Graduate thesis

NTNU
Norwegian University of Science and Technology
Faculty of Humanities
Department of Language and Literature

Brooke Foss

Object Labeling and Categorization: Noun Acquisition in Children

Graduate thesis in English Linguistics and Language Acquisition
Supervisor: Mila Vulchanova
November 2021

NTNU
Kunnskap for en bedre verden

Brooke Foss

Object Labeling and Categorization: Noun Acquisition in Children

Graduate thesis in English Linguistics and Language Acquisition
Supervisor: Mila Vulchanova
November 2021

Norwegian University of Science and Technology
Faculty of Humanities
Department of Language and Literature

NTNU
Norwegian University of
Science and Technology

Abstract

There are many studies that point to children learning and producing nouns earlier than words in other categories because of the nature of language directed at infants by those around them. Children pay attention to things in their environment or have their attention directed to them by others. If a child's attention is repeatedly drawn or directed to a particular object they will most likely learn and produce the word associated with it earlier than a word or object they do not interact with as often. Since concrete objects and their labels tend to be nouns, this category is among the majority of words that children receive and produce early on. To examine this link between language and categorization and early production of nouns, CHILDES database is used as a dataset. The social aspect of language learning is a factor in nouns being learned early. Adults and peers alike reference objects in the surrounding environment that the child can interact with. This concrete representation of a word leads to learning of the specific word. The nouns received and produced by children aged 12-60 months is examined in comparison to verbs, adjectives, and adverbs. Studies conducted with children aged 12 months to 60 months provided the data for the study. The study data was split into 4 age dyads to compare adult input to child language production as the child progressed in their language development. To examine the development of specific nouns and object labels I pulled words from tests meant to establish child vocabulary skills: the McArthur-Bates CDI, and Pearson CELF-5. The CDI is meant to be used with children up to 30 months, and the CELF-5 for child 60 months and older. A visual survey of the data was conducted on the frequency of production by adults and peers in comparison of the production by children. The visual survey conducted indicates that frequency of input from adults does indeed influence children to acquire words and object labels in the noun category from an early age.

Acknowledgments

This thesis has been more academically challenging that I imagined it would. I would like to take the time to thank everyone who made this thesis possible.

First, I would like to thank the academics who helped me during this process: Mila, Giosuè, Evelyn, Lindsay, and Brittany. The collective encouragement and guidance I received made this whole process seem more manageable and less daunting. Each one of you contributed something to this thesis and I am grateful.

Second, I am very grateful for the Women in Language Science Club and my classmates. The club was a wonderful place to connect with other women who felt the same struggles I had. I met some great friends there who have continued to support me even though the club has dissolved, and some have graduated and moved on to other things. Thank you to Evelyn and Isabella for creating that space for us. I do not know how classmates in other programs interact, but I felt like mine were very kind and helpful to one another. Always eager to share and connect over our work. A very special thanks to Laura, Natalie, Akvilè, Miriam, Anya—I am so happy we became friends and could share our thesis woes together.

Third, I would be remiss if I forgot those who mean the most to me. To my friends and family who supported me from afar, and when I had to unexpectedly move home. I definitely would not have even considered moving to another country to pursue my degree without your support. So, mom and dad, Zoe, Naomi, AJ and UJ, Rushele, Jenn, and Neena thank you so much for the encouragement and help while I was in Trondheim and missing you all. I cannot forget my Norwegian family and friends. Thank you Harald for helping your American cousin to navigate Norway. To Miriam and Mikhail thanks for all the long talks and game nights. To everyone else I have neglected to mention thank you. Without the support of all of you this would not have been possible.

Table of Contents

1. Introduction

Think back to what your first words were. Were they possibly *mommy* or *daddy*, or the name of a favorite toy? Odds are if you grew up in a monolingual English speaking home your first words were most likely object labels that would be categorized as nouns. Infants are exposed to a stream of sound that they learn to segment into individual words. Once a word has been established, the child then works to identify the concept behind the word. Simultaneously, the child begins to disambiguate meaning of words and intent of speaker, and place words into different categories. There are many studies that have been conducted that show that children first produce more nouns than any other category—with the next closest category being verbs (Tomasello & Olguin, 1993; Tomasello et al., 1997; Tomasello et al., 2007; Waxman & Booth, 2001; Goodman et al., 2008; Twomey & Westermann, 2018; Laing & Bergelson, 2020; Ackermann et al., 2020). The conclusions as to why this category is generally acquired early have been varied. Some believe it to merely be a coincidence of commonality—caregivers draw an infant's attention to objects consistently. Others believe that an infant just naturally pays more attention to objects in their immediate environment. These studies do not all seek to answer why children acquire nouns more quickly and easily than other categories, but they do confirm similar reasons as to why nouns seem to be so easily acquired by infants:

i. language is socially oriented
ii. input from fluent language users at or above the child's production ability is necessary
iii. the situation and/or environment influences the type of input children receive
iv. nouns are learned and produced early because they are easily assimilated to the noun category through object labels

The focus of this study is the acquisition of words in the noun category only. I chose this category because they are produced accurately and consistently by infants before any other category and will therefore be easier to find in the CHILDES database. I will not be comparing the acquisition of nouns to any other category, such as verbs which are also among the early words children produce. I am not attempting to prove that nouns are among the first words children acquire. Instead, this study is an examination of the frequency of nouns children hear as input and the subsequent frequency of child production of nouns. The acquisition of nouns

cannot explain every aspect of child language development, but they are an important steppingstone to the establishment of word to category link mapping that guides learning (Waxman & Lidz, 2007, p. 303).

Language is socially oriented in that children pay attention to what those around them are saying and doing. Every sound and gesture are registered and catalogued for future use. Initial assumptions around the meaning or intent of sounds and gestures are either confirmed or discarded as incorrect. In this regard input has an essential role in the cognitive development of language in children. On their language journey children need receptive vocabulary (input) to facilitate their cognition and comprehension of language, which then leads to their production of language. As a child begins to produce language, they receive feedback from other speakers. This may take the form of direct feedback—it's not I *goed,* it's *I went*—or indirect feedback—yes, we *went* to the zoo. Regardless of the method of feedback children absorb it and language comprehension advances in language production. Caregivers, siblings, other adults, and peers tend to reference and direct the infant's attention to objects or people in the immediate vicinity, which leads to the infant connecting real world objects to specific words. Conversation directed to the child tends to be tailored around specific objects. Since nouns tend to be connected to objects, they are the first category of words that children are easily able to conceptually grasp. Once a child can name an object, they can begin to categorize it. These developments in language and consequent word learning are the result of a child being around other language speakers with a degree of fluency at or above their own. Consequent social situations—communication with caregivers or teachers, interactions with peers—allow children to explore the workings of language through copious amounts of fluent, but flawed, input from adults. The most studied source of input is the fluent adult language users around children, but another often overlooked source is other children. Peer input is just as important as that of caregivers (Hoff, 2006; Russell et al., 1998; Stanford, 2008). Children need to socialize within various dyads and various environments to practice and strengthen their knowledge of language and its uses. No matter the source, input is influenced by environment, which provides necessary context to child language development.

Environment is not especially where the child happens to be at a given moment, though that does play a role. Rather it looks more like joint attention of child and speaker, attention getting or directing of attention by speech partners, and interest in speech partners or objects in the environment. There is so much that goes into an infant learning even their first word. Input is necessary for language to begin to take shape in the mind of a child. Word learning occurs as an infant identifies relevant objects being referenced in their environment. Simultaneously they parse the relevant sound segment from the speech stream, and then establish a connection between the object and the sound. Even this simple task of listening and watching is not so simple. Speech directed at children is not clear or direct. Parents, siblings, other adults, and peers draw the infant's attention to specific events and objects, but these episodes of directed attention still tend to be ambiguous. For example, a mother might draw her baby's attention to a stuffed bear but talks about things that are not the bear. It might have a hat, there might be another toy next to it, or the mother might make the bear part of a narrative that does not especially have anything to do with what's directly in front of the infant. The infant will take in all the audio and visual cues but does not yet have all the skills and tools to disambiguate and categorize the incoming information. Yet over time the infant will begin to recognize the word *bear* and that it refers to the brown fuzzy object. Once the infant has established the word and the real-life object it refers to the child can begin to build further language structures onto the word.

By 11 months infants can broadly grasp nouns as a group of objects, and by 14 months they can distinguish the category of nouns from the category of adjectives (Waxman & Lidz, 2007, p. 310). As children progress beyond single words, they begin to build sentences that begin with elementary utterances such as *Mama, that my bear*. Eventually, these simple sentences will evolve to a more syntactically correct sentence. Early word learning—seems like a simple enough task—children are great imitators of adult behavior, so why should they not learn language quickly and easily? Within their first year typically developing children should have learned around 100 words in at least one language. Once they reach this base level of language knowledge infants usually have a vocab "burst" and by 18 months are producing around 50 words and able to understand another 200 (He & Arunachalam, 2017, p. 1). The

production of a child's first word is a significant marker of progress in the child's language learning. All typically developing children are capable of becoming fluent speakers of their native language, understanding all its nuances and cultural understandings. During the course of this thesis, I used the CHILDES database to conduct a longitudinal study of the progression of noun learning in children aged 12 months to 60 months. Through the data collected in CHILDES I examine the amount and type of input children receive from their parents and other adults, as well as siblings and other peers.

1.1 Research Questions

This study proposes a combination of these two points—nouns are acquired early by children because language is a social activity and children are heavily influenced by their environment. That is—children acquire nouns and object labels early because of the frequency of nouns they are exposed to in the oral and visual input they receive. This is most likely because caregivers tend to draw children's attention to objects in the environment with a high frequency (i.e., joint attention). Note, I will only be analyzing data for oral input as there is not a consistent way to measure the visual input children receive for this database. Data exploring this hypothesis will be collected from the CHILDES database, as it provides a wealth of data on child language. Not only does it contain data on child produced speech, but it also has input from various sources that the child is exposed to.

The intent of this study is to purvey not only the ability of children to correctly parse and understand nouns and object labels in receptive vocabulary, but to also produce the same nouns and object labels correctly. In order to do so it is necessary to conduct a longitudinal study of the receptive and productive vocabulary of children included in the CHILDES database. Firstly, I am interested in how the initial frequency of nouns produced by adults increases or decreases as the child gets older and begins to produce nouns themselves. Secondly, I am interested in how the frequency adult production of vocabulary affects the acquisition and use of the same vocabulary in children. Specifically, which nouns adults frequently produce when a child is around 12 months, and which nouns the child begins to produce as they approach 60 months. In other words I want to look at frequency of input as a longitudinal predictor of child language production. Thirdly, I am interested in frequency of nouns produced by a child's peers

and the influence they have on each other's language development. As well as whether or not the target child and their peers' receptive and productive vocabulary is more similar at 60 months than at 12 months.

1.2 Thesis Structure

The organization of this paper will proceed as follows. The second chapter will provide theoretical background of how children receive and make use of input from caregivers through two common language learning theories—knowledge-based and usage-based acquisition and why I choose to follow the usage-based theory. Chapter three will focus on word learning. Specifically, through the processes of disambiguation of input—both speech and referents, categorization of words, and the social influences of language development. Chapter four will discuss the relevancy of using a database such as CHILDES as a data source for study of language development and why I chose to use it. Chapter five will introduce my hypothesis and predictions about the data collected from the CHILDES database. Chapter six will review the methods I used to extract and analyze data from the CHILDES database. Chapter seven will present the data collected from CHILDES, discuss the results of the data analysis discovered during the course of this study. Chapter eight will include a summary of the limits of this study, what might be done in future studies and the conclusion of this current study. Following the final chapter will be a list of references and appendices cited within this study.

2. Foundations of Language Acquisition

The trajectory of child language learning for typically developing children generally follows this path: children receive input from other language speakers around them, then gradually parse individual words from the input, which leads to links being established between words and concepts. This chapter will explore the theories behind child language development which seek to explain the why and how children parse words from input and cognitively connect words with concepts. In addition, this chapter will provide background on the social aspects that encourage an infant to use words to communicate. In the examination of language acquisition there is much debate on just what the mental processes are that drive language acquisition. The main argument between theories of child language acquisition focuses on what language abilities are intrinsic to the learner as they begin to acquire language, and what role their environment plays as they learn (Waxman & Lidz, 2006, p. 304). There are two leading theories that linguists use to study child language development: native-linguistic and usage-based. Both theories seek to explain how infants acquire knowledge of language that falls outside of what they have encountered and what entails the intrinsic language abilities, the intent and function of communication, and what influence the environment. The knowledge that they gain is apparent in the developing child's ability to produce and comprehend sentences they have never heard before, interpret ambiguous sentences, and know what is permissible grammatically and structurally in their language, (Lidz & Gagliardi, 2015, p. 335). Where the theories differ is the processes by which a child gains this knowledge. The remainder of chapter 2 will examine the processes of language acquisition and chapter 3 will examine the social contexts of language acquisition.

2.1 General Language Learning

Babies communicate by making noises that do not especially have any form. By the end of their first year infants begin to make determined attempts at communication and produce their first words (Brown, 2014, p. 22). This is accomplished through imitation of what they have heard around them. It is not uncommon for a baby to display some adult like behavior. Behavior such as sitting with a play phone pressed firmly to their ear while they babble animatedly, or bash laptop keys, or frantically scribble away at a piece of paper. This is an example of how much

children pay attention to the adults around them. Infants learn certain behaviors by watching and listening to what their caregivers do. If young children are easily able to pick up cultural and social behaviors within their first year of life, they most definitely are able to acquire language in the same manner. Around 18 months of age children are able to make two- and three-word sentences. Language productivity and comprehension only increases from there with more word combinations being produced with increased complexity (Brown, 2014, p. 22). By the age of three, toddlers are capable of producing questions, negations, and generate creative—but not always correct—utterances. No matter what language they end up speaking, children do not learn language without input. The why and how children acquire language seems like it should be easy to answer. They interact with parents who are presumed to be fluent speakers of a particular language. Siblings, or other children of varying ages might also be supposed to be in a language learner's sphere of influence. If a child is surrounded by language users of any degree of fluency, the child is receiving input, which they then analyze and make hypotheses about. These analyses lead children to form concepts that "capture the relations among the objects and events that they encounter" and simultaneously parse individual words from the language they are immersed in (Waxman & Lidz, 2007, p. 299). Before children reach the point of purposefully imitating adult behavioral and speech patterns, they need to understand the intention and function of communication. There are many developmental steps that occur during this time, such as the individual words an infant has learned then evolving into entire utterances during pretend play events. How does the mind acquire language? The everyday life of a young child seems like it should not be too taxing. Eating, playing, sleeping seem to be the extent of a child's early years—but there is so much going on behind the scenes developmentally. Every visual and audio cue is creating an entire internal world of language—its forms and concepts. Not only is a child learning what individual words are, they are also connecting them to real life objects. As language learning progresses, they begin to map meaning onto words and language forms. During this process children are simultaneously developing concepts around words. By the time a typically developing child reaches one and half years of age, they should have a vocabulary of around one-hundred words. Once they have reached this milestone in vocabulary a child has a working base of language knowledge that will

support them during their language learning (Goodman et al., 2008, p. 525). Children build continuously on their language knowledge and ability until they become fluent language users themselves.

Not every society has the same position on child directed language. Many cultures direct speech toward their children from birth. However, some cultures do not speak directly to their children until they start producing language themselves (Hoff, 2006, p. 59). In a situation like the latter, the child will only receive ambient, or indirect, language input. Regardless of what type of input children receive, it still guides and shapes their perception of language. As long as the child is typically developing (no learning disability, etc.) they should by 5 years of age have a good grasp of language and how to use it effectively. The source and frequency of input raises an important question: how do children acquire language so effectively without formal instruction? This question has yet to be answered definitively by linguists. There are two main theories that seek to explain whether language acquisition is a result of nature or nurture. Study of child language development generally takes two approaches. First, language is the result of the conceptualization of mental processes that are formed and shaped by environmental input, which leads to production and understanding of language, which is the nativist approach (Hoff, 2006, p. 56). Second, language is less the internal processes of language development, and is more the result of socially oriented contexts, which is the usage-based approach (Hoff, 2006, p. 56). These two theories will be discussed in more detail below. Some linguists have attempted to converge the internal process and external factors of these two approaches of child language learning into a method which provides a combined effort to explain the acquisition process as a whole. I will not attempt to do so during this study. Ultimately, these two approaches raise the questions of how "the mind acquires language and of how the social context shapes language development" (Hoff, 2006, p. 56).

2.2 Nativist Approach

On the nature side of the argument is the nativist approach. This method of studying language acquisition was helmed by Noam Chomsky. In this theory children have an innate grammatical knowledge and syntactical structure which they use in tandem with the language they hear to guide the acquisition grammar of the target language (Lieven & Tomasello, 2008, p. 168). This

innate grammatical knowledge is known as universal grammar. Universal grammar, or UG is a "blueprint that all languages follow that forms part of the child's innate capacity for language learning" (Fromkin et al., 2014, p. 13). This innate grammar constrains what is possible for a child to construct in a given language. These boundaries on what is and is not possible in a language are what allow children to acquire language quickly and effortlessly despite differences in input and circumstances (Crain & Thornton, 2011, p. 189). Children must have a UG constraining what they think to be possible to produce in a langue because they could never have enough input to learn everything about a language—this is poverty of the stimulus. Poverty of the stimulus is an argument made by nativists that "children's linguistic experience under-determines what they come to know about human language" (Crain & Thornton, 2011, p. 190). However, the input a child does receive is filtered through UG and UG allows the child to extract relevant information through knowledge of structure and guides them to correctly interpret sentences they have never heard before (Lidz & Gagliardi, 2015, p. 336). According to Lisa Pearl (2019), the input children receive is insufficient due to the ambiguity it presents and children being incapable of resolving said ambiguity (p.2). Ambiguity and resolution will be discussed in more detail in chapter 3. Real world language experiences select features of this grammar that are relevant to the language the child is exposed to most often. As a child is exposed to language, they begin to make assumptions about representations and how they can be constructed. UG gives them the ability to correctly infer what it is about an utterance that is grammatical. Over time as the child participates in more communication events their assumptions about language are confirmed or disconfirmed through further evidence. Their grammatical knowledge is updated to include these changes and development continues (Lidz & Gagliardi, 2015, p. 337). Eventually the child will have adult-like grammar. However, not every linguist follows the idea that a child's language learning is driven by an innate mechanism.

The seemingly innate ability of a child to acquire the grammar of a language quickly and efficiently under the nativist theory has definite appeal as it "helps the child overcome the (possible) under specification of language structure in the input" (Behrens, 2008, p. 384). In the nativist perspective, input is mapped onto linguistic categories, but because the categories or principles of core syntax are innate, they do not have to be learned (Behrens, 2008, p. 384).

However, innate structure alone cannot account for the acquisition of the lexicon of a language, or other properties of language such as the inflectional morphology of German or Dutch (Behrens, 2008, p. 384). Input, for both native and usage theories, as well as the ability of language learners to generalize must, and does, play a role in the acquisition of language-specific properties (Behrens, 2008, p. 384). "Domain specific representations provide the foundation for generalization beyond experience" and experience allows children to select what is relevant to those representations (Lidz & Gagliardi, 2015, p. 334). Abstract representations of language are in the child's grasp from the beginning, and they explore the oddities of language through this lens (Lieven & Tomasello, 2008, p. 168). Once children have segmented a word in the speech stream, they begin to make a connection between it and the real world.

2.3 Usage-Based Approach

On the nurture side of language acquisition is the usage-based theory. According to Tomasello & Lieven, children can only learn language through their experiences (p. 168, 2008). Infants first learn a word or morpheme and over time build more complex and abstract constructions around the word or morpheme. Initially there is no understanding of the internal structure of a construction, rather the infant uses it as a language "chunk" with a specific meaning (Lieven & Tomasello, 2008, pp. 168-169). Over time, and through accumulation of input, infants are able to establish patterns of, and relations between, constructions with increasing complexity. This occurs through the infant's ability to identify individual items in the speech stream and categorize them into specific constituents, such as verbs and nouns (Lieven & Tomasello, 2008, p. 169). Once an infant has established patterns of co-occurrence observed in their environment they are stored for later retrieval. These stored patterns allow children to produce and comprehend sentences that they have not previously encountered (Lidz & Gagliardi, 2015, p. 334). Infants display the ability to quickly acquire language through specific and general patterns of varying complexity in visual and auditory cues (Naigles, 2002, p.157). Infants track patterns of co-occurrences, such as the word *bear* with a fuzzy brown object and retain them in a summarized format (Lidz & Gagliardi, 2015, p. 334). Since these experiences are stored and

easily retrieved, the infant is able to produce and comprehend language beyond what they have experienced (Lidz & Gagliardi, 2015, p. 334).

This theory is far more cognitively based and allows for cognitive abilities to be applied across multiple domains and "form the basis for understanding learning and social interactions more generally" (Lidz & Gagliardi,2015, p. 334). In the usage-based perspective of language development it is the nuances of language that draw children's attention to abstract representations. Initially the child does not have an understanding of the grammar, or internal construction, of language rather they need to be exposed to a large quantity of it over a length of time. When children begin using language they do so "conservatively", meaning they only repeat what they have heard produced by others (Tomasello et al., 1997, p. 374). As they experience more of language children start to perceive patterns and accordingly construct categories and schemas (Tomasello et al., 1997, p. 374). Once those patterns, categories, and schemas have been established, children can begin to communicate in meaningful ways. Their language takes on the patterns of the adults around them. According to Tomasello et al. (1997) this means children are able to produce nouns effectively and correctly because of their association with object labels (pp. 375-376).

2.4 Which Theory?

At first glance it does not seem to be important to differentiate, or choose, between the two theories. The seemingly innate ability of a child to acquire the grammar of a language quickly and efficiently under the nativist theory has definite appeal as it "helps the child overcome the (possible) under specification of language structure in the input" (Behrens, 2008, p. 384). In the nativist perspective, input is mapped onto linguistic categories, but because the categories or principles of core syntax are innate, they do not have to be learned (Behrens, 2008, p. 384). Innate structure alone cannot account for the acquisition of the lexicon of a language, or other properties of language such as the inflectional morphology of German or Dutch (Behrens, 2008, p. 384). Input, as well as the ability of language learners to generalize, must and does play a role in the acquisition of language-specific properties (Behrens, 2008, p. 384). Essentially, linguists in the nativist camp believe children are born with an innate ability to learn language that is

simply refined as they progress. Whereas usage-based linguists believe that children establish an ability to use language as they progress.

Though I will tend toward the usage-based theory of acquisition of language for this thesis there is beginning to be a shift in linguist perspective of how children acquire. It is not delineated between the above discussed views of nativist and usage-based theories, but rather it is something in between. More recent study of child language acquisition has begun to present two additional views: preferential attachment and preferential acquisition (Fourtassi et al., 2020, p. 2). In their article Fourtassi et al. present an idea of language development known as preferential attachment and preferential acquisition. Preferential attachment is the idea that highly connected words that a child knows will "attract more words over time, in a rich-get-richer scenario" (Fourtassi et al., 2020, p. 2). This is called internally driven learning because it is motivated by the child's internal lexicon. Preferential acquisition is the external representation of language. Children learn new words through known words being highly connected to other known words. In other words the organization of a child's language knowledge guides them to make relevant connections as they progress (Fourtassi et al., 2020, p. 2). High frequency of nouns leads to higher connectivity and earlier acquisition. Experiencing the words in various environments from various speakers also aids in child language learning. A second study that looks at a different theory of child language acquisition is the article by Smith & Yu (2008) on statistical learning. In this theory children learn words through making hypotheses about words and their referents and collect statistical evidence that helps them disambiguate the intended connection between the word and its referent (Smith & Yu 2008, p. 1559). This evidence comes through the child experiencing a set of words and referents in multiple speech events. One idea is clear from all these theories—the frequency of input children receive and the environments they receive it in is important in early acquisition.

The purpose of this present study is not to argue for one theory or the other given that both theories require the child to be exposed to input regularly and with intent, as well as to practice the target language. While linguists do not always agree on the mechanism by which children learn language, they do agree on the process by which children are exposed to language. Both theories rely heavily on children receiving input from other language users,

while simultaneously experiencing language used in various situations (Lidz & Gagliardi, 2015, p. 334). Children hear a continuous stream of sound as they are learning language, and over time they begin to segment the speech stream into individual words. Once they have separated sets of sound into words children begin to assign, or map, meaning onto those words. This entails a process of confirming or discarding hypotheses about what a particular word is referring to and its possible meaning. Regardless of which position linguists support it is clear that children must develop general knowledge of language and the specific grammar of their mother tongue.

2.5 Social Language Theories: Interactive Language Development

After perusing theories from the cognitive, or internal, aspect of language learning it is just as important to consider different perspectives on the social and external aspect of language learning. Two foremost names of childhood language acquisition and its motivation are Jean Piaget and Lev Vygotsky. Piaget considers the way children view and use language as being very self-oriented and not easily influenced by environmental or outside factors. Vygotsky believes children to be socially oriented in their language use and as such they are heavily influenced by outside factors (Santiago-Delefosse, & Oderic Delefosse, 2002). Both Piaget and Vygotsky believed that language starts internally, only to start being expressed and shared in communication with others by at least the age of 6. Language is first internal and “egocentric”, reflecting the development of internal thought structure. Once this structure is established then the child can turn outward and begin using language to engage with others in meaningful ways. The social world

> acts as a constraint that orients the development of the child’s thought. Externalized language is not a mere addition of sound to internalized language, it represents a dynamic transformation derived from the restructuring of language, a unique transformation of the syntactic, semantic, and phonetic structures of language. (Santiago-Delefosse & Oderic Delefosse, 2002, pg.737)

Egocentrism is strongly supported by Piaget, but are children really that self-focused in their early language development? Garvey & Hogan (1973) state that while it may appear that

children are egocentric in their first forays into language use, it is not for selfish reasons, rather they believe children are learning to navigate social language and have no skills or abilities in the very beginning. Over time children will gain adult-like skills and understanding of how social communication occurs, and quickly become proficient users. Going from "egocentric" language to social language means the child must go from only having meaning for oneself to sharing meaning with others in successful communication. Perhaps it is easy to see the nativist and usage-based theories reflected here. Since Vygotsky's views on child language acquisition reflect those of the usage-based theory, his will be the leading idea—namely that children use language for social purposes and seek social situations to use language. Chapter 3 will address the different social situations children experience and how those situations will affect their language development.

Jo Ann Farver (1992) explores how children begin to use social language during play with peers. Through play with peers, children are presented with a unique opportunity to practice successful communication. In her paper Farver states that during play children practice integrating new ideas and words into coherent narratives. Play demonstrates children's need to comprehend what others mean, and in turn how to be understood themselves. Social gestures are practiced, as well as elaborate language structure. Throughout play children must manage and monitor communication through conflicts and negotiation. One of those adult-like skills that must be acquired is coherency in discourse. Children achieve all this through play, where they start to share, understand, and integrate each other's imagined scenarios (Farver, 1992, p. 502). According to Keenan & Klein (1975) this includes the unspoken "rules" of conversation: meanings of utterances, felicity conditions, conversational maxims, and joint attention. These authors argue that children are capable of following context and participating in a coherent conversation, but that they often cannot respond in proficient or fluent ways. The child may not have the full grasp of language as an adult, but they can, and do, respond appropriately as they are able. Edward Mueller explored this capability of children to understand others and be understood by others in his 1972 study of children's maintenance of verbal exchanges. He argues that social understanding of language use starts much earlier than many believe it does. In natural language settings, he found that very young children were able to make themselves

understood and solicited appropriate replies. Thus, social intent of language and maintenance of communication is a skill acquired from a very young age. Between the ages of 16-19 months infants are sensitive to their caregivers eye gaze, head-posture, and where the caregivers are directing their voice (Trueswell et al., 2014, p. 119).

