國立臺灣大學文學院語言學研究所

碩士論文

Graduate Institute of Linguistics
College of Liberal Arts
National Taiwan University

Master Thesis

測量華語兒童早期詞彙成長:以語料庫為本之研究

Measuring Early Vocabulary Growth in Mandarin-Speaking Children: A Corpus-Based Study

楊靜琛

Jing-Chen Yang

指導教授：張顯達 博士

謝舒凱 博士

Advisor: Hintat Cheung, Ph.D.

Shu-Kai Hsieh, Ph.D.

中華民國 104 年 7 月

July, 2015

國立臺灣大學碩士學位論文
口試委員會審定書

測量華語兒童早期詞彙成長:以語料庫為本之研究
Measuring Early Vocabulary Growth in Mandarin-Speaking Children: A Corpus-Based Study

本論文係 楊靜琛 君（學號 R98142001）在國立臺灣大學語言學研究所完成之碩士學位論文，於民國 104 年 7 月 21 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

張顯達 （簽名）

（指導教授）

呂佳蓉

謝舒凱

誌謝

這是一篇碩論，但是對在人生的旅途中迷路的我而言，這一路走來相當漫長，途中甚至因為迷惘、猶豫不決而無法堅持下去決定暫時放下這篇論文。雖然現在一樣在迷惘中，但是為了讓自己不要在未來覺得遺憾，我再次打開論文的 word 檔，一步一步將其完成。

這過程中要感謝許多人。首先要感謝兩位指導老師，張顯達老師和謝舒凱老師，不斷的給我鼓勵、以及論文的方向和各種建議。感謝他們能在百忙之中抽空指導我這篇難產的論文。也要感謝呂佳蓉老師在大家都很忙碌的暑假期間來擔任我的口試委員，讓我不至於找不到口試委員無法完成口試。感謝指導老師們與口試委員在口試時給予的各種建議，讓這篇論文能稍稍再美好一點。

另外還要感謝前所長宋麗梅老師與萬能的美玲助教，謝謝她們在我茫然的米蟲階段時，正好提供我工讀的機會，讓我不用為生活費而煩惱，因為這樣一個新的契機讓我有了重新接觸人群並調整生活步調的機會，甚至因此覺得，不如就把論文寫完吧！

還要感謝我所有的老朋友、新朋友、語言所認識的大家、以前大學的好朋友們。因為太多人，怕有遺漏就不一一列舉，感謝你們在我茫然、焦慮、苦惱、厭煩時，願意聽我碎碎念，願意跟我一起吐槽，願意給我一些建議。有時候或許只是一句話，但是每一句話累積起來的能量，支持著我一路完成這篇論文。甚至要感謝我的新室友林妹妹和她的貓，因為有個相當新鮮的新鮮人讓我意識到至少要完成一件事，給自己一個交代，而她的貓的存在本身就是一種正能量，讓我能有好心情。

也要感謝我的父母。或許他們看不慣我的決定，也許不支持，但是至少也不反對，任由我自行蹉跎人生，慢慢摸索人生方向。甚至在我經濟拮据時，仍願意支援我這個還不能賺錢補貼家用的女兒。

再次感謝我生活中遇到的許多人，因為有你們，我才能堅持下來！

中文摘要

本文旨在測量臺灣地區華語兒童早期詞彙成長，以語料庫為語料來源，分析兒童詞彙成長與詞彙組成。詞彙成長的測量項目為總詞彙量、各個詞類之詞頻與其所佔的比例、名詞動詞比率、相異詞比率、以及 D 數值。詞彙組成的測量項目為各個語義範疇和概念層次中名詞的數量。計算結果顯示詞彙在上述測量項目中呈現隨著年齡增加而發展之趨勢。

詞彙量隨年齡增長，各個詞類之詞頻與所佔比例顯示名詞與動詞較其他詞類早習得。名詞動詞所佔比例隨著年齡增加而遞減，其他詞類所佔比例則是遞增。相異詞比率表明所有年齡層的兒童皆習得較多名詞與動詞，同時也較常使用這兩個詞類。量詞和法向詞的相異詞比率遞減，表示這兩個詞類在較大的年齡層中的使用率逐漸增加。名詞動詞比率顯示出在 19-24 個月時有名詞偏向，之後的年齡層中則有動詞偏向。D 數值的增長呈現兒童逐漸增加的詞彙多樣性。

名詞的語義範疇經分析後得知最具體、最親近兒童生活的名詞最早習得，且最常使用，如人物名稱、工具類、動物類、食物類和交通工具類。較抽象、離兒童生活較疏遠的名詞，比較晚習得且使用率較低，如數字、形狀、顏色、以及自然現象類。名詞的概念層次經分析後可知兒童最先習得基本層次詞彙，其後才是上位層次詞與下位層次詞。而習得上位層次詞與下位層次詞的時間點則因語義範疇而不同。基本層次詞的使用率較上位層次詞與下位層次詞的使用率高。

在這些測量兒童詞彙成長的計算方法中，其中幾個測量項目或許可以在後續深入研究後，成為測量詞彙成長之指標。

關鍵詞：詞彙習得、名詞偏向、基本層次範疇、華語習得、詞彙多樣性、語料庫

English Abstract

This study aims to measure early vocabulary growth in Mandarin-speaking children from Taiwan with a corpus-based method. Children's vocabulary growth was examined in aspects of vocabulary growth and vocabulary organization. Vocabulary growth was measured by computing vocabulary size, frequency and proportion of parts-of-speech, noun-verb ratio, type-token ratio, and D measure. Vocabulary organization was measured by computing the number of nouns in semantic categories and conceptual levels. Results of the measures showed a developmental trend with increasing ages.

Vocabulary size increases with age. Frequency and proportion of lexical categories suggest that nouns and verbs were acquired before other word classes. The proportion of nouns and verbs decreases with increasing age, while that of other lexical categories increases. Type-token ratio implies that children acquired more nouns and verbs and also used them very often in all stages. The declining TTR of classifiers and modals indicate that they were used more frequently in later stages. Noun-verb ratio reveals that a weak noun bias was found in 19-24 months, and a verb bias was found in later stages. D values suggest that children's lexicon became increasingly diverse.

Analysis of semantic categories reveals that the most concrete nouns and nouns which were the closest to children's life were acquired earlier, such as people, tools, animals, and food, and vehicles. Nouns in these categories were also used frequently. Nouns which were abstract and far away from children's life were acquired later and used with a lower frequency, such as numerals, shapes, colors, and nouns of natural phenomena. Analysis of conceptual levels has shown that basic-level words were acquired first, followed by subordinate words and superordinate words. The timing of acquiring superordinate level and subordinate level words varies in different semantic categories. Basic-level words were used with a higher frequency than the other two levels.

All of the computations about vocabulary growth and vocabulary organization in this study have revealed general developmental trends of children's early Mandarin vocabulary. Some of the measured items may become an index of measuring of vocabulary growth after further studies.

Key words: vocabulary acquisition, noun bias, basic-level category, Mandarin acquisition, lexical diversity, corpus

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1. Background

Language acquisition is the process that people not only acquire the ability to perceive and comprehend language, but also learn to produce and use language to communicate. The ability of successfully using language to communicate needs knowledge of phonology, morphology, syntax, semantics, and a large amount of vocabulary. The process of acquiring language and knowing how to use language has received a lot of attention in linguistics research as well as in the field of children's development. There have been studies about how children produce sounds, how they learn vocabulary, how they construct sentences, and how they understand adult's language. Among all topics of language acquisition, vocabulary growth is a major milestone to be charted in children's early language development. It can serve as a predictor of children's performances in later language skills (Hao, Shu, Xing, & Li, 2008). Hence, many researchers have investigated the acquisition of vocabulary from various aspects including lexical categories, semantic categories, and basic-level effect in early vocabulary.

Studies about the development of lexical categories in children's language have revealed that nouns and verbs are acquired before other word classes. Some studies further found a universal noun bias in early vocabulary acquisition (Gentner, 1982; Haryu et al., 2005; Hsu, 1996). Noun bias means that nouns are acquired first and the predominant words in early vocabulary. On the contrary, some researchers argue that noun bias is not universal and for Mandarin early vocabulary growth, a verb bias is found instead (Choi & Gopnik, 1995; Dhillon, 2010; Tardif, Gelman, & Xu, 1999;

Tardif, 1996). The issue on noun-or-verb bias in early vocabulary is still not settled.

Studies about semantic categories of nouns have shown a different development tendency. Children produced device nouns frequently, due to the frequent inputs in children's daily life (Peng & Chong, 2010). Sheng et al. (Sheng, Deng, Zhang, Liang, & Lu, 2012) investigated vocabulary using MCDI, and they found that children acquired more onomatopoeia words than words in other categories (i.e. nouns, verbs, adjectives/adverbs, conventional game words, numeral words, quantifiers, interrogatives, pronouns, directional words, and time words). Kuo, Tsay, and Peng (2005) and Jiang (2000) have found that objects which children contact most often were usually most frequently produced words in children's language.

In addition, categorization is an important cognitive process in which people group different objects together and form a class. A hierarchy of levels is found during the categorization process, and there is a common level of categorization, which is called basic-level category (Rosch, 1978). The theory of basic-level category was applied to the categorization of lexicon. Studies about basic-level effect on children's early vocabulary have concluded that basic-level words were acquired earlier and used more frequently in our speech than superordinate words and subordinate words (Jiang, 2000; Zeng & Zou, 2012; Li, 2014).

Previous studies in vocabulary development in Chinese investigated a specific issue or a particular theoretical aspect of vocabulary acquisition by using language samples they collected. A comprehensive study of Chinese children's vocabulary development is yet to be found. Moreover, there has been no corpus-based report on early vocabulary growth in Mandarin-speaking children from Taiwan. Therefore, this study aims to use corpus data to examine the findings in previous researches, and to provide a complete scope of early vocabulary development of Mandarin-speaking

children before 4 years old in Taiwan.

1.2. Purpose of the study

The purpose of this study is to examine the development of children's early Mandarin lexicon with a corpus-based method, and then to find out reliable measurement items of young children's early vocabulary. In this study, children's vocabulary development will be examined in two aspects. One is vocabulary growth which is to examine the amount of vocabulary. The other one is vocabulary organization which is to examine the contents of children's early vocabulary. This study tries to provide evidence to support or oppose the findings in past studies of children's early vocabulary acquisition. Thus, the two research questions are:

(1) What are the major changes in early vocabulary development of Mandarin-speaking children?

(2) How can early vocabulary development of Mandarin-speaking children be indexed?

The following analyses and discussions are supposed to provide evidence for or against the findings in previous studies, as well as to find out one or more items which can be used to measure children's early Mandarin vocabulary. The results would reveal an estimated performance of Mandarin-speaking children's early vocabulary acquisition. A complete scope of early vocabulary development of Mandarin-speaking children before 4 years old is expected. It may provide information for other studies of the same scale and offer a direction for future studies in children language acquisition.

1.3. Organization of this study

The thesis consists of five chapters. Chapter 1 provides the background of this study, and it introduces the purpose of the study and the research questions. Chapter 2 is literature review, introducing methods of observing children language development in past studies and previous findings of children's early vocabulary acquisition. Chapter 3 presents the methodology used in this study, which is a corpus-based analysis. In Chapter 3, the corpus used in this study is introduced first, including the reason of choosing this corpus, information about language samples examined in this study, the definitions of lexical categories, semantic categories, and conceptual levels used in the study. Besides, the items of measuring vocabulary growth and vocabulary organization are introduced. Chapter 4 presents the results of the measurements mentioned in chapter 3 and builds up the profile of children's early lexicon. A general observation of the whole corpus is provided first, and then detailed results of all measured items are provided. Chapter 5 provides a general discussion based on the results, implication of this study, suggestion for future studies and conclusion. Abbreviations used in this study are listed below.

CHILDES: Child Language Data Exchange System

MCDI: MacArthur-Bates Communicative Development Inventory (Liu & Tsao, 2010)

NVR: Noun/Verb ratio; N/V ratio

POS: Part-of-speech

TCCM: Taiwan Corpus of Child Mandarin

TTR: Type/Token ratio

Chapter 2

Literature Review

2.1. Methods of exploring children's early vocabulary

Various methods have been adopted for the study of children's vocabulary development. In some studies, a wordlist was first constructed and validated with parents' input (Fenson, Dale, Reznick, & Bates, 1994; Dale & Fenson, 1996). Parents were asked to check the wordlist to record words and expressions that children have acquired. Researchers also designed experiments to examine children's performance of specific language knowledge (Haryu et al., 2005; Mervis & Crisafi, 1982; Zeng & Zou, 2012). Some other researchers also recorded children's speeches on audio files, transcribe the collected data, and then analyze it (Hsu, 1996; Huang, 2009; Lee, 2014; Tardif, 1996; Yeh, 2009). In this section, three methods of investigating children's language are reviewed: parental report, experiment, and corpus-based study using CHILDES.

Parental report

Vocabulary development is an essential milestone in children's early language development. It can be used to predict children's performances of later language skills (Hao et al., 2008). Due to the importance of children's early vocabulary, researchers have built norms for vocabulary measurements, among which the parental report has been the most widely used measurements. Parental report is less time consuming and easy to conduct, so researchers can easily collect a large number of children's early words in a short time. Parental report needs parents to judge whether their children can understand or produce the words in a given wordlist. Fenson et al. (Fenson et al., 1994)

and Dale and Fenson (Dale & Fenson, 1996) were the first to do this study and provided a norm for the parental report which was called MacArthur-Bates Communicative Development Inventory (MCDI). It was used to investigate American children's comprehension and production vocabulary between the ages of 0;8 (8 months) and 2;4 (2 years 4 months). Based on the American English version of MCDI, researchers have established MCDI in other languages and linguistic variants, such as British English (Hamilton, Plunkett, & Schafer, 2000), Spanish (Jackson-Maldonado, Thal, Marchman, Bates, & Gutierrez-Clellen, 1993), and Mandarin Chinese (Liu & Tsao, 2010).

