected to be acquired earlier than long words. In order to show how the word length affects early lexical acquisition, we aimed to compare infant I's child directed speech with her own words. Therefore, we divided result section into two parts; child directed speech which includes the words the child was exposed and child speech which includes the words uttered by the infant.

5.2.1. Child Directed Speech Analysis

The problem of whether infants treat all the surrounding speech as a language learning environment, or whether they prefer focusing more on the speech addressed to them is a debated topic. We included all the surrounding speech in our phonological complexity analysis. We didn't exclude the utterances which were between the caregiver (mother) and recorder (as we did at ambiguity analysis[13]). However, we prefer to visualize just the speech addressed to infant I at Table 5.2.

[13] Details of this problem were given at Chapter 4

Table 5.2. Word Length Analysis for Child Directed Speech

	Length in Characters																		
Speakers	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	*Mean*
BY1-CHI	-	44	59	14	45	7	11	5	4	2	0	0	0	0	0	0	0	0	3,984293194
BY2-CHI	-	1	16	4	3	1	3	0	0	0	0	0	0	0	0	0	0	0	3,857142857
FAT-CHI	-	91	117	92	112	52	38	17	10	5	4	2	0	0	0	0	0	0	4,412962963
GRA-CHI	-	144	411	99	213	54	45	24	13	3	3	1	1	0	0	0	0	0	3,974282888
MAN-CHI	-	2	1	4	4	0	1	0	0	0	0	0	0	0	0	0	0	0	4,166666667
MOM-CHI	-	99	185	73	103	55	22	12	12	1	5	0	3	4	0	0	0	0	4,195121951
MOT-CHI	-	1716	2513	2284	3001	1554	1102	501	412	201	118	68	25	20	5	3	2	1	4,788555375
NEI-CHI	-	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	5,8
REC-CHI	-	2256	3836	3102	5140	2364	2208	1170	901	536	279	196	78	46	21	15	2	3	5,160429739
WoM-CHI	-	15	15	3	17	5	2	0	0	0	0	1	0	0	0	0	0	0	3,931034483
SUM	-	4368	7154	5675	8639	4093	3433	1730	1352	748	409	268	107	70	26	18	4	4	4,282531454

It shows merely the results of word length analysis of the conversations between the people around and infant I According to the results, it can be said that the words that infant I was directly exposed to had *4,283* letters in average.

This shows that infant I was exposed to generally short words. Moreover, among the people with whom Infant I mostly entered into dialog, the recorder uttered words that had *5,35* letters and the mother uttered words that had *4,72* letters, averagely. The most important finding of this analysis is that as it can be observed from Table 5.2 most of the words that infant I heard had between 2 and 7 letters. Since Turkish is an inflectional language, there was a considerable amount of long words which had more than ten (+10) letters. However, the longer words frequency rate was lower than shorter words.

Table 5.3 shows the results of a comparison between word length analysis and frequency analysis. It can be seen that the nouns which were frequent in child directed speech were also phonologically short; and low frequency nouns were phonologically longer. As for the verbs, shorter verbs were high frequency and longer verbs were low frequency. In our data, there was not a long noun whose frequency rate is higher than a short noun or a short verb whose frequency rate is lower than a long verb.

Table 5.3. Frequency Word Length Comparison for Child Directed Speech

	Nouns		*Verbs*	
Number of Letters	*Low Freq.*	*High Freq.*	*Low Freq.*	*High Freq.*
2	-	*ad*	-	*al, aç, ol*
3	-	*kız*	-	*bak, gel, ver*
4	-	*abla, baba, anne*	-	*otur*
5	-	*bebek*	-	*diyor, yapma*
6	*kamera*	-	-	*geldin, uyandı*
7	*oyuncak örümcek*	-	*tutmamış*	-
8	*karanlık*	-	-	-
9	*şifonyer*	-	*şımarayım*	-
10+	*yaramazlık, üniversite*	-	*ısınıyorduk, çıldırıyorum*	-

5.2.2. Child Speech Analysis

Table 5.4 shows the results of word length analysis of the conversations between infant I and the people around. According to the results, it can be said that the words that infant I uttered had *3, 74* letters in average. This shows that infant I uttered generally short words. Moreover, among the people with whom infant I mostly entered into dialog the recorder uttered words that had 4,*346* letters and the mother uttered words that had *4,239* letters averagely. The most important finding of this analysis is that as it can be observed from table 5.4 most of the words that infant I uttered had between 2 and 6 letters.

Table 5.4. Word Length Analysis for Child Speech

	Length in Characters																		
Speakers	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	*Mean*
CHI	-	821	498	523	418	274	116	83	71	27	12	27	2	19	5	10	0	3	4,2571330
CHI-BY1	-	34	17	14	20	6	9	3	0	1	0	0	0	0	0	0	0	0	3,923076
CHI-BY2	-	2	6	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	3,272727
CHI-FAT	-	45	31	40	23	19	6	7	3	2	1	0	0	0	0	0	0	0	4,112994
CHI-GRA	-	23	27	35	25	24	6	1	6	0	0	0	0	0	0	0	0	0	4,353741
CHI-MOM	-	6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2,571428
CHI-MOT	-	584	366	636	377	184	203	81	49	33	15	7	1	5	0	1	0	0	4,322187
CHI-REC	-	2455	2058	1910	2065	1237	859	533	312	107	51	27	3	3	4	1	2	1	4,447884
SUM	-	3970	3003	3159	2930	1745	1199	708	441	170	79	61	6	27	9	12	2	4	3,74

Table 5.5 shows the results of a comparison between word length analysis and frequency analysis. It can be seen that the nouns which were frequent in child directed speech were also phonologically short; and nouns of which frequency rates were lower were phonologically longer. As for the verbs, shorter verbs had high frequency and longer verbs were low frequency.