Hamo, Blum-Kalka and Hacohen (2004) see child peer language as a necessity for negotiation within social circumstances as well as a way to develop language resembling that of adults. One interesting point these authors bring up is how children's language with everyone, including their peers, needs to be naturally occurring and that the context of utterances highly important to understanding the why's and how's of children's communication in specific instances. While talking to children it is normal for the speaker to employ pointing, gesturing and eye-gaze to direct the attention of the child to the object(s) in question. Infants begin paying attention to and following such cues when they are around 10 months of age (Yu & Smith, 2016, p. 1235). As infants age and progress in their language skills their ability to sustain attention on a particular object or topic begins to lengthen and less to prone to distraction (Yu & Smith, 2016, p. 1235). As will be expounded upon in chapter 3 for word learning, this embodied or sustained attention is essential to children's ability to cognitively map referent to object. Gesture, pointing, and attention getting will also help the child disambiguate meaning of referents as well as the word attached to a particular object. Joint attention by both the adult and the child in the conversation is necessary for the child to make connections between words and objects. The desire for language to be social and communicative will drive the child to try to understand others, as well as be understood (Santiago-Delefosse & Oderic Delefosse, 2002, p. 734). Once cognitive connections have been established between referent and object, children's vocabulary will begin to grow, and language understanding will begin to develop.

3. Word Learning

Imagine a scene for a moment. A mother sits on the floor with her baby. She has a stuffed bear in her hands, and she is moving it around, talking to her baby as she does so. The baby follows her movements, and most important hears what his mother is saying. To the baby his mother's utterances are just a stream of sound. However, with continuous input it will become easier for him to parse out individual sounds. Over time, specific patterns of sound will become recognizable as words, and as the words are encountered in various situations the words begin to take on meaning related to the speakers' intent. When the child begins producing words himself, he will have an understanding of what it means when he says *bear*. As well as how to communicate intent and meaning when he says something like *bear go home*. With the input of his mother and other fluent speakers, the baby seems to be able to quickly, effortlessly learn both words and their meaning. Over time the child will also be able to place *bear* and other object labels into the correct categories. Children may be exposed to significant amounts of input but, is it enough to learn everything there is to know about the rules of their language? Under the usage-based theory of language acquisition—no it is not, but the input children do receive is enough to establish patterns that they can apply to language as a whole. Each word learned, each connection made between a referent and an object is another step to becoming a fluent language user.

3.1 General Word Learning

This chapter will bring the theory of usage-based language development and social theories of word learning together in an examination of the process of language learning. Language acquisition processes are clearly intertwined with word learning, as both language learning and word learning go hand in hand. As a child is exposed to language in different situations, they learn not only the structure of their language but the individual words that make up the language. According to Bybee (2010) the usage-based perspective encompasses several domain-general cognitive processes that impact how children acquire and develop linguistic structure:

I. Categorization: identifying tokens as an instance of a particular type
II. Chunking: the formation of sequential units through repetition or practice

III. Rich memory: the storage of detailed information from experience
IV. Analogy: mapping of an existing structural pattern onto a novel instance, and
V. Cross-modal association: the cognitive capacity to link form and meaning
(p. 7)

Learning words and their meaning is a complicated process that involves more than just understanding that a specific sequence of sounds makes up a particular word. Adding a word to the lexicon entails more than just hearing it repeated hundreds of times. Though hearing a word, or words, many times from caregivers does lend to the likelihood that a child will learn the word and use it sooner and with more frequency (Dale & Li, 2008, p. 517). Early word learning research often focuses on noun and verb acquisition. Nouns and verbs comprise much of a child's early receptive and expressive vocabulary (He & Arunchalam, p. 2). Categorization and learning of nouns will be discussed in the next section. He & Arunchalam (2017) name three learning mechanisms by which children acquire new vocabulary. First, children must segment input from the speech stream into individual words and assign them to grammatical categories. Second, after children have segmented words from speech, they assign (map) meaning to the word and start identifying the concept behind the word. Initially, children do not assign the full meaning of a word (including cultural understanding) immediately to their newly acquired words. Third, children begin to form hypotheses about the meaning of a word. In order to map the meaning(s) of a word, children will require additional exposure to the word in multiple situations with multiple speakers of varying levels of fluency.

More specifically, to learn a word an infant must

I. Identify the relevant item from a stream of activity
II. Parse the relevant pattern of sound from the stream of speech
III. Establish a mapping between the item and the sound pattern
(Waxman & Lidz, 2006, p. 300)

Before an infant can parse a word, much less place the word into a category, they must receive enough input from fluent language users. Input is the stream of sound that children are exposed to deliberately and peripherally. In most western societies speech is directed to the

child, and the child is involved in communication events. While in other societies the child is only ever exposed to ambient input and are only involved in communication events when they begin producing language on their own (Hoff, 2006, p. 59). Through input from the fluent adult language users around them, children begin to learn words and map meaning onto them. The input a child receives may come from siblings and peers as well as the caregivers and teachers around them. Peer input is just as critical for language development and communication skills as that from adults. More on peer input will be discussed later in this chapter. The quantity and quality of input perhaps unsurprisingly affects how children process language and develop vocabulary (Mani & Ackermann, 2018, p. 253). Children who are spoken to often by their parents have larger vocabularies at a younger age. This is demonstrated in a study by Cartmill et al. (2013) where they recorded 50 parents interacting with their children at home. Their conclusion was that children around 14 months who are not only talked to by their parents, but also involved in the topic and objects in the immediate environment had significant vocabulary gains around the 50 month mark (pp. 11279). Repetition of words related to objects and significant social interaction leads to greater mapping of not only sound to word, but also word to object (Cartmill et al., 2013, p. 11282).

Input provided to the child contains statistical regularities that aid in distinguishing word boundaries (He & Arunachalam, 2017, p. 2). By 8 months infants are able to track these regularities and segment words in the speech stream, and they are beginning to detect the linguistic structure of language through patterns of speech (Twomey & Westermann, 2018, p. 62). Though syllables do occur with greater frequency within words themselves, infants have to constantly evaluate input and correct hypotheses they have about words and their boundaries to weed out misunderstandings (He & Arunachalam, 2017, p. 2). As a result, when children typically begin producing words around one year of age they do not do so because they have memorized the words and their possible meanings (Goodman, Dale & Li, 2008, p. 516). Rather, according to the usage-based theory, word learning and hypotheses about meanings of words are driven by what the child hears (Crain & Thornton, 2012, p. 197). Mapping meaning onto word forms also begins around 6 months of age, but it is not entirely successful until after the

12 month mark, but even this success may be due to children learning to read environmental cues better (He & Arunachalam, 2017, pp. 2-3).

Over the course of their first year children are acquiring the structure and patterns of their native language by being immersed in it (Waxman & Lidz, 2007, p. 304). Allowing for variation among individuals children should have at least 10 words in their vocabulary by 12 months, 80 by 18 months, and 240 by 24 months (Mani & Ackermann, 2018, p. 253). It is not just repetition of a word that helps children to parse the word from the speech stream, but there is some other motivation for them to attach the word to a specific object. A number of studies (Yu & Smith, 2016; Mani & Ackermann, 2018; Ackermann et al., 2020; Laing & Bergelson, 2020) in some way point to interest in objects in the environment to be a contributing factor of which words children learn first. These studies also point to some visual element being responsible for where children direct their interest. Indeed, it is the co-occurrence of a word and an object is what helps a child break through the ambiguity of many potential referents in their environment (Yu & Smith, 2012, p. 244). For North American English-speaking children, nouns or object labels, make up the majority of first words.

3.2 Noun and Object Label Learning

How do children acquire object labels? or common nouns as they are often termed. Input from adults plays a very important role in object label and noun acquisition; however, it is not just the amount of input that aids a child in adding the word to their lexicon. The context in which the child is introduced to the word is equally important. In the example of the mother and infant playing with a stuffed bear, the mother says the word *bear* and directs her baby's attention to the object. While hearing the word, the infant can see the bear, touch the bear, and perhaps even taste the bear. All of these senses working in tandem allow an infant to connect the word with the object they are interacting with (Cartmill et al., 2013). During different speech events the baby will, ideally, hear repetition of a specific word in connection to a specific object. In the given example the word is *bear*. The baby will begin to form hypotheses about meaning of this specific word. As the baby continues to encounter this word in different situations with different speakers, the more data the baby will have on the possible meaning of the word: the soft, fuzzy object that the baby's mother is holding and talking about. The baby

will acquire this word naturally as it is part of the lexicon of their mother tongue. This is the statistical part of learning suggested by Smith and Yu (2008). The more a child hears a word in conjunction with a referent, the more evidence they will have of the connection between the two. Additionally, caregivers provide a feedback loop, where they respond to babble and draw attention to object(s) within the baby's line of sight (He & Arunachalam, 2017; Laing & Bergelson, 2020).

Once children have segmented a word out of the speech stream, they begin to make a connection between it and the real world. There is an interaction between the internal language processing and the external (environmental) processing that occurs as children take in audio (or visual) input and apply it to the objects around them (Yu & Smith, 2012, p. 245). Children need to experience words in multiple situations and in various contexts to map meaning between the word and the intended referent. For instance, learning nouns may be aided by visual cues provided by an adult. Much like example used previously where the mother is pointedly drawing attention to the bear and speaking to her baby about it. Perhaps the mother will point out the same bear in a different room at another time. Eventually through repeated exposure the baby will understand that the noun *bear* refers to that specific toy (Bunce & Scott, 2017, p. 672). Though it will require repeated iterations of the noun and toy being presented at the same time, as object-referent mapping can be rather ambiguous. Being able to see and possibly touch the object being referred to is another aspect of word learning. When a child is able to see cues from a caregiver and follow their eye gaze (and/or pointing) they are more likely to map meaning of a referent correctly (Trueswell et al., 2014, p. 119). Conversely, Cartmill et al. (2013) state that children do not require profuse amounts of exposure to a new word to acquire it. Rather, if the child is exposed to the word in a significant circumstance—in which the intended referent of the utterance is very clear—the child may only need to encounter the word once to acquire it.

Early child language use reflects the categories of words they hear most often, and for English monolingual children, it is nouns or object labels (Goodman et al., 2008, p. 516). Many studies have established this fact. Particularly well known is the Wug test by Berko (1980). As well as a study by Tomasello and Olguin (1993) on the productivity of noun morphology by

children between the ages of 16 months and 26 months. In their 1993 study, Tomasello and Olguin found that children appear to exhibit knowledge of nouns, noun types, and morphology of new nouns, far better than they do of verbs. Just to reiterate the introduction, this study is not intending to prove that children do indeed acquire nouns first among categories. Rather it is a survey of what nouns children in the CHILDES database produce between the ages of 12 months and 60 months. As previously stated, it is a possibility that children classify words in the way they relate to real world experiences—verbs mean action, nouns label objects, etc. The external regularities of these occurrences guide children to confirm or discard hypotheses they might have about a word and its referent. Or children use distributional regularities to place words in different categories based on morphological markers like tense or aspect belongs with verbs and plurals and possessives belongs with nouns (Tomasello & Olguin, 1993, p. 461). These are the language internal cues that guide children to begin establishing the grammar of each word class.

As children progress in their language learning they will begin to encounter situations where the object of reference is not entirely clear. Ambiguity of references will begin to crop up as children begin to encounter more diverse and complex situations. In the next section I will discuss current theories on how children observe and resolve ambiguity of what those around them are referring to.

3.3 Ambiguity and Resolution

Infants appear to be very good at picking out patterns in seemingly abstract data. Once an infant has picked a word out of the speech stream, they can then start finding commonality in observed uses of the word. By observing the word in different non-linguistic contexts, the child can begin to form different hypotheses about possible meaning(s) the word has. Children can also constrain possible meanings of a word through distributional, phonological, and syntactic information to place the word in a grammatical category (Bunce & Scott, 2017, p. 651). Constraint on a words' meaning can begin in its placement in a sentence. Twenty-four-month-olds who hear a novel word used as a noun are capable of assuming it refers to an object, or if the same novel word is used as a verb, then they are capable of assuming it refers to an action (Bunce & Scott, 2017, p. 651). After determining the type of sentence they are hearing the child

must then establish what is being referenced. Potential referents for infants should ideally be situated in their immediate vicinity to reduce ambiguity.

During word learning, infants are constantly hypothesizing about not only word meaning but intent and reference. What a speaker is referring to is not always clear. Take for instance the well-known *gavagai* example where two individuals observe a rabbit running through a field. One observer points and exclaims "gavagai". The other observer, not a speaker of the same language, does not know whether the first speaker is referring to the type of creature, the color of the rabbit, or some part of the rabbit itself (Quine et al., 2013, p. 25). Referent resolution may not just be about discerning which object a speaker is referring to. Rather, it may also include interpreting intent of a speaker.

Early in a child's language learning journey, there seems to be less ambiguity about object labeling, as parents talk about what their child is attending to (Trueswell et al., 2014, p. 118). The child's environment is also much smaller when they are infants. They only attend to what they can see such as the toys on the floor with them that they can focus on and manipulate, thus reducing ambiguity (Yu & Smith, 2012, p. 245). As the child gets older and progresses they are exposed to more complex visual cues, such as gesture, gaze tracking, and following object movement (Trueswell et al., 2014, p. 118). Context must also be considered – what is relevant to the referent and what is not. These are all skills that continue to be used into adulthood, however instead of being used for word learning they are transferred to disambiguating vague speech or referents.

Once children have learned enough about language to begin producing words, Awareness of social stimuli occurs from very early on. Infants are able to follow the eye gaze of adults and focus on the same object as the adult closer to the end of the first year (Keenan & Klein, 1975). While approaching the end of their first year, infants begin to direct adult's attention toward objects of interest and inquire about them. It is during such negotiations that children start to learn that they need to produce speech in a coherent and relevant manner. Whether the child is speaking with an adult or with a peer of similar age, the child will need to

ensure that they both are attending to the same object and addressing their responses in a relevant manner (Keenan & Klein 1975, p. 371).

Co-occurrence of the novel word and the referent object is the clearest way for a child to disambiguate reference and meaning. This ideal learning situation does not often occur naturally outside of a language laboratory. It is more likely to occur when the speaker and the child have joint attention on the referent (Trueswell et al., 2014, pp. 118-119).

3.4 Social Situations and Language Alignment

As children progress in their language abilities and begin using words more, they learn to align their speech to those around them. they should have a grasp of conversational rules, which means they should have some social understanding as well. This does not include just cultural and linguistic rules but involves the universal ability to use language relevantly and appropriately (Keenan & Klein, 1975, p. 366). It is apparent in speech directed to adults, but it is also present in speech directed to peers or siblings. When children begin to produce speech, they need to be able to monitor not only what they themselves are saying for comprehensibility—but they also need try to understand others and gather clues about language through context and other means. Children start tracking where a speaker is looking and inferring what they intent of the speaking is as young as 18 months (Najnin & Banerjee, 2018, p. 2). Joint attention is important for children as they begin to map meaning from referent-object (Naigles, 2002; Trueswell et al., 2014: Bunce & Scott, 2017; Najnin & Banerjee, 2018; Taxitari et al., 2020). Social contexts are important for child language acquisition. Children need to encounter a variety of individuals in various social situations to practice and learn the many aspects of language use (Pelligrini et al., 2002, p. 379). Input from adults gives children the necessary framework of language, and peer interaction allows them the necessary freedom to explore said framework. In addition to practicing language forms, children also need social encounters to learn the shared meaning of language (Farver, 1992, p. 501). Varied language learning environments also help children learn the semantics, syntax and social-communicative cues that are necessary to decoding and using language (He & Arunachalam, 2017). When studying the efficacy of child language development their socio-economic status (SES) and parental education level is often used considered as factors in the success (or lack

thereof) of developing skills. Determining these as viable factors is not the goal of this study. Here I am simply interested in how often children receive input and how they respond to it.

3.5 Adult Input and Child Language Development

Adult input is important for many reasons. The CHILDES database is mostly situations where children interact with adults. This is important for data collection because children need fluent language input from adults; they need it frequently and in large amounts (Hoff, 2006, p. 71). Teaching children social cues, such as turn taking, and how to respond appropriately to others is essential to communication. When children are addressed by adults, they are propelled into a language learning situation. Obviously, the adult (should) know more about language and its functions than the child, which results in scaffolding and ought to lead to more significant vocabulary gain (Perry et al., 2018, p. 17).

As stated previously not every culture speaks to children in the same manner. Many societies in countries such as South America, Africa and Asia do not consider speaking to young children a necessity (Hoff, 2006, p. 59). Differences in languages aside, children who are not spoken to directly do not speak in single word increments, rather they "produce large memorized chunks of input, which they later analyze into component words" (Hoff, 2006, p. 59). Preverbal children who are spoken to prior to producing speech, begin with single words that they later begin combining into larger segments. Despite the different approaches of deliberate vs peripheral input, children are able to become fluent speech users with full grasp of social cues. My study focuses on North American English speakers, where caregivers (most often mothers) speak directly to their preverbal children and draw their attention objects in the environment.

As mentioned above, SES is often used as a variable in child language development studies. The reason being that parents who provide more input to their children will result in the child developing vocabulary early and will add to their lexicon more quickly (Goodman et al., 2008, p. 517). Perry et al. (2018) report that children from lower SES situations will hear around 30 million fewer words than children of higher SES situations (p. 1). While I do not intend to examine this variable in depth, I bring it up to point out the importance of significant

adult input to language development. In the same study Perry et al. (2018), find that once these children begin school, they quickly gain language skills as well as increased ability in social cue reading (p. 25).

By interacting with adults, children are introduced to situations where they hear words referring to a specific object in a specific context. While adults may introduce the object to the child, more often the adult addresses what they perceive the child to be attending to (Trueswell et al., 2014, p. 118). Indeed, in the same study they state that if parents labeled what their child was currently attending to, the child's vocabulary development was far better than children whose parents who did not follow the same labeling pattern (p.118). Children show attention to social cues, such as eye-gaze, as early as 16 months. Back to the example of the mother and the stuffed bear. If the mother holds the bear in front of her baby and says *bear*, the baby is more likely to follow her mother's gaze to the bear and begin to associate that label with the object. Especially if the baby is then allowed to take the bear and hold it, allowing for continued attention on the object after the labeling event (Trueswell et al., 2014, p. 119). The more frequently a word is used by caregivers, the more likely the child is to produce it early in development.

As children progress into producing speech, parents are then able to introduce other social cues such as turn taking. Indeed, according to Perry et al. (2018) turn taking is a very important structure for adults to use with developing children. During their study, Perry et al. found that when adults initiate conversation and address children in such a way that a response is required, children will speak more. But when turn taking was removed from adult response, or adults did not initiate the conversation, children were less likely to respond (pp. 10-11).

3.6 Peer Input

Most of the studies on child language focus on input children receive from adults, as outlined in section 3.1. However, there has been more study in recent years on how peer interaction is just as important to language development as being exposed to fluent/adult language. Farver (1992) gives four milestones of early peer directed speech development:

1. 18 months: children share meaning of pretend activities with objects

2. 30 months: children engage in pretend play through shared focus on play routines and objects
3. 36 months: children enact social roles and employ story lines during play
4. 48 months: children negotiate and maintain play roles and story lines with each speech participant contributing to event

   (p. 502)

Shared play, or social pretend play, as it is often called in the literature, provides children with a situation in which they can connect with peers around their same level of language ability and start linking real-life labels for items to mental associations. Free, or social, play and significant relationships are important to child language development. Mutual play motivates them in their language use (Pelligrini et al., 2002, p. 387). Shared play is spontaneous and subject to instantaneous change. While structure of play, such as a pretend trip to the store, may remain the same, what occurs during such a speech event can differ drastically. Children quickly learn that they need to pay attention to speech and visual cues to maintain mutual understanding of events. Social speech requires them to create mutual understanding with their speech partners through monitoring understanding in others and sustaining their shared attention (Farver, 1992). In order to maintain shared attention, children must respond to their peers, explain themselves, ask questions for clarification and give relevant answers (Fekonja et al., 2005, p. 106).

There are two studies in the CHILDES database that present material of peers interacting together, which I used for comparing noun production in children with peers versus adults. This section will focus on peer-to-peer speech and the potential it has for fluent language development. Child to child interactions are just as important as adult to child interactions. While adults provide "proper" language and literacy, a child's peers—be they siblings or otherwise—are a necessary tool for practicing and learning language skills (Blum-Kulka & Snow, 2004, p. 292). Peer language has only been a serious topic of interest since the 1970's (Blum-Kulka & Snow, 2004, p. 294). Currently, most work on child-peer interactions is still mostly within the realm of socio-cultural and psychological fields of study (Blum-Kulka & Snow, 2004, pp. 292-293). This is perhaps because from a social and psychological perspective peer language

and interactions is very relevant to their study of social and psychological development. Childhood years are very formative in how children will develop cognitively as well as how they will respond to stimuli and behave in their environment as adults (Russell et al., 1998, pp. 314-315). For linguistics it is important to study the development of social language and discourse skills. discourse skills such as turn-taking, pragmatics (i.e., Grice's maxims), which would include attention to the topic at hand and remaining relevant. Conversational skills are mostly practiced with peers (Blum-Kulka & Snow, 2004, p. 293). Interactions with peers allow children the opportunity to be both a learner and a teacher. Especially in a classroom setting.

Perry et al. (2018), in their study of a low socio-economic status (SES) background classroom, found that children need ample amounts of feedback from peers just as much as they needed conversational input from teachers (pg. 17). The study used cameras to record and observe children's natural interactions with each other and teachers. Through their study Perry et al. (2018) found that children need the structured input from adults during focused activities that were meant to build vocabulary and strengthen language structure (p. 7). Then those children who, after the activities, went and pointedly interacted with their peers rapidly improved their language understanding and production (Perry et al., 2018, p. 16). Children are constantly on the move. They migrate from the adults where they observe "formal" language use, to peers where they "test" their knowledge by verbally negotiating with children of around the same language skill. If children do not understand each other, they will either try to work it out or give up and move on to another speech partner. This can depend on the relationship the child has with their peer and whether they perceive it to be worth the effort to pursue a conversation.

There can be many reasons why a child might struggle to make themselves understood. These factors can lead to isolation and minimal peer interaction. Two most common reasons are that they are a second language learner of the shared language of the classroom, or they have some significant disability that prevents them from successfully interacting with their peers. What might prevent a child from seeking out speech partners might be their desire to connect with others. In a 1998 study of and ESL classroom by Rebekah Fassler, she noticed an underlying theme that might drive a child to engage in conversation with their peers, or

conversely, avoid any interaction with their peers. Those children who were more motivated to seek out speech partners and engage in conversational turn-taking were more successful in improving their language skills (Fassler, 1998, p. 401). Those in her study who displayed this motivation seemed to not have a fear of making mistakes in a new language. Those who were not as comfortable in seeking out speech partners, or not as confident in their speech skills, did not develop as quickly (Fassler, 1998, p. 400). Generally, this was attributed to their fear of being misunderstood or shyness of other students. Fassler's study also found that children needed the structured and focused input from the teacher, as much as they needed the unstructured time to interact after teacher input in order to practice the new skill with their peers (Fassler, 1998, p. 404). A more recent study of a classroom with non-native speakers of a language by Timm Albers (2013) found the same results. Albers found in his survey of German Kindergartens that if the child was unable to communicate themselves clearly due to limited L2 skills they were less likely to establish relationships and thus less likely to progress in their L2 (2013, p. 22). Disabilities can contribute to a child not being able to establish relationships with their peers. However, this is not a factor that will be addressed during this study as the studies selected from CHILDES do not have any participants who have a disability of one form or another. Socio-economic status will not be addressed here either, nor will there be any distinction made between male and female children.

As children begin to interact together more and more, they begin to establish relationships. This is a very important aspect of language acquisition. Friendships develop and a rapport begins to build. In friendships each sees the other as an equal in their language abilities (Jones, 2002, p. 64). Even if they were not close friendships children who socialize with their peers in meaningful ways soon begin to make meaningful linguistic developments. Increasingly they are able to create shared meaning during spontaneous play through mutual understanding of everyday activities. Also, they are able to create more complex play schemes by responding to and building on their playmates' contributions to the event (Farver, 1992, p. 513). Often during such events one child will have a slightly higher language ability than the other. The other child should benefit from this discrepancy as they try to align their speaking ability to their peers' they will gain competence (Justice et al., 2011, p. 1769).

4.Corpora as Data

There are many different methods that are used to study language. Using corpora as a data source being one of them. Compiling raw data into a corpus is a time consuming and arduous task. First you have to gather raw data from either a laboratory setting, or a natural setting. Second, you have to transcribe and annotate the data if it is from a spoken source, or if it is a written source, just annotation is required. There are many existing corpora available for researchers to use as data for language phenomena, with the CHILDES data base from TalkBank being one of them. Using corpora to study child language acquisition is not a new phenomenon. Before detailing why I chose this as a methodology for my study, a bit of background on corpora is necessary. Merriam Webster (2021) defines a corpus as

> a collection or body of knowledge or evidence *especially*: a collection of recorded utterances used as a basis for the descriptive analysis of a language

A corpus can be assembled from any number of sources: newspapers, blogs, social media, recorded speech events; basically, anywhere natural language takes place. Behrens (2008) divides corpora into 3 categories

- Diary studies: specifically, entries made by parents of the utterances their children make. Have been known as a study tool since the turn of the 1800's. The early diaries recorded developmental data as well as linguistic.
- Large sample studies: data from a large sample of participants.
- Longitudinal studies: data gathered over a significant span of time from one or multiple participants.

  (pg. XI-XII)

Corpus linguistics is a methodology that allows linguists to search for specific linguistic elements within a bank of recorded and annotated language. As well as make generalizations about the data contained in a corpus. It also allows researchers to look at language development, or change, over a period of time—or rather how it stays the same (McEnery & Hardie, 2011, p. 51).

4.1 Why a Corpus Study?

Recording and implementing data from children has been used as a study tool for analyzing conversational data for a very long time. Initially such data existed as diary recordings by researchers who had children and has evolved to include audio and video recordings that can be transcribed and annotated (Behrens, 2008, pp. xii-xiv). Frequency analysis on corpus data can be conducted in systemic manner allowing for objective verification of results (McEnery, & Wilson, 2001, p. 15). As stated in the previous paragraph corpora may be constructed from written, spoken, or visual language. Usually whatever the corpus is compiled from is a large set of tokens—at least 30,000 words—that is far too massive to be analyzed without the aid of a computer (McEnery & Wilson, 2001, p. 2). If the collected data is then transcribed and annotated, it is machine (computer) searchable. This lends itself to lexical, semantic, and syntactic analysis quite nicely. Though, according to McEnery & Wilson (2001) it does need to be framed in the following context:

- Machine-readable form: implicit data is annotated with explicit linguistic information
- Finite size
- Sampling and finiteness: broad range of participants and situations to provide an accurate representation of population sampled
- Standard reference: limited variation between studies due to a continuous base of data

  (pp. 29-32)

Once a corpus meets the above statements it presents qualitative and quantitative data that can be manipulated in many different ways to answer a variety of questions. There are three methods used to sort corpus data according to Gries (2009):

- Frequency lists: a list of the words contained within a corpus and the frequency with which they occur—the most common way to analyze a corpus
- Lexical co-occurrence (collocations): when words occur together—can search for words left/right occurring of the target word

- Concordances: often called KWIC (key word in context)—what contexts a particular word occurs in
  (pp. 12-17)

Corpora allow language researchers to take hypothesis about language and study them in written form. Properly annotated digital corpora allow researchers to quickly search thousands, millions of tokens from thousands or millions of people in one sitting. I chose to use a corpus study because it would easily allow me to look at frequency data of children acquiring nouns and object labels, and the average age they acquire them.

4.2 CHILDES Corpus

For data collection to support my thesis I used the TalkBank project: Child Language Data Exchange System, or CHILDES as it is commonly known, to gather data. Within the database are many types of studies: longitudinal, large sample, bilingual, clinical, as well as data from many languages other than English. All the data within the database is available for public use, and indeed a few studies presented here make use of or refer to corpora as data sources (Goodman et al., 2008; Taxitari et al., 2020; Fourtassi et al., 2020). For the purposes of this study, I was not interested in the bilingual or clinical corpora. I also excluded studies, or portions of studies, that focused on data from children over the age of 60 months. The methods I chose to use to search this extensive database will be discussed in section 6.

5. The Present Study

The study by Goodman, Dale, and Li (2008), which has been referenced briefly, examines the CHILDES database looking to answer questions very similar to my own. They set out to determine if the frequency of parental input influenced the lexical acquisition in children. Their findings determined that yes indeed the frequency of parental input does seem to influence which words children acquire initially (Goodman et al., 2008, p. 528). However, Goodman et al. only looked at parental input, whereas I included data from parents, teachers, grandparents, and peers. In addition, Goodman et al. (2008) searched for acquisition of individual words in specific categories: common nouns, people words, verbs, adjectives, closed class, and other (pg. 521). However, in my study I only looked for common nouns—which are referents of objects, substances, people, events, and locations (Goodman et al., 2008, p. 521). Table 1 shows how they categorized words.