Experiments

Some researchers may choose to conduct experiments to examine children's language performance. The advantage of an experiment approach is that researchers can test an assumption by controlling variables. In this way, researchers can know which assumption can be applied to children's acquisition. However, the disadvantage is that children's language samples are elicited, and the elicitation data might be different from children's natural speech. Since there are advantages and disadvantages of different research methods, researchers would use different methods or different subjects to repeatedly test an assumption.

For instance, in answering the question of a possible noun bias in early vocabulary acquisition, Gentner (1982), Hsu (1996), and Tardif (1996) counted the numbers of words in children's spontaneous speech. On the other hand, Haryu and colleagues (2005) used an experimental method to investigate the noun bias in early acquisition. They conducted an experiment to investigate whether there is a "Noun Bias" in Mandarin Chinese, Japanese, and English preschoolers of three- and five-year-old. They compared the ease of fast-mapping novel nouns to a novel object and novel verbs to a novel action.

Their findings supported the universal "Noun Bias" view.

In order to exploring the basic level effect in children's vocabulary, Jiang (2000) and Lee (2014) collected a number of spontaneous speech samples from children, and counted the number of words. On the other hand, Mervis and Crisafi (1982) conducted an experiment, and found that basic-level terms were more advantageous to children's lexicon learning than superordinate terms and subordinate terms. Furthermore, Zeng and Zou (2012) have worked on their study with three methods, including controlled experiments, to explore the early development of category levels of Mandarin-speaking children. Their results showed that basic-level words were dominant in comprehension and production of early vocabulary.

Corpus-based method

A corpus is a collection of a large number of language transcriptions. Corpus database provides children's spontaneous speech, which is exactly recorded from their real daily speech. The most well-known corpus of child language is Child Language Data Exchange System (CHILDES) (MacWhinney, 2000). It provides the transcription format, CHAT format, and analyzing tools, CLAN tools (MacWhinney, 2000). There are some advantages of using CHILDES. Researchers can conduct their researches with available data when it is difficult to have native speakers of other languages to be their subjects. They can make cross-linguistic comparisons to find out a general tendency of language development. On the other hand, the disadvantage of a corpus-based approach is that researchers cannot test certain assumptions particularly. What corpus data contains is a sample of children's everyday language use. The collected language samples may be restricted to some conversation topics. Researchers interested in other topics which are not available in CHILDES still need to collected language samples by

themselves.

2.2. Performance of children's early vocabulary

2.2.1. Vocabulary size

Past studies of English-speaking children have found that few children produce any words before age one. Most children produce their first recognizable words in 15 months or so. They have approximately 100 to 600 distinct words at 2 years old. They have about 14,000 words in comprehension and fewer in production by 6 years old. These numbers imply that children acquire words between 2 to 6 years old at a rate of nine to ten words a day (Clark, 1993). Hsu (1996) has reported the vocabulary size of Mandarin-speaking children before 6 years old in his study. Children had 260 words before 2 years old, 634 words before 3 years old, 771 words before 4 years old, 808 words before 5 years old, and 895 words before 6 years old. Yeh's study (2009) has shown a mean of nouns and verbs: 166 words before 2 years, 402 words before 3years, and 478 words before 4years. Tsay and Cheng (2011) reported the vocabulary size of 8 children speaking Taiwanese Southern Min from approximately one and a half years old to 4 years old. The total vocabulary size of these children is smaller or just a little bit larger than 2000. The vocabulary growth rate of two children was also reported and was compared with Clark's results. Clark (1993) discussed two young children's vocabulary growth. One child Keren (reported in Dromi 1987) produced up to 337 new words by 1;5 while the other child Damon produced up to 337 new words by 1;9. As for Tsay and Cheng's study (2011), they found that a child reached 337 new words between 1;5 and 1;6, and another child reached 337 new words between 1;6 and 1;7.

2.2.2. Noun bias or Verb bias

A universal "Noun Bias" in young children's vocabulary development has been debated heatedly for years. Gentner (1982) proposed Natural Partitions Hypothesis, stating that there is a preexisting perceptual-conceptual distinction between concrete concepts (nouns) and predicative concepts (verbs), and the distinction between nouns and predicate terms is based on this perceptual-conceptual distinction. The Natural Partitions hypothesis also holds that nouns belonging to concrete concepts are conceptually simpler or more basic than verbs and other predicates. Examining cross-linguistic data further evidenced that young children acquire nouns easier and earlier than verbs, and this "Noun Bias" is a universal phenomenon.

However, some studies of Mandarin Chinese and Korean vocabulary development argued against "Noun Bias", suggesting that nouns are not always acquired first (Choi & Gopnik, 1995; Tardif, 1996; Sheng, Deng, Zhang, Liang, & Lu, 2012). Based on the data collected in Beijing, Tardif (1996) argued against the universality of "Noun Bias". Ten Mandarin-speaking children participated in her longitudinal study, but only the second or third recording was analyzed when children's mean age was 21 months. In order to know whether the different definition of nouns and verbs lead to various conclusions, she included several strict and broad definitions of nouns and verbs, object labels and action words, and nominals and predicates. The results of her sliced data suggested that Mandarin-speaking children produce more verbs than nouns in their early lexicon. She further suggested that linguistic and sociocultural input factors accounted for a "Verb Bias". Mandarin verbs occur frequently in adult inputs and verbs are highlighted by occurring in salient positions. Furthermore, morphological simplicity has effects on the bias in children's performance. The morphology of English nouns is simpler than that of English verbs, and this leads to the earlier and easier acquisition of

nouns. Unlike English, Mandarin is morphologically transparent. “Noun Bias” is not reinforced by Mandarin morphology., so nouns are acquired later. The input frequency account was further evidenced in Dhillon (2010) who examined the observational data of English-, Spanish-, and Mandarin-speaking children in the CHILDES database. Children’s age ranged from 1 year and 7 months to 2 years and 11 months. The number of noun types, verb types and other types, the number of noun types versus the number of verb types, and the proportion of nouns divided by the sum of the proportion of nouns and verbs were computed. The results showed that Mandarin-speaking children exhibited a “Noun Bias” in the early stage (1;7 to 2;0) but no “Noun Bias” in the later ages. English- and Spanish-speaking children displayed a “Noun Bias” across all ages. These results have supported Dhillon’s (2010) arguments that argument-dropping in Mandarin would make children receive more verbs from the adults’ inputs than children of other languages. These factors lead to the prediction that children will learn verbs more easily.

Nevertheless, Hsu (1996) has reported that Mandarin-speaking children produce more nouns (60%) than verbs (25%), supporting a “Noun Bias”. Yeh (2009) conducted a study about young children’s acquisition of nouns and verb in Mandarin Chinese in Taiwan. The results indicated that nouns are the major words which children acquire in their early ages, as well as that younger children use relatively more nouns whereas older children use relatively more verbs. It also suggests that children’s cognitive abilities and social interactions have influences on the early acquisition of nouns and verbs. In addition to observational studies, Haryu and colleagues (2005) had done an experiment to investigate whether there was a “Noun Bias” in Mandarin Chinese, Japanese, and English preschoolers of three- and five-year-old. They compared the ease of fast-mapping novel nouns to a novel object and novel verbs to a novel action. One of

the conditions was “bare verb condition” in which novel verbs were presented with no argument. There are morphological affixes in English and Japanese but not in Mandarin, thus making the condition a “bare word condition” in Mandarin. The results showed that both the 3- and the 5-year-olds in three languages could fast-map a novel noun to a novel object, but they could not fast-map a novel verb to its meaning properly until five years old. Moreover, the results of bare word condition in Mandarin showed that children tended to map a novel word to a novel object. These findings supported the universal “Noun Bias” view. They concluded that the difficulty of Mandarin verb learning was resulted from the lack of verb morphology and the argument-dropping property of Mandarin. When children encounter a novel word, they need clues to decide whether to map the novel word to an object or an action. However, argument-dropping is allowed in Mandarin, and Mandarin verbs are not morphologically inflected. The linguistic properties of Mandarin imply that the linguistic clues to decide to map a word to an object or an action are not always available. As a consequence, children rely on the universal “Noun Bias” to map a novel word to a novel object.

In a cross-linguistic corpus-based study, Liu and her collaborators (2008) examined the noun versus verb (N/V) ratio in types in English, Cantonese and Mandarin, found different results. They conducted two studies examining 13- to 60-months-old children’s data in the CHILDES database (MacWhinney, 2000). The first study selected a total of 72 files, and the second study used all data available from the corpora. Their results revealed that when averaged over all ages, English-speaking children showed a stronger “Noun Bias”, but Cantonese-speaking children had a relatively weak “Noun Bias” and Mandarin-speaking children even had no “Noun Bias”. The same pattern was observed in adults’ inputs, and Mandarin-speaking adults even displayed a “Verb Bias”. Interestingly, children of all languages displayed “Noun Bias” before 24 months old.

The authors believed that adults' inputs might have influence on children's vocabulary learning. As what Tardif (2006) and Tardif, Gelman, and Xu (1999) have stated, Mandarin-speaking parents focused on verbs when talking to their children and the verb use was highly specific in Mandarin. On the contrary, English-speaking parents focused on nouns and used more general purpose verbs when talking to their children.

The controversy of the "Noun Bias" in Mandarin remains after continuous efforts. Previous studies investigating this issue in Mandarin by using the CHILDES database found the "Verb Bias", challenging the universal "Noun Bias" view. To clarify the dialectal variation in the so-called "Noun Bias" in Mandarin, this study examines the use of nouns and verbs in young Mandarin-speaking children in Taiwan. Moreover, Liu and her colleagues (2008) and Tardif (1996) concluded that there is no "Noun Bias" for Mandarin children using data from the CHILDES database. It is possible to reach a different conclusion using a different corpus. Besides, although Tardif (1996), Dhillon (2010), and Haryu and colleagues (2005) illustrated the biases in terms of morphology and argument structure of Mandarin, their explanations were different. Therefore, if there is a "Noun Bias" observed in this study, it supports Haryu and colleagues' explanation (2005); on the contrary, if there is no "Noun Bias" observed, it supports the explanation of Tardif (1996) and Dhillon (2010).

2.2.3. Lexical complexity in acquisition

Word frequency provides general information about the collected language samples. Word type frequency represents the number of different words children know, and word token frequency represents the total number of items children produced. However, producing many types is not equal to using all the types frequently; meanwhile, producing many tokens does not mean producing many various word types.

Thus, vocabulary diversity needs to be examined, and measurements of vocabulary diversity are frequently used in language research. One of the measurements based on the ratio of the number of different words (Types) to the total number of words (Tokens) is known as the type-token ratio (TTR). A high ratio of TTR may represent a rich vocabulary diversity which means that children produce relatively more different types in their speech. On the contrary, a low ratio of TTR may represent a weak vocabulary diversity which means that children produce fewer different types or repeatedly produce the same types in their speech.

Many measures of vocabulary diversity have been based on the type-token ratio. Unfortunately, Heaps's law (Heaps, 1978) predicts that the more words (tokens) a sample has, the less possible it is that new words (types) will show up. That is to say, the first few tokens in a sample are likely to be new types, but later words are likely to be types that have been used before. Thus, measures based on TTR are likely to be affected by the sample size. The TTR values are lower in samples with more tokens and vice versa (Tweedie & Baayen, 1998).

In order to fix the sample size problem of TTR, another measure of lexical diversity was invented. A program called *vocd* was developed to calculate D (McKee, Malvern, & Richards, 2000). This method depends on the analysis of the probability of a new word appearing in longer and longer samples. The analysis leads to a mathematical model of how TTR interacts with token size. By comparing the mathematical model with empirical data in a transcript, it provides a measure of lexical diversity which is called D. The formula of the model is the following equation.

$$\mathrm{TTR} = \frac{D}{N}\left[\left(1 + 2\frac{N}{D}\right)^{\frac{1}{2}} - 1\right]$$

N is the number of tokens. TTR is type-token ratio. D is a part of this formula to produce a theoretical curve which most closely fits the empirical TTR curve in the random samples. This equation yields a family of curves with different values of D, as depicted in Figure 2.1. The calculation of D-measure is still based on the TTR, but it prevents the flaw of TTR. It avoids being highly affected by the token size of samples. Besides, it uses all the words produced by the participants.

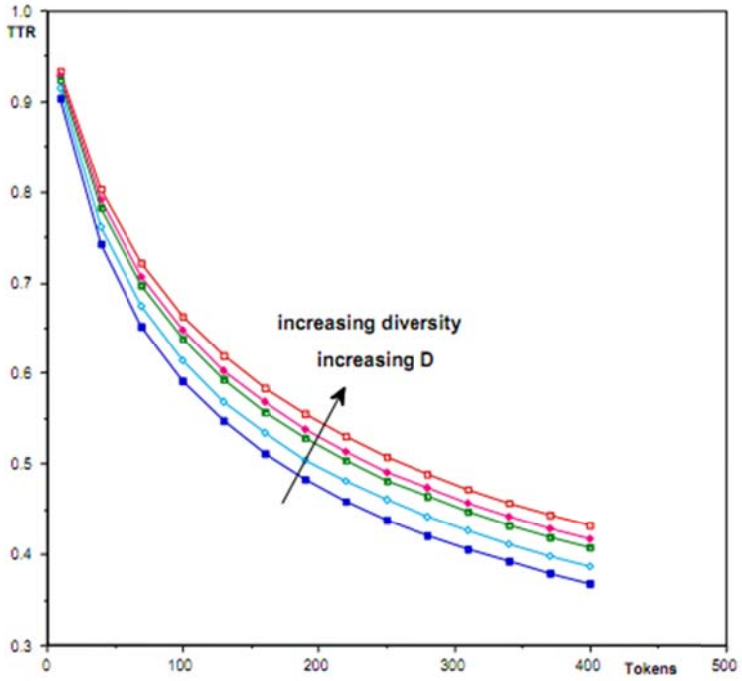


Figure 2.1: Family of curves of increasing diversity with increasing D values (McKee et al., 2000).