Table 5.5. Frequency Word Length Comparison for Child Speech

	Nouns		*Verbs*	
Number of Letters	*Low Freq.*	*High Freq.*	*Low Freq.*	*High Freq.*
2	-	-	-	*al, aç*
3	-	*muz, çiş, pil*	-	*bak, ver, yap*
4	-	*anne, abla*	-	*olur*
5	-	*kaset, kalem*	-	*almış, geldi*
6	*öpücük*	*sesini*	*seyret, saklan*	-
7	*tuvalet, penguen*	-	*fırçala*	-
8	*merdiven*	-	*başlıyor*	-
9	*kütüphane*	-	*çalışmıyor*	-
10+	*televizyon*	-	*çıkartacağım*	-

5.3. Discussion

The goal of this study was to measure the role of computational complexity by which short words are expected to be acquired first. To address this question we introduced the comparison of word length analysis with frequency analysis. We used a longitudinal corpus starting at 8 months to 36 months. We founded that the shorter words had higher frequency rates in both child directed and child speech. The reason of this is that shorter words are easy both for acquisition and production.

As Storkel & Rogers (2000) points out, phonological variables affect the formation of the mental lexicon and facilitate lexical acquisition. These phonological variables are prosody, word length and syllable structure. Word length, in our study, affects an infant's ability to extract and represent the phonological form. Therefore, shorter words are easier to acquire due to the facilitation at the sub-lexical level whereas longer words are difficult both for acquisition and production.

CHAPTER VI

GENERAL DISCUSSION

The results of this preliminary study showed that there couldn't be any noun preference in terms of first words in Turkish; nor was any complexity or difficulty for verbs. Instead, computational complexity assumption demonstrated that frequency, ambiguity and word length were important parameters for early word learning. Firstly, most frequent words did not prove any noun predominance; in place of this, the rate of verbs were equal to nouns and sometimes many more than that. Furthermore, caretakers did not have any bias to nouns or object names but there was a preference to verbs during two word stage. The high rate of verbs in child speech also challenges noun-first view. Secondly, ambiguity analysis showed that social and attentional cues in the natural language's context were important factors for word learning. Therefore, disambiguated words in the context were the most frequent words in child speech. Thirdly, phonological complexity analysis indicated that word length affected the infant's ability of word learning. Thus, short words were more advantageous when compared to long words. However, ambiguity as well as word length seemed to be entangled with word frequency.

To some studies from the first words onward there is a universal noun preference from the point of early words produced and a universal noun bias from the point of early frequencies (Nelson, 1973; Huttenlocher, 1974; Greenfield & Smith, 1976; Goldin-Meadow, Seligman, Gelman 1976; Gentner, 1982; Gentner & Boroditsky, 2001; Snedeker & Gleitman, 2004). On the contrary, to some there is not any noun preference in the sense of early words nor any noun predominance in terms of frequency rates (Choi and Gopnik, 1993; Tardif et al., 1996 and 1997; Brown, 1998; de Leon, 1999; Bornstein et al., 2004; Childers et al., 2007; Tardif et al., 2008;). The lack of such an alleged universal noun bias necessitates a somewhat general explanation which should be independent of the specific languages. Our goal with this preliminary study was to address the issue from a cross-linguistic view. We studied Turkish which has different complex typological properties than English.

Computational complexity assumption on early lexical acquisition claims that frequent word types in child directed speech; unambiguous word types in the context of natural language; and phonologically short words will be acquired early.

The results of this preliminary study showed that universal noun bias can not be an alternative to early vocabulary learning, nor can any verb bias be. There can be a third position which claims that language structure emerges from language use. This usage-based model of language acquisition, first proposed by Tomasello (2003), attributes the acquisition of language to the understanding of how others use language. The primary focus for the child is the specific communicative events in which people learn and use language by

observing the actual language use. One of the main claims of usage-based approach, which is parallel with this study, is that early word learning is data driven and grammar structure emerges from patterns of use (Tomasello, 2003). The computational complexity assumption on early word learning is also a data-based approach in which the child's task can be eased via computational biases (frequency, ambiguity and shortness). Therefore, frequent word types in child directed speech are seen in child speech as do unambiguous and short words.

Gentner & Boroditsky (2001) and Snedeker & Gleitman (2004) proposed that the noun-verb distinction is not a distinction between the noun or verb category but between concrete concepts and abstract concepts. Clark (1993) and Goldberg (1998) have made another distinction between light and heavy verbs and proposed that light verbs (go, do and make) are acquired earlier than heavy verbs. These recent views about the distinction between words types is also a biased approach in that according to explanations, most of the nouns fall under the category of concrete concepts and most of the verbs to abstract concepts. Instead of a distinction between concrete and abstract concepts, we propose a distinction between unambiguous and ambiguous concepts which is more reliable. For example, is the word *gel* "come" an abstract word or concrete word? The computational complexity assumption is giving the answer: If it is disambiguated by the context, then it is an unambiguous word; if it is not disambiguated in the context, then it is an ambiguous word.

In the acquisition of Turkish, infants are exposed to child directed speech with a high frequency of verbs. It means that most of the utterances contain just verbs, and this creates a situation in which verbs have a remarkable positions in the language. Due to the advantaged positions of verbs in Turkish, verbs are learned relatively early. Thus, the children in this language adapt their early lexicon to the surrounding adult language from the beginning, gradually. As infants heard more verbs than nouns, in the child speech it is expected to produce more verbs. When we observed the utterances infants produced, it can be said that from early on there were verbs and nouns. The fluctuations of nouns and verbs during the development is the result of complex but regular verb morphology in Turkish. Therefore, one can hypothesize that infants must have somehow solved this problem on the basis of the data. And we saw that this variance, until the second year of age, is finished with the mastering of verb morphology. According to Aksu-Koç & Slobin (1985), Turkish speaking children's first productive morphology appears as early as fisrt year but not productively used after second year. We can see in our data that although noun types are dominant between the fifteenth and twent forth months, as children start to use language productively, verbs are used more. It could be that before dropping nouns (in accord with the Turkish grammar) the children first have to pass a stage where they produce them overtly. It also means that children become competent enough to use verb inflections. When the children start to speak more adult like at the end of the second year, verb types outnumbers nouns but TTR stay close. Since our explanations for the development of word categories is based on three infants and it is a fact that children vary in the way they acquire language, these results must be in doubt. However, this is a natural restriction of a longitudinal research.