TABLE 1. *Categories and numbers of words analyzed in the present study, based on parent report from the CDI and parental production in CHILDES transcripts*

Lexical category	Example words	Number of words
Common nouns	ball, frog, juice	256
People words	doctor, girl, mommy	21
Verbs	bite, hug, take	90
Adjectives	big, happy, tired	55
Closed class	that, in, some	68
Others	please, lunch, park	72
Total		562

(Goodman et al., 2008, p. 522)

Goodman et al. (2008) use the CHILDES database to examine input frequency in relation to age of acquisition. The lexical information Goodman et al., (2008) used for their study came from the McArthur-Bates Communicative Developmental Inventory for norming data on age of acquisition for 562 words from children aged 8 months to 30 months (2008, p. 515). The CHILDES database provided frequency that parents use words with their children from ages seven months to seven and half years. During their study, Goodman et al. found that high parental frequency early on is associated with earlier acquisition of specific words (p. 524). Though comprehension of words in relation to parental frequency only showed significant

impact with the acquisition of common nouns (Goodman et al., 2008, p. 515). Common nouns are objects and substances; with events and locations being delineated as other categories by Goodman et al. (2008, p.521). Additionally, Goodman et al. found that individual nouns were produced very infrequently, but they were acquired the earliest as displayed in the following table (2008, p. 523).

TABLE 2. *Correlations between age of acquisition in production and selected estimates of word frequency*

Category	*N*	*r* (parental frequency)	*r* (Kucera–Francis)	*r* (Thorndike–Lorge)
Common nouns	256	0·55**	0·21**	0·23**
People words	21	0·52*	−0·05	−0·04
Verbs	90	0·22*	0·00	0·05
Adjectives	55	0·28*	−0·15	−0·11
Closed class	68	0·24*	0·00	0·01
Others	72	0·34**	0·11	0·16

* $p<0·05$, ** $p<0·01$.

During their study Goodman et al. found that the mean age at which children produce 100 words was around 17 months. They chose to look at 100 words because increasing vocabulary acquisition occurs between 50 and 100 words, which means that children are learning words more quickly and variable affects are more poignant than when acquisition began (Goodman et al., 2008, p. 525). Their findings are displayed in the figure labeled table 4.

TABLE 4. *Correlations between age of acquisition in production and selected estimates of word frequency for words acquired early (in first 100) or later (after first 100)*

Category	Words acquired in first 100: *r* (*N*)	Words acquired after first 100: *r* (*N*)
Common nouns	0·19 (53)	0·43** (203)
People words	0·85* (5)	0·60** (16)
Verbs	0·26 (4)	0·14 (86)
Adjectives	0·36 (3)	0·27* (52)
Closed class	0·82 (4)	0·38** (64)
Others	0·13 (10)	0·14 (62)

* $p<0·05$, ** $p<0·01$.

(Goodman et al., 2008, p. 526)

This table shows the correlation between the frequency of parental input and the age at which children began to produce in the target 100 words and after the 100 words had been successfully learned. As seen in table 4 the frequency and learned words correlation for common nouns only occurred after the child had acquired the 100 words. However, closed class words have a strong correlation between frequency and production early on, but this correlation does not continue (Goodman et al., 2008, p. 526). Goodman et al. explain this as being a result of the small number of closed-class words that are learned prior to 20 months and produced frequently by children (2008, p. 526).

I did not set out to replicate the study by Goodman et al. (2008) but as demonstrated above it bears a few similarities to what I wished to investigate in my own study. As previously mentioned in the study by Fourtassi et al. (2020) the frequency of tokens leads to higher connectivity of words, and thus aids in a child's vocabulary acquisition (p. 2). Also, concrete nouns that are easily connected to items in the immediate environment during various contexts allow children to discern meaning more easily and quickly (Fourtassi et al., 2020, pp. 2-3). Both of these studies look at frequency of tokens—specifically nouns—and how it affects early child language development. I wanted to look at how frequency of nouns spoken to the target child around 12 months affects their vocabulary as they approach 60 months. It is not achievable with the data available, but it would be interesting to discover if children are able learn nouns earlier not only because they hear them the most often from fluent language users, but also because they are the nouns are highly connected categorically and environmentally. Meaning that children hear nouns spoken in conjunction with some sort of representation of the object referenced (toy, book, picture, etc.). While many of the studies included in the CHILDES database collected data in a lab setting where the child and others interacted with toys and books, it is not entirely feasible to include such factors in this study. Additionally, I wanted to determine if during play, children will be able to orient themselves to their peers and begin to produce nouns and object labels that were previously unfamiliar to them. However, for the same reasons that I did not include interactions with items in the environment I cannot include situations such as playtime in this study. Chapter 6 will describe the methods I did use to examine not only these questions but also the following hypotheses.

5.1 Hypotheses

Based on the theoretical background detailed above and in conjunction with the research questions in section 1.1 I present the following hypotheses to be tested:

- Novel nouns and object labels frequently produced by caregivers when children are 12 months will be more quickly mapped to referents and produced with greater frequency by 60 months
- Peers may not produce the same nouns and object labels around 12 months because of differing language abilities among the dyads, but through scaffolding by 60 months they will have gained enough social language skill that they will orient to each other and produce similar nouns and object labels

To reiterate the hypotheses I have for this study are: firstly, the data will show caregivers produce tokens—specifically nouns—with greater frequency when the target child is 12 months of age than when the target child is 60 months of age. Secondly, the adults will produce nouns and object labels frequently when target child is around 12 months of age. Which will lead to the target child producing those nouns and object labels with increasing frequency until they are 60 months of age. Thirdly, when it comes to peer interactions, I predict that children will not produce many nouns and object labels with each other when they are 12 months of age but will produce more of the same nouns and object labels as they approach 60 months. Children will begin to produce relevant nouns and object labels with greater frequency after receiving frequent input from adults and peers and being exposed to varying social situations. They will also be able to navigate communication situations better with peers who may be slightly ahead or behind linguistically when they are closer to 60 months (i.e. will produce more similar vocabulary).

6. Methods

These last two years, as many researchers found, gathering data from participants in a classroom or laboratory setting, proved rather difficult. As a result, it seemed most logical to turn to a data source that already existed and would contribute answers to the questions I wished to answer. For this reason, I chose to use the CHILDES database, since it is an assemblage of corpora collected by many researchers from child participants of different ages in various situations with various conversation partners in environments that include, classroom, laboratory, and home settings. Here I will discuss the specific methods I chose to collect my target data. In the previous chapter I referenced the study by Goodman et al. (2008), which also pulled data from the CHILDES corpus in relation to noun and object labels and parental input frequency. And as previously stated I am not duplicating their study, nor am I building on it, but I am interested in some of the same questions as they were. Which means that I am going to use a few of the same methods as they did. Namely, the frequency of tokens from adult speakers as well as children. Specifically what Goodman et al. (2008) label as common nouns, or words that refer to objects and substances (p. 521).

One massive benefit of using databases for research is that they are already annotated and congruous in the way they are formatted. The data in CHILDES was collected and annotated by the individual researchers who submitted their studies to the database and formatted according to the CHILDES standard template. It is a large database with 61 studies in just the North American English portion that I needed to sort and search the corpora for nouns and object labels. But searching such a large database is difficult without a search mechanism to handle the large amount of data. The CHILDES TalkBank project has its own software to search all the corpora contained within its database. This software is called *CLAN* (Computerized Language ANalysis). However, I found it to be difficult to use and the data, once collected, had to be exported to another platform for analysis and visual representation. Using statistical software like RStudio made more sense to use as a data-gathering tool, since it is capable of analyzing data in many ways. For analysis of collected data, RStudio allows the compilation of said data into visual representations and comparisons in a much more convenient manner. Sanchez et al. (2019) wrote a package for R that is a database-formatted mirror of CHILDES

called *childes-db*. This R package allows for CHILDES data to be imported and searched directly within R. It is also designed to make examining and analyzing the large CHILDES database slightly easier to use for researchers that are not especially competent at digital research methods. Initially data was collected from CHILDES using *childes-db*, once I had collected all tokens that represented nouns, I analyzed the data and made visual representations with R Studio tools.

The data that exists in CHILDES was accumulated and assembled by other researchers for their respective studies. Generally, the researchers' data was gathered through the use of audio recordings of children, or audio recordings with supporting video. The settings where the children were recorded varied from study to study as well. Some studies were simply recordings of activities of the child, or children, as they went through daily routines and interactions with people of varying ages. Others were recordings conducted at the child's school where they interacted with peers and teachers. Still other studies were conducted in controlled lab environments with structured stimuli and tasks. These structured tasks involved the mother guiding their child through an activity such as reading a book, or eating a snack, interacting with toys, or media tools provided by the lab. The data from all of these studies is available in CHILDES, and all the data from the individual studies was transcribed and annotated by the respective researchers prior to being uploaded to CHILDES. Before uploading the data, the researchers followed a set template given by TalkBank so that the data would be uniform and easily searchable. The annotation of the data is very thorough and is separated into many categories and variables. These include the name, age, and sex of the target child, who the speaker of the utterance was, part of speech, utterance type, and number of morphemes in the utterance. However, not all of these variables were relevant to my study. I chose eight variables to work with consistently: gloss (word), part of speech (verb, noun, adj, etc.), utterance type (question, declarative), corpus name (name of study data came from), speaker role (mother, father, teacher, etc.), target child name, target child age, and target child sex.

6.1 Participants and Age Groups

An initial pull of data returned a large number of tokens at 9,252,642 in 2,466,930 utterances. This is not the total number of tokens or utterances in the entire database as I only pulled data

from the 12–60-month range of target child ages from the North American English collection. As stated above the data was separated into eight categories: gloss, part of speech, utterance type, speaker role, corpus name, target child name, target child age, and target child sex. For this study I pulled data from 1,037 children in 57 corpora by searching all tokens in the CHILDES database. Initially I pulled tokens from the 57 corpora which amounted to 9,252,642 individual tokens. I used data from the following 43 CHILDES corpora: Bates, Bernstein, Bliss, Bloom, Bohannon, Braunwald, Brent, Brown, Clark, Davis, Demtras1, Detmetras2, Feldman, Gelman, Gleason, Gopnik, Haggerty, Hall, Higginson, HSLLD, Kuczaj, MacWhinney, McCune, McMillan, Morisset, Nelson, NewEngland, NewmanRatner, Peters, PetersonMcCabe, Post, Providence, Rollins, Sachs, Snow, Soderstrom, Suppes, Tardif, Valian, VanHouten, VanKleeck, Warren, and Weist. I did not include data from all 61 studies because a few were specifically created with children who had some sort of language development delay or other speech impairment to investigate what impact the disorders had on the child's language acquisition.

Since I was not interested in all occurrences of tokens, I filtered the results for occurrences that had been labeled as noun, which resulted in 1,201,433 tokens remaining. As a comparison to how many nouns were spoken by participants in the individual studies. I also filtered data for verbs, adverbs, and adjectives into separate data sets. The data collected showed results from speakers that had been divided into 24 categories. These categories included parents, siblings, relatives, friends, teachers, researchers conducting the lab sessions, as well as words produced by media during lab language sessions. I was only interested in data collected from individuals who were labeled as Mother, Father, Teacher, Grandmother/father, Adult, Target Child, Brother, Sister, Playmate, Child, and Sibling. The other categories were excluded either because they were not relevant to my study, or because they did not provide a significant amount of data. Once I had only tokens from parents, peers and target children, I split it into age groups. The age range of target children I chose to work with was from 12-60 months. I split the data into four groups: 12-24, 24-36, 36-48, and 48-60. Each group was split around the year mark. Note, the term target child will be used from here on to refer to any child in the studies who were labeled as Target Child and does not especially mean the same child participated in all 43 studies listed above.

Once data had been collected for each age group from each category of speaker, I eliminated any data set that had less than 10 tokens. There were also a few groups who did not have data for all ages. Namely, the father(s) for the 36-48 month range, friends for 48-60 months, and fellow student(s) for the 12-24 and 24-36 months range. I eliminated these because they obviously did not contribute any data. Once all unnecessary or empty data had been eliminated, I moved onto searching for nouns spoken by each speaker category divided by age group or dyad. But first I needed to ascertain what a normal vocabulary looked like for each age group.

6.2 Age of Acquisition of Nouns and Object Labels

In order to establish a base of what vocabulary a child should have acquired by certain ages I used the McArthur-Bates Communicative Development Inventory (CDI) which includes specific nouns and object labels that should be acquired by children before 30 months of age. Goodman et al. (2008), use CDI to establish norms for individual nouns, and this is why I decided to do the same for my study. The CDI comes in two sections—infant and toddler. The first section covers children aged 8-16 months and includes 396 words in 19 semantic categories. The second section covers toddlers aged 16-30 months with 680 words in 22 semantic categories. I did not make a distinction between these two divisions in the CDI vocabulary, instead I split the nouns and object labels into themed groupings which will be discussed further on.

When the CDI is administered during studies it is normally given to parents who are asked to rate their child's comprehension and production abilities (Fenson et al., 1994, p. 16). First, parents are asked to rate their child's comprehension of words. Second, they are asked to rate their child's ability to comprehend the words that they do produce. The reason the CDI makes a distinction between comprehension and production and comprehension is that children are capable of producing words that they do not especially understand (Fenson et al., 1994, p. 16). The reason this occurs does not especially stem from a learning or language disorder as some typically developing children also produce words they do not fully understand.

Since the CDI is meant for children ages 8 months to 30 months, and my study is on children from 12 months to 60 months, I also included vocabulary from the Pearson CELF-5 that

was meant to test children and young adults aged 60 months to 21 years. I only used vocabulary from the section that was meant to test children up to 60 months. The Pearson CELF-5 is a clinical test for diagnosing possible language disorders in children, but it is also useful for establishing a baseline of what a child comprehends as well as what they produce. Instead of parents self-reporting their child's language abilities the test is administered by a teacher, or an administrator qualified to evaluate language development. The goal of the test is to establish a baseline of a child's language ability through their everyday language use, what is difficult for them, where their strengths and interests lie (Wiig et al., 2021, p. 4). CELF-5 assesses language through sentence comprehension of increasing length and complexity, linguistic concepts through logical connections between verbal directions and manipulation of objects (Wiig et al., 2021, pp. 6,7). Children's ability to properly apply morphology and semantic features to words and then structure the words into a complete grammatical sentence is also determined by the CELF-5 (Wiig et al., 2021, pp. 10,12,16). Clearly, I could not run every aspect of the CELF-5 assessment related to perception and critical thinking on data I collected from CHILDES, but the vocabulary used in the CELF-5 for the appropriate age dyad would help to assay the ability of children in my data.

I collected the words from both the CDI and the Pearson CELF-5 and only selected the nouns and object labels from both to be used in my study. The tokens in the category of noun were split into groups: animals, anatomy—parts of the body, toys, furniture, people, clothing, food, vehicles, miscellaneous everyday items, and environment—words related to the child's surroundings. Once I had finished with my target words list, I needed to pull data in which to compare them. Participant data was pulled from the American English portion of the CHILDES database. Age of participants was limited to 12 months-60 months for this study. Despite this limitation there was still the previously listed 43 studies that could be analyzed. There are many existing studies that surveyed their participant results through the filters of SES and gender that is not something I am currently interested in, though I did leave gender as a category when surveying data out of curiosity. I was interested in adult and target child interactions, and peer and target child interactions. There were 39 parent/teacher-child studies and 4 child-peer studies that I used as the source of my data.

6.3 Input and Production Frequency

As previously stated, I searched the database for every token that was labeled as a noun. I filtered the results to reflect what tokens were spoken by adults. This included utterances from adults labeled as: mother, father, grandmother, grandfather, adult, investigator, and teacher. Since, I was also interested in how children influence each other's language I included a few peer groups as well which were: brother, sister, sibling, playmate, friend, and child (utterances from a child other than the target child). These are not all of the speakers in the database, but they are the ones that I was specifically interested in, and they had the most tokens recorded. To get an idea of how many nouns total were uttered by each speaker I searched for anything labeled as *n* (noun) in the database. This gave me an idea of how often each speaker used nouns and object labels outside of the data norming lists.

In order to establish the base line of production of nouns and object labels produced by all the speakers in the database, I searched for the vocabulary from the CDI and Pearson norming lists. There were 380 nouns on the CDI norming list and 47 nouns from the Pearson CELF-5 norming list. The reason the Pearson list is so much shorter is that there were a number of nouns that appeared in both lists. I decided to leave those duplicates on the CDI noun list and kept the unique ones on the Pearson list. Since the Pearson CELF-5 was meant for children 60 months and older anyway, it would be interesting to see whether children younger than 60 months produce those unique nouns. As stated above, I initially searched for both individual tokens and utterances in the database. Once, I had the utterances from the speakers I intended I needed to see which nouns and object labels they spoke and how frequently. I grouped the nouns from CDI into the separate noun categories I established. The nouns in the Pearson list were very similar to those in the CDI except for a couple of categories, which included: labels for places related to the outdoors, and nouns related to the marking of time. The established lists of nouns were compared to those in the data to determine which, if any, lexical items were produced by adults, peers, and the target child. Markdown of the transcripts in the database includes indications of sounds that are not especially words—either they were a child's babble or unintelligible half formed words that carried no meaning. These were eliminated as they are unnecessary to this study. Words that were not nouns or object labels on a specified vocabulary

list were also eliminated. Since I was trying to determine if the target children were achieving significant vocabulary, I only included the nouns and object labels determined by CDI and Pearson to be milestones in vocabulary acquisition. I did not include any nouns or object labels that did not appear in either list.

First, I compared the CDI and Pearson lists with all of the tokens and utterances together to see what the overall data looked like. Second, I looked at the noun and object label lists in relation to the separate groups of speakers. Once I had which nouns and object labels were uttered by which speakers and organized by age of target child, I could begin looking at frequency of utterances by age. I repeated the same procedure for each adult category, determining the frequency of each noun on the lists by the age of the target child. the same process was conducted on the data from the peer's categories. Once I had completed that, I made histograms for each category according to the designated age dyads. A longitudinal survey of frequency of input was done by comparing data from the 12 months to 24 month dyad with the data from the 24 months to 36 months dyad, the 36 months to 48 months dyad, and the 48 months to 60 month dyad. This was to assess what nouns and object labels are spoken frequently to children at 12 months and increasingly on to 60 months—not only from adults but also from the target child's peers. Concurrently, the frequency of which words children produce frequently from around 12 months, on through the previously listed dyads, and progressively up to 60 months was also analyzed. In chapter 7 the results of these frequencies and comparisons will be discussed.

7. Results and Discussion

Before reviewing the results of the survey of the CHILDES database I need to first reiterate the research questions, and hypotheses I made in chapters 1 and 5. To begin with I am interested in how the initial frequency of nouns produced by adults increases or decreases as the target child gets older and themself begins to produce nouns and object labels. Secondly, I am interested in the frequency of specific nouns produced by adults when a child is around 12 months and how that affects the child's ability to produce those same nouns around 60 months. Specifically, which nouns adults frequently produce when a child is around 12 months, and which nouns the child begins to produce as they approach 60 months. In other words I want to look at frequency of input as a longitudinal predictor of child language production. Thirdly, I am interested in the frequency of nouns produced by a child's peers and the influence they have on each other's language development. As well as whether or not the target child and their peers' receptive and productive vocabulary is more similar at 60 months than at 12 months.

The hypotheses I had for the results of this study were not very complex. I predicted that the data would show caregivers producing nouns and object labels frequently when children are around 12 months of age. This should lead to the child producing those nouns and object labels with increasing frequency until they are 60 months of age. Related to peer interactions, I predicted that children will not produce many nouns and object labels with each other when they are 12 months of age. Children should begin to produce relevant nouns and object labels with greater frequency after hearing them repeatedly by adults in the 12-36 month range. The target child should also be influenced by peers who are slightly ahead or behind linguistically when they are closer to 12 months. This should be reflected in the target child producing vocabulary similar to that of their peers as the target child approaches 60 months. I will be investigating these research questions through the lens of frequency of tokens. The cumulative number of tokens in the data were organized by speaker role. Total tokens produced in the data by the target child were: 6,348. Total token production in the data for the adult group was as follows—mother: 573,164; father: 60,990; teacher: 14,462; adult: 10,391; investigator: 84,372 and grandparents: 6,769. Total token production in the study for

peer groupings were as follows—sibling: 4,301; brother: 7,539; sister: 10,486; child:15,069 and playmate: 458.

7.1 Input and Production Frequency

In the first analysis of my data set, I looked at the frequency of noun tokens produced by adults starting when the child was around 12 months and the progression until the child reached 60 months. Before narrowing the scope to the specific nouns children learn after exposure, I wanted to establish what the general input of nouns from adults and peers was. The ages of the target child in the data set were split into 4 dyads and organized by speaker role. To begin with I organized the distribution of the dyads for adults and peers by the age the target child started to receive input from adults, peers. Then, for the target child I arranged the dyads by what age they started producing nouns and object labels. For the adult data the mean age they started producing input for the target child was 32.46 months. The median age was 29.8 months with the first quantile being 22.59, the second 29.79, and the third quantile 40.69. The minimum age the parent started producing input for the target child in the data was 12 months, and the maximum age was 59.92 months. In the peer data the mean age they started producing input for the target child was 48.57 months, and the median age was 57 months. For the peer data the first quantile was 41.92, the second and the third quantile were the same at 57 months. The minimum age peers started interacting with the target child in the study was 12.69 months and the maximum age was 59.92., The mean age of production for the target child data while interacting with both peers and adults was 35.21 and the median age was 32.53. The first quantile of target child data is 25.66 months, and the second was 32.52 months, and the third quantile was 44.23months. Minimum target child age was 12 months, and the maximum age was 59.92 months. For a visual of how the age of the target child data is distributed in this study, the boxplot in Figure 6 displays the data by the age of target child participants.

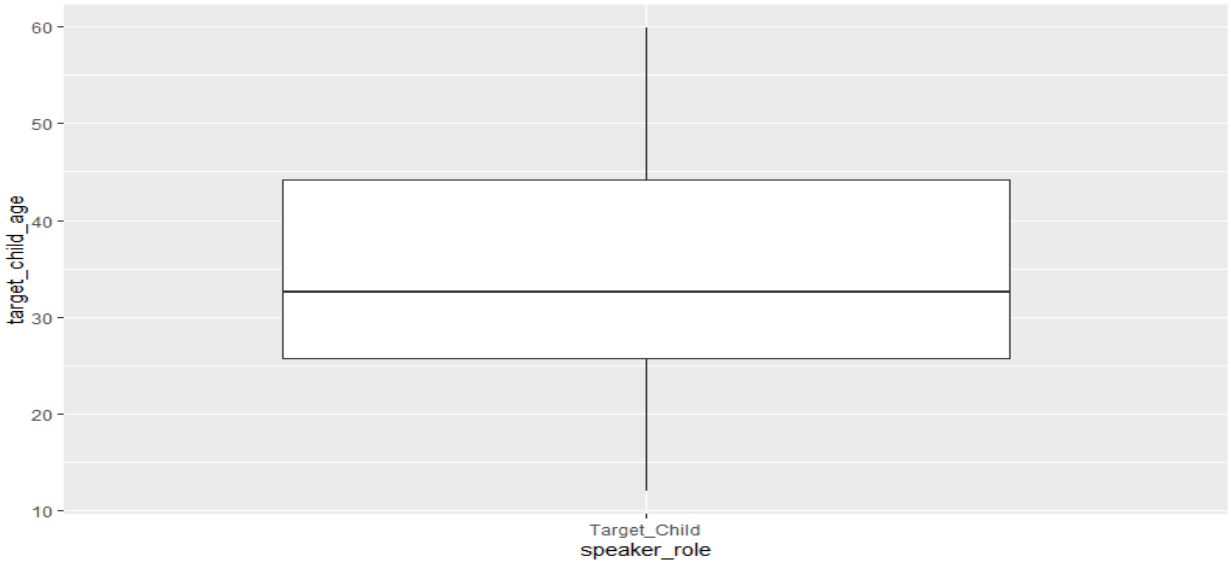


Figure 1 Boxplot of Target Child Data by Age

The height of the line on the boxplot shows the recorded target child age from the minimum to the maximum. The top of the box designates the top quartile of child data, and the base of the box is the base quartile. While the line that divides the box is the median age. There are no outliers in this data set. The data in the following study includes tokens that were spoken directly to the child as well as those that were uttered to other people in the vicinity of the child—ambient input. The following histogram shows the frequency of utterances by the adults either to the target child, or in their vicinity, from 12 months to 60 months.

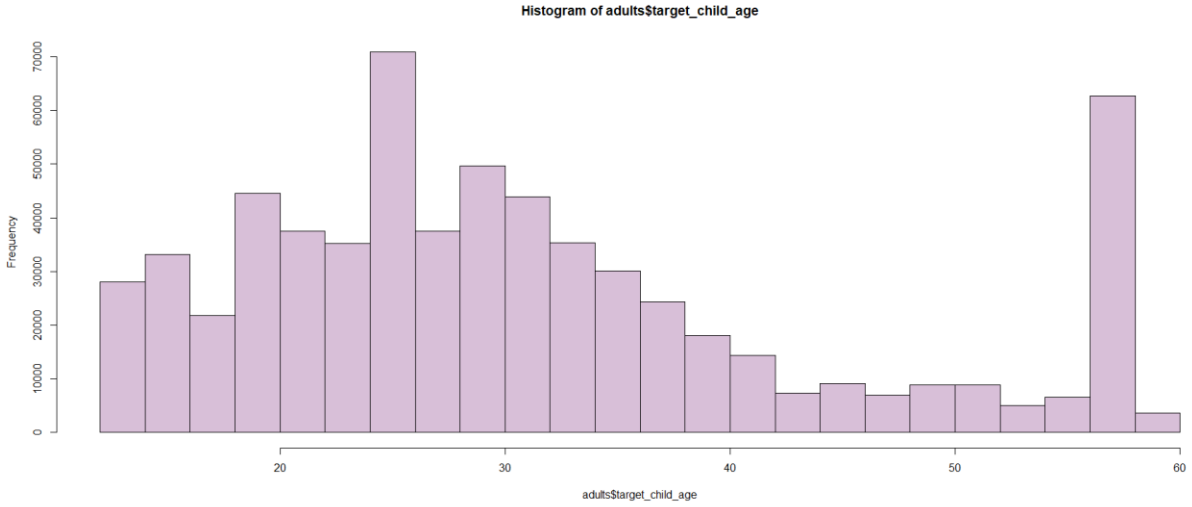


Figure 2 Adult Token Production

Histograms are a form of data visualization that measures how data occurring over a period of time is dispersed. Frequency is measured by how many times an adult speaks to the target child or someone else near them in the data. As Figure 2 shows, the frequency by which children are exposed to input from adults steadily increases from 12 months to about 30 months when it tapers off to an even level around 40 months of age. The spike of frequency closer to 60 months of age appears to be an anomaly. Reason being on closer inspection of the individual studies in the CHILDES database there appears to be more studies conducted with children in this age range—and therefore is not an accurate description of frequency of adult utterances for this age dyad. This speculation on the spike is further supported by it also showing up in peer, and child data. Which lends to it being an anomaly weighted to the amount of data available for that age range.

Figure 3 displays the count of tokens from each speaker role selected for the adult input. The column for count of tokens from the mother has by far the most input compared to the other speakers. In Figure 3 the age range for the target child is 12 to 60 months.