D-measure can be calculated by using *vocd* command provided by the CLAN program (MacWhinney, 2000; McKee et al., 2000). The following steps are the procedure of how the program *vocd* calculates D values in a transcript. First, *vocd* extracts many subsamples of varying sample sizes with random sampling (without replacement). The sample size is from 35 tokens to 50 tokens (N = 35 to N = 50), and 100 subsamples are extracted for each sample size. Second, the TTR of each subsample

is calculated, and the average TTR for each sample size is calculated to represent that point of the curve of TTR of that transcript. Third, the software finds the best fit of this empirical curve based on the TTR and the token size, and obtains the value of D-measure. The average D value of subsamples of varying sample sizes is calculated, and the best-fit D values is obtained with the least square difference method. Finally, repeat step 1 to step 3 for twice. The average best-fit D value is the best D value of that transcript (see McKee et al., 2000, for a complete flow chart). A high value of the D-measure reflects high lexical diversity, and a low value means low lexical diversity. The calculation of D-measure takes different sample sizes into consideration, so it is proven to be a more valid and reliable measure of lexical diversity (MacWhinney, 2000; McCarthy & Jarvis, 2007; McKee et al., 2000).

Liu and her colleagues (2008) used D-measure to quantify vocabulary diversity of speech samples in their cross-linguistic study about early lexical development. They found both language and age had significant main effects on lexical diversity. Mandarin-speaking children had their mean D value of 44.21. They also found that differences between any two age groups are significant, except for the difference between 25-36 months old and 37-48 months old. The results reveal that children's speech becomes more diverse with increasing ages.

2.2.4. Performance in semantic categories

Peng and Chong (2010) have conducted a study of early acquisition of nouns from two Mandarin-speaking children age from 1 to 3 years old. They have categorized nouns into kinship terms, organs, clothing, device terms, vehicles, food, animals, natural objects, and colors. Their study has indicated that children's early noun acquisition is restricted by children's cognitive ability and language inputs. Children acquire the

category of device nouns, such as 燈 (dēng, lamp), 筆 (bǐ, pen) and 球 (qiú, ball), more frequently than nouns of other categories. On the other hand, the production of nouns in categories like kinship terms, organs, natural objects and colors is relatively few. The high frequency of producing device nouns is due to the frequent inputs in children's daily life. Besides, children's cognitive ability also has effects on their language acquisition, so children acquire color terms later until three years old.

Sheng et al. (2012) examined Chinese young children's Mandarin development in Nanjing, using a revised Chinese version of MCDI (Liang et al, 2001). The study was done with 326 toddlers of two age stages, 14-16 months and 24-26 months. They have found that toddlers at stage 1 could express 42.09 words (SD = 64.43), and they acquired more onomatopoeia words than words in other categories (i.e. nouns, verbs, adjectives/adverbs, conventional game words, numeral words, quantifiers, interrogatives, pronouns, directional words, and time words). Besides, they could express more verbs than nouns. The results of stage 1 further showed that the age of 15 months was found to be the critical period of producing nouns, verbs, adjectives/adverbs, conventional game words and onomatopoeia.

Likewise, Lee (2014) has found in her study that animals, tool, and food contributes the most proportion on nouns in early Hakka vocabulary acquisition. She also concluded that the process of acquiring nouns, such as personal pronouns, body parts, clothing, tools, food, and animals, coincided with cognitive development. Additionally, children acquire nouns which are familiar and close to them first, and acquire nouns which are unfamiliar and far away later. Besides, children even invented new words for colors, for example, the strawberry color.

Kuo, Tsay, and Peng (2005) also classified Taiwanese nouns into 22 categories which were adapted from Jiang's (2000) classification. Kuo et al. (2005) compared their

results with Jiang's (2000) results, and they have found that categories in which nouns occurred most frequently were overlapped in two studies, including animal, food, body parts, natural phenomenon or substances, fruits and vehicles. Thus, it seems that objects which children contact most often are usually most frequently produced in children's language no matter in which language.

2.2.5. Basic level effects

Human beings can classify objects of similar features into one category in order to make the world less complex, and this process is called "categorization" (Bancroft, 1995). Based on personal experiences and social background, people categorize objects differently. Generally speaking, objects can be categorized at different hierarchical levels, and there exists a common basic-level at which human beings tend to represent objects in the world. The common level of categorization is called basic-level category (Rosch, 1976; Rosch, 1978). Besides, Rosch et al. (1976) proposed a three-level categorization: basic level, superordinate level, and subordinate level based on the prototype theory (Rosch, 1973; Rosch & Mervis, 1975).

(1) **Superordinate category**: objects which are the most general and abstract, with fewer common features and lower resemblance among the members.

(2) **Basic-level category**: objects which are concrete and share more common features among members. Members in one class have similar shapes and can be distinguished from those in other classes with other features.

(3) **Subordinate category**: objects which are the most specific and concrete. Members have higher within category resemblance and many individual features to distinguish them from the contrasting subordinate category.

According to Rosch et al. (1976) and Rosch (1978), objects in the basic level

category are the most concrete objects, the shortest lexicon, the most useful level of classification, the most frequently used terms, and the first acquired terms for children. Rosch and her colleagues (1976) also found that children as young as three years old have already mastered the basic-level category. Even two-year-old children have had basic-level categories, although the members of basic-level are different from adult's concepts. The discrepancy between children and adults becomes smaller when children increase their lexicon and social experiences. In the experiment of Mervis and Crisafi's study (1982), they found that basic-level terms are more advantageous to children's lexicon learning than superordinate terms and subordinate terms.

Besides, Jiang (2000) extended the basic level effect to examine Mandarin nouns in adults' and children's lexicon. She uses several criteria to examine the basic-level effect in Mandarin, such as basic-level lexicon should be at the central level, concrete, shortest, highly frequently-used, high derivative, and early acquired. Based on these criteria, she analyzed word derivativity, word frequency and the learning order of words. She found that children as young as 11 to 24 months old had 70% lexicon which was basic-level words, followed by subordinate words and superordinate words. Basic-level words were the most derivative and highly frequently used than the other two levels. She further found that basic-level words which were related to children's daily life had much higher derivativity, and much higher frequency of use, such as basic-level words about animals, food, toys and transportation vehicles. In terms of toys and fruits, superordinate words were even more frequently used than basic-level words.

However, Jiang (2000) took only eight semantic categories into consideration and only selected one word to represent each category of each level when discussing the derivativity and the frequency of children's use. The eight categories are animals, plants, food, fruits, clothing, toys, furniture, and transportation vehicles. Her analysis was not

comprehensive enough, so more semantic categories are included in this study to explore the basic level effect on Mandarin vocabulary from young children.

There is a study of basic-level effect on Mandarin vocabulary with various research methods. Zeng and Zou (2012) have conducted their study with three methods: controlled experiments, longitudinal case data, and Zhou's corpus data to investigate the early development of category levels of Mandarin-speaking children. Their results showed that basic-level words were dominant in comprehension and production of early vocabulary. Basic-level words were acquired first, followed by subordinate words and superordinate words. They further found a correlation between the development of categorization and language development. The development of categorization ability promotes children's early language development.

A study of basic-level effect on another language, Hakka, also reached the same conclusion. In Lee's study (2014), her results have shown that a large number of Hakka nouns were basic-level words, subordinate words were fewer than basic-level words, and superordinate words were the least. In terms of the frequency of use, basic-level words were used with a higher frequency than superordinate words and subordinate words. Adopting Jiang's theory of the ability of deriving new words (Jiang, 2000), basic-level nouns had the highest ability of deriving new words. Basic-level nouns of the category of tools had the best ability to derive new words. Besides, she also discovered that the timing of acquiring Hakka nouns of the superordinate level and subordinate level categories differs according to the types of nouns.

Chapter 3

Methodology

This chapter introduces the source of language samples, and how they were analyzed. The language samples were archived in Taiwan Corpus of Child Mandarin (Cheung, Chang, Ko, & Tsay, 2011). The analyses focused on examining vocabulary growth and vocabulary organization. Vocabulary growth is defined as how vocabulary grows in the aspect of quantity. It was measured by computing the number of words, type/token ratio, D measure, and noun/verb ratio. Vocabulary organization is defined as how vocabulary grows in the aspect of word meanings. It was measured by computing the number of noun types in different semantic categories and the three levels of categorization.

3.1. About the TCCM corpus

3.1.1. Taiwan Corpus of Child Mandarin

The language samples examined in this study are from the Taiwan Corpus of Child Mandarin (Cheung, Chang, Ko, & Tsay, 2011), hereafter TCCM[1]. TCCM is a recently-built corpus which contains scripts transcribed from Mandarin-speaking children's language samples from past studies conducted in Taiwan. TCCM is a language data exchange system which aims to provide an open access for general public and researchers, and to use a standardized format to transcribe and code the language samples collected in Taiwan. This corpus follows the standards of the Child Language

[1] TCCM website: http://taiccm.org/

Data Exchange System (CHILDES) (MacWhinney, 2000), the most important child language system in the world, using its CHAT format and CLAN tools (MacWhinney, 2000). Besides, TCCM is a browsable database, so researchers can directly do analyses on the website.

TCCM has contained 362 files. Children's age ranges from one year and five months (1;5) to eight years and five months (8;5). There are several different types of language samples in this corpus for different research interests. It includes language samples not only from typically developing children but also from children with specific language impairment. Besides, children's language samples are not limited to only daily conversation. The data includes samples from parent-child book-reading, child story-reading, and children's spontaneous speech. In TCCM, the HTC01 (Cheung, 1998) database provides data from children's spontaneous speech; the HTC02 (Cheung, 2003) database contains samples from children with specific language impairment, but this database has not been accessible yet; the CJC01 (Huang, 2009) database provides data from parent-child book-reading conversation; the CJC02 (Chang, 2003) database provides data from adult-child conversation. In order to prevent possible effects from genre types and effects from the difference between typically developing and impaired children, the language data examined in the present study was limited to only one specific type. That is to say, only language samples from spontaneous conversation produced by typically-developing children, the HTC01 database, were examined.

The transcription has been processed. The utterances in the transcripts have been segmented and part-of-speech has been tagged onto words. The standards of word segmentation and part-of-speech tags are mainly based on *Segmentation Principle for Chinese Language Processing* (Chinese Knowledge Information Processing Group, 1998b), and a technical report about Sinica Corpus provided by Chinese Knowledge and

Information Processing Group in Academia Sinica Institute of Information Science (CKIP Group, 1998a; CKIP Group, 1998b). It also refers to *Elementary School Children's Common Words Report* (National Languages Committee, 2000), *The POS Guidelines for the Penn Chinese Treebank* (Xia, 2000a), *The Segmentation Guidelines for the Penn Chinese Treebank* (Xia, 2000b). It concerns the properties of Mandarin-speaking children's language development as well. Thus, the processed language data in TCCM may provide researchers with much clearer data and a more convenient way for searching or follow-up analyses.

3.1.2. Use TCCM rather than CHILDES

In the field of corpus-based studies about child language, CHILDES has been the most important and well-known corpus, but it is not the best choice in the present study. The three disadvantages of using CHILDES are regional difference of Mandarin, inconsistent transcription system, and insufficient parts-of-speech information.

The first disadvantage of using CHILDES is the regional difference of Mandarin. Although Mandarin database is also provided in CHILDES, most of the data are recorded in Beijing and Nanjing, China, only a few data are from Hsinchu, Taiwan. As a result, many corpus-based studies about child language development of Mandarin have been based on data in the Beijing Mandarin database in CHILDES. For example, Tardif (Tardif, 1996) has done her study about the issue of the predominance of noun or verb in early language development with the Beijing Mandarin. Liu and her collaborators (Liu et al., 2008) conducted a cross-linguistic corpus-based study, using the Beijing Mandarin database as well. However, there are some dialectal variations among Mandarin from different regions. CHILDES may be a choice for those who are interested in Beijing Mandarin or Nanjing Mandarin, but it provides little help for

researchers investigating Mandarin spoken by children in Taiwan. On the other hand, TCCM collects language data from Mandarin-speaking children in Taiwan. Therefore, researchers interested in Mandarin in Taiwan now have a choice of using TCCM.

Secondly, using CHILDES has the problem of inconsistent transcription system. The Mandarin transcripts are transcribed either in Chinese characters or in Pinyin system, the Standard Mandarin Romanization system (ISO 7098, 1982). Using Pinyin, sounds of Chinese characters are represented in Roman letters. However, a serious problem exists in the scripts transcribed in the Pinyin system. Pinyin system only records sounds; as a result, homophones are represented identically. Since there are more characters than sounds in Chinese, it is difficult to differentiate homophones when Chinese characters are transcribed in Pinyin system. For example, the sound *zuò* could represent 做 (do/make), 坐 (sit), and 座 (seat). A researcher could hardly determine which of several characters should be applied when a single character in the form of Pinyin exists without its context in a sentence. In addition to the difficulty of determining the appropriate semantic meaning of the homophone, a researcher has to deal with the problem of searching for a homophone of a specific meaning. Therefore, investigators using Mandarin database in CHILDES have to rewrite the transcripts into Chinese characters. On the contrary, the language samples in TCCM have been transcribed in Chinese characters rather than in Pinyin system. An investigator can examine the transcripts directly without the step of converting Pinyin system into character system. Likewise, they can easily recognize the characters and the semantic meanings of the homophones with the help of Chinese characters.