The second major result is that children's TTR mirror adult distributions from early on in the acquisition process. Children, after two years, use verbs and nouns frequently. There is not any noun predominance because of any morphologically difficult verbs. Another factor that could explain the mirror effect is the number of prompting contexts (Stoll et al., 2011) where adults both draw children's attention to objects and actions. This correlates with a close TTR of nouns and verbs. The precise context effect to which children are highly sensitive is also important. According to Goldfield (1993) while playing with toys more nouns occurred than verbs whereas during playing without toys, more verbs occurred than nouns. This context effect is also shown in many studies (Tardif et al., 1999; Choi, 2000; Ogura et al., 2006). With the light of our data, we can also say that the precise context is an important factor in

Turkish children's early word learning, too. In the video sessions where toy play was observed, infants exposed more nouns whereas during just roaming or eating snacks, more verbs occurred.

Turkish children start out with verbs that are accompanied with social and attentional cues. For example, verb meanings (of *bak*) like "specifically where to look" often in company with ostension and gestures. This case entails a great amount of information about the verb uttered by their caretakers and this could decrease TTR for nouns. Of course, we do not suggest a verb first alternative for early lexical acquisition nor does Brown (1998). There are nouns which are accompanied by social and attentional cues in Turkish, too. In English and Hebrew, where the TTR for nouns is higher, children start out with verbs that have very general meanings such as *get, make, and do* (Clark, 1993; Ninio, 1999). The point is the disambiguation by the natural language context through social and attentional cues. We proposed a preliminary semantic analysis for nouns and verbs in child directed speech which uses contextual cues for detecting the meaning of a word. The results show that words disambiguated by the context are the words that a child uses productively no matter what their category is.

In addition, a number of factors contribute to the impact of semantic analysis and the most important two are the frequency and phonological complexity. Because once frequency is controlled, it is not accurate for verbs with more general meanings are acquired later or earlier than verbs with specific meanings (Theakston, Lieven, Pine & Rowland, 2004). Therefore, we compared our semantic analysis's results with the frequency analysis's results. We observed that an infant's most of the frequent words (which means they are acquired) are disambiguated by the context. We controlled these three variables' effect on early word learning and we saw that they were much related. Most of the time both frequency and ambiguity or both frequency and shortness were the determinant factors.

Phonological variables such as prosody, word length and syllable structure have important effects on the formation of the mental lexicon. These variables can ease the task of word learning and facilitate lexical acquisition. In this study, we looked at the word length effect on early word learning superficially. Word length has effects on an infant's ability to extract and represent the phonological form in our Turkish sample. Therefore, shorter words are assumed to be acquired due to the facilitation at the sub-lexical level whereas longer words are assumed to be acquired later. The facilitation effect of word length is true both for acquisition and production. In the acquisition side, there were many examples in our data such as most frequent words were always short words independent of their category. The most frequent noun across the three infants' speech has four letters (anne "mom") and the most frequent verb has three letters (bak "look"). On the production side, most of the mothers observe that their children shorten words like *ka* for kaplumbağa "turtle" in Turkish. This kind of production shows that word length effect is an important factor for production also and children prefer to shorten long words. Even for the most frequent verb *bak* which has just three letters, children, at early periods prefer to use *bı.* It may also be a habit coming from babbling.

CHAPTER VII

CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

The present study aimed to investigate the acquisition of the early lexicon and its underlying mechanism from computational complexity perspective by which short, frequent and unambiguous words are supposed to be acquired first; and long, ambiguous or infrequent words (including nouns) are not predicted to be acquired early. It was found out that frequent word types in child directed speech were also frequent in child speech. Moreover, since Turkish infants were exposed more verb types, they produced more verbs. From the one word stage to the two word stage, Turkish infants produced more nouns compared to verbs. Therefore, it would be informative to investigate which underlying social/cognitive mechanism makes Turkish infants produce more nouns during this process and which contextual factors support the development of these noun predominance. Furthermore, the lexical type token ratio analysis's findings are interesting since in child directed speech there were more nouns and in child speech, after two years, nouns become dominant for one Infant. Since this study is a preliminary one, we did not conducted lexical type token ratio for other two children. However, further research is necessary to investigate lexical TTR analysis for more infants.

In the present study's ambiguity analysis part, we did not code the surrounding speech as language learning environment, we just coded the speech addressed to them. Yet, it would be interesting to investigate the role of surrounding speech against the speech addressed to them. Both of them should have different contributions on the development of the early lexicon. Also, it was very interesting that according to our ambiguity analysis, the word *baba* "father" which was a frequent word both in child directed speech and child speech, was not clearly disambiguated by the context. The reason of this is that because father was not seen in any of the sessions, we did not code it. There must have been many words like this and they were not frequent. Therefore, for future researches, the video coding criteria must include the words that cannot be directly observed in the context. Moreover, for the disambiguation of words we used some social and attentional cues and found that *kid eyes* attentional cue was the most important cue for children. If the ambiguity analysis was done for the other two infants and results were turned a database to which some machine learning algorithms were applicable, it would be interesting to investigate which social and attentional cues would be important for which word category.