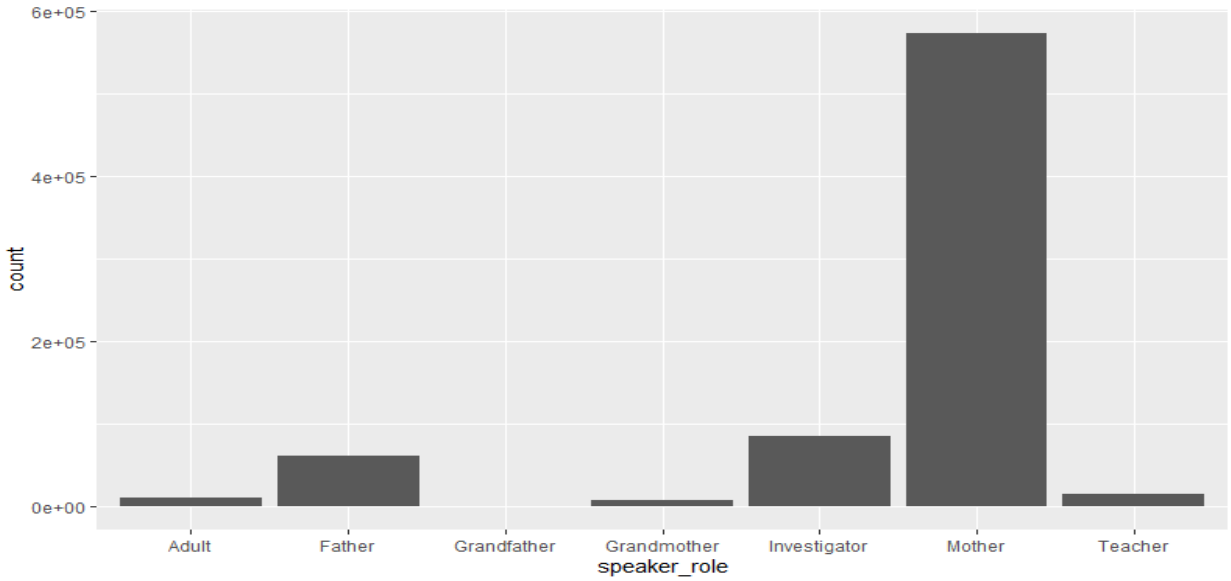


Figure 3 Adult Token Production by Role

A large portion of the studies in the CHILDES database are mother and child studies, so it stands to reason that the most input recorded in the database comes from mothers. Most usually children receive the majority of input from their parents and are oriented to their input from

very early on (Taxitari et al., 2020, p. 1). With that in mind the visual display of data in Figure 2 is a fairly accurate demonstration of how much of the input children in the 12 to 60 month age range receive from their parents (Goodman et al., 2008, p. 520). Since American children younger than 60 months are not yet in school and children around 60 months are just beginning school there is not much data from teachers for this study. Investigators are the individuals who collected data for the various studies in CHILDES. As the Figure 2 shows, investigators often provided input to the target children in the study as well as to others around them. Interestingly, there is enough data from grandparents to incorporate data from them in the study. I included data from speakers other than parents to demonstrate that children do receive significant input from adults other than their parents.

Previously I have alluded to children needing input from peers as well as adults. As the next figure demonstrates, peers do indeed contribute to input that young children receive. This is especially true if the child has older siblings who interact with them often.

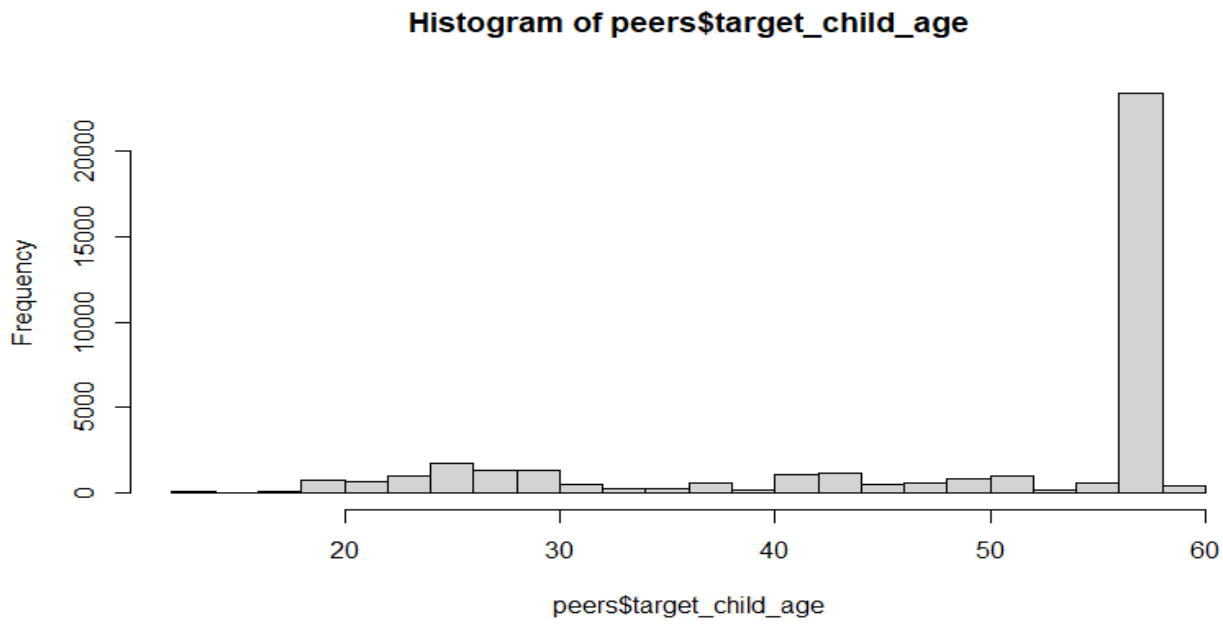


Figure 4 Frequency of Peer Tokens

Figure 4 is a histogram for peer utterances and similar to the adult data in Figure 2 it conveys the frequency of tokens directed to the target child. The data in Figure 4 shows there is not

much input of significance from peers until the target child is closer to 20 months old. From there the data is steady in frequency until 30 months, where it drops until 40 months. It picks up a little from 40 months to 50 months, and then there is that peculiar spike closer to 60 months. This most likely not a reflection of true to life examples, but rather an expression of the data recorded in the CHILDES database. Despite the sparse representation of peer input this histogram does show that peers have a steady influence on each other during their language development. Figure 5 displays the amount of input from each speaker group of peers in the 12 to 60 month age range of the target child.

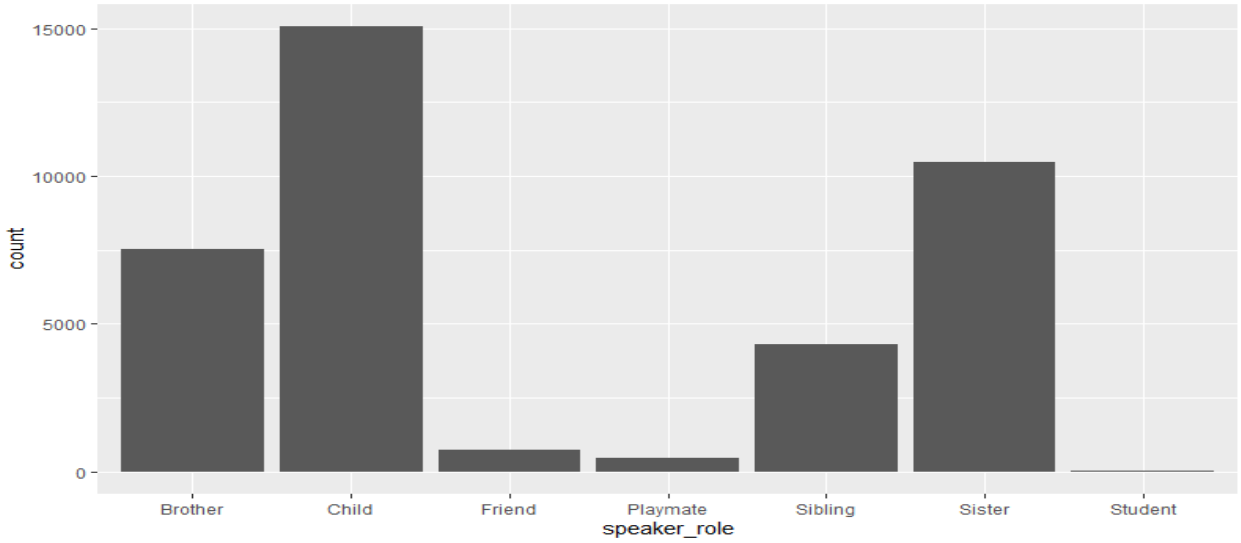


Figure 5 Peer Tokens by Speaker Role

Again, the data in figure 5 is not anywhere near what the adult input is, but it is still significant. The individual speakers are split into roles CHILDES designates as sibling groups and peers outside of the family. The majority of these studies were conducted in either the child's home or in a laboratory setting. The input from siblings is quite large when you combine Brother, Sister, and Sibling. The speaker role of Child, Friend, and Playmate are all interchangeable in the studies as a child who is not a sibling or schoolmate of the target child. There is not much data from playmates or other students because children in this age range have not yet started school. Regardless of the closeness of relationship of the source, children need to interact with peers to practice input they have received from adults. Learning to negotiate social situations with peers gives children an opportunity to start piecing together language and acquire

vocabulary. Peers act as a sounding board for children to practice their newfound language skills and negotiate social situations. Referring back to Figure 4, the target child's interaction with peers is not high in frequency, but it is fairly steady from 12 to 60 months. Even the few interactions children have with peers are essential to their language development. Both input and interaction with adults and peers is necessary for language acquisition.

Adults and peers as demonstrated above speak frequently to children, but what does the child's production look like? As shown in Figure 6 typically developing children slowly but steadily begin to speak with greater frequency.

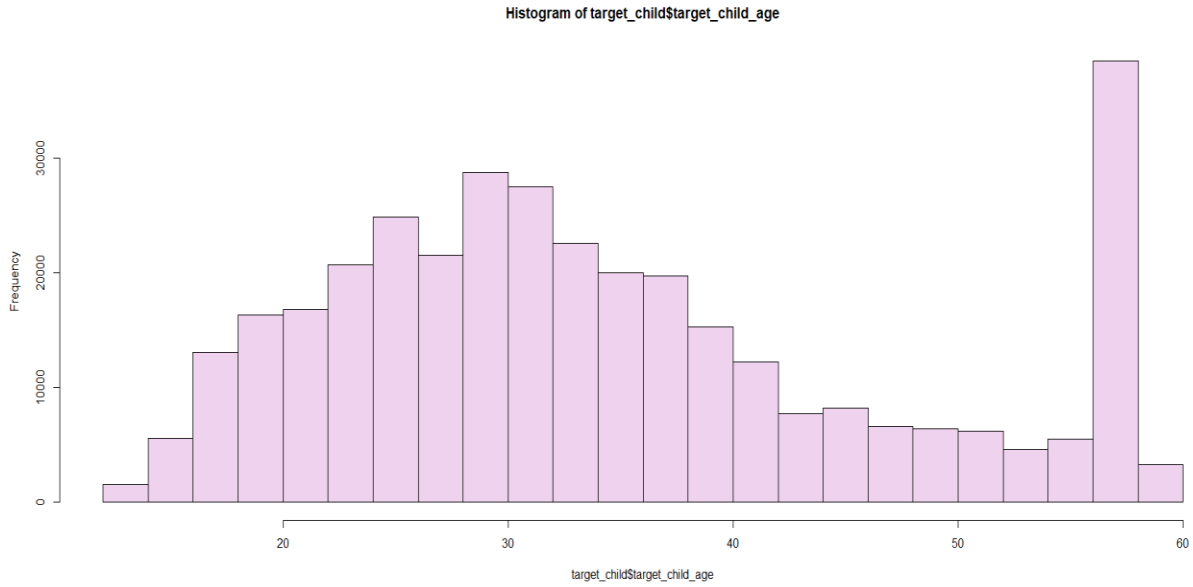


Figure 6 Frequency of Target Child Tokens

The tokens produced by children in Figure 6 more closely follows the data in figure 2 from the adults. The frequency of tokens produced by children begins to increase before 20 months and steadily gains until around 30 months where it starts to level off. Once again, the odd spike of frequency around 60 months makes an appearance.

By reviewing the visual representation of frequency of input it seems that children need to hear a word repeatedly to gain vocabulary and acquire language. The majority of input comes from adults who interact with the children. Figure 7 gives the distribution of adult input to children in the 12 to 60 month age range.

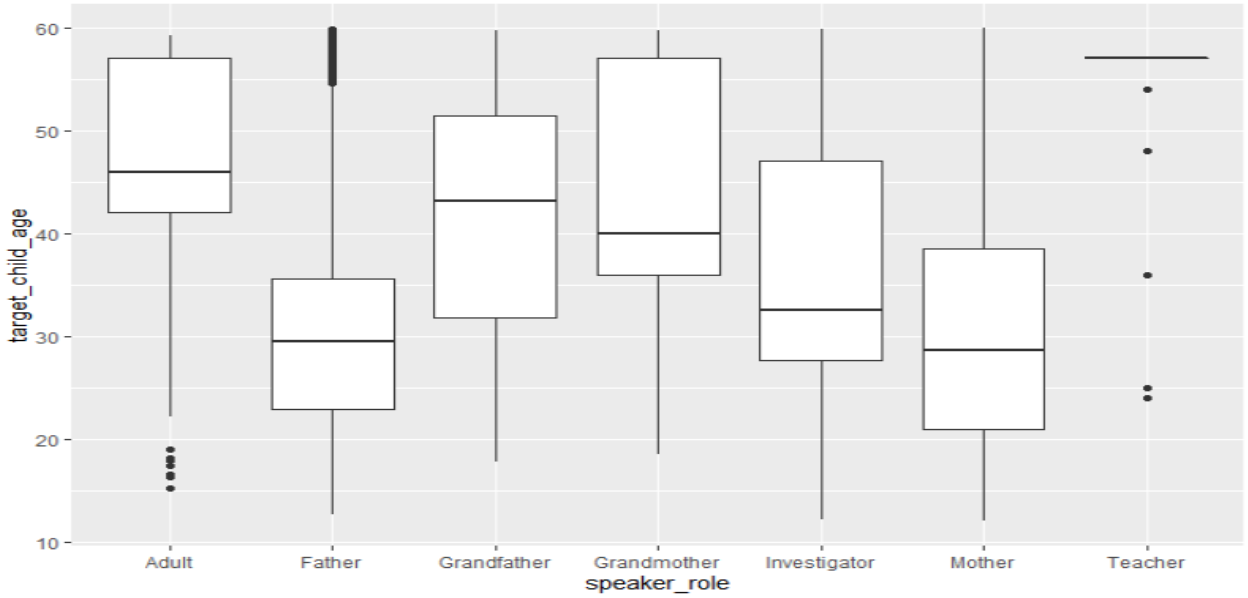


Figure 7 Adult Input by Target Child Age

Just by looking at the frequency data in Figure 7 versus that of Figure 8 below it is not hard to see that the adult influence over children's language development is greater than that of the target child's peers.

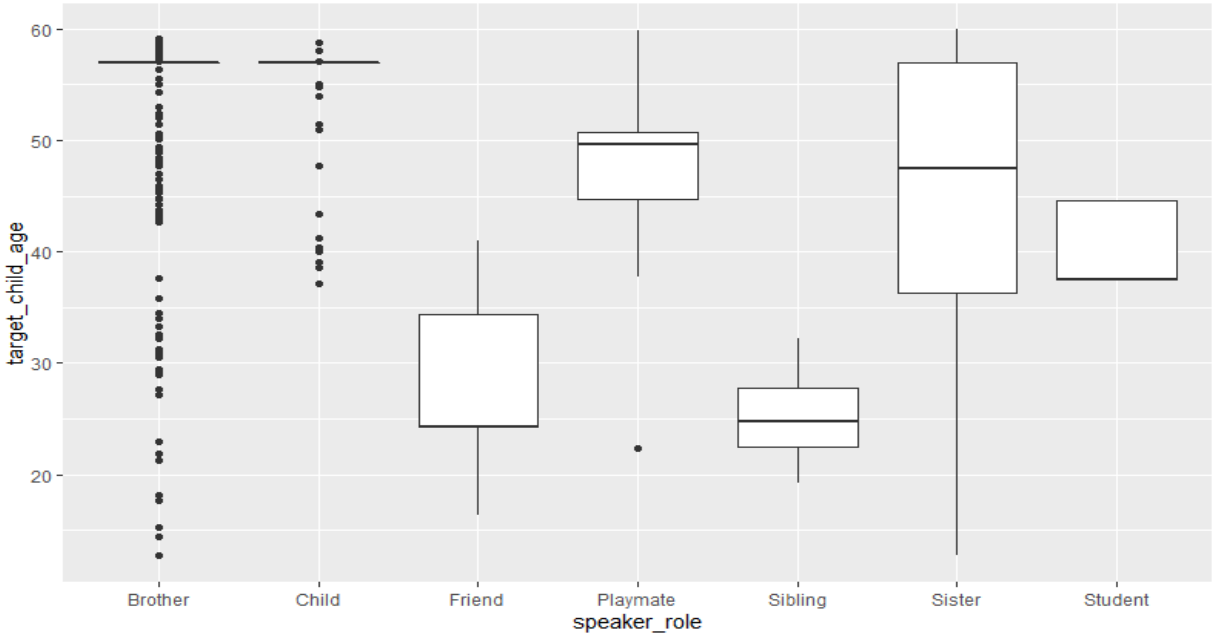


Figure 8 Peer Input by Target Child Age

The box and whisker plot in Figure 8 correlates the token data from peer speakers in CHILDES by target child age. Peer data for this age group, at least from the CHILDES database, is not significant enough to say if it makes a large impact on language development or not. Perhaps a study of children who are of school age and in a classroom where they are more likely to be in peer situations, as well as being more capable of communicating meaningfully, would show a more significant impact of peer speech. These box and whisker plots show frequency well enough, but they do not show what the individual tokens are that adults and peers are using with the target child. In section 7.3 the data related to individual nouns used in the study will be discussed. As Figures 7 and 8 show, the data is more correlated to the relationship between adults and children than between peers and children. The data above shows that children get plenty of input from 12 to 60 months and that they start producing language with some consistency fairly early.

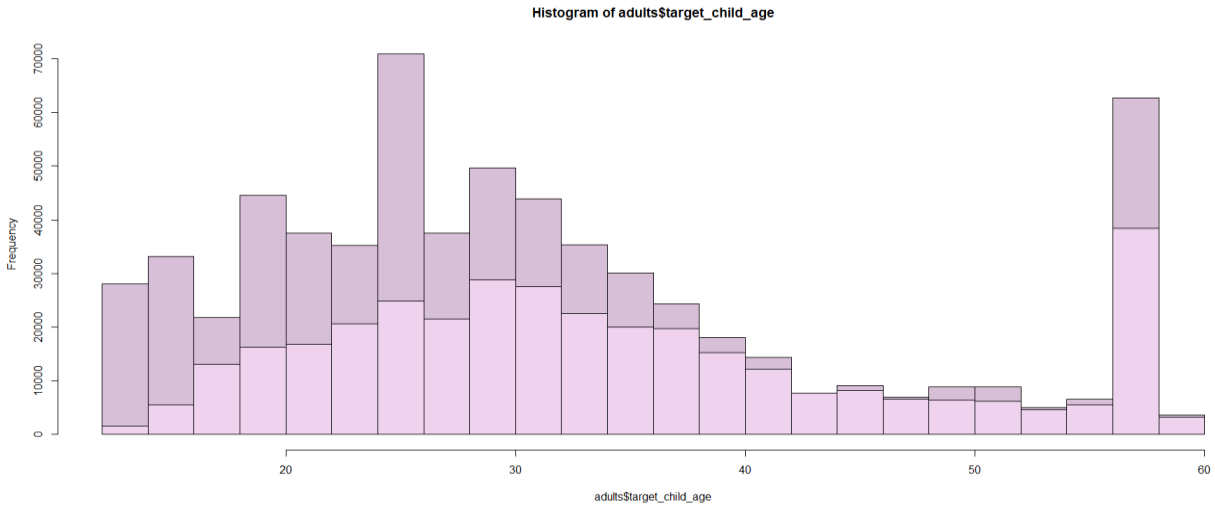


Figure 9 Adult and Child Production

Figure 9 is a comparison of the adult production data with that of the target child. The dark color is the adult data, and the lighter color is the child data.

The word clouds in figures 10, 11, and 12 demonstrate the frequency of token noun production from children in comparison to adults and peers.

Figure 10 Frequent Target Child Words Ages 12-60 Months

Figure 11 Frequent Peer Words Target Child Ages 12-60 Months

Figure 12 Frequent Adult Words Target Child Ages 12-60 Months

The tokens in the data have been all nouns, but the individual words in this category has yet to be examined. The word clouds in the figures above represent the frequency of total nouns and object labels in the study produced by adults, peers, and the target child. Word clouds display words in the data with the largest words being the nouns with the highest frequency from any group with baby, ball, box, car, and mom being among most spoken nouns from all three groups. With graphs like word clouds there are not any numbers associated with frequency of production, but it does give a good visual of frequency. This shows that children's production starts shadowing that of the adults from a young age.

These word clouds indicate that there are many of the same words produced by adults, peers, and the target child. It would appear from the word clouds that frequency does influence children's acquisition of nouns. Not only frequent input from parents but also from the target child's peers. Peer input may not be influential from the beginning, but as the child reaches 60 months and starts school this may prove to be more significant. If the target child has older siblings the data may also prove to be different, but there were not many studies in CHILDES

that had sibling groups to examine the influence siblings may have on each other from a young age. The words *here* and *there* have a high frequency of occurrence. This indicates that objects and things in space were being referred to often enough to be a substantial part of daily interactions.

Until this point token noun input was examined as a whole from a frequency perspective. In section 7.3 individual nouns and the correlation between frequency of hearing it and when the child begins produce specific nouns and object labels were being examined. This was accomplished by using the norming lists from Pearson and CDI to establish which nouns children should know and produce by specific ages.

7.2 vocab lists

Having established the overall frequency of receptive and productive vocabulary in the data in the previous section, it was time to establish the ability of the target children against the vocabulary from the norming lists. The vocab lists were run against individual speakers and age dyads. This was so as to ascertain what vocabulary was spoken to children at what age and at what age they began to produce it themselves. The vocabulary from CDI was for children aged 8 months to 30 months, and the Pearson vocabulary was for children 60 months to 21 years. Nouns and object labels from each test were easily placed into their respective groupings mentioned in section 7.2. Since I was interested in how nouns and object labels produced by adults, and peers, influenced children as they progressed in the language development from 12 months to 60 months, I looked at vocabulary produced by parents, peers, and target children in that age range. Some noun groupings had higher frequencies of production that others. To begin this survey of noun production I looked at the names of animals and labels of toys from the CDI list.

7.2.1 CDI Vocabulary

The frequency of adult production of those animal words can be seen in Figure 12.

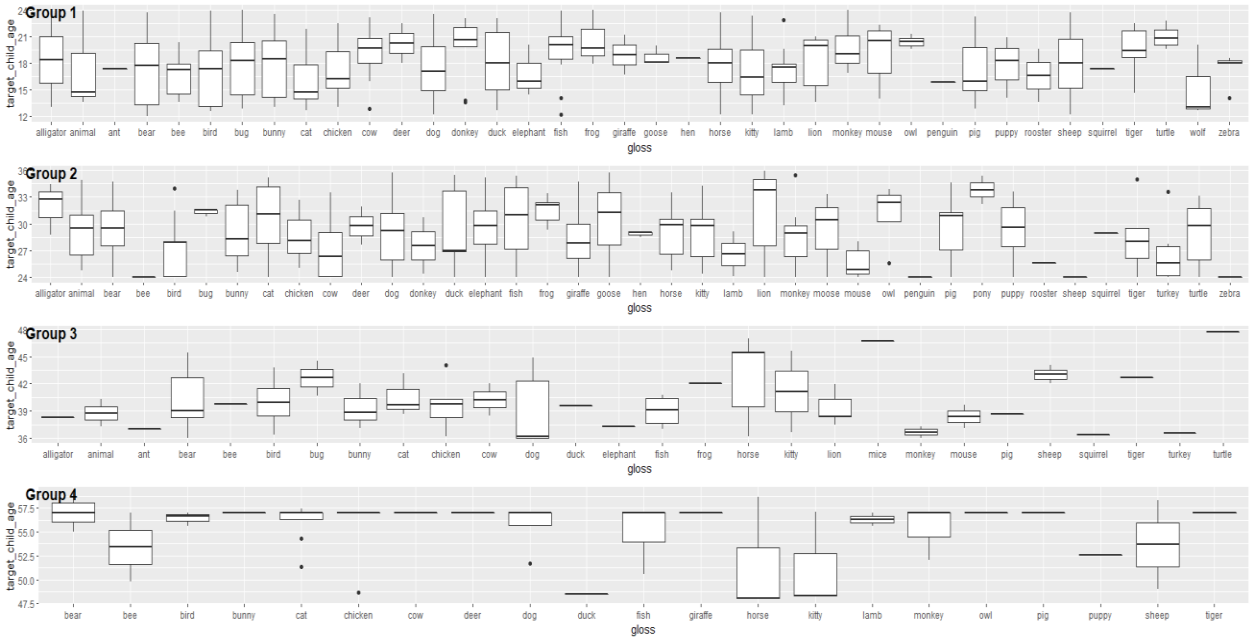


Figure 13 CDI Animal Vocabulary Parent Production

Figure 13 displays the frequency adult production of animal nouns in the four age dyads. There are a few things quite apparent from these graphs. First of all, it is quite apparent that adults use a wide variety of animal words with young children in the 12 month to 36 month age range. Secondly, there are only a few words that are spoken consistently from 12 months to 60 months: *bear, bird, cat, cow, duck, fish, horse, kitty*, and *sheep*. There are a few outliers produced around 12 months: *bird, dog, elephant*, and *pig*. *Kitty* has the lowest median age of production, and *elephant* has the highest median age of production for the 12 month to 24 month range. Per Figure 4 in the previous section, the data in this age group for peer input was

not very significant, and indeed there was not enough for group 3 to create a graph.

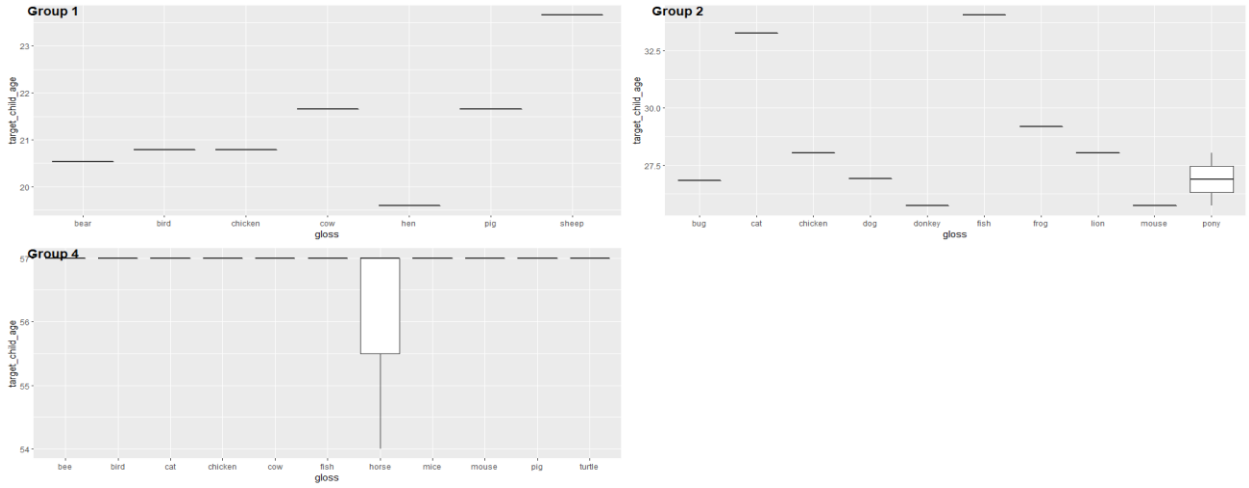


Figure 14 CDI Animal Vocabulary Peer Production Child Ages 12-36,48-60 Months

Peer production of animal vocabulary is not very significant or broad, but the target child is still hearing it from sources other than adults. In Figure 15 is the target child production of the same CDI animal vocabulary as the adult production in Figure 13.