Lastly, researchers who have to take lexical category into consideration in their study would find that the Mandarin database in CHILDES does not provide sufficient parts-of-speech information. Any transcribed language samples have to be preprocessed

before analyses. Sentences have to be segmented, and POS information has to be added onto each word. Nevertheless, some of the transcripts in the Mandarin database in CHILDES do not provide parts-of-speech information. Investigators have to tag POS onto each word by themselves before their analyses. Rather, people using TCCM do not have to do any preprocess. All of the language samples in TCCM have already been segmented, and POS tags have been added onto all words of all scripts as well.

In brief, choosing CHILDES would have to deal with three disadvantages, while choosing TCCM would not. The most important of all, the purpose of the present study is to investigate language acquisition of Mandarin-speaking children in Taiwan, and thus TCCM is better than CHILDES.

3.1.3. Data used in this study

In order to explore language development from a longitudinal observation, language samples of children's spontaneous speech in this study were retrieved from the HTC01 database (Cheung, 1998). The files were collected from 10 young children. Children's age range was from 1;5 (1 year and 5 months) to 4;3 (4 years and 3 months) (mean = 32.76 months, SD = 8.25 months). The following Table 3.1 is the information of each child in the HTC01 database, including their age range, recording sessions, number of child utterances, adult utterances and all utterances. These language samples were audio-recorded at the child's home once a month. The data was transcribed with CHAT format using the CLAN analyzing tools (MacWhinney, 2000). Most children have more than 10 recording sessions for more than one year. According to the CHAT transcription format, utterances should end with an utterance terminator. The basic utterance terminators are the period, the exclamation mark, and the question mark. Thus, each line in a transcription contained one utterance which was also one intonation unit.

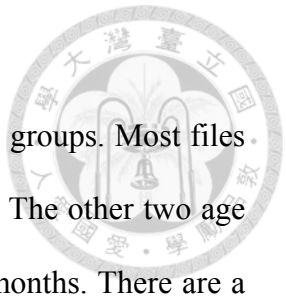
Table 3.1: Children's information in the HTC01 database.

Child	Age range	Sessions	Child utterance	Adult utterance	Total utterance
CHENG	3;1－3;11	11	3564	5246	8810
CHOU	2;1－3;4	16	5253	9131	14384
CHW	3;6－4;3	9	2685	3649	6334
JC	2;2－3;5	14	5244	8738	13982
Pan	1;7－3;9	19	3700	7028	10728
WANG	2;5－3;4	12	3192	4471	7663
WU	1;7－2;10	12	2854	8835	11689
WUYS	2;7－3;10	10	1879	5704	7583
XU	1;6－2;5	11	2726	3948	6674
YANG	1;5－2;9	13	2745	7131	9876

In the HTC01 database, there are 127 files of spontaneous speech from 10 typically-developing children. In order to observe the developmental trend of different age stages, children's language samples were divided into 7 groups with six months as a scale. Children of 25 months to 42months old contribute the most files. Three file are from children younger than 19 months old, and another three files from a child older than 48 months old. These six samples are insufficient to form groups of 13-18 months and 49-54 months as other age groups with more than 10 samples, so these six files were not included in the following analyses. Therefore, there are a total of 121 files examined in the study. The following Table 3.2 provides the number of samples analyzed in each age group, the number of children in each age stage, and the number of utterances of children and adults.

Table 3.2: The information of samples in each age group.

Age	Samples	Child	Child utterance	Adult utterance	Total utterance
19-24 months	19	4	4345	10957	15302
25-30 months	31	7	9537	19167	28704
31-36 months	28	7	6183	11682	17865
37-42 months	29	7	8360	14000	22360
43-48 months	14	4	3589	4812	8401
19-48 months	**121**	**10**	**32014**	**60618**	**92632**



As seen in Table 3.2, the files are not equally distributed in age groups. Most files are in the group of 25-30 months, 31-36 months, and 37-42 months. The other two age groups consisted of 19 files in 19-24 months and 14 files in 43-48 months. There are a total of 32014 child utterances and 60618 adult utterances in these 121 files. Children's mean age and standard deviation in each age group are provided in Table 3.3.

Table 3.3: The mean age and standard deviation in each age group.

Age range	Mean age	SD
19-24 months	21.4	1.83
25-30 months	27.6	1.78
31-36 months	33.2	1.79
37-42 months	39.1	1.62
43-48 months	45.3	1.49

3.1.4. Lexical categories

Lexical category is also called part-of-speech. The standard of tagging part-of-speech (POS) in the TCCM database is mainly adapted from *Segmentation Principle for Chinese Language Processing* provided by Chinese Knowledge and Information Processing Group of the Academia Sinica (CKIP Group, 1998a; CKIP Group, 1998b). There are 28 POS tags used in coding the language samples. The 28 POS tags and the word class they represent are listed in Table 3.4 below. In subsequent analyses, two lexical categories which are not Mandarin lexicon are excluded: foreign words (CS), most of which are Taiwanese Southern Min, and unknown words (CHM). Word frequency of each POS is presented in Table 1 in Appendix 3. A total number of 4026 word types and 75739 word tokens are included in the analysis.

Table 3.4: The 28 POS tags.

Tag	Part of Speech	Tag	Part of Speech
ADV	Adverb	**Npro**	Pronominal
ADJ	Adjective	**Nt**	Time noun
ASP	Aspect marker	**NEG**	Adverb of negation
CONJ	Conjunction	**ONM**	Onomatopoeia
DE	Nominal marker, Adverbial marker, Complement marker	**PREP**	Preposition
DT	Determiner	**QN**	Quantifier
INT	Interjection	**SFP**	Sentence final particle
CS	Foreign words	**CHM**	Unknown word
IDM	Idiom	**Va**	Stative verb
CL	Classifier/measure word	**Vc**	Shi−word
MOD	Modal	**Vi**	Intransitive verb
Nn	Common Noun	**Vt**	Transitive verb
Nloc	Localizer	**Vr**	Reduplicated verb
Nppn	Proper noun	**WH**	Interrogatives “WH−”

In subsequent analyses, one of the measures of vocabulary growth is based on noun-verb ratio. However, as what Tardif (1996) stated, the inclusion of nouns and verbs is too broad and controversial. The classification of nouns and verbs may be different due to different standards in various studies. Thus, the definitions of these two word classes in TCCM are illustrated in detail. As for other parts-of-speech, the corresponding examples are provided in the Appendix 1. Nouns and verbs can be divided into some subcategories according to their meanings and functions. Noun category in this corpus includes common nouns, localizers, proper nouns, pronominal, and time nouns. Verb category is composed of stative verb, shi-word, intransitive verb, transitive verb and reduplicated verb.

Noun: Nouns are tagged as **Nn**. Nouns includes all countable and uncountable concrete nouns, abstract nouns, collective nouns, and children’s unique forms. Example (3-1) lists the common nouns. Example (3-2) lists two children’s unique forms.

(3-1) 桌子 (zhuōzi, table), 水 (shuǐ, water), 香氣 (xiāngqì, aroma)

(3-2) 蛛蛛車 (zhūzhūchē, spider car), 吸玻璃魚 (xībōlíyú, absorbing glass fish)

Localizer: Localizers are tagged **Nloc**, including location words which indicate the relative position of objects. A noun with a postposition is viewed as localizer as well. The following example (3-3) lists location words. Example (3-4) demonstrates that the postpositions 裡 (lǐ) and 上 (shàng) are attached to a noun to indicate the location. The whole phrase functions as a localizer.

(3-3) 中間 (zhōngjiān, middle/center), 西北 (xīběi, northwest)

(3-4) 杯子裡 (bēizilǐ, in the cup), 地上 (dìshàng, on the ground)

Proper noun: Proper nouns are tagged as **Nppn**, including all proper names such as the name of an individual person, place, etc., except for time-related names. Example (3-5) lists the name of a person and the title of a book. In addition, country names and city names are also proper nouns, as seen in example (3-6).

(3-5) 余光中 (yúguāngzhōng, Yu Kwang-Chung), 詩經 (shījīng, Book of Songs)

(3-6) 西班牙 (xībānyá, Spain), 台北 (táiběi, Taipei)

Pronominal: Pronominals are tagged as **Npro**, including personal pronouns, reflexive pronouns, demonstrative pronouns, and generic pronouns. The examples of pronominals are listed below.

(3-7) personal pronouns: 我 (wǒ, I/me), 你 (nǐ, you), 我們 (wǒmen, we/us)

(3-8) reflexive pronouns: 自己 (zìjǐ, oneself)

(3-9) demonstrative pronouns: 這裡 (zhèlǐ, here), 這邊 (zhèbiān, here/this side)

(3-10) generic pronouns: 之 (zhī, it/them), 其 (qí, that)

Time noun: Time-related nouns are tagged as **Nt**, including the name of a historical time and a repeated time. The name of a historical time means the name of a dynasty and the reign title, as seen in example (3-11). The name of repeated time includes dates, seasons, and festivals, as seen in example (3-12).

(3-11) 清朝 (qīngcháo, the Qing Dynasty), 西元 (xīyuán, Anno Domini/ A.D.)

(3-12) 春天 (chūntiān, spring)

Stative verb: Stative verbs are coded as **Va** which are defined as describing states of affairs, rather than actions. Stative verbs usually have only one argument. The items in the example below are stative verbs. It shall be noted that sometimes stative verbs such as 好 (hǎo, good), 新 (xīn, new), and 漂亮 (piàoliàng, beautiful) should be classified into adjective when it precedes a noun.

(3-13) 大 (dà, big), 高 (gāo, high), 浪漫 (làngmàn, romantic),辛苦 (xīnkǔ, toilsome), 豐富 (fēngfù, abundant)

The coding of stative verbs in the TCCM database is different from the coding in the CHILDES database. In the TCCM database, some stative verbs, such as 漂亮 (piàoliàng, beautiful), are tagged as Va, while it is viewed as an adjective in the CHILDES database. This coding standard is applied to both Mandarin and Cantonese language samples in the CHILDES database. That is, what is coded as Va in the TCCM database is tagged as an adjective in the CHILDES database, and such a difference may lead to inconsistent results of some analysis such as computing the N/V ratios.

Shi-word: Shi-words are tagged as **Vc**, including the word 是 (shì, copula verb be), 像 (xiàng, resemble), and 等於 (děngyú, equal). It is noteworthy that the word 像 (xiàng) is polysemy. 像 (xiàng) or 好像 (hǎoxiàng) also functions as a preposition or an adverb. Thus, 像 (xiàng) is only viewed as a shi-word, when expressing the meaning of equal or having the function of the be verb.

Intransitive verb: Intransitive verbs are tagged as **Vi**, including verbs which cannot take a direct object, including intransitive dynamic verbs, motion verbs, verbs in

verb-object construction with locations as the object, and weather verbs. The following examples are intransitive verbs.

(3-14) 開會(kāihuì, to hold a meeting), 跑 (pǎo, run), 逛街 (guàngjiē, go shopping), 下雨 (xiàyǔ, rain)

In addition, some verbs need two arguments semantically, but the objects do not follow the verb directly. Instead, the objects need to follow prepositions or to be fronted. This kind of verbs is also classified as intransitive verbs. For instance, 求婚 (qiúhūn, propose) and 心動 (xīndòng, touch one's heart) are also viewed as intransitive verbs.

Transitive verbs: Transitive verbs are tagged as **Vt**, including verbs which usually need two arguments semantically and can take a direct object. Examples of transitive verbs are listed below. Besides, the verb 要 (yào, want) and the existential verb 有 (yǒu, have) are also viewed as transitive verbs.

(3-15) 打 (dǎ, hit), 進 (jìn, enter), 吃 (chī, eat), 買 (mǎi, buy), 使用 (shǐyòng, use), 放 (fàng, put), 去 (qù, go), 猜想 (cāixiǎng, guess)

Reduplicated verb: Reduplicated verbs are tagged as **Vr.** The reduplicated verbs occur in some patterns. The patterns and examples are listed below.

(3-16)

AA: 走走 (zǒuzǒu, walk), 看看 (kànkàn, look at)

A 來 A 去: 跑來跑去 (pǎoláipǎoqù, run around)

A 上 A 下: 跳上跳下 (tiàoshàngtiàoxià, jump up and down)

A 一 A: 敲一敲 (qiāoyìqiāo, knock again and again)

越來越 A: 越來越胖 (yuèláiyuèpàng, fatter and fatter)

一面 A 一面 B: 一面說一面笑 (yímiànshuō yímiànxiào, saying and smiling)

一邊 A 一邊 B: 一邊走一邊吃 (yìbiānzǒu yìbiānchī, walking and eating)

The coding of reduplicated verbs is the same as other POS, attaching the tag onto each word directly. For example, 走走 (zǒuzǒu, walk) is coded is "Vr|走走"; 跑來跑去 (pǎoláipǎoqù, run around) is coded as "Vr|跑來跑去"; 跳上跳下 (tiàoshàngtiàoxià, jump up and down), is coded as "Vr|跳上跳下". However, it is noted that the "A 一 A" pattern is coded as "V|A Vr|一 A" and the "越來越 A" pattern is coded as "Vr|越來越 V|A". For example, 敲一敲(qiāoyìqiāo, knock and knock) is coded as "Vt|敲 Vr|一敲"; 越來越胖 (yuèláiyuèpàng, fatter and fatter) is coded as "Vr|越來越 Va|胖". As for more complicated word phrase, the phrase is segmented and coded separately as seen in the following example.

(3-17)

From file "HTC01_CHW402_08.cha"

*WAN: 你 **一邊坐** 他 有 沒有 **一邊叫** ?

nǐ yìbiā zuò tā yǒu méiyǒu yìbiānjiào

Is he barking at you while you sit down?