In addition to frequency and ambiguity, the present study presented valuable information on the effect of phonological complexity on early word learning. We showed that short words in child directed speech were acquired easily and children's most frequent words were all short words. We conducted phonological complexity analysis on one infant and if we used other two children, results would be more explanatory. This is a promising research topic and further studies can explain whether what we found is a robust finding and, to what extent, it might be generalized the other languages. Moreover, when looking at the

phonological complexity's effect on word learning, we had a general view point to word types. Therefore, we didn't count the letters of nouns or verbs or any other category individually both for child directed speech and child speech. Counting the letters of nouns and verbs both for the language heard and produced would be time consuming but at the same time very useful.

Lastly, for further research, assembling a data according to our computational complexity assumptions would be very precious and beneficiary for human computer learning mechanisms and a basis for future artificial intelligence studies.

In order to answer long-standing developmental questions with the principles of our study, collaborative efforts of researchers in several different disciplines and sub-disciplines such as language sciences (theoretical linguistics, psycholinguistics and language acquisition), computer science, psychology and cognitive science is required. We have tried to approach the problem of early lexical acquisition in young Turkish infants from an interdisciplinary perspective.

REFERENCES

Aksu Koç, A. and Slobin, I.D. (1985). "The Acquisition of Turkish" The Cross linguistic Study of Language Acquisition. Lawrence Erlbaum Associates Inc. Publs. Hillsdale, New Jersey s.839-878.

Ambridge, B., Pine, M. J., Rowland, F. C., Young, R. C. (2008a). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgments of argument structure overgeneralization errors. Cognition 106 (1), 87–129.

Ambridge, B., Rowland, F. C., Pine, M. J. (2008b.) Is structure dependence an innate constraint? New experimental evidence from children's complex question production. Cognitive Science 32 (1), 222–255.

Arunachalam, S., & Waxman, S. R. (2011). Grammatical form and semantic context in verb learning. Language Learning and Development, 7, 169–184.

Bachman, L. (1990). Fundamental considerations in language testing. Oxford: Oxford University Press.

Bachman, L., & Palmer, A. (1996). Language testing in practice. Oxford: Oxford University Press.

Baldwin, D. A. (1991). Infant contributions to the achievement of joint reference. Child Development 62, 875±890. [Reprinted in P. Bloom, ed., Language acquisition. Cambridge, Mass.: MIT Press, 1994.

Baldwin, D. A., Markman, E. M., & Melartin, R. L. (1993). Infants' ability to draw inferences about nonobvious object properties: Evidence from exploratory play. Child Development, 64(3), 711-728.

Bassano, D., Maillochon, I., & Eme, E. (1998). Developmental changes and variability in the early lexicon: a study of French children's naturalistic productions. Journal of Child Language 25, 493–531.

Bassano, D. (2000). Early development of nouns and verbs in French: exploring the interface between lexicon and grammar. Journal of Child Language 27, 521–559.

Bates, E., Bretherton, I., & Snyder, L. (1988). From first words to grammar: individual differences and dissociable mechanisms. Cambridge: Cambridge University Press.

Bates, E., & Carnevale, G.F. (1993). New directions in research on language development. Developmental Review, 13, 436-470.

Bates, E., P. S. Dale, and D. Thal. (1995). Individual differences and their implications for theories of language development. In P. Fletcher and B. MacWhinney, eds., Handbook of child language. Oxford: Blackwell.

Benedict, H. (1979). Early lexical development: Comprehension and production. Journal of Child Language 6, 183±201.

Bloom, L., Lifter, K., & Hafitz, J. (1980). The semantics of verbs and the development of verb inflections in child language. Language, 56, 386–412.

Bloom, L., Tinker, E., & Margulis, C. (1993). The words children learn: Evidence against a noun bias in early vocabularies. Cognitive Development 8, 431–450.

Bloom, P. (1994b). Possible names: The role of syntax-semantics mapping in the acquisition of names. Lingua 92, 297±329. [Reprinted in L. R. Gleitman and B. Landau, eds., The acquisition of the lexicon. Cambridge, Mass.: MIT Press.

Bloom, P. (2002). Mindreading, communication, and the learning of the names for things. Mind and Language, 17, 37-54.

Bornstein, M., Cole, L., Maital, S., K., Park, S. Y., Pascual, L. (2004). Cross linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean and American English. Child Development, 75, 1115–1140.

Bozşahin, C. (2012). Combinatory Linguistics. Berlin/Boston: Mouton de Gruyter.

Brent, M. and Siskind, J. (2001). The role of exposure to isolated words in early vocabulary development. Cognition, 81:B33–B44.

Brown, R. (1957). Linguistic determinism and part of speech. Journal of Abnormal and Social Psychology 49, 454±462.

Brown, P. (1998). Children's first verbs in Tzeltal: evidence for an early verb category. Linguistics 36, 713–753.

Bruner, J. S. (1978). From communication to language: A psychological perspective. In I. Markova, ed. The social context of language. New York: Wiley.

Camaioni, L., & Longobardi, E. (2001). Nouns versus verb emphasis in Italian mother-to-child-speech. Journal of Child Language 28, 773–785.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. Monographs of the Society for Research in Child Development, 63 (4, Serial No. 255)

Carroll, J.B. (1964). Language and thought. Englewood Cliffs, NJ: Prentice Hall.

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. Cognitive Development 10, 159–199.

Caselli, C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. Journal of Child Language 26, 69–111.

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. Developmental Psychology, 38, 967–978.

Childers, J. B., Vaughan, J., & Burquest, D. A. (2007). Joint attention and word learning in Ngas-speaking toddlers in Nigeria. Journal of Child Language 34, 199–225.

Choi, S., & Gopnik, A. (1993). Nouns are not always learned before verbs: an early verb spurt in Korean. In E. V. Clark (Ed.). The proceedings of the 25th annual Child Language Research Forum (pp. 96–105). New York, NY: Cambridge University Press.

Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in Korean: a cross-linguistic study. Journal of Child Language 22, 497–529.