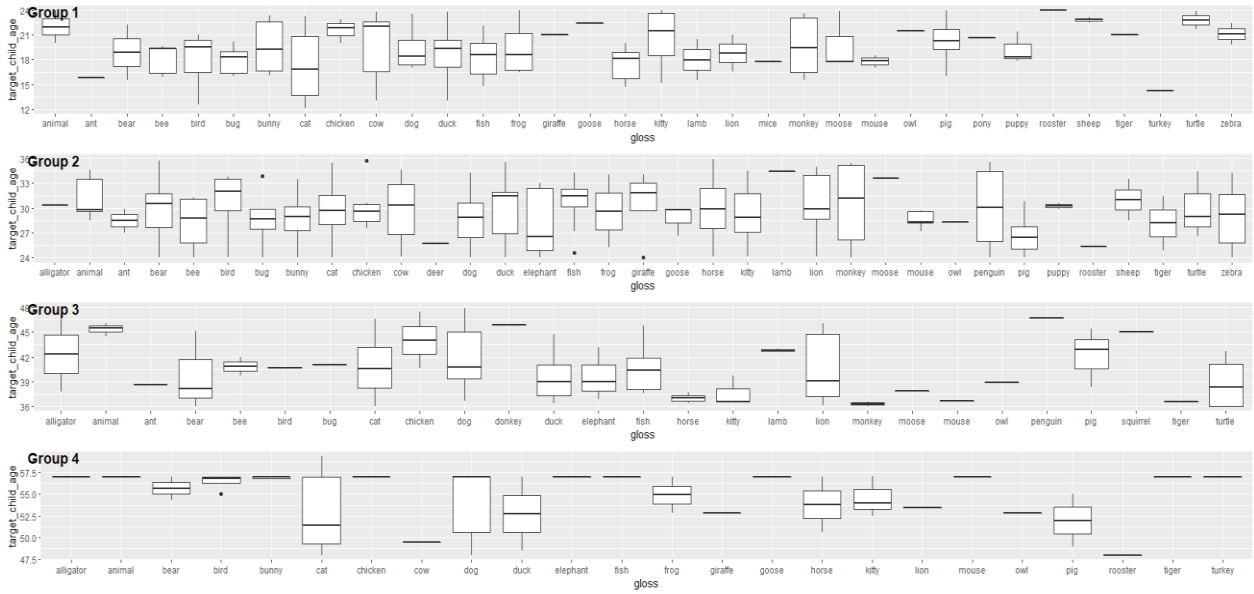


Figure 15 CDI Animal Vocabulary Child Production

With the exception of *cat* and *bird* the target child does not produce much of the CDI animal vocabulary until closer to 14 months. The median age of production is closer to 20 months for the majority of the words in the 12 months to 24 months group. *Cat* has the lowest median age

of production for the same age group. Also, not every word produced by the adults is present in the expressive vocabulary of children in this age group. Notably, *alligator, chicken, deer, donkey, penguin, squirrel,* and *wolf*. The differences between adult production and child production in the 12 month to 36 month age group are quite different. Children do not produce all of the words that they receive in input from adults or peers. Nor do they produce them soon after hearing them. Rather they seem to need to hear them for some time before they begin producing animal names. The other group that had high occurrences of CDI object labels was toys.

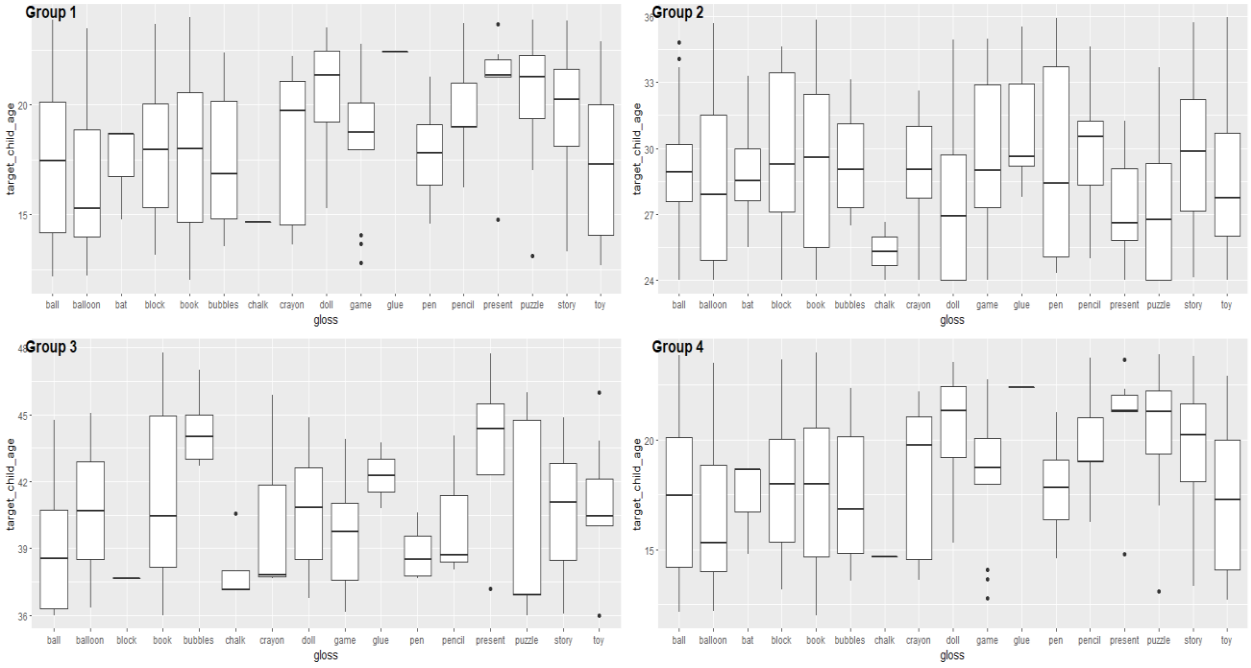


Figure 16 CDI Toy Vocabulary Adult Production

The vocabulary size in this group is not as large as the animal one, but they are produced slightly earlier and more consistently from 12 to 60 months. In addition to adult production of toy related nouns there are a few instances of peer production as well

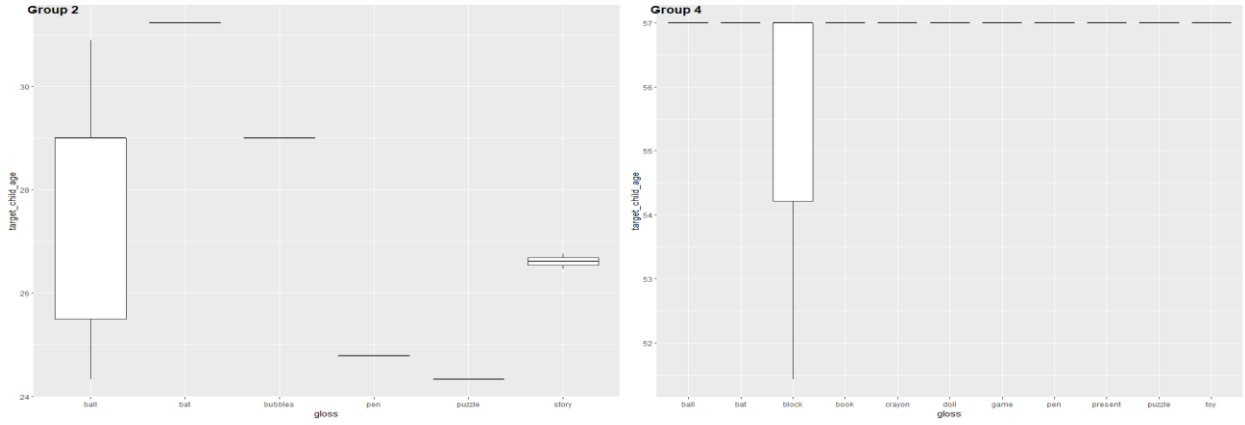


Figure 17 CDI Toy Vocabulary Peer Production Target Child Ages 24-36, 48-60 Months

Peer production in figure 17 is again not significant as there was not enough data in group 1 or 3 to graph. However, it is still interesting to see the interaction peers have with each other. the two words *ball* and *block* are among the most frequent so maybe in play peers are able to orient together to toys in their environment. In figure 18 the target child production again does not follow the adult production in figure 16.

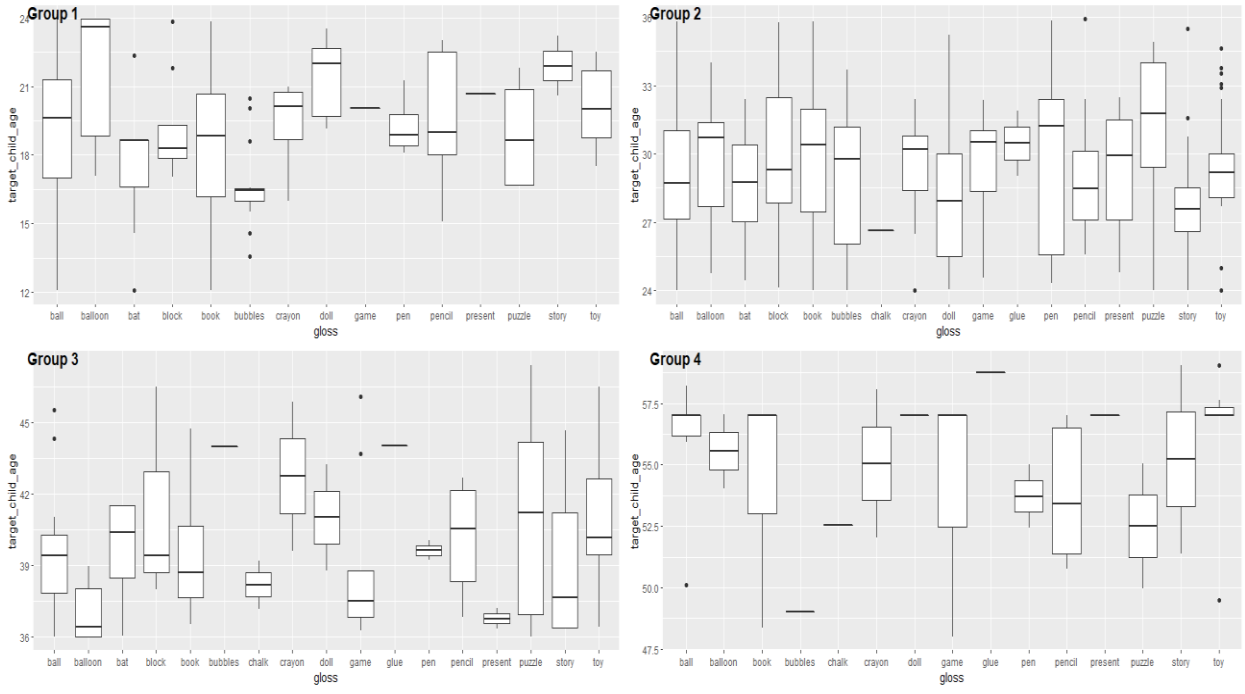


Figure 18 CDI Toy Vocabulary Target Child Production

The target child production of toy labels matches adult production better than that of the animal names. Though, production of the object label does not occur immediately after the

target child hears it spoken for the first time. As noted earlier it takes time for a child to process and begin to speak the word. For instance the word *chalk* is first produced by the adult when they target child is 14 months old, but the target child does not produce it until they are 52 months old. 12 to 24 months is still fairly young to be producing vocabulary, but it is expected that a child, after repeatedly hearing a word for some weeks, or months, should begin to produce it with some consistency.

Comparing the adult production figures with the target child production in the above figures, there is quite a difference between it the adult and target child production of animal names. In the target child production of animal vocabulary is missing many of the animals seen in the adult production. Also, the target child production of the remaining animals is considerably lower than the adult output. This trend continues for toy labels in Figures 14 and 16. For all graphs relating to the target child and adult production it appears as though output drops off in the 48 to 60 month range. If adult production of the animal and toy vocabulary words has dropped off, it could be because target child production as simultaneously increased, and the adult support is no longer needed. However, interestingly target child production appears to have dropped off as well in the 48 to 60 month range. Though the target child appears to still have not acquired all of the vocabulary that was introduced in the 12 to 24 month range. Indeed, the target child seems to avoid speaking some words. Compare the production of adult production of *doll* in Figures 14 and the target child production in Figure 16. Adult production of the word *doll* starts when the target child is around 18 months and continues quite steadily until the study cut-off at 60 months. Whereas the target child in Figure 15 starts to produce the word around 20 months and uses it frequently until 43 months. However, in group 4 of figure 15 there is hardly any utterances of the word by the target child. This seems like it should be unlikely with the frequency the target child hears it uttered from the adults. Perhaps this is an anomaly in the data, or specifically one of the target children.

A comparison of a larger corpus of peer data might aid in seeing where the target child is in their language development. As stated before, there was not enough peer data to graph two of the groups, and even Figure 25 shows that the toy vocabulary produced by peers does not signify much. Indeed, very few toy labels are produced by peers, and *doll* is not one of

them. The lack of peer data in the animal and toy noun groupings does not help to answer my study questions about peers orienting to each other during social situations. Thus far there has not been enough data to show what influence peers might have on the target child. Indeed, after trying to graph the frequency of peer nouns in the 12 to 60 month age range there is not sufficient data to show what sort of influence peers might have over each other. Perhaps if this study were conducted with school age children in a classroom setting there would be more evidence of children engaging in social situations and communicating with one another.

Animal names and toy labels are not the entirety of the data selected for the study there were still other noun groupings to be explored. Unsurprisingly the food nouns had a large variety of production by both the adults and target children.

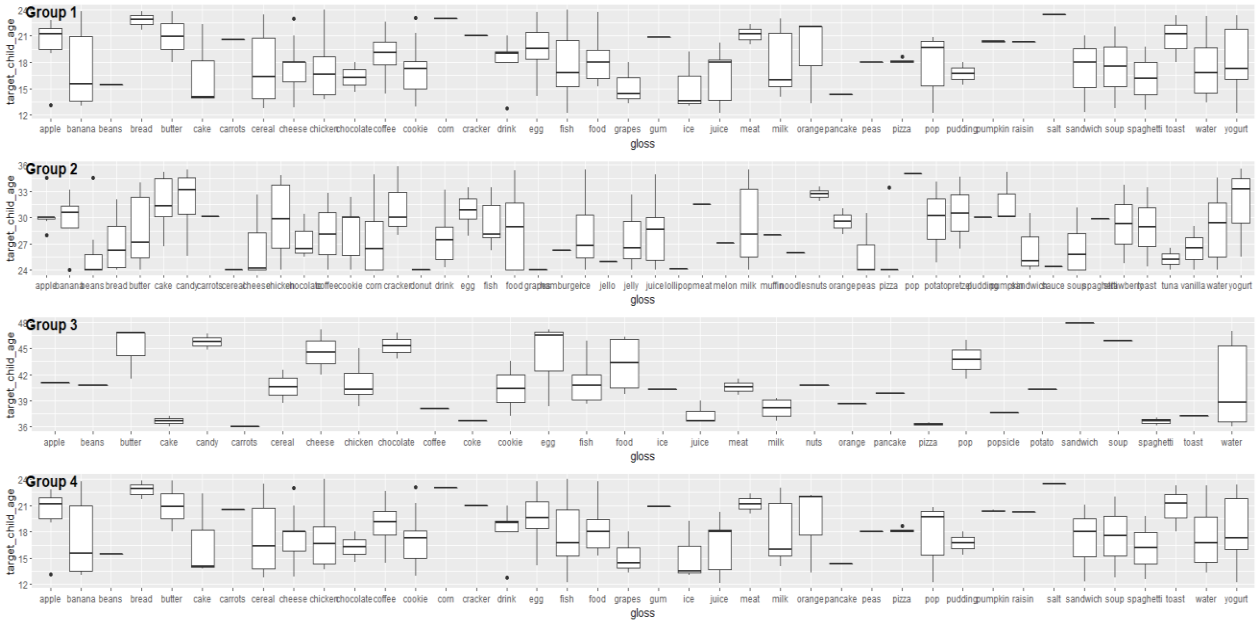


Figure 19 CDI Food Vocabulary Adult Production

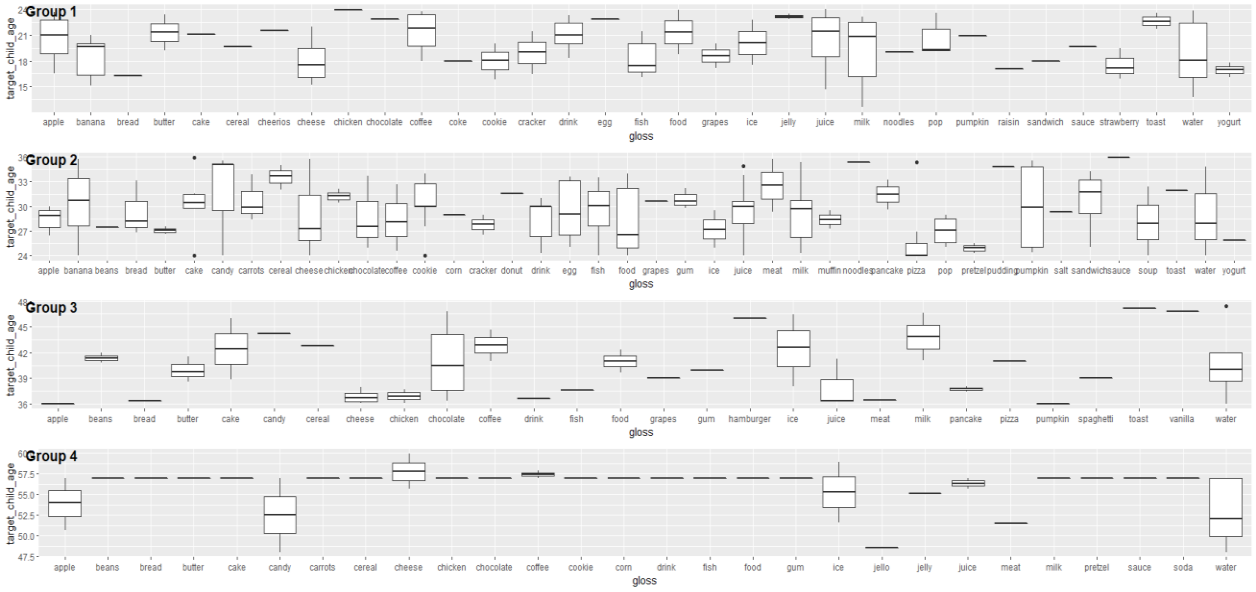


Figure 20 CDI Food Vocabulary Target Child Production

The adults appear to produce a large quantity of food names, especially in groups 1, 2, and 4. However, the target child is not as productive, nor do they consistently produce food names. For instance *milk* is produced by the adult through all four groups regularly but the target child seems to only produce it frequently in group 1, but then not at all in group 4. Peer input was again rather sparse. There was only enough data for group 1 and 4 to graph.

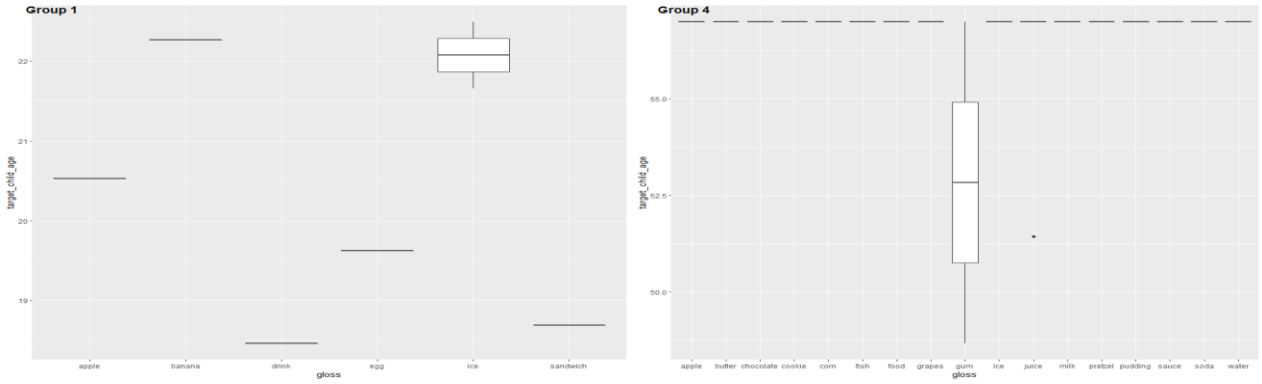


Figure 21 CDI Food Vocabulary Peer Production Target Child Ages 12-24, 48-60 Months

Returning to the word *milk* peers did not produce it until group 4, but only when the target child was around 60 months. Perhaps other noun groups will contain more peer data then these have so far.

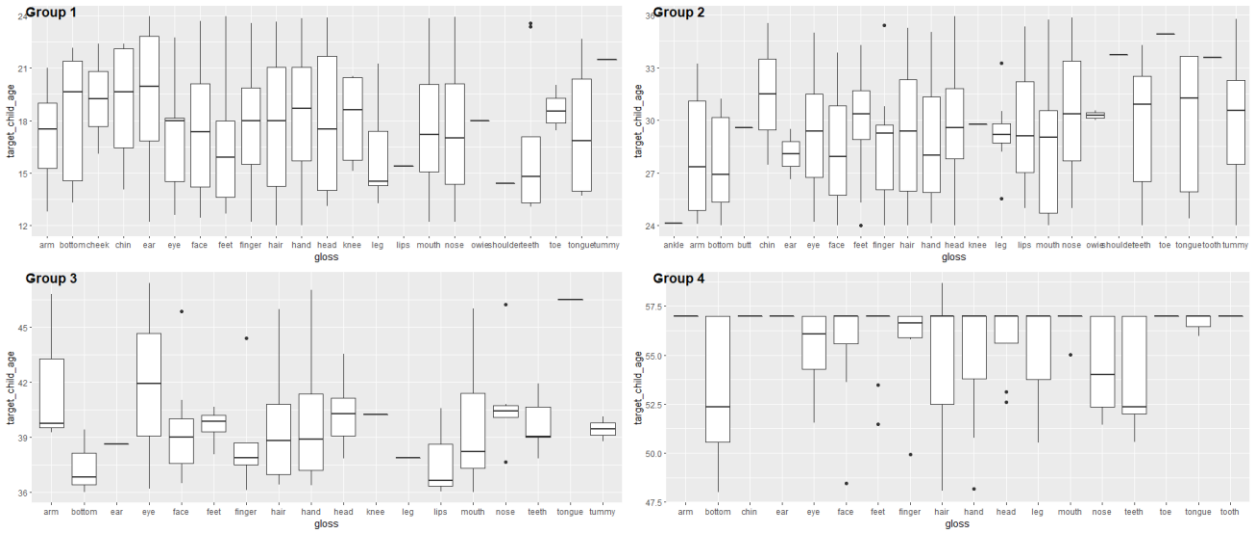


Figure 22 CDI Anatomy Vocabulary Adult Production

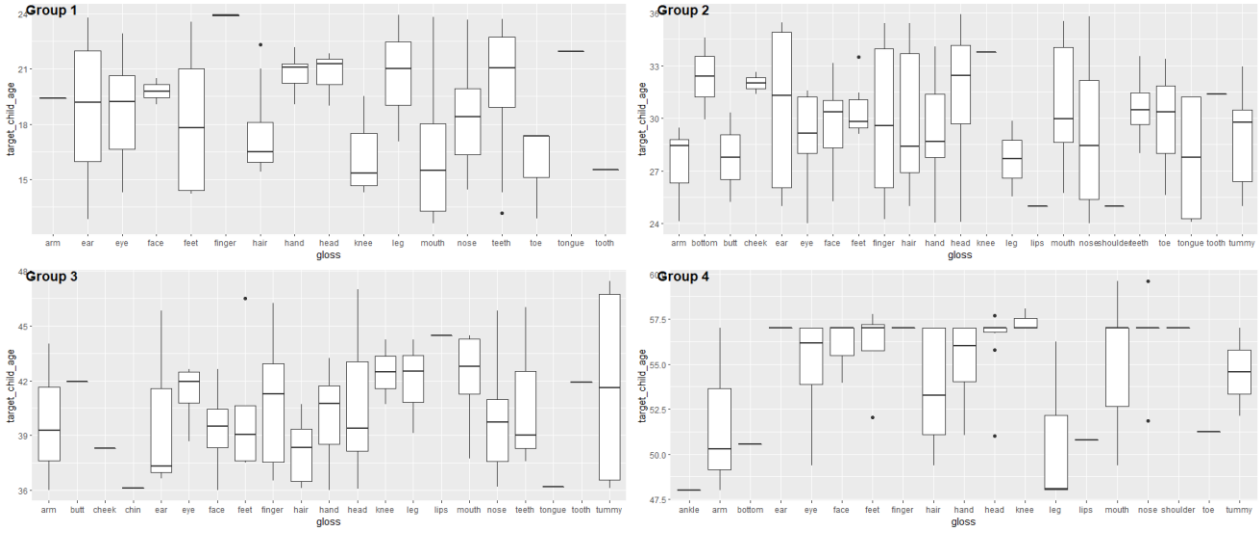


Figure 23 CDI Anatomy Vocabulary Target Child Production

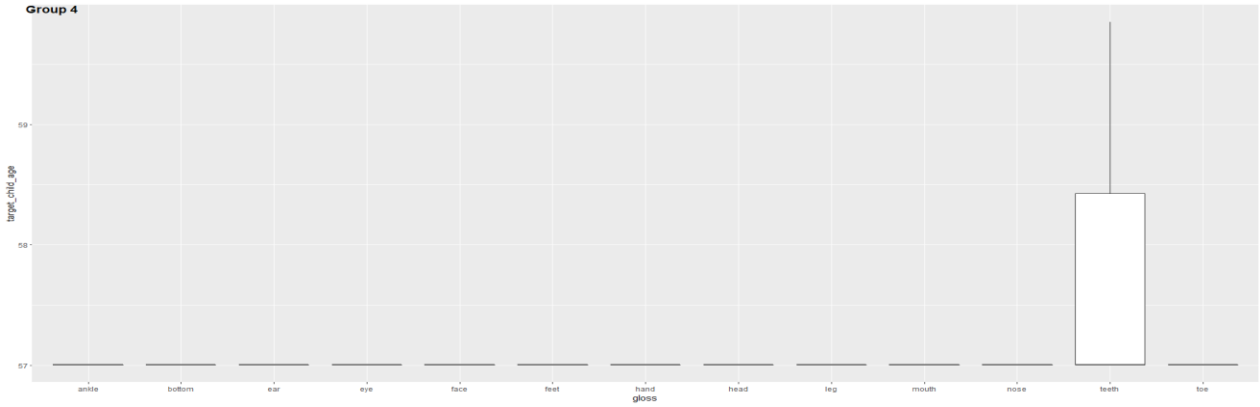


Figure 21 CDI Anatomy Vocabulary Peer Production Target Child Ages 48-60 Months

Figures 22, 23, and 24 are the production of anatomy names produced by adults, peers, and target children. Again the adults produce the most tokens, and peers the least. The word *ankle* is produced briefly by adults when the target child is 24 months old, and it makes an appearance in the target child production around 47 months of age.