%mor: Npro:2sg|你 **Vr|一邊坐** Npro:3sg|他 Vt|有 NEG|沒有 **Vr|一邊叫** ?

The following Table 3.5 provides the POS tags and examples of nouns and verbs used in the TCCM corpus. Because different inclusion of nouns and verbs may lead to debates on noun bias or verb bias in early Mandarin acquisition, a broad and strict inclusion of nouns and verbs are taken into consideration when calculation N/V ratio. Broad noun includes all subcategories of nouns; while strict noun includes only common nouns Nn. Broad verb includes all subcategories of verbs, while strict verb excludes stative verbs Va.

Table 3.5: POS tags and examples of nouns and verbs used in the TCCM corpus.

Tag	POS	Example		
Nn	Noun	桌子	水	公園
		zhuōzi	shuǐ	gōngyuán
		table	water	park
Nloc	Localizer	中間	地上	水裡
		zhōngjiān	dìshàng	shuǐlǐ
		middle	on the ground	in the water
Nppn	Proper noun	台北市	小明	雙魚座
		táiběishì	xiǎomíng	shuāngyúzuò
		Taipei City	XiaoMing	Pisces
Npro	Pronominal	你們	他	自己
		nǐmen	tā	zìjǐ
		you	he	self
Nt	Time noun	春天	現在	早上
		chūntiān	xiànzài	zǎoshàng
		spring	now	morning
Va	Stative verb	高	大	辛苦
		gāo	dà	xīnkǔ
		high	big	toilsome
Vc	Shi-word	是	像	
		shì	xiàng	
		copula	(look) like	
Vi	Intransitive verb	跑	下雨	逛街
		pǎo	xiàyǔ	guàngjiē
		run	rain	go shopping
Vt	Transitive verb	吃	打	「有」字句
		chī	dǎ	yǒu
		eat	hit	*you3*sentence
Vr	Reduplicated verb	看看	走來走去	敲「一敲」
		kànkàn	zǒuláizǒuqù	qiāo “yìqiāo”
		look at	walk around	knock and knock
		「越來越」胖		
		“yuèláiyuè” pàng		
		fatter and fatter		

3.2. Measuring vocabulary growth

In the present study, vocabulary growth was measured by computing the cumulative vocabulary, the frequency of POS, the proportion of POS, noun/verb ratio, the type/token ratio of POS, and D measure.

3.2.1. Vocabulary size

In order to get vocabulary size, type frequency in each age group was computed, and it represented the number of distinctive words children produced in that period. Besides, the number of cumulative vocabulary was also calculated and it would show the number of new words in a stage. The procedure is to count the cumulative vocabulary of each child first. Then, the mean cumulative vocabulary in each age stages was computed. The calculation was done with the *freq* command in the CLAN tools, and used the code below to get the total number of different word types in the transcript.

The *freq* code:

> freq +t*CHI +t%mor +s"ADV*" +s"ADJ*" +s"ASP*" +s"CL*" +s"CONJ*" +s"DE*" +s"DT*" +s"IDM*" +s"INT*" +s"MOD*" +s"NEG*" +s"Nloc*" +s"Nn*" +s"Nppn*" +s"Npro*" +s"Nt*" +s"ONM*" +s"PREP*" +s"QN*" +s"SFP*" +s"Va*" +s"Vc*" +s"Vi*" +s"Vr*" +s"Vt*" +s"WH*" @

The POS was specified in the code to prevent polysemy words from being counted as one type when POS was not specified. For example, the number of types of the child CHENG's language sample in 3;1 was 195 when the POS was specified, and 189 when it was not specified. To count cumulative vocabulary, +u was added to the code to count the merged total number of different word types from successive transcript files, to prevent counting the same word type in several files repeatedly.

3.2.2. POS frequency

Word frequency is calculated in terms of word types and word tokens respectively. Counting the number of types makes us know how many different words children produced, while counting the number of tokens makes us know the total number of words children produced and which words they used most frequently. In this study, the mean frequency and total frequency were calculated. The frequency calculation was

done with the *freq* command in the CLAN program. For example, use the following code to compute the frequency of words tagged as "ADV" in one children's language sample file, and the result like Figure 3.1 is obtained.

The *freq* code: freq +t*CHI +t%mor +s"ADV*" @

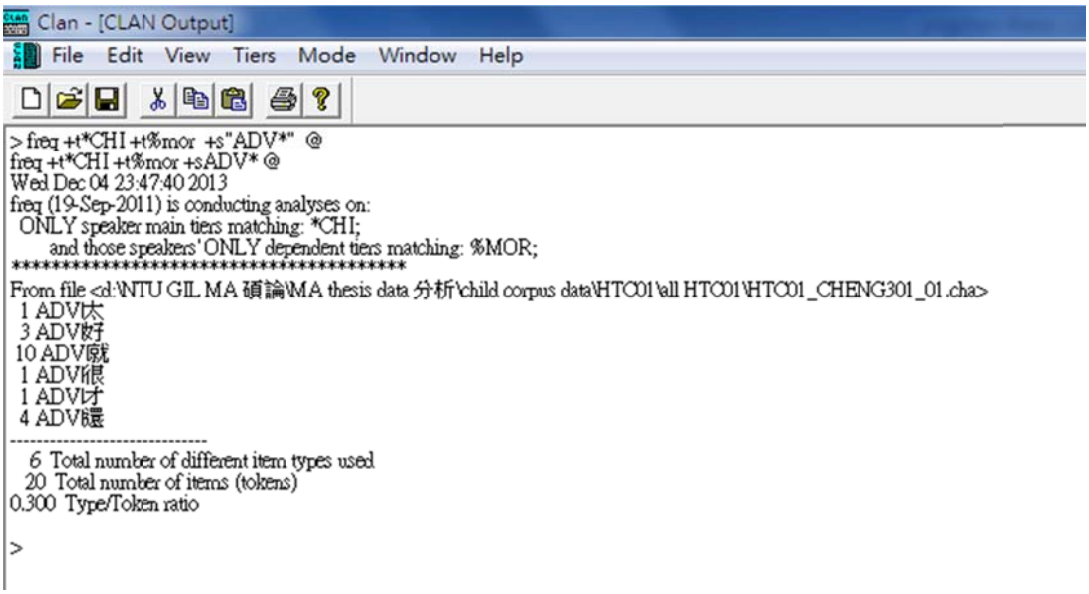


Figure 3.1: The output of *freq* command.

The output contains the list of the words of that POS, the type frequency (Total number of different item types used), token frequency (Total number of items) and TTR (Type/Token ratio). The result shows that there are six words tagged as ADV in that file, so the total number of different item types is six. The number in front of each item is the frequency of that word; sum them up and get 20 which is the total number of tokens. The type/token ratio of ADV is 0.30 in that file.

Mean word frequency of each age stage was computed. First, read all files of a stage into CLAN and use the *freq* command to calculate the frequency of each POS in each file. Then, add +f in the code to save the results as output files. One input file got one output result file. The output files listed the words of that POS, and the type frequency, token frequency and TTR. The next step was to calculate the mean frequency

and its standard deviation in each age group. The total word frequency in each age group was calculated as well, and this was provided to know the total number of words produced by all children in a stage. To calculate the total number of words in each age group, +u is used in the code, and it automatically merged the files and computed the total word frequency.

However, one important thing to note is that the mean word type frequency calculated by the above steps (calculation A) does not equal to the total word type frequency divided by the number of files in an age group (calculation B). The discrepancy between two calculations resulted from the repeated counting of one type from different files in calculation A. When the word frequency of each file was counted separately in calculation A, it was possible to count a word type as one type in file A and count it again in file B. On the contrary, when the total word frequency was computed, it merged all files first and then counted the word frequency. A word type was counted as one type in this method of calculation. Consequently, the total word type frequency divided by the number of files in an age group (calculation B) is an average number of word types of knowing, whereas the mean type frequency (calculation A) is more like an average number of word types of using.

3.2.3. Type/token ratio

The mean token frequency reveals the number of total tokens children may produce in a sliced time, and the mean type frequency shows the number of different words children may use. However, a child may know many different words but only frequently use some of them, so type/token ratio needed to be examined. TTR of each POS in each file was computed and then the average TTR in each age group was calculated. The formula for TTR is:

$$\text{TTR} = \frac{Number\ of\ word\ typs}{Number\ of\ word\ tokens}$$

Since the result of *freq* command consists of the number of word types, tokens and TTR, as Figure 3.1 has shown, thus, the calculation of TTR is done with the *freq* command in the CLAN program as well.

3.2.4. POS proportion

Since the different sample size has influence on the word frequency of each age group when comparing between parts-of-speech, the frequency need to be transferred into percentage when comparing the proportion of POS. First, get the word frequency of a POS and the total number of words in each file. Second, get the proportion of that POS: dividing the total number of words in each file by the word frequency of that POS in that file. The calculation formulation is listed below. Third, repeat the two steps above for all POS in all files. Last, calculate the average proportion of a POS in an age group. The POS type proportion and token proportion are calculated respectively.

$$\text{POS Proportion} = \frac{Number\ of\ words\ of\ a\ POS}{Number\ of\ all\ words}$$

3.2.5. Noun/verb ratio

In order to know whether children in the TCCM showed a tendency of noun bias or verb bias, the noun/verb ratio (NVR) in types was computed. The calculation procedure is as the following steps. First, the number of the noun types and verb types in each file was calculated. Second, the ratio of nouns versus verbs in each file was the number of noun types divided by the number of verb types, as the formulation below shows. Third, calculate the average N/V ratio of each age group.

$$\text{NVR} = \frac{Number\ of\ Noun\ types}{Number\ of\ Verb\ types}$$

When the number of noun types within a child's vocabulary outweighs verb types, the obtained N/V ratio will be higher than 1, and it will be viewed as "Noun Bias". On the contrary, when the number of verb types within a child's vocabulary outweighs noun types, the obtained N/V ratio will be smaller than 1, and it will be viewed as "Verb Bias". In addition, there has been a controversy about noun bias or verb bias in early Mandarin acquisition, and one of the reasons of this controversy might be caused by the different definition of nouns and verbs. Thus, a broad and strict inclusion of nouns and verbs are taken into consideration when calculation N/V ratio. Four kinds of N/V ratios will be obtained under four situations, as Table 3.6 below shows. Broad nouns include all noun categories, common nouns (Nn), localizers (Nloc), proper nouns (Nppn), pronominals (Npro), and time nouns (Nt), while strict nouns include only common nouns (Nn). Meanwhile, broad verbs include all verbs, stative verbs (Va), Shi-words (Vc), intransitive verbs (Vi), transitive verbs (Vt), and reduplicated verbs (Vr), whereas strict verbs exclude stative verbs (Va), since stative verbs are viewed as adjectives in some studies.

Table 3.6: Noun/Verb ratios of different inclusion of nouns and verbs.

Noun/Verb ratio	Broad Noun (all noun categories)	Strict Noun (only common noun)
Broad Verb (including stative verb)	NVR 1	NVR 3
Strict Verb (no stative verb)	NVR 2	NVR 4

3.2.6. D-measure

In this study, D values are calculated to examine the lexical diversity of children in

different age groups. D-measure is calculated by using *vocd* command provided by the CLAN program (MacWhinney, 2000; McKee et al., 2000). First, read in the transcript into the CLAN program. Second, use the *vocd* command to compute the D value of that transcript. An example of the output of *vocd* is provided in Appendix 2. Third, calculate the average D value of each age group. The mean D value should provide some indications. As children grow up, their speech should become more complex and diverse. Thus, an increasing D value with increasing age is expected.

3.3. Measuring vocabulary organization

Vocabulary can be classified into different groups according to their concept meanings or their conceptual levels. Although children may acquire more vocabulary with increasing ages, words they acquire may be from the same or different categories. Vocabulary organization is defined as children's language competence in different conceptual levels and in different semantic categories. In this study, vocabulary organization was measured by computing the number of nouns in different semantic categories and conceptual levels. Children are expected to know more words from various semantic categories in their later age stages than in earlier age stages. Moreover, they are also expected to acquire more basic-level words in their earlier life and more superordinate and subordinate words in later age stages.

3.3.1. Classification of semantic category

The classification of nouns is different in various studies. The semantic categories of nouns in this study is based on MacArthur-Bates Communicative Development Inventory (MCDI) (Taiwan version) (Liu & Tsao, 2010), Jiang (2000), and Lee (2014).

The MCDI (2010) has two forms for children of different ages. The infant form

was designed for 8- to 16-month-old children, and the toddler form was designed for 16- to 36-month-old children. There were 17 semantic categories in the infant form and 19 categories in the toddler form. Since the language samples in the present study were from children of 19 months to 48 months old, categories in the toddler form were adopted. The categories were: animals, body parts, vehicles, food and drink, clothing, toys, small household items, furniture and rooms, outside things, places to go, games and routines, people or pronouns, words about time, prepositions and locations, action words, connecting words, descriptive words, demonstratives and quantifiers, and question words. Only categories about nouns were examined, so action words, connecting words, descriptive words, and question words were excluded.

In Jiang's (2000) study, she used more categories to classify nouns. Her classification was based on the meanings of words and how people used it in their daily lives. There were 26 categories: animals, food, body parts, toys, vehicles, clothing, fruits, natural phenomena and natural materials, tools, stationery, household appliances, insects, vegetables, plants, colors, shapes, furniture, tableware, cleaning supplies, medicine, musical instruments, kinship terms, places and architecture, numerals, people, and abstract nouns. In her study, the categories of kinship terms and people were viewed as proper names which were excluded from her analyses. On the other hand, Lee (2014) used 13 categories in her study about the acquisition of Nouns of Hakka. The categories were: people, body parts, clothing, tools, food, animals, natural phenomena and materials, toys, colors, shapes, names of locations and buildings, numerals, and abstract nouns.