Chomsky, N. (1975). The logical structure of linguistic theory. New York: Plenum.

Chomsky, N. (1981). Lectures on government and binding. Dordrecht: Foris.

Chomsky, N. (1986). Knowledge of language: Its nature, origin, and use. New York: Praeger.

Christophe, A., and E. Dupoux. (1996). Bootstrapping lexical acquisition: The role of prosodic structure. The Linguistic Review 13, 383±412.

Chukovsky, K. (1968). From two to five. Berkeley: University of California Press, 1968.

Clark, E. (1993). The lexicon in acquisition. Cambridge: Cambridge University Press.

Clark, E. (2003). First language acquisition. New York: Cambridge University Press.

Çöltekin, Ç. and Bozşahin, C. (2007). Syllable-based and morpheme-based models of Bayesian word grammar learning from CHILDES database. In Proc.of the 29th Annual Meeting of Cognitive Science Society. Nashville, TN.

de Léon, L. (1999). Verb roots and caregiver speech in early Tzotzil (Mayan) acquisition. In B. Fox, D. Jurafsky, & L. Michaelis (Eds.) Cognition and function in language (pp. 99–119). Stanford: CSLI.

de Villiers, J. G. & de Villiers, P. A. (1985). Acquisition of English. In D. Slobin (Ed.), The cross linguistic study of language acquisition: Vol. 1. The data (pp. 27-140). Hillsdale, NJ: Lawrence Erlbaum.

Dewey, J. (1894). The psychology of infant language. Psychological Review, 1, 63-66.

Diessel, H., Tomasello, M., (2001). The acquisition of finite complement clauses in English: a corpus-based analysis. Cognitive Linguistics 12, 97–141.

Dromi, E. (1987). Early lexical development. Cambridge: Cambridge University Press.

Ekmekçi, F. Ö. (1982). Language development of a Turkish child: a speech analysis in terms of lenght and complexity. Journal of Human Sciences, 2, 103-112.

Elman, Jeffrey. (1990). Finding structure in time. Cognitive Science, 14:179–211.

Elman, Jeffrey. (1996). Rethinking Innateness: A Connectionist Perspective on Development. Cambridge, MA: MIT Press.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D., & Pethick, S. J. (1994).Variability in early communicative development. Monographs of the Society for Research in Child Development, 59 (5, Serial No. 242).

Fletcher, P. (1985). A child's learning of English. Oxford: Blackwell.

Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. Language Learning & Development, 9, 1–24.

Frawley,W. (1992). Linguistic semantics. Hillsdale, NJ: Lawrence Erlbaum.

Gallivan, J. (1988). Motion verb acquisition: Development of definitions. Perceptual and Motor Skills, 66, 979–986.

Gelman, S. A., & Tardif, T. (1998). A cross-linguistic comparison of generic noun phrasesin English and Mandarin. Cognition, 66, 215–248.

Gentner, D. (1978a). On relational meaning: The acquisition of verb meaning. Child Development, 49, 988-998.

Gentner, D. (1978b). A study of early word meaning using artificial objects: What looks like a jiggy but acts like a zimbo? Papers and Reports on Child Language Development, 15, 1-6, Stanford University. (Reprinted in J. Gardner (Ed.), Readings in developmental psychology, (2nd ed., pp. 137-142). Boston: Little Brown).

Gentner, D. (1981). Some interesting differences between nouns and verbs. Cognition and Brain Theory, 4, 161–178.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj II (Ed.), Language development: Vol. 2. Language, thought, and culture (pp. 301–334). Hillsdale, NJ: Lawrence Erlbaum.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity and early word learning. In M. Bowerman & S. C. Levinson (Eds.), Language, culture, and cognition: Vol. 3. Language acquisition and conceptual development (pp. 215–256). New York: Cambridge University Press.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. Cognition, 73, 135–176.

Gleitman, L. R. (1990). The structural sources of verb meaning. Language Acquisition 1, 3±55. [Reprinted in P. Bloom, ed., Language acquisition. Cambridge, Mass.: MIT Press, 1994.]

Gold, E. M. (1967). Language identification in the limit. Information and Control, 16:447–474.

Goldberg, A. E. (1998). Patterns of experience in patterns of language. In M. Tomasello (Ed.). The new psychology of language (pp. 203–218). Mahwah, NJ : Lawrence Erlbaum.

Goldfield, B. A. (1993). Noun bias in maternal speech to one-year-olds. Journal of Child Language 20, 85–99.

Goldin-Meadow, S., Seligman, M. E. P., & Gelman, R. (1976). Language in the two-year-old: Receptive and productive stages. Cognition, 4, 189–202.

Golinkoff, R. M., & Hirsh-Pasek, K. (2006). The emergentist coalition model of word learning in children has implications for language in aging. In E. Bialystok, & F. Craik (Eds.), Lifespan cognition: Mechanisms of change (pp. 207-222). New York, NY: Oxford University Press.

Gopnik, A. (1981). Development of non-nominal expressions in 1-2-year-old: why the first words aren't about things. In P. S. Dale, & D. Ingram (Eds.) Child language - an international perspective (pp. 93–104). Baltimore, Maryland: University Park Press.

Gopnik, A. (1988). Three types of early words: the emergence of social words, names and cognitive-relational words in the one-word stage and their relation to cognitive development. First Language 8, 49–69.

Gopnik, A. & Choi, S. (1990). Do linguistic differences lead to cognitive differences? A cross-linguistic study of semantic and cognitive development. First Language, 10, 199-215.

Gopnik, A. & Choi, S. (1995). Names, relational words and cognitive development in English and Korean Speakers: Nouns are not always learned before verbs. In M. Tomasello & W. Merriman (Eds.) Beyond names for things: Young children's acquisition of verbs. New Jersey: Erlbaum.

Göksel, A. and C. Kerslake (2005). Turkish, A Comprehensive Grammar. London: Routledge.