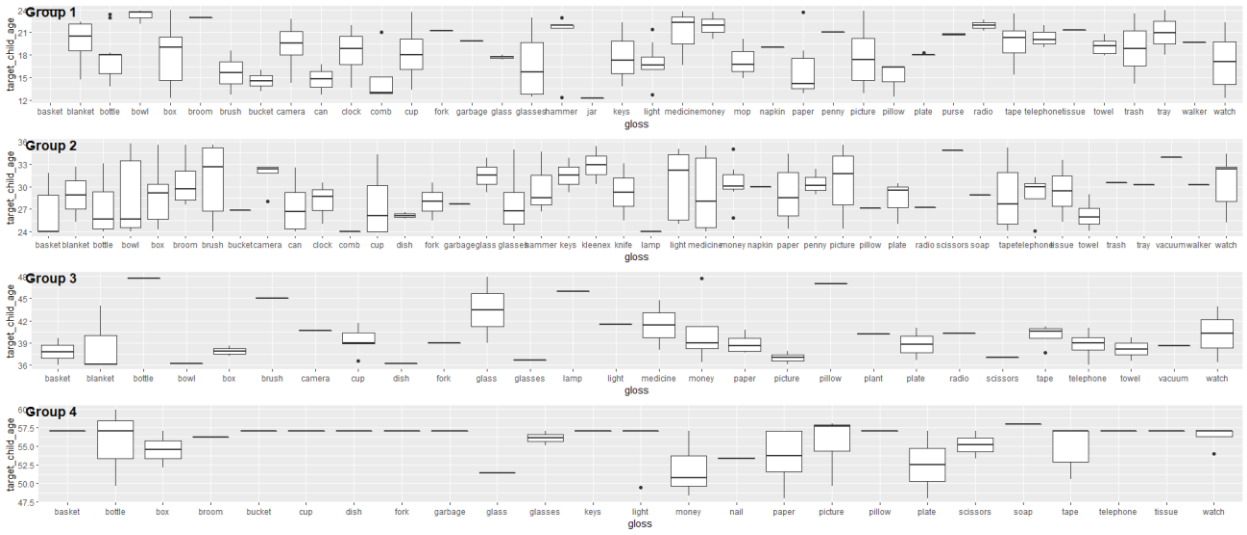


Figure 25 CDI Item Vocabulary Adult Production

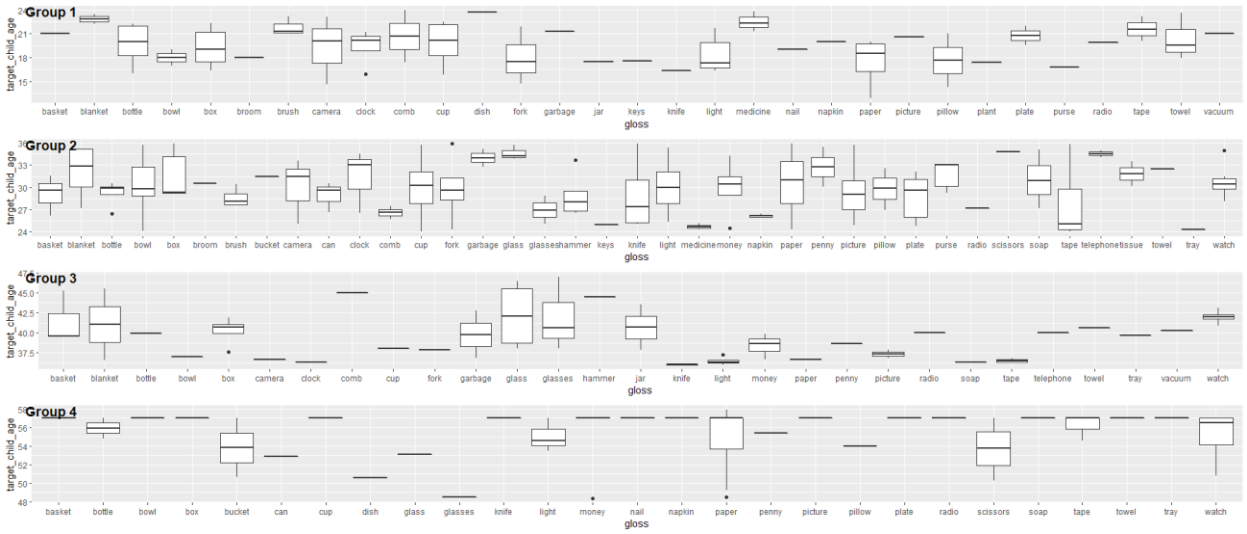


Figure 26 CDI Item Vocabulary Target Child Production

In graph 25 the adult produces *vacuum* briefly when the target child is 32 months and 38 months, but it appears in the target child's production around 40 months. Seems a difficult word for a child to produce, especially without much input and feedback on production. *Watch* was produced frequently by the adult when the target child was 12 to 42 months and tapers off until after 52 months. The target child begins to produce it around 30 months and increases use until the end of the study. There was not any peer data to include for this group of nouns.

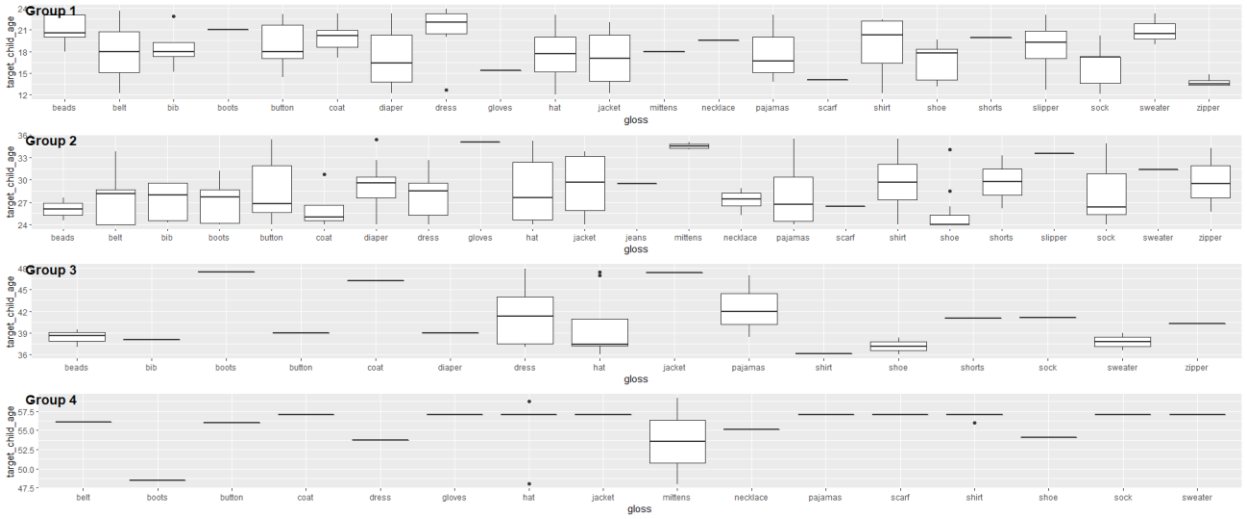


Figure 27 CDI Clothing Vocabulary Adult Production

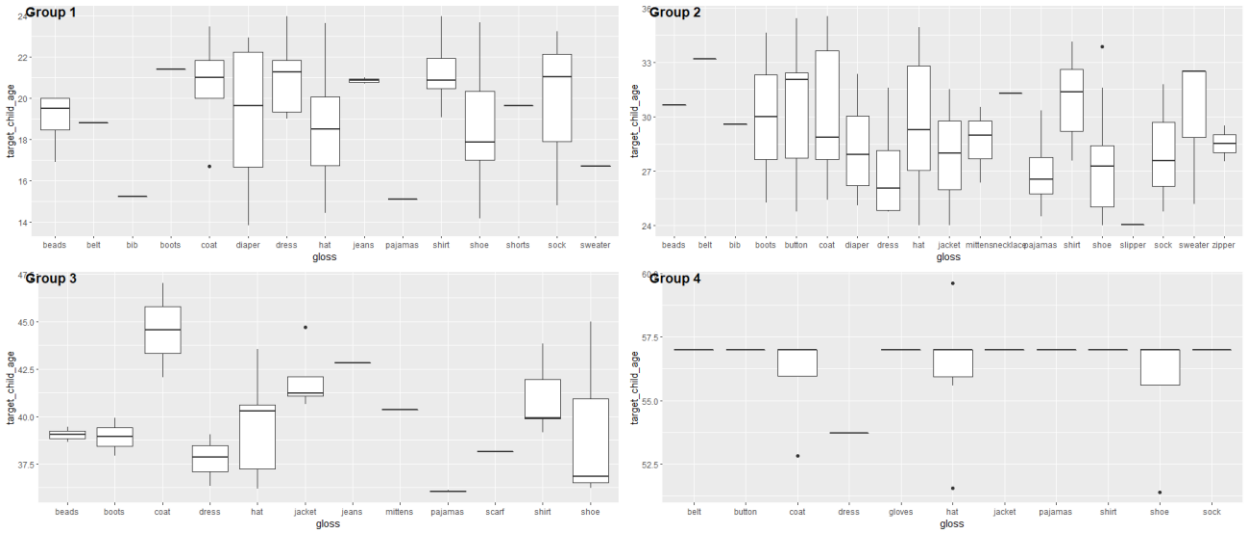


Figure 28 CDI Clothing Vocabulary Target Child Production

The child data in figures 27 and 28 show that the target child followed the adult production quite closely. *Beads* and *zipper* drop off the list at the same time, but the target child appears to learn the word *mittens* without much adult input. Again, there was not substantial peer input for this particular noun group to graph. The next vocabulary group does finally show some measurable peer input.

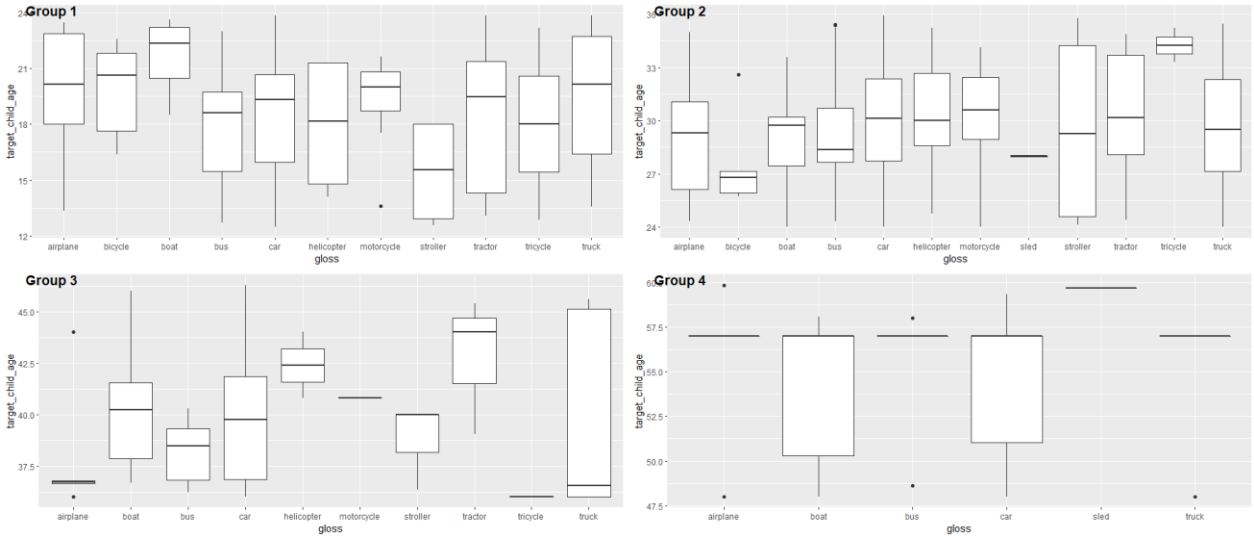


Figure 29 CDI Vehicle Vocabulary Adult Production

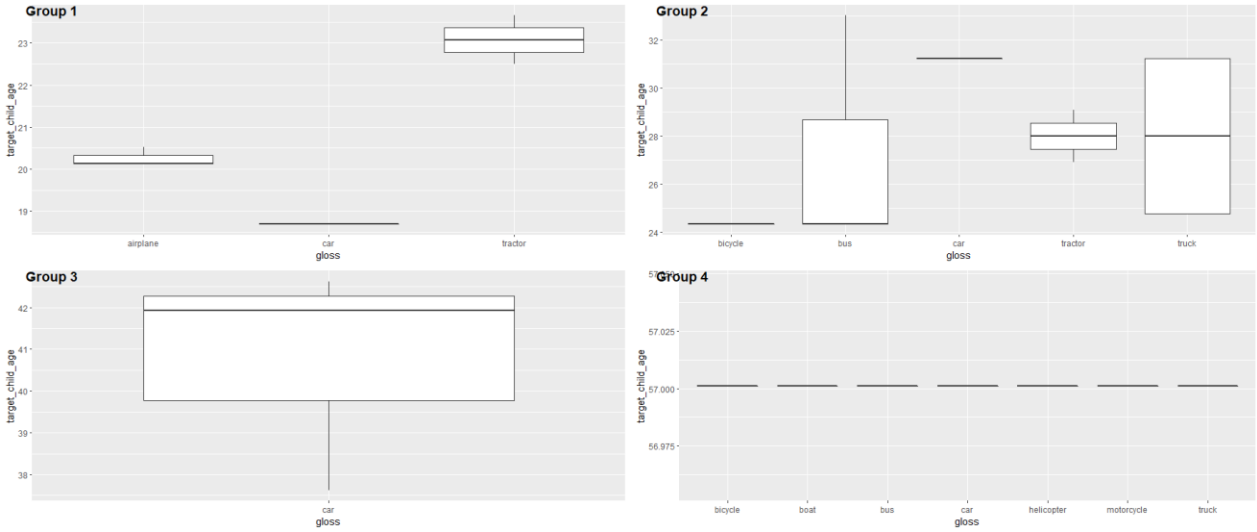


Figure 30 CDI Vehicle Vocabulary Peer Production

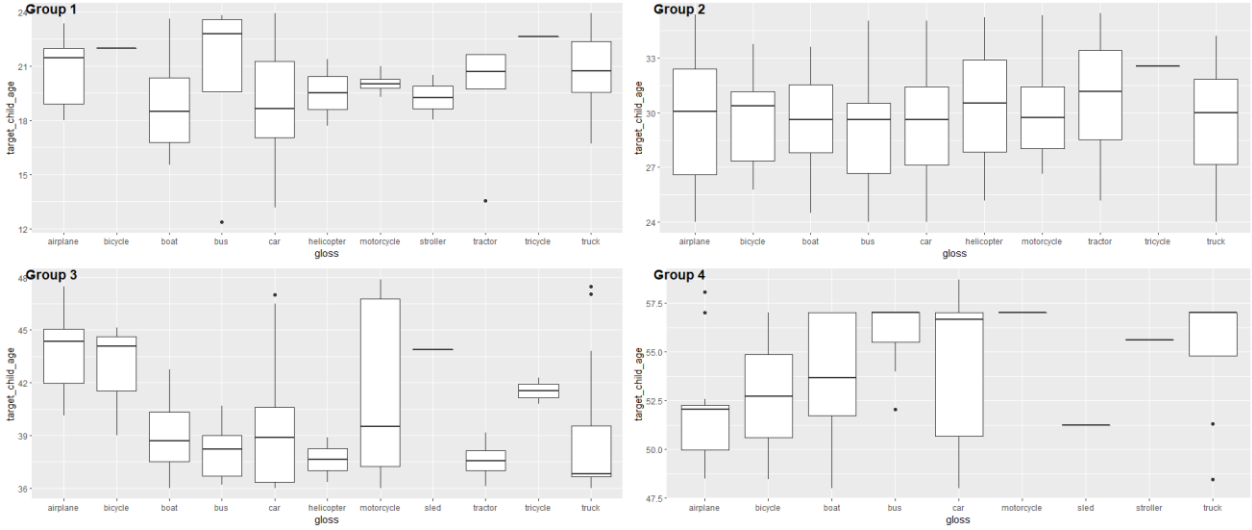


Figure 31 CDI Vehicle Vocabulary Child Production

For once there was enough peer data to graph all 4 groups. Interestingly, the target child production data picks up in group 4 where the adult production drops off. For example the adult produces *truck* often until group 4 where it barely makes an appearance. Though *truck* makes more than an appearance in group 4 of the target child production. *Truck* is in the peer data in group 2 and 4, perhaps playing with toys has more of an influence on the continued production by the target child. Target children appear to receive quite frequent input of the CDI vocabulary from adults in the 12-60 range. In the next section the vocabulary from the Pearson CELF-5 test will be examined.

7.2.2 Pearson Vocabulary

Having looked at the vocabulary from the CDI test, which is only for children up to 30 months, I then ran the same analyses with the Pearson vocabulary which is meant for children 60 months and older. Despite the nouns from the Pearson CELF-5 being intended for children around 60 months I ran it against all four age dyads, meaning I looked at how frequently children 12 months up to 60 months used vocabulary meant to test older children.

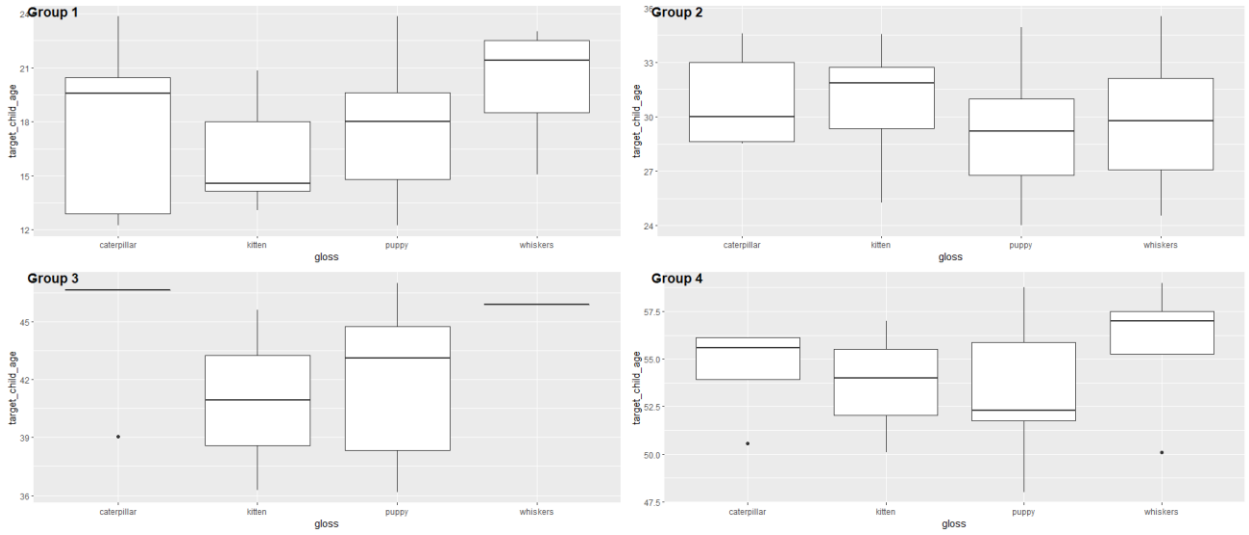


Figure 32 Pearson Animal Vocabulary Adult Production

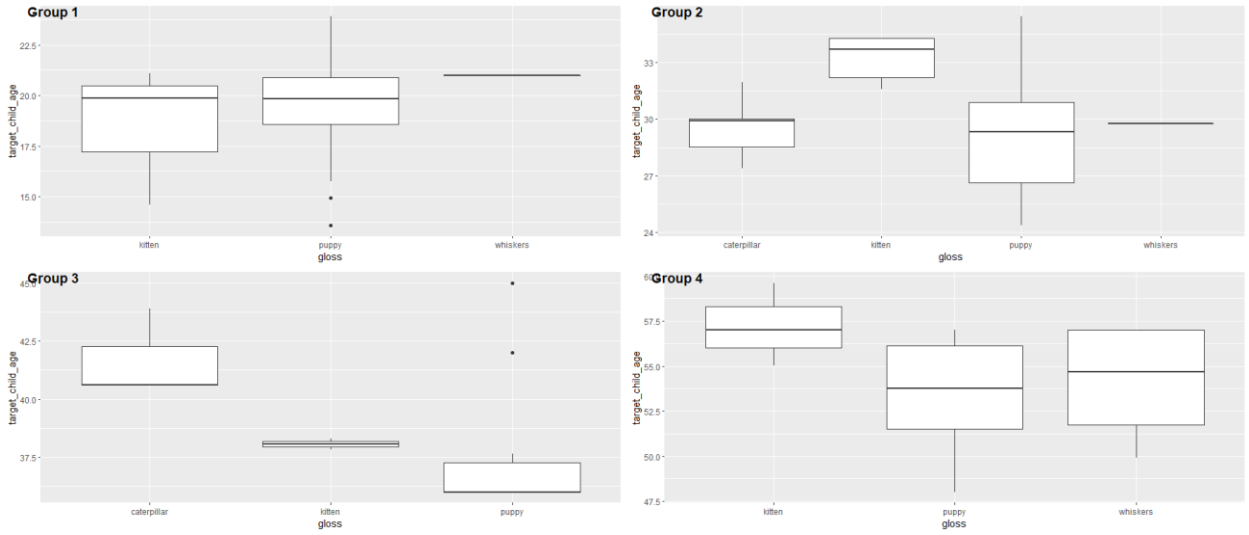


Figure 33 Pearson Animal Vocabulary Child Production

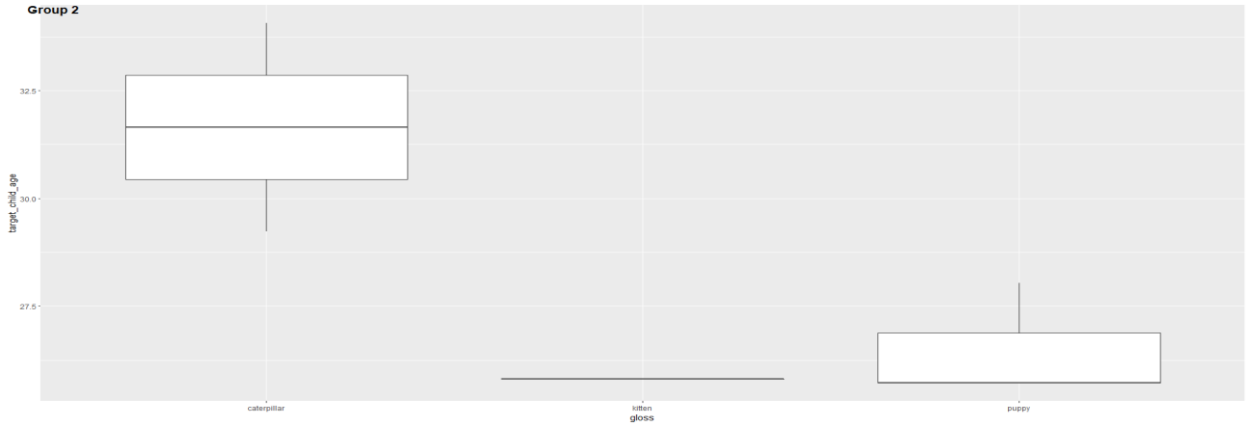


Figure 34 Pearson Animal Vocabulary Peer Production Ages 24-36 Months

It is odd to see a word like caterpillar pulled from a test meant for 60 months and older appeared in a 24 to 48 month old's productive vocabulary. Also of note the fact that it is not produced by adults or target children in the 48 to 60 month age range. Interestingly peers spoke it to, or near to as the case may be, the target child in the 27 to 33 month range.

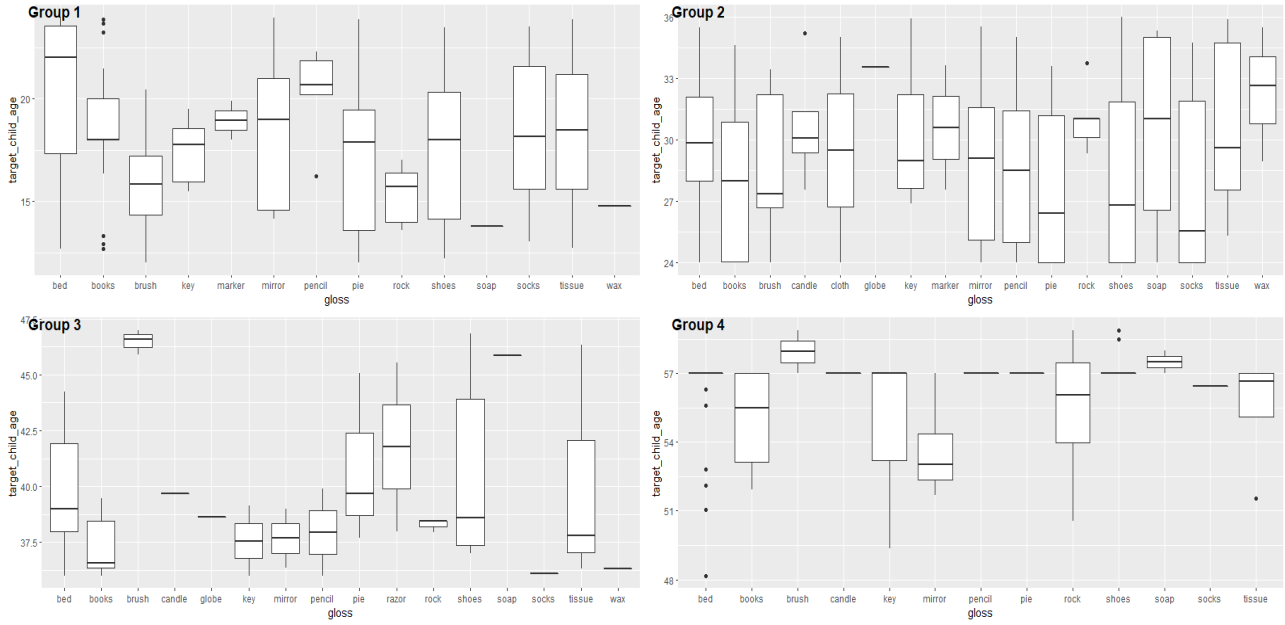


Figure 35 Pearson Item Vocabulary Adult Production

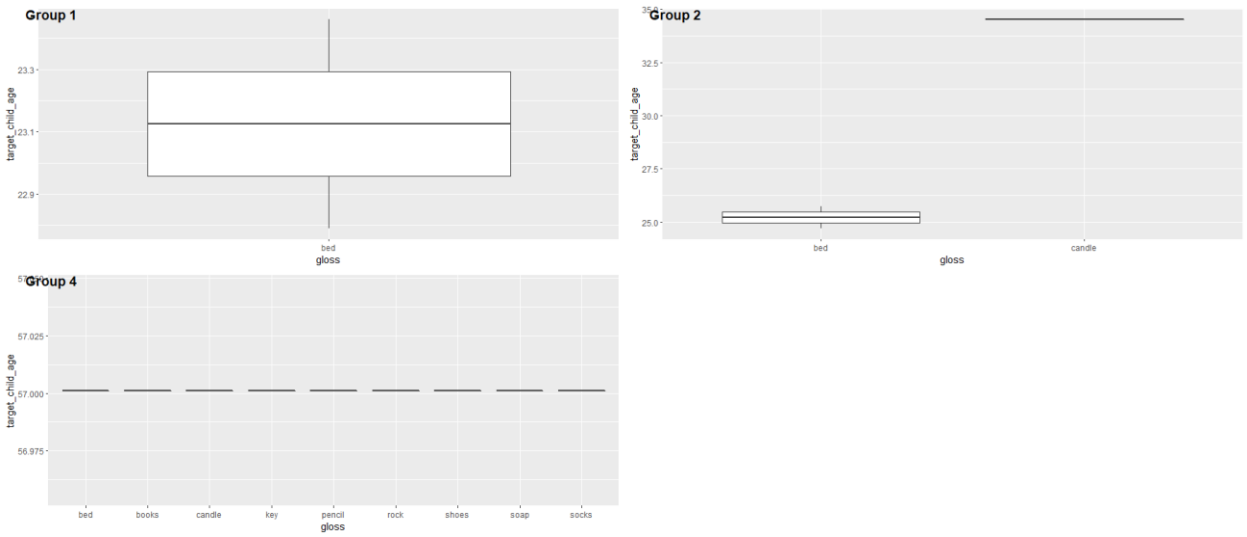


Figure 36 Pearson Item Vocabulary Peer Production Target Child Ages 12-36, 48-60

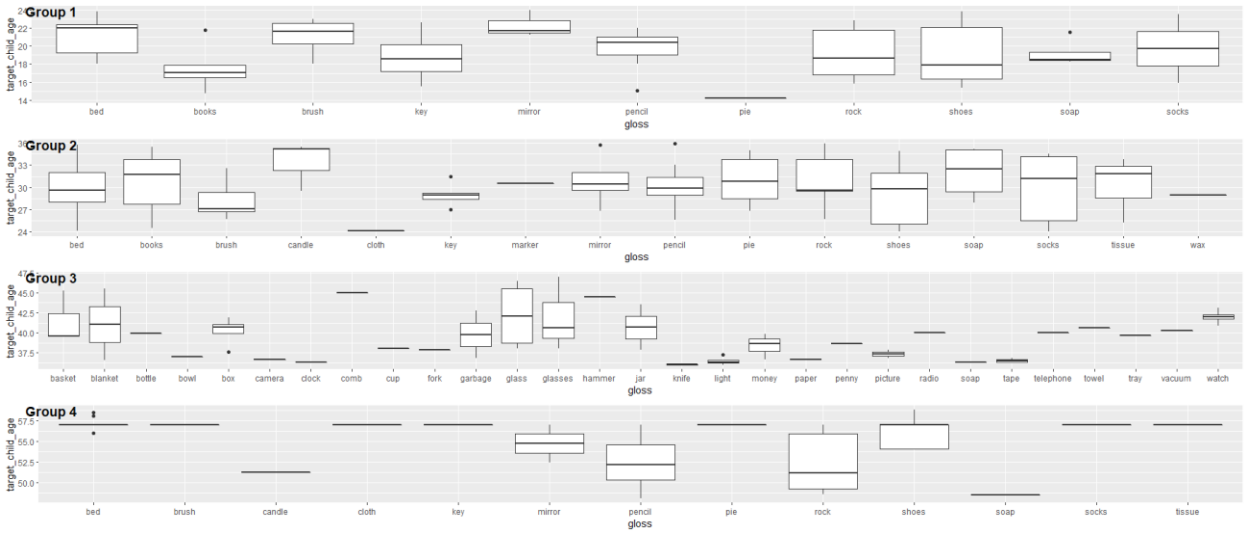


Figure 37 Pearson Item Vocabulary Target Child Production

Again something odd in this data set. The target child begins producing the word *basket* in group 3 around 40 months, without any production from adults or peers prior to the occurrence in target child speech.