Table 3.7 provides the comparison of classifications among the studies mentioned above. The classification of nouns in the present study was mainly based on the categories of Lee (2014), adopting the advantages of classifications of MCDI (Liu &

Tsao, 2010) and Jiang (2000), and formed 17 categories:

(1) People: 哥哥 (gē ge, brother), 小馬叔叔 (xiǎomǎ shúshu, Uncle Ma), 醫生 (yīsheng, doctor)

(2) Pronouns: 我 (wǒ, I), 這些 (zhè xiē, these), 別人 (bié rén, someone else)

(3) Body parts (including body-related phenomena): 嘴巴 (zuǐ bā, mouth), 手 (shǒu, hand), 鼻水 (bí shuǐ, runny nose)

(4) Clothing: 外套 (wài tào, coat), 帽子 (mào zi, hat), 褲子 (kù zi, pants)

(5) Vehicles: 卡車 (kǎ chē, truck), 汽車 (qì chē, car), 火車 (huǒ chē, train)

(6) Tools: 剪刀 (jiǎn dāo, scissors), 手錶 (shǒu biǎo, watch)

(7) Furniture: 床 (chuáng, bed), 桌子 (zhuō zi, table), 門 (mén, door)

(8) Food: 果汁 (guǒ zhī, juice), 蘋果 (píng guǒ, apple), 飯 (fàn, rice)

(9) Animals (including body-parts of animals): 大象 (dà xiàng, elephant), 小鳥 (xiǎo niǎo, birs), 尾巴 (wěi bā, tail)

(10) Natural phenomena and materials: 太陽 (tài yang, sun), 樹 (shù, tree), 花 (huā, flower)

(11) Toys: 球 (qiú, ball), 娃娃 (wá wá, doll), 小提琴 (xiǎo tí qín, violin)

(12) Colors: 紅色 (hóng sè, red), 綠色 (lǜ sè, green)

(13) Shapes: 圓形 (yuán xíng, circle), 三角形 (sān jiǎo xíng, triangle)

(14) Locations and buildings: 動物園 (dòng wù yuan, zoo), 學校 (xué xiào, school), 台北 (tái běi, Taipei), 橋 (qiáo, bridge)

(15) Spatial words: 右邊 (yòu biān, right side), 家裡 (jiā lǐ, at home)

(16) Numerals: 五 (wǔ, five), 三點 (sān diǎn, three o'clock)

(17) Abstract nouns: 故事 (gù shì, story), 音樂 (yīn yuè, music)

Table 3.7: Semantic categories in different studies.

	MCDI (Liu & Tsao, 2010)	**Jiang (2000)**	**Lee (2014)**	**This study**
1	people or pronouns	kinship terms	people	people
2		people		pronouns
3	body parts	body parts	body parts	body parts
4	clothing	clothing	clothing	clothing
5	vehicles	vehicles	tools	vehicles
6	household items	tools		tools
		Stationery		
		household appliances		
		Tableware		
		cleaning supplies		
		medicine		
7	furniture and rooms	furniture		furniture
8	food and drink	food	food	food
		Fruits		
		Vegetables		
9	animals	animals	animals	animals
		insects		
10	outside things and natural phenomena	natural phenomena and natural materials	natural phenomena and natural materials	natural phenomena and natural materials
		plants		
11	toys	toys	Toys and musical instruments	toys
		musical instruments		
12	descriptive words	colors	colors	colors
13		shapes	shapes	shapes
14	places to go	places and architecture	names of places and architecture	locations
15	prepositions and locations			spatial words
16	demonstratives and quantifiers	numerals	numerals	numerals
17	words about time	abstract nouns	abstract nouns	abstract nouns

To explore the distribution of children's lexicon in semantic categories, the present study calculates the sum frequency of each category in each age group. It will provide a general scope about how children's speech organizes. Vocabulary which has an unidentifiable meaning or was not completed will be excluded from the analysis. For example, 寶 (bǎo, treasure) is to be 寶劍 (bǎojiàn, treasured sword) as shown in (3-18). The child did not complete it, so it was excluded.

(3-18) From file "HTC01_WUYS208_02.cha"
*EXA: 這 個 叫 什麼 名字 ？
zhè ge jiào shénme míngzì
What's the name of this?
*CHI: 這 有 劍 .
zhè yǒu jiàn
There is a sword.
*EXA: 蛤 ？
há
What?
*CHI: 寶 +...
bǎo
Treasured…
*CHI: 布袋戲 有 劍 寶劍 .
bùdàixì yǒu jiàn bǎojiàn
There is a sword, a treasured sword in the glove puppetry.
*EXA: 寶劍 .
bǎojiàn
A treasured sword.
*EXA: 阿 他 的 寶劍 呢 ？
ā tā de bǎojiàn ne
Ah, where is his treasured sword?

Another one is 呼呼 (hū hū) as seen in the following example (3-19). It had an unclear meaning, so it was also excluded from the analysis either.

(3-19) File "HTC01_JC302_11.cha": line 2469. Keyword: 呼呼
*EXA: 這 是 什麼 時候 買 的 ？
zhè shì shénme shíhòu mǎi de
When was this bought?
*CHI: 我 拿 那 個 來 .
wǒ ná nà ge lái
I take that here.
*EXA: 好 .

hǎo
OK.
*CHI: 我 拿 那 個 呼呼 .
wǒ ná nà ge hūhū
I take that huhu.
*CHI: 你 不 要 拿 喔 .
nǐ bù yào ná ō
Don't take it.
*EXA: 好 .
hǎo
OK.

3.3.2. Conceptual levels

Nouns were classified into three conceptual levels: superordinate level, basic level, and subordinate level. In this study, the categorization followed the principle provided by Rosch et al. (1976). The steps of classifying words into conceptual levels are:

(1) Superordinate nouns were the most abstract. Basic-level nouns were the words which could give a concrete image of an object. Subordinate nouns were the specific names or types of a basic-level category.

(2) "X is a kind of Y." If Y was a superordinate word, X was viewed as a basic-level noun. If Y was a basic-level word, X was viewed as a subordinate noun (Miller, 1990). For example, "dogs are a kind of animal," so dogs were classified as basic-level nouns. "Chihuahuas are a kind of dog," so Chihuahuas were classified as subordinate nouns.

(3) Nouns in the category of animals were not classified by the biological taxonomy. A noun was classified as a basic-level noun when the animal it represented was far from the prototype and people seldom used the general basic-level noun to name it. For example, 企鵝 (qì é, penguin) is a kind of 鳥 (niǎo, bird), but it is not a prototype of a bird. Thus, 企鵝 (qì é, penguin) was classified as a basic-level noun.

(4) Refer to the E-HowNet (CKIP Group, 2011), which is a lexical knowledge base consisting of definitions for lexical senses and an ontology, to find a suggested level when it was difficult to decide the conceptual level a noun belonged to.

Table 3.8 provides examples of conceptual levels. The total number of types of each conceptual level in each age group was calculated. Moreover, the total number of types of each conceptual level in each semantic category was also computed. Noun lists of semantic categories and conceptual levels in all age stage are provided in the Appendix 5.

Table 3.8: Examples in conceptual levels.

Conceptual level	**Example**		
Superordinate	工具	動物	玩具
	gōng jù	dòng wù	wán jù
	tool	animal	toy
Basic-level	筆	馬	球
	bǐ	Mǎ	Qiú
	pen	horse	ball
Subordinate	鉛筆	斑馬	籃球
	qiān bǐ	bān mǎ	lán qiú
	pencil	zebra	basketball
	鋼筆	太空馬	棒球
	gāng bǐ	tàikōng mǎ	bàng qiú
	fountain pen	space horse	baseball
	彩色筆		保齡球
	cǎi sè bǐ		bǎo líng qiú
	color pen		bowling ball

Chapter 4

The Profile of Children's Early Lexicon

4.1. Analysis of children's vocabulary growth

4.1.1. Vocabulary size in all stages

Children's vocabulary size is reported in this section. The frequency distribution was examined by two dimensions, POS and age stage. The following table provides the word frequency and proportion of main POS in the analyzed data: Noun, Verb, ADV, CL, QN, ADJ, CONJ, PREP, WH, and MOD. Other parts-of-speech were classified into the category called others, because they may not be able to reflect children's language development, such as idioms (IDM), sentence final particles (SFP), interjections (INT), and onomatopoeia words (ONM). POS which had fewer than 10 types was also classified into others, such as determiners (DT), aspect markers (ASP), nominal markers (DE), and adverb of negation (NEG).

Vocabulary size of each POS

Vocabulary sizes of all parts-of-speech are illustrated here. Table 4.1 presents the word frequency and proportion of each POS in the language samples. Table 4.1 shows that nouns and verbs are the main POS. In the aspect of word tokens, the database contains mainly nouns and verbs, about 58%, followed by classifiers 4.8% and adverbs 4.1%. Quantifiers, WH-words, and modals are about 5.2% of all tokens. The number of tokens is fewer than 1000 in ADJ, CONJ and PREP. Others are about 25% of all tokens, mainly from SFP, INT, DT, ASP, DE, and NEG. However, the type frequency reveals that 1892 noun types and 1444 verb types contribute about 83% among all word types. The third main POS is adverb, 109 different types. The remaining 7 parts-of-speech

contain less than 100 different types in each of them. The 7 word classes account for about 6.0%. The POS in the group of others accounts for 8.1% of all word types.

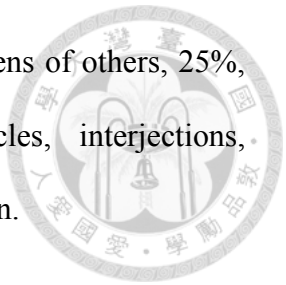
Table 4.1: The total number of main POS.

POS	Word Token	Percentage	Word Type	Percentage	TTR
Noun	20543	27.1%	1892	47.0%	0.092
Verb	23513	31.0%	1444	35.9%	0.061
ADV	3098	4.1%	109	2.7%	0.035
CL	3616	4.8%	64	1.6%	0.018
QN	1344	1.8%	60	1.5%	0.045
ADJ	554	0.7%	49	1.2%	0.088
CONJ	811	1.1%	28	0.6%	0.035
PREP	790	1.0%	21	0.4%	0.027
WH	1453	1.9%	17	0.4%	0.012
MOD	1121	1.5%	14	0.3%	0.012
Others	18896	24.9%	328	8.1%	0.017
SUM	**75739**	**100%**	**4026**	**100%**	**0.053**

The high TTR reveals that the vocabulary diversity is higher in nouns, 0.092, verbs, 0.061, and adjectives, 0.088. Referring to the number of types and tokens, the high vocabulary diversity means that children know many different words of nouns, verb and adjectives. On the contrary, the low TTR indicates that the diversity is lower in WH-words, 0.012, modals, 0.012, POS of others such as determiners, aspect markers, nominal markers, and adverb of negation, 0.017, and classifiers, 0.018. Referring to the number of types and tokens, the low diversity means that almost all types of WH-words, modals, determiners, aspect markers, nominal markers, and adverb of negation are frequently used in children's speech.

The vocabulary size provides some indication about 10 children's spontaneous speech from 1.5 to 4 years old. First of all, nouns and verbs are the major types of children's vocabulary, accounting for more than 80%. Children also frequently use nouns and verbs, accounting for about 60%. They also frequently used adverbs and

classifiers in their speech, about 9%. Besides, the proportion of tokens of others, 25%, indicated that children frequently used sentence final particles, interjections, determiners, aspect markers, nominal markers, and adverb of negation.



Word frequency and TTR in each age stage

The frequency distribution in different age stages was also examined and is presented in Table 4.2, including type frequency, token frequency and TTR in each age group. Data in 19-24 months contains a total number of 5657 word tokens; 19726 words in 25-30 months, 16669 words in 31-36 months, 23431 words in 37-42 months. Data in the 43-48 months contains 10256 word tokens. In terms of the number of word types in an age stage, 815 word types were collected in 19-24 months, 1774 word types in 25-30 months, 1668 word types in 31-36 months, 2028 word types in 37-42 months, and 1138 word types were collected in 43-48 months. Among 4026 word types in the language samples, data in 19-24 months accounts for 20.2%; data in 25-30 months and 31-36 months accounts for more than 40% words; data in 37-42 months even accounts for more than 50%. Data in 43-48 months accounts for 28.3% word types.

Table 4.2: The total word frequency and TTR in each age stage.

Frequency	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
Type	815	1774	1668	2028	1138
	20.2%	44.1%	41.4%	50.4%	28.3%
Token	5657	19726	16669	23431	10256
	7.5%	26.0%	22.0%	30.9%	13.5%
TTR	0.144	0.09	0.1	0.087	0.111

As for TTR, the ratio in 19-24 months is the highest, 0.144 which means vocabulary diversity is comparably higher in children's earlier stages when they have been acquiring new words in these stages. Thus, the possibility of children producing a

new word in their speech is higher than later age stages. The TTR value decreases to 0.09 and 0.1 in the following two age stages. Later on, the value in 37-42 months is the lowest, 0.087. The ratio increases to 0.111 in 43-48 months. The declining vocabulary diversity may indicate that children have got familiar with the words that they acquired in the earlier stages and begin productively using the vocabulary in their speech. The slight increasing TTR in 31-36 months and 43-48 months may imply that children learned some new words which they were able to understand or newly encountered in the two stages.

A figure of word frequency is plotted, and an M-shaped distribution is shown in Figure 4.1. The M-shaped distribution of word frequency may be influenced by the different sample size in the age groups. As the previous section has mentioned, the collected files are not equally distributed. Most files were from 25-30 months, 31-36 months, and 37-42 months. There are more tokens in an age group with more files, and vice versa, an age group with fewer files contains fewer word tokens. However, it is also possible that children do produce more words in the two stages, due to their development rather than sample sizes.