Greenfield, P., & Smith, J. (1976). The structure of communication in early language development. New York: Academic Press.

Haspelmath, M. (2002). Understanding Morphology. London: Arnold.

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R., Brand, R., Brown, E., Chung, H. L., et al. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. Monographs of the Society for Research in Child Development, 65(3).

Huttenlocher, J. (1974). The origins of language comprehension. In R. Solso (Ed.), Theories in cognitive psychology. Potomac, MD: Erlbaum.

Huttenlocher, J., Smiley, P., & Charney, R. (1983). Emergence of action categories in the child: Evidence from verb meanings. Psychological Review, 90, 72–93.

Imai, M., Haryu, E., & Okada, H. (2005). Mapping novel nouns and verbs onto dynamic action events: Are verb meanings easier to learn than noun meanings for Japanese children? Child Development, 76(2), 340–356.

Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R., & Shigematsu, J. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. Child Development 79, 979–1000.

Inhelder, B., and J. Piaget. (1964). The early growth of logic in the child. London: Routledge and Kegan Paul.

Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E., & Gutierrez-Clellen, V. (1993). Early lexical development in Spanish-speaking infants and toddlers. Journal of Child Language, 20, 523–549.

Jakobson, R. (1968). Child language, aphasia and phonological universals. The Hague: Mouton.

Johnson, M., Demuth, K., & Frank, M. C. (2012). Exploiting social information in grounded language learning via grammatical reduction. Proceeding of the Association for Computational Linguistics

Kachergis, G., Yu, C., & Shiffrin, R. M. (2013). Actively Learning Object Names Across Ambiguous Situations. Topics in Cognitive Science, 5(1), 200-213.

Kauschke, C., & Hofmeister, C. (2002). Early lexical development in German: a study on vocabulary growth and vocabulary composition during the second and third year of life. Journal of Child Language 29, 735–757.

Kelly, B. F. (2006). The development of constructions through early gesture use. In E.V. Clark and B.F Kelly (Eds.). The Acquisition of Constructions. CSLI Press. Pp1-11.

Kersten, A.W., & Smith, L. B. (2002). Attention to novel objects during verb learning.Child Development, 73, 93–109.

Kim, M., McGregor, K., & Thompson, C. (2000). Early lexical development in English-and Korean-speaking children: language-general and language-specific patterns. Journal of Child Language 27, 225–254.

Kucera, H. & Francis, W. (1967). Computational analysis of present day American English. Providence, RI: Brown University Press.

Landau, B., & Gleitman, L. R. (1985). Language and experience: Evidence from the blindchild. Cambridge, MA: Harvard University Press.

Langacker, R.W. (1987). Foundations of cognitive grammar: Vol. 1. Theoretical prerequisites. Stanford, CA: Stanford University Press.

Lieven, E. V. M. (1978a). Conversations between Mothers and Young Children: individual differences and their implications for the study of language learning. In N. Waterson & C. Snow (Eds.), The Development of Communication (pp.173-187). London: John Wiley.

Lieven, E. V. M. (1994). Cross linguistic and cross-cultural aspects of language addressed to children. In C. Gallaway & B. J. Richards (Eds.), Input and Interaction in Language Acquisition (pp. 56-73). Cambridge: CUP.

Locke, J. (1983). Phonological acquisition and change. New York: Academic Press.

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

Macnamara, J. (1982). Names for things: a study of human learning. Cambridge, MA: Bradford Books.

MacWhinney, Brian (2000). The CHILDES Project: Tools for Analyzing Talk. Mahwah, NJ: Lawrence Erlbaum Associates. ISBN 0-8058-1005-6. Retrieved 1 May 2009.

Markman, E. M. (1994). Constraints children place on word meanings. In P. Bloom, ed., Language acquisition. Cambridge, Mass.: MIT Press. [Reprinted from Cognitive Science 14, 57±77, 1990.]

Matthews, D., Lieven, E., Theakston, A., Tomasello, M. (2005). The role of frequency in the acquisition of English word order. Cognitive Development 20 (1), 121–136.

McMurray, B. (2007). Defusing the childhood vocabulary explosion. Science, 317(5838), 631

Mehler, J., P. Jusczyk, G. Lambertz, N. Halsted, J. Bertoncini, and C. Amiel-Tison. (1988). A precursor of language acquisition in young infants. Cognition 29,144±178.

Miller, J. F. (1981). Assessing language production in children: experimental procedures. London: Edward Arnold.

Moon, C., Lagercrantz, H. and Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: a two-country study. Acta Paediatrica, 102: 156–160. doi: 10.1111/apa.12098

Morgan, J. L. (1986). From simple input to complex grammar. Cambridge, Mass.: MIT Press.

Naigles, L., Hoff-Ginsberg, E., (1998). Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. Journal of Child Language 25 (1), 95–120.

Nelson, K. (1973). Structure and strategy in learning to talk, vol. 38 of Monographs of the Society for Research in Child Development. London: Blackwell.

Nelson, K., Hampson, J., & Shaw, L. (1993). Nouns in early lexicons: evidence, explanations, and extensions. Journal of Child Language 20, 61–84.

Ninio, A. (1999). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. Journal of Child Language 26, 619–653.

Ogura, T., Dale, P., Yamashita, Y., Murase, T., & Mahieu, A. (2006). The use of nouns and verbs by Japanese children and their caregivers in book-reading and toy-playing contexts. Journal of Child Language 33, 1–29.

Oviatt, S. L. (1980). The emerging ability to comprehend language: An experimental approach. Child Development 51, 97±106.

Peters, A. M. (1983). The units of language acquisition. Cambridge: Cambridge University Press.

Pickett, J. P., (Eds.). (2000). The American Heritage Dictionary of the English Language (4th ed.). Boston: Houghton Mifflin.