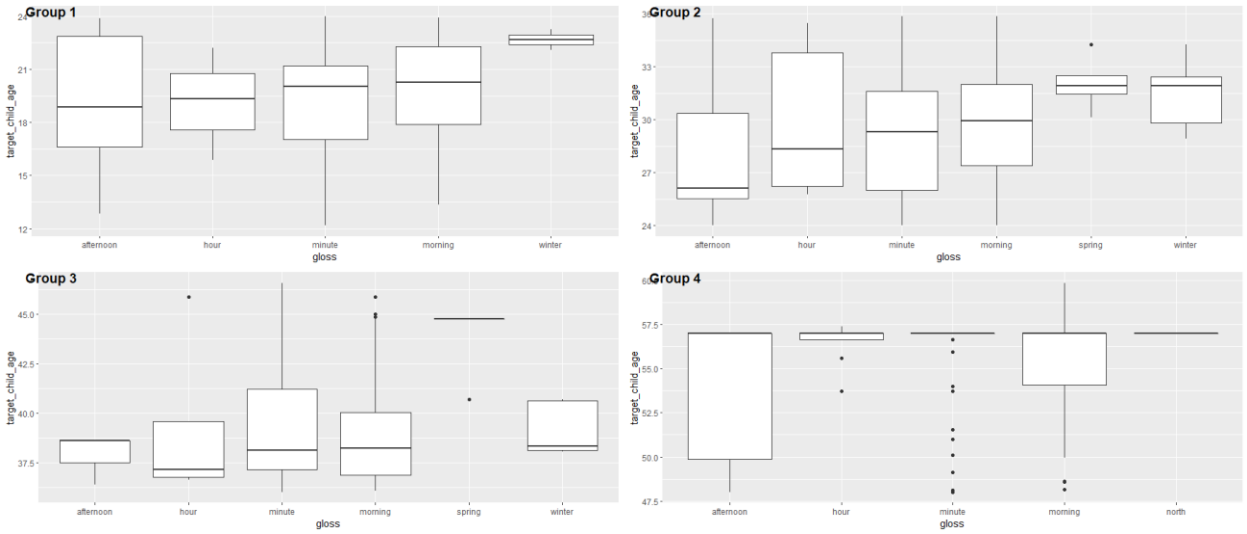


Figure 38 Pearson Time Vocabulary Adult Production

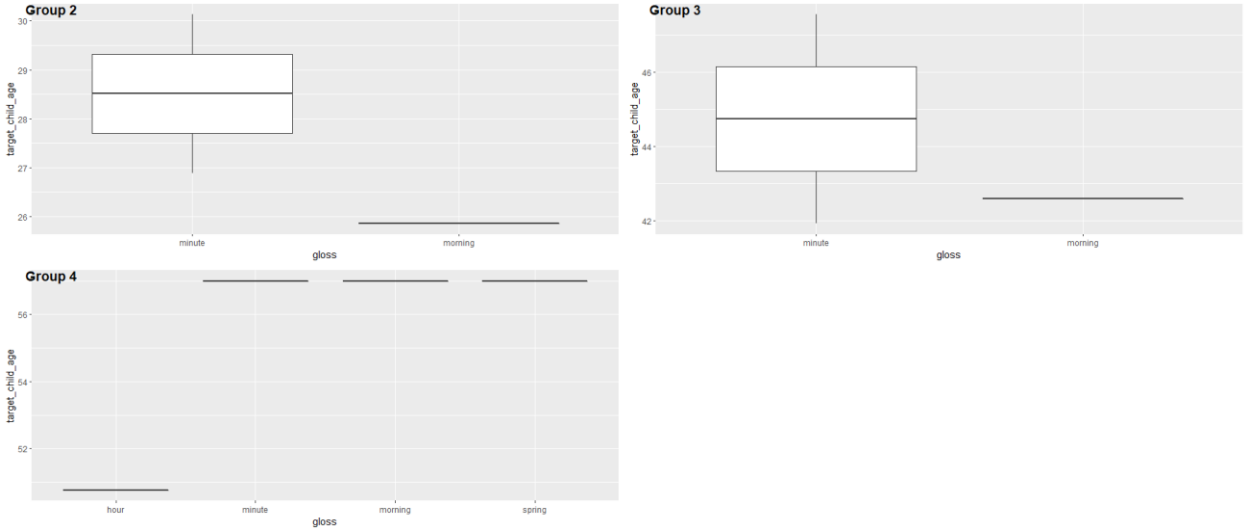


Figure 39 Pearson Time Vocabulary Peer Production Target Child Ages 24-60

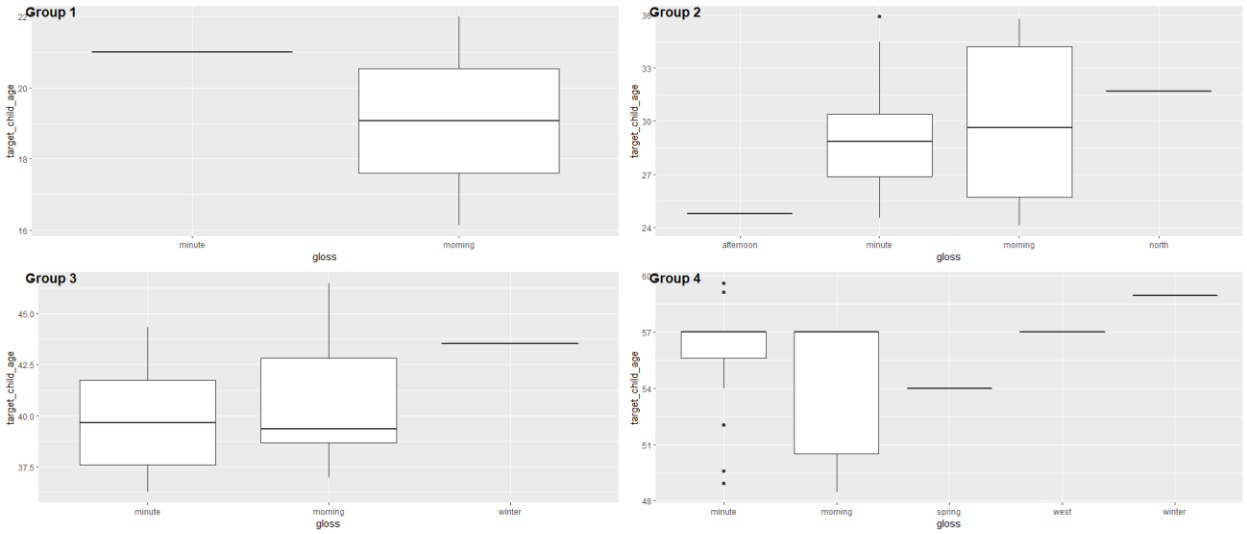


Figure 40 Pearson Time Vocabulary Target Child Production

This group of nouns is interesting as they are connected to more abstract concepts and are not as easily connected to concrete objects. The target child seems to pick up *minute* early on and increase in use. As well as *morning*. Perhaps because they are the most easily connected to a specific time frame in conjunction with high frequency production from adults. Those two words are also the most consistent productions from peers for this category.

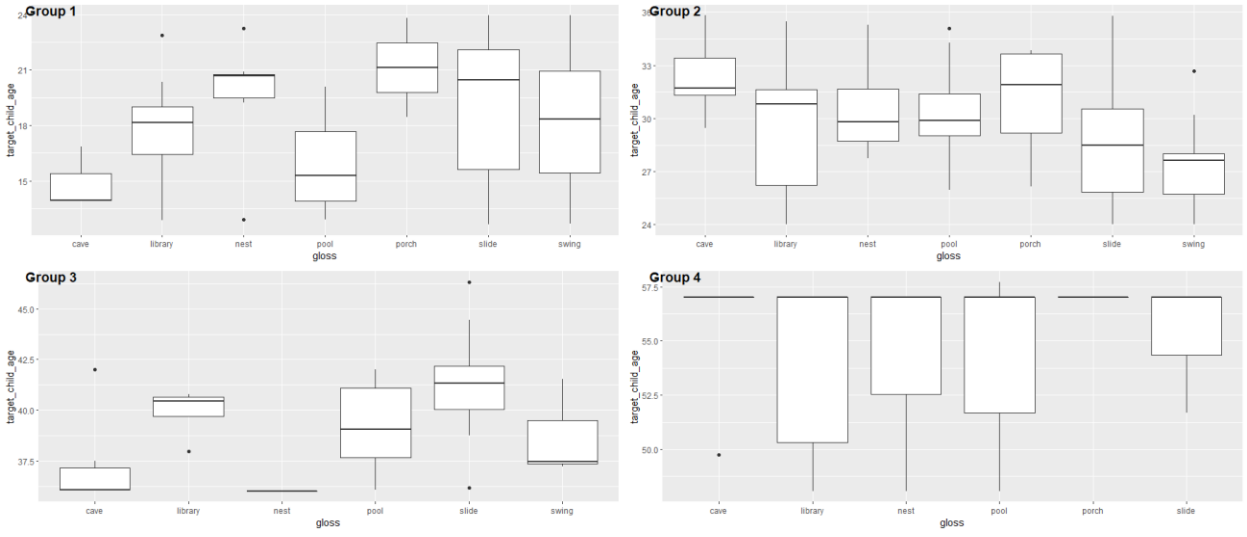


Figure 41 Pearson Place Vocabulary Adult Production

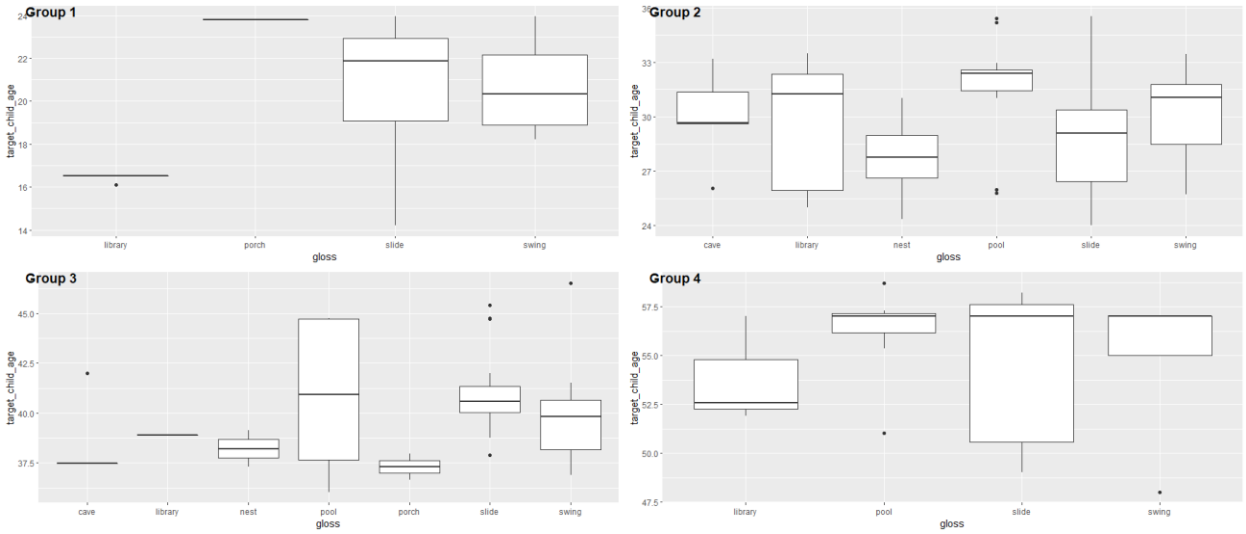


Figure 42 Pearson Place Vocabulary Target Child Production

Cave is produced quite early by the target child. It first appears in the target child production around 26 months with adult production occurring around 12 months. Though it does not appear at all in the 48 to 60 month age range, maybe because unlike the other locations in this noun group the target child may not have a clear picture of what a cave might be. *Slide* and *swing* are present in the target child's vocabulary from 14 months and 19 months respectively. They are abundant in the adult productive vocabulary, which could be the main reason the child produces it so soon and frequently, but it could also be because of positive associations with the objects themselves.

7.3 Discussion

The analyses in the previous sections looked the frequency of both noun tokens and age specific noun acquisition respectively in order to answer the research questions and hypotheses. Having presented the visual break down of my data I will now summarize the general patterns and implications of the results. This visual survey of child acquisition of nouns and object labels was meant to analyze the frequency of child production of specific nouns and object labels. Which was compared to adult production of the same tokens. Frequency of input from adults and how it affects children's long term language development was the main focus of the study.

First question was whether the frequency of adult production of nouns increased or decreased over the target child age range of 12 months to 60 months. Both this research

question and the following were inspired by the Goodman et al. (2008) article on frequency of parental input. In figure 2 the adult input frequency is displayed and in figure 6 is the target child frequency. Figure 2 shows that initially adult input begins with high frequency that steadily increases from 12 months to 30 months after which it decreases slightly then begins to level off. Target child frequency in figure 6 starts low around 12 months and steadily increases until 30 months, where again it slowly tapers down. The histogram in figure 2 that displays the number of tokens in caregiver input per child age, the data are skewed positively. Meaning there is less production when the target children are older. This either means that parents speak more when the target child is younger, or simply, that there are fewer studies in the CHILDES database that pertain to children under 60 months. These results also address the hypothesis I had of adults frequently producing nouns and object labels when the target child was around 12 months, which would then influence the target child's ability to produce the same nouns and object labels when they approach 60 months. This is true for the majority of the nouns, but there were a number of nouns as noted in section 7.3.2 that the target child experienced in the 12 to 24 month range, but they still did not produce it when they reached 60 months.

In section 7.2 I used word clouds to showcase the words of highest frequency in the data. There were a number of nouns and object labels that appeared in the adult, peer, and target child data. The appearance of the words *here* and *there* in the word clouds indicated that parents, peers, and target children refer to objects and things in space often, which suggests that environment plays a role in language acquisition. Demonstratives such as these demand joint attention by the child and give them a frame of reference for the co-occurrence of the object and its label (Diessel & Coventry, 2020). Because *here* and *there* occur with high frequency it further attests to the social aspect of language acquisition—you must be paying attention to the speaker as they indicate the object in order to follow the conversation.

To address the second question of which nouns adults produce when the target child is around 12 months and what the relation is with the nouns the child produces when they are around 60 months section 7.3 analyzed the data in light of two vocabulary lists. Analysis of the data does indeed show that children around 60 months begin to produce vocabulary that is

produced frequently by the adults around them when they are 12 months of age. This is in line with the findings from Goodman et al. (2008), Ferguson & Waxman (2017), Yu & Smith (2012), and many others that show frequency of input from adults, particularly parents, plays a significant role in child vocabulary development. Through the visual representation of each vocabulary group it is easy to track which nouns and object labels are spoken to children early in the study which words they speak. Frequency of nouns and object labels here may truly reflect how often caregivers reference them, or they may possibly be the result of the stimuli used in the individual studies that make up the data. For instance the frequency of CDI animal words in the adult input may be simply due to these animals being overwhelmingly present in story books and toys available for children. One other conclusion that emerges is that toys are possibly referred to earlier than other items, which suggests an embodied account: children are embedded in their environment (with toys), and later parents start using books and refer to animals (also possibly outside the house). The embodied idea is also supported by the presence of food terms (child daily routines). A final note on words which do not appear in input to target child, but are present in production: the records we have simply do not record each and every instance of the child being exposed to a certain word

The third question focused on peer influence on vocabulary development. However, the CHILDES database focused on adult input and frequency of tokens over that of the target child's peers. As a result there is not enough data for this study to measure the influence peers might have over each other on their language development. Since the data for this study came from children aged 12 months to 60 months who are not yet in school, and the children were not often recorded interacting with siblings, this makes sense that there would not be enough data to show what, if any, influence peers have on each other's vocabulary acquisition. Either the target child's peers do not speak much near or with them, or the data from CHILDES does not contain enough natural interactions to measure the influence they might have over each other. A separate study would need to be conducted specifically searching for peer interactions and how they influence language development.

8. Summary and Conclusion

Children learn language through hearing it spoken by those around them. Over time what they hear as a stream of sound begins to form patterns. Those patterns become words, words start to form strings of meaningful communication, and a child is well on their way to becoming a fluent speaker of a language. By using a database of child language to study the frequency of adult and peer input it should be easy to start mapping the journey children take to acquisition. What I hoped to discover in this study was how the early input of adults shaped the language of children as they progressed in language development. I also hoped to demonstrate that peers do have some influence on each other when it comes to language learning.

As demonstrated in chapter 7 adults do have the most influence on children's language acquisition in the 12 month to 60 month age range. For the most part the target child's production of the CDI and Pearson vocabulary does match the input they receive from the adults, with a few exceptions. Unfortunately, there is not enough peer data in the CHILDES database to say what, if any, influence peers have on each other's language acquisition and vocabulary development.

8.1 Study Matters

There are some, Noam Chomsky for one, who do not think that corpus data are valid methods of language research. However, there are many who think that corpus data is the best tool to analyze language phenomena because corpora are limited in what they contain. Using a corpus can be very useful for studies. There is an entire branch of linguistics that focuses wholly on finding linguistic phenomena through data in corpora. Though there are a few drawbacks to using corpora for study. For instance, the CHILDES database I used is slightly older data mostly from the 80's. Most of the studies in the corpus are conducted in lab settings where the researchers were looking for specific phenomena that would skew the data toward their study rather than natural language use in general. Also, unless the transcript has a video accompanying it, it is hard to infer where the child's attention is focused or being directed. Even if there is an accompanying video, there is not an easy way to incorporate the additional visual data to provide further context. Text corpora miss out on this important fact of child language development unless they are annotated in a way to convey visual and gestural cues.

There are also a number of other flaws that showed up in the data, which have been referenced throughout. The data visualization shows a spike in utterances between 58 to 60 months for the adults, peers, and target children, but it is not a reflection of more interaction between participants in that age range. Rather it is because there are many more studies included in the CHILDES database that include children around that age—it is due to the larger amount of data available for that age that the spike in utterances appears. There are also gaps in data from different speakers because there simply is no data for them from any study included in CHILDES. Namely, there is no data for the speaker Father for children 36-48 months, and speakers from the peer group are never consistently recorded. There are no outliers, or anomalies, in the database data either. This is more than likely because that data is a compilation of studies conducted by other researchers. It is not common to keep outliers in studies as they are usually an individual anomaly. Not having outliers is not a problem per se, but in a study conducted in lab rather than through a database would normally have at least one.

8.2 Future Study

For myself future study on the topic of noun and object labels would focus more on peers and their influence on each other. This was my original plan for this study, but as stated in chapter 6, covid-19 interrupted my plans and I had to alter my study to match the circumstances. Since the CHILDES database covered more of adult and child interactions it made sense to switch to adult influence on young children's acquisition of nouns and object labels—rather than focus on peer interactions. There are many studies, which have been referenced herein, that indicate peers do play a significant role in child language acquisition. This is an important part of language development that needs more exploration in future studies.

Goodman et al. (2008) used CHILDES as a way to back up their findings in a separate study conducted with parents, that parental frequency aids in early acquisition of certain nouns (p. 523). After conducting my own study, using only CHILDES for data and no other source, the method used by Goodman et al. seems to be a better solution. The data contained in the CHILDES database tends to be skewed toward specific age brackets, and specific speakers. This appears to be because of the number of studies in the database that include data from certain

ages and speakers. Future study might be aided by only using the specific age brackets and speakers that have the most data so as to avoid the data being weighted to those groups.

References

Ackermann, L., Hepach, R., & Mani, N. (2020). Children learn words easier when they are interested in the category to which the word belongs. *Developmental Science, 23(3), e12915–n/a.* , https://doi.org/10.1111/desc.12915.

Albers, T. (2013). Peer interaction and language acquisition in German kindergartens. *Journal of Educational and Social Research, 3(3)*, 11-24.

Behrens, H. (2008). *Corpora in language acquisition research: History, methods, persepctives.* Amsterdam: John Benjamin Publishing Company.

Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics, 47(2)*, 383–411. https://doi.org/10.1515/LING.2009.014.

Blum-Kulka, S. & Snow, C.E. (2004). Introduction: the potential of peer talk. *Discourse Studies, 6(3)*, 291-306. https://doi.org/10.1177/1461445604044290.

Bunce, J. P. & Scott, R. M. (2017). Finding meaning in a noisy world: exploring the effects of referential ambiguity and competition on 2.5 year-olds' cross-situational word learning. *Journal of Child Language, 44*, 650-676. https://doi.org/10.1017/S0305000916000180.

Bybee, J. (2010). *Language, Usage and Cognition.* Cambridge University Press. https://doi.org/10.1017/CBO9780511750526.

Cartmill, E.A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S. & Medinac, T. N. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences - PNAS, 110(28)*, 11278–11283. https://doi.org/10.1073/pnas.1309518110.

Crain, S. & Thornton, R. (2012). Syntax acquisition. *Wiley Interdisciplinary Reviews. Cognitive Science, 3(2)*, 185-203. https://doi.org/10.1002/wcs.1158.

Diessel, H. & Coventry, K. R. (2020). Demonstratives in spatial language and social interaction: an interdisciplinary review. *Frontiers in Psychology, 11,* , 555265–555265, https://doi.org/10.3389/fpsyg.2020.555265.

Farver, J. (1992). Communicating shared meaning in social pretend play. *Early Childhood Research Quarterly, 7(4)*, 501-516. https://doi.org/10.1016/0885-2006(92)90047-3.

Fassler, R. (1998). Room for talk: Peer support for getting into english in an ESL kindergarten. *Early Childhood Research Quarterly, 13(3)*, 379-409. https://doi.org/10.1016/S0885-2006(99)80047-7.

Fekonja, U.; Umek, L.; & Kranjc, S. (2005). Free Play and other daily preschool activities as a context for child's language development. *Studia Psychologica, 47(2)*, 103-117.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E.,Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B. & Stiles, J. (1994). Variability in early communicative development. *Monographs fo the Society for Research in Child Development, Vol. 59, No. 5*, i-185. https://doi.org/10.2307/1166093.

Ferguson, B. & Waxman, S. (2017). Linking language and categorization in infancy. *Journal of Child Language, 44*, 527-552.

Fourtassi, A., Bian, Y. & Frank, M.C. (2020). The Growth of Children's Semantic and Phonological Networks: Insight From 10 Languages. *Cognitive Science, 44(7)*, e12847–n/a. https://doi.org/10.1111/cogs.12847.

Fromkin, V., Rodman, R., & Hyams, N. . (2014). *An introduction to language.* New York: Cengage Learning.

Goodman, J.C., Dale, P.S. & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language, 35*, 515-531. https://doi.org/10.1017/S0305000907008641.

Gries, S. T. (2009). *Quantitative corpus linguistics with R: a practical introduction.* New York: Taylor & Francis.

He, A. X., & Arunachalam, S. (2017). Word Learning Mechanisms. *Wiley Interdiscip Rev Cogn Sci. 8(4)*, e1435–n/a. https://doi.org/10.1002/wcs.1435.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26(1)*, 55-88.

Jones, I. (2002). Social Relationships, Peer Collaboration and Children's Oral Language. *Educational Psychology, 22(1)*, 63-73.

Justice, L.; Petscher, Y.; Schatschneider, C.; & Mashburn, A. (2011). Peer Effects in Preschool Classrooms: Is Children's Language Growth Associated with Their Classmates' Skills? *Child Development, 82(6)*, 1768-1777. https://doi.org/10.1111/j.1467-8624.2011.01665.

Keenan, E. O. & Klein, E. (1975). Coherenecy in Children's Discourse. *Journal of Psycholinguistic Research, Vol. 4, No. 4*, 365-380. https://doi.org/10.1007/BF01067065.

Laing, C. & Bergelson, E. (2020). From babble to words: Infants' early productions match words and objects in their evironment. *Cognitive Psychology, 122*, 101308–101308. https://doi.org/10.1016/j.cogpsych.2020.101308.

Lidz, J. & Gagliardi, A. (2015). How nature meets nurture: Universal grammar and statistical learning. *Annual Review Linguist, 1(1)*, 333-353.

Lidz, J., Waxman, S., & Freedman, J. (2003). What infants know about syntax but couldn't have learned: experimental evidence for syntactic structure at 18 months. *Cognition, 89(3)*, 295-303. https://doi.org/10.1016/S0010-0277(03)00116-1.

Lieven, E. & Tomasello, M. (2008). Children's First Language Acquisition from a Usage Based Perspective. In P. J. Robinson, *Handbook of cognitive linguistics and second language acquisition* (pp. 168-196). Routledge.

MacWhinney, B. (2000). The CHILDES Project: Tools for analyzing talk. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Mani, N. & Ackermann, L. (2018). Why do children learn the words they do? *Child Development Perspectives, 12(4)*, 253-257. https://doi.org/10.1111/cdep.12295.

McEnery, T. & Wilson, A. (2001). *Corpus Linguistics: An Introduction.* Edinburgh: Edinburgh University Press Ltd.

McEnery, T., & Hardie, A. (2011). *Corpus Linguistics: Method, Theory, and Practice.* Cambridge University Press. https://doi.org/10.1017/CBO9780511981395.

Merriam-Webster. (n.d.). *Corpus in Merriam-Webster.com dictionary*. Retrieved from March 22, 2021 https://www.merriam-webster.com/dictionary/corpus.

Naigles, L. R. (2002). Form is easy, meaning is hard: resolving a paradox in early child language. *Cognition, 86(2)*, 157-199. https://doi.org/10.1016/S0010-0277(02)00177-4.

Najnin, S., & Banerjee, B. (2018). Pragmatically Framed Cross-Situational Noun Learning Using Computational Reinforcement Models. *Frontiers in Psychology, 9*, 5–5. https://doi.org/10.3389/fpsyg.2018.00005.

Pearl, S. (2021). Poverty of the stimulus without tears. *Language Learning and Development,* , 1–40. https://doi.org/10.1080/15475441.2021.1981908.

Pelligrini, A.D., Melhuish, E., Jones, I., Trojanowska, L., & Gilden, R. (2002). Social contexts of learning literate language: The role of varied, familiar, and close peer relationships. *Learning and Individual Differences, 12*, 375-389.https://doi.org/10.1016/S1041-6080(02)00044-4.

Perry, L.K., Prince, E.B., Valtierra, A.M., Rivero-Fernandez, C., Ullery, M.A., Katz, L.F. Laursen, B., & Messinger, D.S. (2018). A year in words: The dynamics and consequences of language experiences in an intervention classroom. *PLoS ONE, 13(7)*, e0199893–e0199893. https://doi.org/10.1371/journal.pone.0199893.

Quine, W. V. O., Churchland, P. S., & Føllesdal, D. (2013). *Word and Object.* Cambridge: MIT press.

Russell, A., Pettit, G. S., & Mize, J. (1998). Horizontal qualities in parent-child relationships: parallels with and possible consequences for children's peer relationships. *Developmental Review, 18*, 313-352. https://doi.org/10.1006/drev.1997.0466.