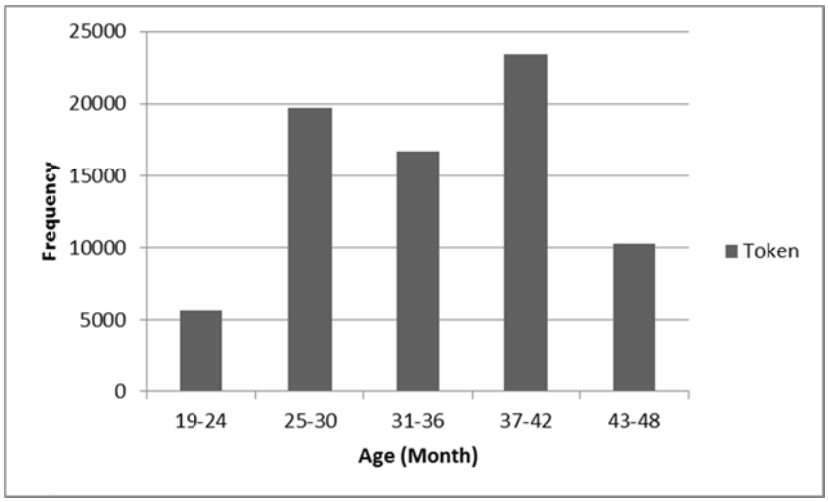


Figure 4.1: The distribution of total word token frequency.

The frequency distribution of each POS and the corresponding TTR in every age group are provided in Table 2, Table 3 and Table 4 in Appendix 3. Generally, the total frequency of POS in each age group is also an M-shaped distribution. But some parts-of-speech with increasing types and tokens rather than an M-shaped distribution may imply a development tendency with age, such as ADV, CL, QN, ADJ, CONJ, PREP, and WH.

Cumulative vocabulary

Cumulative vocabulary of each child and the mean vocabulary size in each age stage are provided in Table 4.3. As seen in Table 4.3, children's mean cumulative vocabulary is 290 in 19-24 months, 552 in 25-30 months, 741 in 31-36 months, 868 in 37-42 months, and 882 in 43-48 months. The rate of vocabulary growth is higher in earlier stages before 42 months. However, the rate of vocabulary growth slows down in 43-48 months. Figure 4.2 shows that the vocabulary growth before 42 months is a steep distribution, while the growth after 42 months is a flat distribution.

Table 4.3: Children's cumulative vocabulary.

Child	Age	Session	19-24 months	25-30 months	31-36 months	37-42 months	43-48 months
XU	1;7－2;5	10	284	474	—	—	—
YANG	1;7－2;9	11	222	444	623	—	—
WU	1;7－2;10	12	420	811	912	—	—
CHOU	2;1－3;4	16	—	772	1058	1369	—
WANG	2;5－3;4	12	—	472	871	1085	—
JC	2;2－3;5	14	—	444	604	951	—
CHENG	3;1－3;11	11	—	—	—	776	1146
CHW	3;6－4;0	6	—	—	—	352	646
Pan	1;7－3;9	19	233	444	694	916	982
WUYS	2;7－3;10	10	—	—	425	625	752
Mean			**290**	**552**	**741**	**868**	**882**

Note: The symbol "—" means no data was recorded.

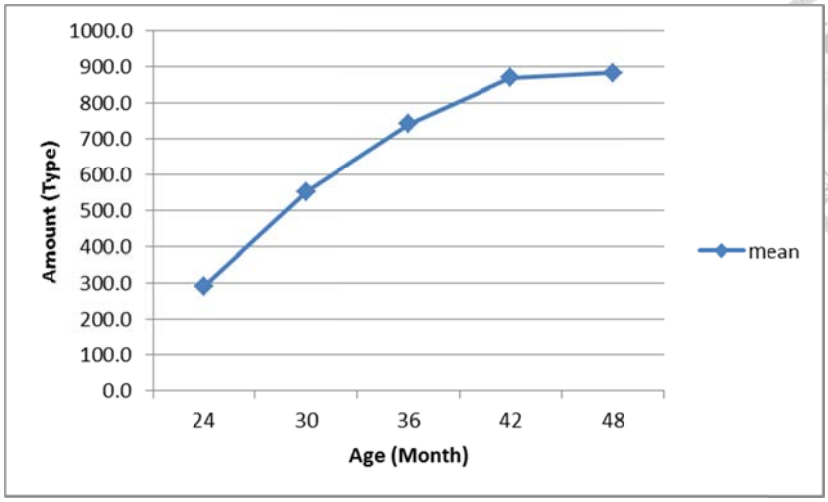


Figure 4.2: The mean of children's cumulative vocabulary.

4.1.2. POS frequency in all stages

POS frequency in different age stages is reported in this section. Table 4.4 is the average type frequency of POS in each age group. As seen in Table 4.4, the average total types increase from 89.3 to 209.6 word types before 37-42 months. The total word types decrease to a mean of 189.6 types in 43-48 months, but it is still the second highest word frequency. Since the word frequency was averaged, the decline may not result from a fewer sample size in this stage.

Table 4.4: The mean type frequency of POS in each age stage.

POS	19-24 months	25-30 months	31-36 months	37-42 months	43-48 months
Noun	32.3	56.4	50.8	67.7	59.1
Verb	32.3	64.8	65.1	76.2	67.0
ADV	2.4	7.1	9.8	13.5	11.3
CL	1.3	2.6	3.9	3.9	3.0
QN	1.6	2.4	3.5	5.2	4.4
ADJ	1.0	2.1	1.9	2.3	2.3
CONJ	0.4	1.7	2.9	3.8	3.6
PREP	0.7	2.3	2.8	3.8	3.2
WH	1.2	2.8	3.3	3.7	3.8
MOD	0.8	1.9	2.2	2.6	2.4
Others	15.3	22.4	22.8	26.9	29.5
Total	**89.3**	**166.4**	**168.8**	**209.6**	**189.6**

Examining the mean word types of each POS in five age stages, almost all POS show a similar increase pattern as the total word types: the word frequency is the lowest in 19-24 months, the word frequency keeps increasing stage by stage until 37-42 months, and the frequency declines in 43-48 months. Figure 4.3 presents the line graph of word frequency of POS in different age stages. The figure shows that the frequency distributions of nouns, verbs, and adjectives are an M-shaped distribution, while the distributions of other parts-of-speech are in a steep raising distribution before 42 months and a decreasing distribution after 42 months.

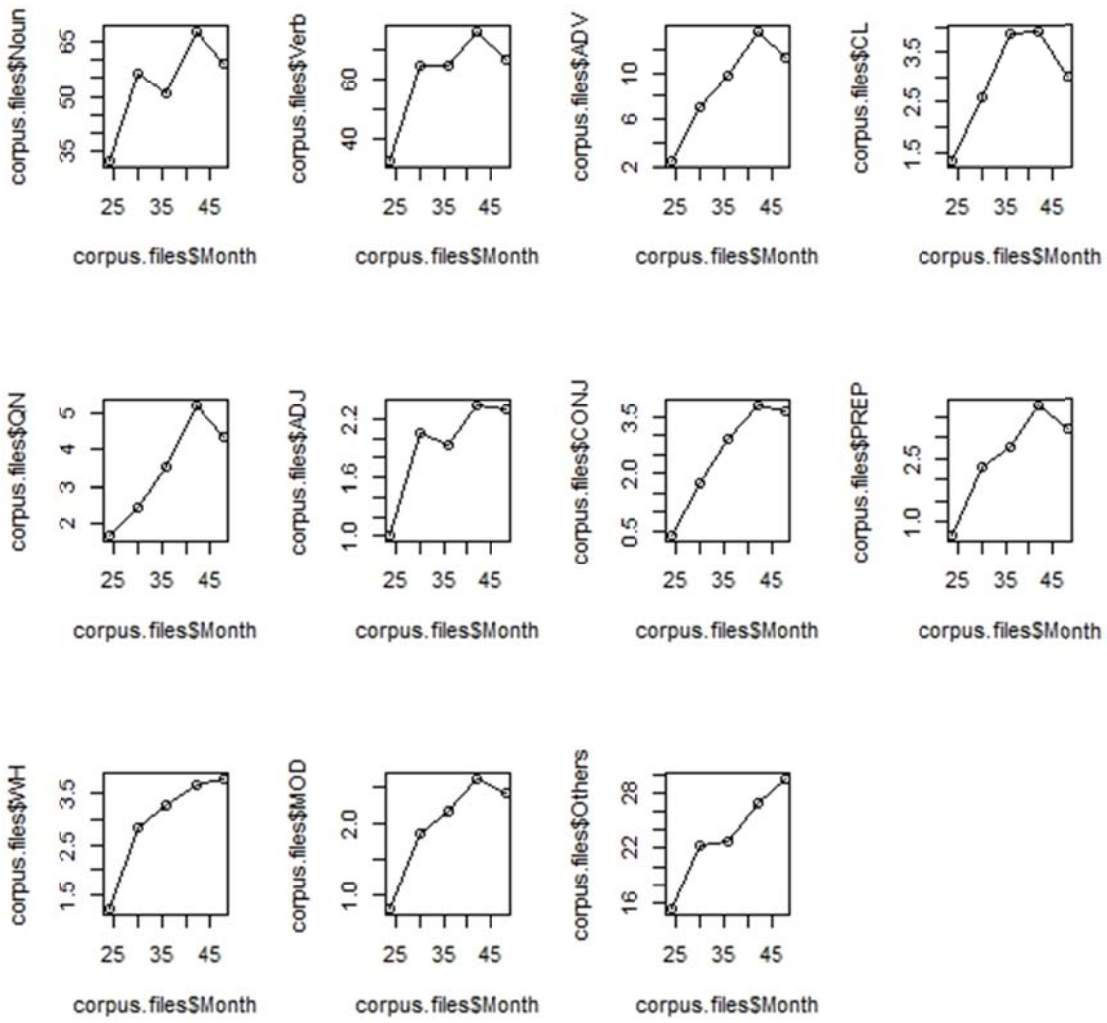


Figure 4.3: The line graph of word types in POS.

The number of word tokens of each POS was also computed. Table 4.5 is the average word token frequency of POS in age stages. As seen in Table 4.5, the mean total word tokens is 297.7 tokens in 19-24 months, 636.3 tokens in 25-30 months, 595.3 tokens in 37-42 months, and 732.6 tokens in 43-48 months. The frequency distribution is an M-shaped distribution. There was a tremendous increase in word tokens in 25-30 months and 37-42 months, but a decrease in 31-36 months and 43-48 months. Since the word frequency was an average value, the M-shaped pattern may not result from different sample sizes in the stages; instead, it may be children's real performance before they were 4 years old.

Table 4.5: The mean token frequency of POS in each stage.

POS	19-24 months	25-30 months	31-36 months	37-42 months	43-48 months
Noun	93.0	188.6	154.3	207.3	185.6
Verb	97.7	217.1	183.1	237.3	208.6
ADV	5.3	21.3	25.6	41.0	30.6
CL	11.7	24.3	31.9	39.9	42.1
QN	5.4	7.7	13.9	15.4	11.9
ADJ	1.8	5.9	3.9	5.8	4.1
CONJ	1.0	3.8	8.0	12.2	6.9
PREP	1.6	5.9	7.6	9.3	6.8
WH	4.4	10.9	13.4	14.3	17.4
MOD	1.1	9.1	8.0	14.3	12.6
Others	74.7	141.8	145.6	211.1	205.9
Total	**297.7**	**636.3**	**595.3**	**808.0**	**732.6**

As seen in the above table, children produced more nouns and verbs than other POS in all stages. They produced about a mean of 93 nouns and 97 verbs in 19-24 months, but they produced about more than 100 nouns and verbs in later stages. The frequency distributions of nouns, verbs, adjectives, and modals are also an M-shaped distribution. The declined token frequency in 31-36 months and 43-48 months may imply that children learn to use other methods to express their ideas, such as using more vocabulary in other word classes. This seems to be the case that the number of word

tokens keeps increasing in ADV, CL, QN, CONJ, PREP, WH, and others until at the 37-42 months. Figure 4.4 shows the line graph of word tokens of POS in each stage.