Pinker, S. (1989). Learnability and cognition. Cambridge, Mass.: MIT Press.

Popper, K. (1963). Conjectures and Refutations, Routledge, London.

Quine, W.V. (1960). Word and object. Cambridge, MA: MIT Press.

Richards, B. (1987). Type/Token Ratios: what do they really tell us?. Journal of Child Language, 14, pp 201209.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), Cognitive development and the acquisition of language. New York: Academic Press.

Rosch, E. (1975). Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 1975, 104, 192-233.

Rowland, C.F., Pine, J.M., (2000). Subject-auxiliary inversion errors and wh-question acquisition: 'what children do know!' Journal of Child Language 27 (1), 157–181.

Siskind, Jeffrey. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. Cognition, 61:39–91.

Slobin, D. I. (1975). The more it changes . . . On understanding language by watching it move through time. Papers and Reports on Child Language Development (Stanford University), 1975, 10, 1-30.

Slobin, D. I. (2001). Form-function relations: How do children find out what they are? In M. Bowerman & S. C. Levinson (Eds.), Language acquisition and conceptual development (pp. 406–449). Cambridge: Cambridge University Press.

Snedeker, J., & Gleitman, L. (2004). Why is it hard to label our concepts? In G. Hall & S.Waxman (Eds.). Weaving a lexicon (pp. 603–636). Cambridge, MA: MIT Press.

St. Augustine. (397/1963). The Confessions of St. Augustine (R. Warner, Ed.). New York, NY: Clarendon Press.

Steedman, Mark, and Julia Hockenmaier. (2007). The computational problem of natural language acquisition. Ms., University of Edinburgh.

Stein, C. (1957). Lectures In America. Boston: Beacon Hill.

Stoll, S., Balthasar, B., Lieven, E., Banjade, G., Bhatta, N. T., Gaenszle, M., et al. (2012). Nouns and verbs in Chintang: children's usage and surrounding adult speech. Journal of Child Language, 39, pp 284-321.

Storkel, H. L., & Rogers, M. A. (2000). The effect of probabilistic phonotactics on lexical acquisition. Clinical Linguistics & Phonetics, 14, 407–425.

Supalla, T., & Newport, E. (1978). How many seats in a chair? The derivation of nouns and verbs in American Sign Language. In P. Siple (Ed.), Understanding Language through Sign Language Research. Academic Press.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), Language typology and the lexicon, Vol. III: Grammatical categories and the lexicon (pp. 57–149). Cambridge: Cambridge University Press.

Talmy, L. (2000). Toward a cognitive semantics: Vol. I. Conceptual structuring systems. Cambridge, MA: MIT Press.

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. Developmental Psychology, 32, 492–504.

Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: a comparison of English, Italian, and Mandarin. Journal of Child Language 24, 535–565.

Tardif, T., Gelman, S.A., & Xu, F. (1999). Putting the "noun bias" in context: a comparison of English and Mandarin. Child Development, 70, 620-635.

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., Marchman, V. (2008). Baby's first ten words. Developmental Psychology 4, 929-938.

Taylor, M., and S. A. Gelman. (1989). Incorporating new words into the lexicon: Preliminary evidence for language hierarchies in two-year-old children. Child Development 59, 411±419.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2002). Going, going, gone: The acquisition of the verb "go." Journal of Child Language, 29, 783–811.

Tomasello, M. (1992). First Verbs: A Case Study of Early Grammatical Development. Cambridge University Press.

Tomasello, M., & Kruger, A. C. (1992). Acquiring verbs in ostensive and nonostensive contexts. Journal of Child Language, 19, 311–333.

Tomasello, M., & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. Cognitive Development 10, 201–224.

Tracy, F. (1893). The language of child hood. American Journal of Psychology. 1893, 6.

Waxman, S. R., & Klibanoff, R. S. (2000). The role of comparison in the extension of novel adjectives, Developmental Psychology, 36, 571–581.

Whorf, B. L. (1956). Language, thought and reality. Cambridge, Mass.: MIT Press.

Yu, C. & Smith, L. B. (2007). Rapid Word Learning under Uncertainty via Cross-Situational Statistics. Psychological Science, 18(5), 414-420.

Zettlemoyer, Luke S., and Michael Collins. (2005). Learning to map sentences to logical form: Structured classification with Probabilistic Categorial Grammars. In Proc. of the 21st Conf. on Uncertainty in Artificial Intelligence. Edinburgh.

APPENDIX

RAW COUNTS OF NOUNS AND VERBS USED PER ADULTS AND PER TARGET CHILDREN FOR EACH RECORDING CYCLE

Child	Age	Child				Adults			
		types		tokens		types		tokens	
		nouns	verbs	nouns	verbs	nouns	verbs	nouns	verbs
1	00;08;02	0	0	0	0	87	103	336	334
	01;00;03	0	1	0	3	47	79	247	471
	01;00;17	0	0	0	0	30	38	97	138
	01;01;15	0	0	0	0	101	121	495	396
	01;02;05	2	1	6	4	51	57	219	238
	01;02;23	3	2	27	13	57	92	210	389
	01;03;09	5	2	23	11	73	86	358	357
	01;03;26	6	2	38	4	70	68	293	277
	01;04;10	0	0	0	0	98	121	460	456
	01;05;11	2	2	6	13	84	86	321	288
	01;06;00	1	1	2	13	84	100	305	386
	01;06;17	0	1	0	2	42	48	74	198
	01;07;00	4	1	17	2	41	52	169	217
	01;07;14	3	0	15	0	41	48	121	188
	01;08;01	6	3	23	7	43	57	149	244
	01;08;16	2	1	6	2	36	46	131	208
	01;10;00	4	4	56	12	42	82	172	370
	01;10;14	1	0	6	0	41	54	82	175
	01;10;28	4	3	8	6	45	59	170	250
	01;11;27	1	2	5	7	67	57	200	217
	02;00;12	2	1	19	5	31	38	103	119
	02;00;26	2	1	17	2	62	58	241	261
	02;01;17	1	0	4	0	42	59	123	211
	02;02;01	2	4	6	18	44	65	182	291
	02;02;14	6	16	34	65	58	67	193	261
	02;02;29	8	10	25	57	26	55	87	216
	02;03;12	15	21	50	78	44	67	146	361
	02;03;28	10	15	106	81	102	97	358	404
	02;04;09	28	38	233	186	94	127	429	622
	02;04;23	19	35	125	145	119	123	445	426
	02;05;05	22	49	133	185	98	138	391	630
	02;05;21	15	22	96	102	98	136	366	612
	02;06;05	22	37	124	209	61	106	237	526