Santiago-Delefosse, M. J. & Oderic Delefosse, J. M. (2002). Spielrein, Piaget and Vygotsky: Three Positions on Child Thought and Language. *Theory & Psychology, Vol. 12(6)*, 723-747. https://doi.org/10.1177/0959354302126001.

Silge, J. & Robinson, D. (2017). *Text mining with R: A tidy approach.* O'Reilly Media, Inc.

Smith, L & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106(3)*, 1558-1568. https://doi.org/10.1016/j.cognition.2007.06.010.

Taxitari, L., Twomey, K. E., Westermann, G., & Mani, N. (2020). The limits of infants' early word learning. *Language Learning and Development, Vol. 16, No. 1*, 1-21.

Tomasello, M. & Olguin, R. (1993). Twenty-Three-Month-Old Children Have a Gategorical Category of Noun. *Cognitive Development, 8*, 451-464. https://doi.org/10.1016/S0885-2014(05)80004-8.

Tomasello, M., Akhtar, N., Dodson, K., & Rekau, L. (1997). Differential productivity in young children's use of nouns and verbs. *Journal of Child Language, 24*, 373-387. https://doi.org/10.1017/S0305000997003085.

Trueswell, J.C., Lin, Y., Armstrong, B., Cartmill, E.A., Goldin-Meadow, S. & Gleitman, L.R. (2014). Perceiving referential intent: Dynamics of reference in natural parent-child interactions. *Cognition, 148*, 117-135. https://doi.org/10.1016/j.cognition.2015.11.002.

Twomey, K. E., & Westermann, G. (2018). Learned labels shape pre-speech infants' object representations. *Infancy, 23(1)*, 61-73. https://doi.org/10.1111/infa.12201.

Waxman, S. R. & Lidz, J.L. (2007). Early word learning. In *Handbook of Child Psychology* (pp. 299-333).

Wiig, E. H., Semel, E. & Secord, W. A. (2021, 10 26). *CELF-5 Test objectives and descriptions.* Retrieved from Pearson Assessments: pearsonassessments,com/content/dam/school/global/clinical/us/assets/celf-5/celf-5-test-objectives-and-descriptions.pdf

Yu, C. & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition. 125(2)*, 244-262.https://doi.org/10.1016/j.cognition.2012.06.016.

Yu, C. & Smith, L.B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology, 26*, 1235-1240. https://doi.org/10.1016/j.cub.2016.03.026.

Appendix A

1: CDI Infant Words

Child's Name _____ Sex _____

Birthdate _____ Today's Date _____

The MacArthur Communicative Development Inventory: Words and Gestures

Proper Mark ● Improper Marks ⊗ ◔ ◒ ⊙

PART I EARLY WORDS

A. FIRST SIGNS OF UNDERSTANDING

Before children begin to speak, they show signs of understanding language by responding to familiar words and phrases. Below are some common examples. Does your child do any of these?

	Yes	No
1. Respond when name is called. (e.g., by turning and looking at source)	○	○
2. Respond to "no no" (by stopping what he/she is doing, at least for a moment).	○	○
3. React to "there's mommy/daddy" by looking around for them.	○	○

B. PHRASES (28)

In the list below, please mark the phrases that your child seems to understand.

	understands		understands		understands
Are you hungry?	○	Don't touch.	○	Open your mouth.	○
Are you tired/sleepy?	○	Get up.	○	Sit down.	○
Be careful.	○	Give it to mommy.	○	Spit it out.	○
Be quiet.	○	Give me a hug.	○	Stop it.	○
Clap your hands.	○	Give me a kiss.	○	Time to go night night.	○
Change diaper.	○	Go get ___	○	Throw the ball.	○
Come here/come on.	○	Good girl/boy.	○	This little piggy.	○
Daddy's/mommy's home.	○	Hold still.	○	Want to go for a ride?	○
Do you want more?	○	Let's go bye bye.	○		
Don't do that.	○	Look/look here.	○		

C. STARTING TO TALK

	Never	Sometimes	Often
1. Some children like to "parrot" or imitate things that they've just heard (including new words that they are just learning, and/or parts of sentences, for example, repeating "work now" after mother says "Mommy's going to work now.") How often does your child imitate words?	○	○	○
2. Some children like to go around naming or labeling things, as though proud of knowing the names and wanting to show this. How often does your child do this?	○	○	○

D. VOCABULARY CHECKLIST

The following is a list of typical words in young children's vocabularies. For words your child understands but does not yet say, place a mark in the first column (understands). For words that your child not only understands but also uses, place a mark in the second column (understands and says). If your child uses a different pronunciation of a word (for example, "raffe" for "giraffe" or "sketti" for "spaghetti") mark the word anyway. Remember, this is a "catalogue" of words that are used by many different children. Don't worry if your child knows only a few right now.

1. SOUND EFFECTS AND ANIMAL SOUNDS (12)

	understands	understands and says		understands	understands and says		understands	understands and says
baa baa	○	○	meow	○	○	uh oh	○	○
choo choo	○	○	moo	○	○	vroom	○	○
cockadoodledoo	○	○	ouch	○	○	woof woof	○	○
grrr	○	○	quack quack	○	○	yum yum	○	○

2. ANIMALS NAMES (Real or Toy) (36)

	understands	understands and says		understands	understands and says		understands	understands and says
animal	○	○	duck	○	○	penguin	○	○
bear	○	○	elephant	○	○	pig	○	○
bee	○	○	fish	○	○	pony	○	○
bird	○	○	frog	○	○	puppy	○	○
bug	○	○	giraffe	○	○	sheep	○	○
bunny	○	○	goose	○	○	squirrel	○	○
butterfly	○	○	horse	○	○	teddy bear	○	○
cat	○	○	kitty	○	○	tiger	○	○
chicken	○	○	lamb	○	○	turkey	○	○
cow	○	○	lion	○	○	turtle	○	○
deer	○	○	monkey	○	○			
dog	○	○	mouse	○	○			
donkey	○	○	owl	○	○			

3. VEHICLES (Real or Toy) (9)

	understands	understands and says		understands	understands and says		understands	understands and says
airplane	○	○	car	○	○	stroller	○	○
bicycle	○	○	firetruck	○	○	train	○	○
bus	○	○	motorcycle	○	○	truck	○	○

4. TOYS (8)

	understands	understands and says		understands	understands and says		understands	understands and says
ball	○	○	book	○	○	pen	○	○
balloon	○	○	bubbles	○	○	toy	○	○
block	○	○	doll	○	○			

5. FOOD AND DRINK (30)

	understands	understands and says		understands	understands and says		understands	understands and says
apple	○	○	chicken	○	○	meat	○	○
banana	○	○	coffee	○	○	milk	○	○
bread	○	○	cookie	○	○	noodles	○	○
butter	○	○	cracker	○	○	orange	○	○
cake	○	○	drink	○	○	peas	○	○
candy	○	○	egg	○	○	pizza	○	○
carrots	○	○	fish	○	○	raisin	○	○
cereal	○	○	food	○	○	spaghetti	○	○
cheerios	○	○	ice cream	○	○	toast	○	○
cheese	○	○	juice	○	○	water	○	○

6. CLOTHING (19)

	understands	understands and says		understands	understands and says		understands	understands and says
beads	○	○	hat	○	○	shoe	○	○
bib	○	○	jacket	○	○	shorts	○	○
boots	○	○	jeans	○	○	sock	○	○
button	○	○	necklace	○	○	sweater	○	○
coat	○	○	pajamas	○	○	zipper	○	○
diaper	○	○	pants	○	○			
dress	○	○	shirt	○	○			

7. BODY PARTS (20)

	understands	understands and says		understands	understands and says		understands	understands and says
arm	○	○	finger	○	○	nose	○	○
belly button	○	○	hair	○	○	owie/boo boo	○	○
cheek	○	○	hand	○	○	tooth	○	○
ear	○	○	head	○	○	toe	○	○
eye	○	○	knee	○	○	tongue	○	○
face	○	○	leg	○	○	tummy	○	○
foot	○	○	mouth	○	○			

8. FURNITURE AND ROOMS (24)

	understands	understands and says		understands	understands and says		understands	understands and says
bathroom	○	○	drawer	○	○	refrigerator	○	○
bathtub	○	○	garage	○	○	rocking chair	○	○
bed	○	○	high chair	○	○	sink	○	○
bedroom	○	○	kitchen	○	○	stairs	○	○
chair	○	○	living room	○	○	stove	○	○
couch	○	○	oven	○	○	table	○	○
crib	○	○	play pen	○	○	TV	○	○
door	○	○	potty	○	○	window	○	○

9. SMALL HOUSEHOLD ITEMS (36)

	understands	understands and says		understands	understands and says		understands	understands and says
blanket	○	○	glasses	○	○	plate	○	○
bottle	○	○	hammer	○	○	purse	○	○
bowl	○	○	keys	○	○	radio	○	○
box	○	○	lamp	○	○	scissors	○	○
broom	○	○	light	○	○	soap	○	○
brush	○	○	medicine	○	○	spoon	○	○
clock	○	○	money	○	○	telephone	○	○
comb	○	○	paper	○	○	toothbrush	○	○
cup	○	○	penny	○	○	towel	○	○
dish	○	○	picture	○	○	trash	○	○
fork	○	○	pillow	○	○	vacuum	○	○
glass	○	○	plant	○	○	watch	○	○

10. OUTSIDE THINGS AND PLACES TO GO (27)

	understands	understands and says		understands	understands and says		understands	understands and says
backyard	○	○	park	○	○	snow	○	○
beach	○	○	party	○	○	star	○	○
church *	○	○	pool	○	○	store	○	○
flower	○	○	rain	○	○	sun	○	○
garden	○	○	rock	○	○	swing	○	○
home	○	○	school	○	○	tree	○	○
house	○	○	shovel	○	○	water	○	○
moon	○	○	sky	○	○	work	○	○
outside	○	○	slide	○	○	zoo	○	○

* or word used in your family

11. PEOPLE (20)

	understands	understands and says		understands	understands and says		understands	understands and says
aunt	○	○	grandma *	○	○	teacher	○	○
baby	○	○	grandpa *	○	○	uncle	○	○
babysitter	○	○	lady	○	○			
babysitter's name	○	○	man	○	○			
boy	○	○	mommy *	○	○			
brother	○	○	child's own name	○	○			
child	○	○	people	○	○			
daddy *	○	○	person	○	○			
girl	○	○	sister	○	○			

* or word used in your family

12. GAMES AND ROUTINES (19)

	understands	understands and says		understands	understands and says		understands	understands and says
bath	○	○	night night	○	○	yes	○	○
breakfast	○	○	no	○	○			
bye or bye bye	○	○	patty cake	○	○			
dinner	○	○	peekaboo	○	○			
don't	○	○	please	○	○			
hello	○	○	shh/shush/hush	○	○			
hi	○	○	thank you	○	○			
lunch	○	○	wait	○	○			
nap	○	○	wanna/want to	○	○			

13. ACTION WORDS (55)

	understands	understands and says		understands	understands and says		understands	understands and says
bite	○	○	help	○	○	show	○	○
blow	○	○	hit	○	○	sing	○	○
break	○	○	hug	○	○	sleep	○	○
bring	○	○	hurry	○	○	smile	○	○
bump	○	○	jump	○	○	splash	○	○
clean	○	○	kick	○	○	stop	○	○
close	○	○	kiss	○	○	swim	○	○
cry	○	○	look	○	○	swing	○	○
dance	○	○	love	○	○	take	○	○
draw	○	○	open	○	○	throw	○	○
drink	○	○	play	○	○	tickle	○	○
drive	○	○	pull	○	○	touch	○	○
eat	○	○	push	○	○	watch	○	○
fall	○	○	put	○	○	walk	○	○
feed	○	○	read	○	○	wash	○	○
finish	○	○	ride	○	○	wipe	○	○
get	○	○	run	○	○	write	○	○
give	○	○	say	○	○			
go	○	○	see	○	○			

14. WORDS ABOUT TIME (8)

	understands	understands and says		understands	understands and says		understands	understands and says
day	○	○	night	○	○	tomorrow	○	○
later	○	○	now	○	○	tonight	○	○
morning	○	○	today	○	○			

15. DESCRIPTIVE WORDS (37)

	understands	understands and says		understands	understands and says		understands	understands and says
all gone	○	○	empty	○	○	old	○	○
asleep	○	○	fast	○	○	pretty	○	○
bad	○	○	fine	○	○	red	○	○
big	○	○	gentle	○	○	scared	○	○
blue	○	○	good	○	○	sick	○	○
broken	○	○	happy	○	○	sleepy	○	○
careful	○	○	hard	○	○	soft	○	○
clean	○	○	hot	○	○	thirsty	○	○
cold	○	○	hungry	○	○	tired	○	○
cute	○	○	hurt	○	○	wet	○	○
dark	○	○	little	○	○	yucky	○	○
dirty	○	○	naughty	○	○			
dry	○	○	nice	○	○			

16. PRONOUNS (11)

	understands	understands and says		understands	understands and says		understands	understands and says
her	○	○	me	○	○	this	○	○
his	○	○	mine	○	○	you	○	○
I	○	○	my	○	○	your	○	○
it	○	○	that	○	○			

17. QUESTION WORDS (6)

	understands	understands and says		understands	understands and says		understands	understands and says
how	○	○	when	○	○	who	○	○
what	○	○	where	○	○	why	○	○

18. PREPOSITIONS AND LOCATIONS (11)

	understands	understands and says		understands	understands and says		understands	understands and says
away	○	○	inside	○	○	there	○	○
back	○	○	off	○	○	under	○	○
down	○	○	on	○	○	up	○	○
in	○	○	out	○	○			

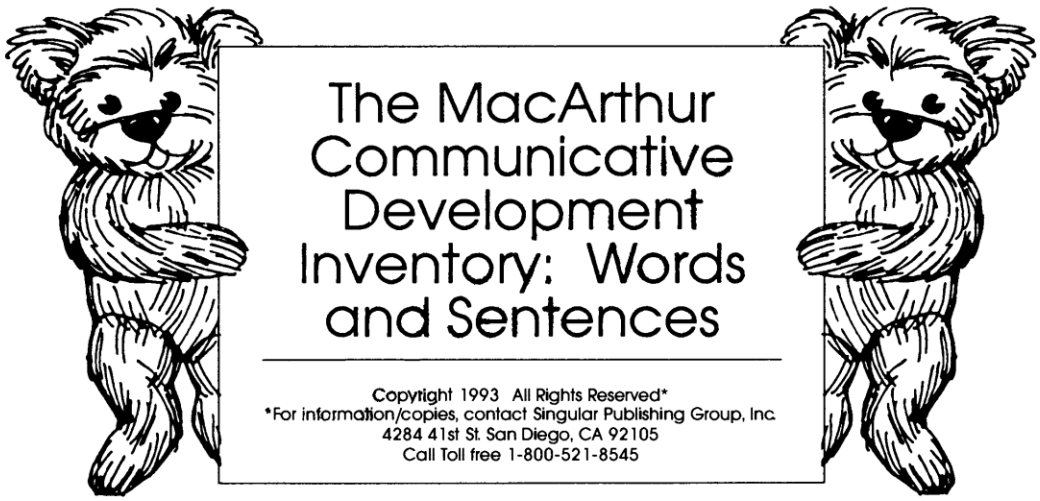
19. QUANTIFIERS (8)

	understands	understands and says		understands	understands and says		understands	understands and says
all	○	○	none	○	○	same	○	○
another	○	○	not	○	○	some	○	○
more	○	○	other	○	○			

2: CDI Toddler Words

Child's Name _____ Sex _____

Birthdate _____ Today's Date _____



The MacArthur Communicative Development Inventory: Words and Sentences

Proper Mark ● USE NO. 2 PENCIL ONLY Improper Marks ⊗ ◔ ⊙ ◒

PART 1 WORDS CHILDREN USE

A. VOCABULARY CHECKLIST

Children understand many more words than they say. We are particularly interested in the words your child SAYS. Please go through the list and mark the words you have heard your child use. If your child uses a different pronunciation of a word (for example, "raffe" instead of "giraffe" or "sketti" for "spaghetti"), mark the word anyway. Remember that this is a "catalogue" of all the words that are used by many different children. Don't worry if your child knows only a few of these right now.

1. SOUND EFFECTS AND ANIMAL SOUNDS (12)

baa baa ○	meow ○	uh oh ○
choo choo ○	moo ○	vroom ○
cockadoodledoo ○	ouch ○	woof woof ○
grr ○	quack quack ○	yum yum ○

2. ANIMALS (Real or Toy) (43)

alligator ○	duck ○	penguin ○
animal ○	elephant ○	pig ○
ant ○	fish ○	pony ○
bear ○	frog ○	puppy ○
bee ○	giraffe ○	rooster ○
bird ○	goose ○	sheep ○
bug ○	hen ○	squirrel ○
bunny ○	horse ○	teddybear ○
butterfly ○	kitty ○	tiger ○
cat ○	lamb ○	turkey ○
chicken ○	lion ○	turtle ○
cow ○	monkey ○	wolf ○
deer ○	moose ○	zebra ○
dog ○	mouse ○	
donkey ○	owl ○	

3. VEHICLES (Real or Toy) (14)

airplane	firetruck	tractor
bicycle	helicopter	train
boat	motorcycle	tricycle
bus	sled	truck
car	stroller	

4. TOYS (18)

ball	chalk	pencil
balloon	crayon	play dough
bat	doll	present
block	game	puzzle
book	glue	story
bubbles	pen	toy

5. FOOD AND DRINK (68)

apple	food	pizza
applesauce	french fries	popcorn
banana	grapes	popsicle
beans	green beans	potato
bread	gum	potato chip
butter	hamburger	pretzel
cake	ice	pudding
candy	ice cream	pumpkin
carrots	jello	raisin
cereal	jelly	salt
cheerios	juice	sandwich
cheese	lollipop	sauce
chicken	meat	soda/pop
chocolate	melon	soup
coffee	milk	spaghetti
coke	muffin	strawberry
cookie	noodles	toast
corn	nuts	tuna
cracker	orange	vanilla
donut	pancake	vitamins
drink	peanut butter	water
egg	peas	yogurt
fish	pickle	

6. CLOTHING (28)

beads	jacket	slipper
belt	jeans	sneaker
bib	mittens	snowsuit
boots	necklace	sock
button	pajamas	sweater
coat	pants	tights
diaper	scarf	underpants
dress	shirt	zipper
gloves	shoe	
hat	shorts	

7. BODY PARTS (27)

ankle ○	feet ○	nose ○
arm ○	finger ○	owie/boo boo ○
belly button ○	hair ○	penis* ○
buttocks/bottom* ○	hand ○	shoulder ○
cheek ○	head ○	tooth ○
chin ○	knee ○	toe ○
ear ○	leg ○	tongue ○
eye ○	lips ○	tummy ○
face ○	mouth ○	vagina* ○

* or word used in your family

8. SMALL HOUSEHOLD ITEMS (50)

basket ○	glassess ○	plate ○
blanket ○	hammer ○	purse ○
bottle ○	jar ○	radio ○
box ○	keys ○	scissors ○
bowl ○	knife ○	soap ○
broom ○	lamp ○	spoon ○
brush ○	light ○	tape ○
bucket ○	medicine ○	telephone ○
camera ○	money ○	tissue/kleenex ○
can ○	mop ○	toothbrush ○
clock ○	nail ○	towel ○
comb ○	napkin ○	trash ○
cup ○	paper ○	tray ○
dish ○	penny ○	vacuum ○
fork ○	picture ○	walker ○
garbage ○	pillow ○	watch ○
glass ○	plant ○	

9. FURNITURE AND ROOMS (33)

basement ○	drawer ○	rocking chair ○
bathroom ○	dryer ○	room ○
bathtub ○	garage ○	shower ○
bed ○	high chair ○	sink ○
bedroom ○	kitchen ○	sofa ○
bench ○	living room ○	stairs ○
chair ○	oven ○	stove ○
closet ○	play pen ○	table ○
couch ○	porch ○	TV ○
crib ○	potty ○	washing machine ○
door ○	refrigerator ○	window ○

10. OUTSIDE THINGS (31)

backyard ○	rain ○	star ○
cloud ○	rock ○	stick ○
flag ○	roof ○	stone ○
flower ○	sandbox ○	street ○
garden ○	shovel ○	sun ○
grass ○	sidewalk ○	swing ○
hose ○	sky ○	tree ○
ladder ○	slide ○	water ○
lawn mower ○	snow ○	wind ○
moon ○	snowman ○	
pool ○	sprinkler ○	

11. PLACES TO GO (22)

beach ○	home ○	school ○
camping ○	house ○	store ○
church* ○	movie ○	woods ○
circus ○	outside ○	work ○
country ○	park ○	yard ○
downtown ○	party ○	zoo ○
farm ○	picnic ○	
gas station ○	playground ○	

*or word used in your family

12. PEOPLE (29)

aunt ○	doctor ○	nurse ○
baby ○	fireman ○	child's own name ○
babysitter ○	friend ○	people ○
babysitter's name ○	girl ○	person ○
boy ○	grandma* ○	pet's name ○
brother ○	grandpa* ○	police ○
child ○	lady ○	sister ○
clown ○	mailman ○	teacher ○
cowboy ○	man ○	uncle ○
daddy* ○	mommy* ○	

*or word used in your family

13. GAMES AND ROUTINES (25)

bath ○	hello ○	shopping ○
breakfast ○	lunch ○	snack ○
bye ○	nap ○	so big! ○
call (on phone) ○	night night ○	thank you ○
dinner ○	no ○	this little piggy ○
give me five! ○	patty cake ○	turn around ○
gonna get you! ○	peekaboo ○	yes ○
go potty ○	please ○	
hi ○	shh/shush/hush ○	

14. ACTION WORDS (103)

bite ○	drive ○	hug ○	read ○	swim ○
blow ○	drop ○	hurry ○	ride ○	swing ○
break ○	dry ○	jump ○	rip ○	take ○
bring ○	dump ○	kick ○	run ○	talk ○
build ○	eat ○	kiss ○	say ○	taste ○
bump ○	fall ○	knock ○	see ○	tear ○
buy ○	feed ○	lick ○	shake ○	think ○
carry ○	find ○	like ○	share ○	throw ○
catch ○	finish ○	listen ○	show ○	tickle ○
chase ○	fit ○	look ○	sing ○	touch ○
clap ○	fix ○	love ○	sit ○	wait ○
clean ○	get ○	make ○	skate ○	wake ○
climb ○	give ○	open ○	sleep ○	walk ○
close ○	go ○	paint ○	slide ○	wash ○
cook ○	hate ○	pick ○	smile ○	watch ○
cover ○	have ○	play ○	spill ○	wipe ○
cry ○	hear ○	pour ○	splash ○	wish ○
cut ○	help ○	pretend ○	stand ○	work ○
dance ○	hide ○	pull ○	stay ○	write ○
draw ○	hit ○	push ○	stop ○	
drink ○	hold ○	put ○	sweep ○	

15. DESCRIPTIVE WORDS (63)					
allgone	○	full	○	orange	○
asleep	○	gentle	○	poor	○
awake	○	good	○	pretty	○
bad	○	green	○	quiet	○
better	○	happy	○	red	○
big	○	hard	○	sad	○
black	○	heavy	○	scared	○
blue	○	high	○	sick	○
broken	○	hot	○	sleepy	○
brown	○	hungry	○	slow	○
careful	○	hurt	○	soft	○
clean	○	last	○	sticky	○
cold	○	little	○	stuck	○
cute	○	long	○	thirsty	○
dark	○	loud	○	tiny	○
dirty	○	mad	○	tired	○
dry	○	naughty	○	wet	○
empty	○	new	○	white	○
fast	○	nice	○	windy	○
fine	○	noisy	○	yellow	○
first	○	old	○	yucky	○

16. WORDS ABOUT TIME (12)					
after	○	morning	○	today	○
before	○	night	○	tomorrow	○
day	○	now	○	tonight	○
later	○	time	○	yesterday	○

17. PRONOUNS (25)							
he	○	me	○	their	○	we	○
her	○	mine	○	them	○	you	○
hers	○	my	○	these	○	your	○
him	○	myself	○	they	○	yourself	○
his	○	our	○	this	○		
I	○	she	○	those	○		
it	○	that	○	us	○		

18. QUESTION WORDS (7)							
how	○	when	○	which	○	why	○
what	○	where	○	who	○		

19. PREPOSITIONS AND LOCATIONS (26)					
about	○	down	○	on top of	○
above	○	for	○	out	○
around	○	here	○	over	○
at	○	inside/in	○	there	○
away	○	into	○	to	○
back	○	next to	○	under	○
behind	○	of	○	up	○
beside	○	off	○	with	○
by	○	on	○		

20. QUANTIFIERS AND ARTICLES (17)					
a	○	each	○	other	○
all	○	every	○	same	○
a lot	○	more	○	some	○
an	○	much	○	the	○
another	○	not	○	too	○
any	○	none	○		

Appendix B

Pearson Vocabulary

Word Classes

Start	Reversal Rule	Repetitions	Discontinue Rule
All ages: Item 1	None	Allowed	Four consecutive 0 scores

Correct responses are in bold. Circle 1 if the student selects both correct words, and 0 if the student selects one or more incorrect words. If necessary, precede each item with **Listen,** to focus the student's attention on the new word series.

Demo	**puppy**	frog	**dog**		
Trial 1	milk	**apple**	**banana**		
Trial 2	**cat**	**whiskers**	nest	SCORE	
1. **cat**	cow	**kitten**		1	0
2. **marker**	**pencil**	strawberry		1	0
3. **foot**	**hand**	belt		1	0
4. stroller	**helicopter**	**plane**		1	0
5. eyes	**socks**	**shoes**		1	0
6. **candle**	apple	**flashlight**		1	0
7. **slide**	**swing**	snowman		1	0
8. alligator	**caterpillar**	**butterfly**		1	0

Say, **Now you will have four words to choose from.** (Turn the Stimulus Book page and say,) **Tell me the two that go together best.**

9. **comb**	soap	**brush**	tissue	1	0
10. pink	**circle**	**square**	7	1	0
11. key	razor	**hammer**	**nail**	1	0
12. **C**	3	**M**	check	1	0

Say, **Now we are going to do some that don't have pictures. Listen and tell me the two words that go together best.** Proceed to Item 13.

13. **running**	**jumping**	eating	hearing	1	0
14. **books**	porch	**library**	cave	1	0
15. bed	**garage**	pool	**car**	1	0
16. **silent**	**quiet**	gentle	tired	1	0
17. rock	**shirt**	pie	**cloth**	1	0
18. park	**candle**	**wax**	mirror	1	0
19. **minute**	decade	**hour**	winter	1	0
20. **north**	Celsius	globe	**west**	1	0
21. spring	**morning**	weekend	**afternoon**	1	0
22. sniff	**eyes**	**see**	hand	1	0
23. dancing	caring	**hearing**	**smelling**	1	0
24. **smooth**	wise	**rough**	heavy	1	0
25. affirming	appointing	**achieving**	**accomplishing**	1	0
26. **equal**	early	size	**alike**	1	0
27. crooked	**connected**	**joined**	rotated	1	0
28. **quest**	quench	**search**	literal	1	0
29. **longitude**	volume	attitude	**latitude**	1	0
30. **enthusiastic**	envious	effective	**eager**	1	0
31. **permanent**	**temporary**	faulty	stereo	1	0
32. disagree	**persuade**	**urge**	compound	1	0
33. **occupied**	relevant	complicated	**vacant**	1	0
34. adventurous	**prosperous**	reliable	**wealthy**	1	0
35. **biography**	lecture	**memoir**	parchment	1	0
36. intellectual	**essential**	gradual	**crucial**	1	0
37. repossess	recycle	**renovate**	**restore**	1	0
38. informal	apathy	**imitation**	**caricature**	1	0
39. **arid**	**parched**	tricky	serene	1	0
40. **assimilate**	**incorporate**	retaliate	articulate	1	0

Raw Score

Word Classes Item Analysis

Category	Item
Semantic Class	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, **11**, 12, 13, **16**, 19, 20, 21, 23, 35, 38
Location	14, 15
Composition	17, 18
Synonym	**16**, 25, 26, 27, 28, 30, 32, 34, 36, 37, 39, 40
Object Function	**11**, 22
Word Opposites	24, 29, 31, 33

Note. Bold items appear in more than one category.

Brooke Foss

NTNU

Kunnskap for en bedre verden