Age		Child				Adults			
		types		tokens		types		tokens	
		nouns	verbs	nouns	verbs	nouns	verbs	nouns	verbs
	02;06;20	10	33	92	154	79	120	283	556
	02;07;01	12	27	37	93	98	101	345	422
	02;07;17	22	33	71	104	109	134	403	580
	02;08;02	21	29	95	120	73	86	320	355
	02;08;28	41	74	175	315	79	121	329	561
	02;09;24	35	43	180	171	87	104	341	411
	02;10;08	37	51	161	211	53	81	187	285
	02;10;22	45	85	192	375	66	112	260	479
	02;11;04	48	63	152	282	62	76	209	276
	02;11;20	24	48	106	194	67	124	307	502
	03;00;03	24	42	76	139	68	104	300	457
2	00;07;28	0	0	0	0	131	127	486	597
	00;08;13	0	0	0	0	99	98	446	476
	00;10;22	0	0	0	0	89	115	451	539
	00;11;22	0	1	0	10	79	104	454	592
	01;00;19	2	1	15	3	35	28	160	157
	01;04;23	4	1	32	4	51	51	308	192
	01;05;08	4	2	9	4	104	110	512	434
	01;07;17	4	6	27	24	110	144	693	697
	01;08;03	8	8	25	28	109	121	519	568
	01;08;17	2	0	4	0	49	51	198	190
	01;09;28	1	3	3	8	88	65	376	263
	01;10;18	11	6	55	34	78	93	451	497
	01;11;12	16	8	51	33	84	47	356	236
	01;11;29	16	20	43	90	78	129	377	597
	02;00;20	33	30	166	167	131	152	711	692
	02;01;20	29	44	137	249	105	142	472	708
	02;03;29	43	63	202	240	92	114	398	531
	02;04;22	28	65	111	253	87	121	367	569
	02;05;10	28	62	129	286	61	134	410	545
	02;06;14	29	69	125	259	79	130	436	548
	02;07;03	31	68	87	262	87	152	356	636
	02;07;20	36	67	189	299	127	126	550	593
	02;08;03	46	86	185	349	62	100	262	381
	02;09;04	63	74	236	245	74	105	322	369
	02;09;19	54	81	303	378	69	127	365	546
	02;10;07	28	44	121	223	114	114	522	599

	Age	Child				Adults			
		types		tokens		types		tokens	
		nouns	verbs	nouns	verbs	nouns	verbs	nouns	verbs
3	00;08;01	0	0	0	0	128	108	782	565
	00;09;10	0	1	0	3	40	47	136	157
	00;09;26	0	0	0	0	55	52	262	283
	00;10;12	0	0	0	0	96	106	438	476
	00;10;28	0	0	0	0	74	113	312	486
	00;11;07	0	1	0	4	70	85	279	403
	00;11;23	0	0	0	0	80	63	337	289
	01;00;08	0	0	0	0	107	128	589	558
	01;00;25	0	0	0	0	104	107	464	565
	01;01;10	0	1	0	3	146	149	809	803
	01;01;22	0	0	0	0	29	41	128	241
	01;02;06	0	0	0	0	129	150	571	792
	01;02;22	0	0	0	0	71	103	334	568
	01;03;14	9	6	52	18	121	118	611	625
	01;03;28	9	4	58	20	147	147	647	797
	01;04;11	13	2	82	7	100	146	426	733
	01;04;24	7	5	47	12	128	151	493	685
	01;05;08	12	8	46	29	111	161	580	697
	01;05;22	16	2	78	9	152	140	672	749
	01;06;08	25	7	130	34	139	147	662	725
	01;06;20	20	5	81	15	130	127	595	695
	01;07;05	27	14	115	43	134	169	726	772
	01;07;19	19	21	101	73	85	105	361	507
	01;08;03	42	20	167	56	164	134	669	654
	01;08;16	19	7	69	18	81	100	432	440
	01;08;21	33	9	124	24	133	155	679	650
	01;09;14	41	21	159	82	116	113	593	604
	01;09;28	43	34	139	99	118	115	433	479

TEZ FOTOKOPİ İZİN FORMU

ENSTİTÜ

Fen Bilimleri Enstitüsü ☐

Sosyal Bilimler Enstitüsü ☐

Uygulamalı Matematik Enstitüsü ☐

Enformatik Enstitüsü ☐

Deniz Bilimleri Enstitüsü ☐

YAZARIN

Soyadı : .....
Adı : .....
Bölümü : .....

TEZİN ADI (İngilizce) : .....
.....
.....
.....
.....

TEZİN TÜRÜ : Yüksek Lisans ☐ Doktora ☐

1. Tezimin tamamı dünya çapında erişime açılsın ve kaynak gösterilmek şartıyla tezimin bir kısmı veya tamamının fotokopisi alınsın. ☐
2. Tezimin tamamı yalnızca Orta Doğu Teknik Üniversitesi kullancılarının erişimine açılsın. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.) ☐
3. Tezim bir (1) yıl süreyle erişime kapalı olsun. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.) ☐

Yazarın imzası ..... Tarih .....