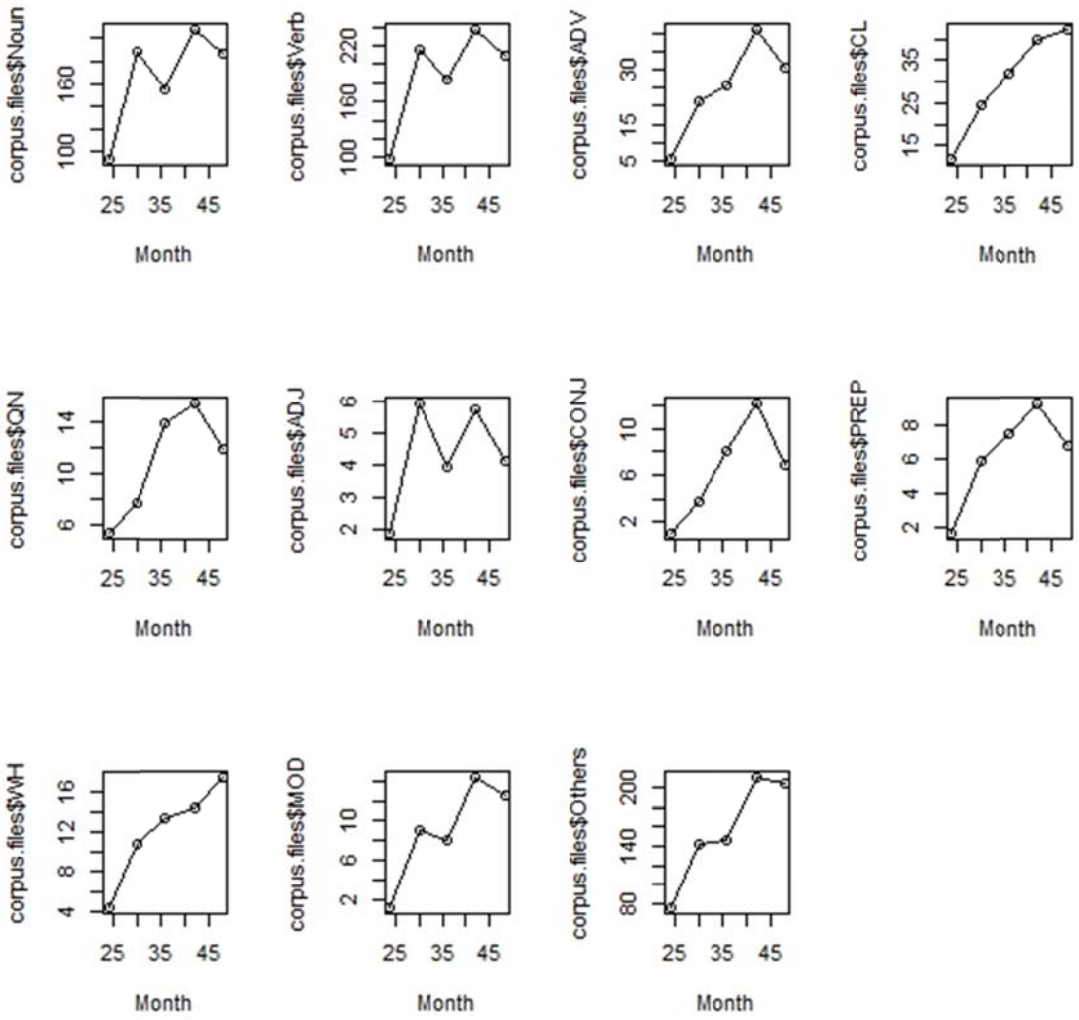
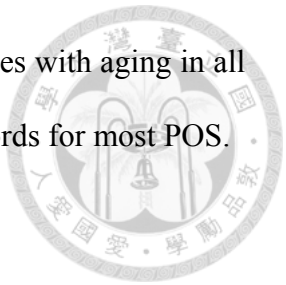


Figure 4.4: The line graph of word tokens in POS.

In summary, there are two stages showing a steep increase of vocabulary, 25-30 months and 37-42 months. The trend keeps steady or decreases in 31-36 months. From the aspect of ages, it shows that the number of nouns, verbs, adverbs and others outnumber the number of rest parts-of-speech in all stages. However, there are more classifier and quantifier tokens than adverbs. From the aspects of POS, it shows that the

fewest number of words are in 19-24 months, but the number increases with aging in all POS. Besides, words in 37-42 months have the highest number of words for most POS.



4.1.3. TTR in all stages

The diversity of children's vocabulary is reported in this section. If children acquire new types, then the TTR will increase; meanwhile, if children acquire no more new types and repeatedly use words they have known, then the TTR will decrease. The Table 4.6 below offers the mean TTR of POS in each age group.

Table 4.6: The mean TTR of each POS in each age stage.

	19-24 months	25-30 months	31-36 months	37-42 months	43-48 months
Noun	0.37	0.33	0.35	0.35	0.33
Verb	0.34	0.31	0.37	0.34	0.34
ADV	0.51	0.41	0.44	0.38	0.42
CL	0.22	0.21	0.15	0.1	0.11
QN	0.27	0.45	0.38	0.39	0.46
ADJ	0.36	0.41	0.48	0.51	0.49
CONJ	0.13	0.47	0.43	0.43	0.69
PREP	0.33	0.4	0.45	0.55	0.59
WH	0.32	0.51	0.45	0.33	0.39
MOD	0.47	0.37	0.37	0.24	0.25
Others	0.24	0.18	0.17	0.14	0.15
All POS	**0.37**	**0.33**	**0.35**	**0.35**	**0.33**

The highest TTR of a POS among five age stages were examined. Nouns, ADV, CL, MOD, and others have a higher TTR in the earliest age group, 19-24 months. The highest TTR of WH-words is in 25-30 months. Verbs have the highest TTR in 31-36 months. Adjectives have the highest TTR in 37-42 months. QN, CONJ, and PREP have the highest TTR in 43-48 months. The result of TTR shows that content words like nouns and verbs have a mid TTR ($0.3 < TTR < 0.4$). Function words and abstract content words (WH, CONJ, QN, PREP, ADV, ADJ, MOD) have a high TTR (> 0.4). Function words like classifiers and others have a low TTR (< 0.3).

Figure 4.5 is the line graph of the TTR of POS in each age stage. Nouns, verbs, and

others have a flat line, which represents that TTR do not change much in different age stages. Previous section reports that children acquire more types of nouns and verbs, and children frequently use them as time goes by. TTR in this section also proves that the TTRs of nouns and verbs have a flat trajectory. The declining TTRs of CL and MOD indicate that children produce some classifiers and modals more frequently in later stages.

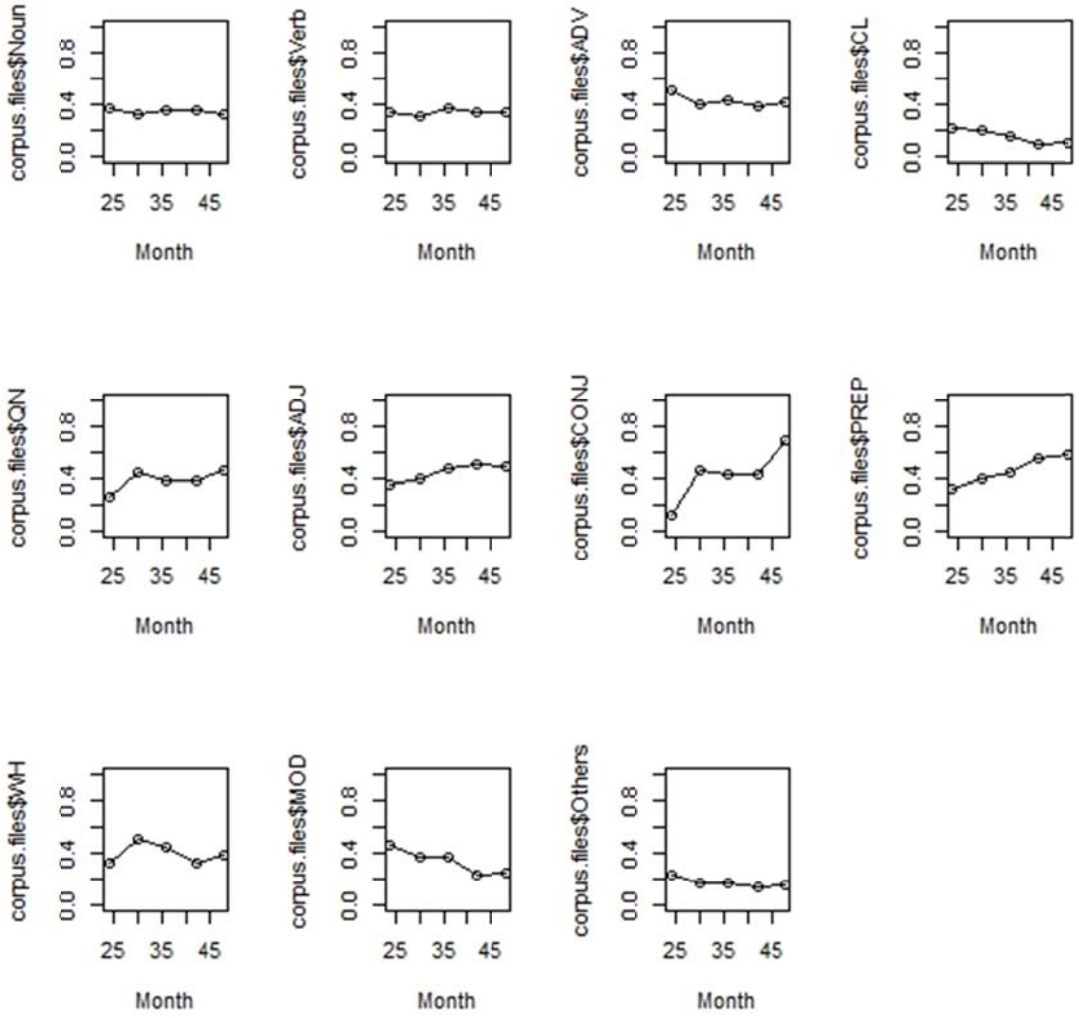


Figure 4.5: The line graph of TTR of POS.

4.1.4. Development of POS across month stages

The result of mean proportion of POS in each age stage is reported here. Table 4.7 and Table 4.8 provide the mean proportion of POS in each age stage. Wordlist of POS except for nouns and verbs are in Appendix 4.

Table 4.7: The mean proportion of POS types in each age stage.

	19-24 months	25-30 months	31-36 months	37-42 months	43-48 months
Noun	36.6%	34.2%	30.2%	32.1%	30.8%
Verb	35.8%	38.2%	38.1%	36.1%	35.6%
ADV	2.4%	4.0%	5.7%	6.5%	5.9%
CL	1.5%	1.5%	2.3%	1.8%	1.7%
QN	1.9%	1.5%	2.2%	2.6%	2.4%
ADJ	1.0%	1.1%	1.0%	1.1%	1.2%
CONJ	0.3%	0.9%	1.7%	1.8%	1.8%
PREP	0.7%	1.4%	1.6%	1.9%	1.7%
WH	1.3%	1.8%	2.0%	1.8%	2.1%
MOD	0.7%	1.2%	1.3%	1.3%	1.3%
Others	17.8%	14.1%	13.9%	13.1%	15.5%

Table 4.8: The mean proportion of POS tokens in each age stage.

	19-24 months	25-30 months	31-36 months	37-42 months	43-48 months
Noun	31.5%	29.6%	25.9%	25.5%	25.5%
Verb	32.8%	34.2%	30.6%	29.6%	28.8%
ADV	1.5%	2.9%	4.1%	5.0%	4.1%
CL	4.5%	3.3%	5.4%	5.0%	5.5%
QN	1.9%	1.1%	2.5%	2.0%	1.8%
ADJ	0.6%	0.9%	0.6%	0.6%	0.6%
CONJ	0.2%	0.5%	1.3%	1.4%	0.9%
PREP	0.5%	1.1%	1.2%	1.4%	1.0%
WH	1.3%	1.9%	2.4%	1.9%	2.5%
MOD	0.3%	1.3%	1.3%	1.9%	1.8%
Others	25.0%	23.1%	24.8%	25.8%	27.4%

Nouns: As seen in Table 4.7, children produced 36.6% noun types in the stage 19-24 months, 34.2% in 25-30 months, 30.2% in 31-36 months, 32.1% in 37-42 months and 30.8% in 43-48 months. The noun proportion decreases gradually before children's 36 months, later increases in the following six months before they were 42 months old,

and decreases again before 48 months old. In the aspect of tokens, a similar trend is shown. The proportion of noun tokens in the 19-24 months is the highest 31.5%, and it decreases in later stages, less than 30%. The proportion is about 25.5% after 31 months.

Verbs: Children produced 35.8% verb types in 19-24 months, 38.2% in 25-30 months, 38.1% in 31-36 months, 36.1% in 37-42 months, and 35.6% in 43-48 months. The verb proportion increases in 25-30 months, and it does not change much in 31-36 months. In later stages, the proportion of verb types decreases gradually. On the contrary, the token proportions of verb before 30 months, 32.8% and 34.2%, are higher than those in the later stages.

Adverbs: Children produced 2.4% types of adverb in 19-24 months, 4.0% in 25-30 months, 5.7% in 31-36 months, 6.5% in 37-42 months, and 5.9% in 43-48 months. The proportion of adverb increases gradually until children were 42 months old, and it decreases after their 42 months. The token proportion of ADV shows a similar trend as type proportion. Children produced 1.5% tokens of ADV, and gradually increased the proportion until 37-42 months, 5%, and then decreased the proportion in 43-48 months.

Classifiers: Children produced 1.5% types of classifier in 19-24 months and 25-30 months, 2.3% in 31-36 months, 1.8% in 37-42 months, and 1.7% in 43-48 months. The proportions of classifiers are the same in the first two stages. After children's 30 months, the proportion increases in 31-36 months, but it decreases gradually in the following two stages, after they were 36 months old. On the contrary, children produced more proportion of tokens than types of CL. They produced 4.5% CL in 19-24 months, 3.3% in 25-30 months, 5.4% in 31-36 months, 5% in 37-42 months, and 5.5% in 43-48 months. Although children produced fewer types of CL, they frequently used CL.

Quantifiers: Children produced 1.9% quantifier types in 19-24 months, 1.5% in 25-30 months, 2.2% in 31-36 months, 2.6% in 37-42 months, and 2.4% in 43-48 months.

The proportion of quantifier types decreases in 25-30 months, but it increases gradually before 42 months. The proportion decreases slightly again in 43-48 months. The proportion of quantifier tokens shows a similar trend, but the highest proportion is in 31-36 months, 2.5%. In the later stages, the proportion decreases to 1.8%.

Adjectives: Children produced 1.0% adjective types in 19-24 months, 1.1% in 25-30 months, 1.0% in 31-36 months, 1.1% 37-42 months, and 1.2% in 43-48 months. The proportion of adjectives does not change much in five age stages. The token proportions of adjectives are lower than 1%. The highest token proportion is in 25-30 months, and those in the rest age stages have no difference.

Conjunctions: Children produced 0.3% conjunction types in 19-24 months, 0.9% in 25-30 months, 1.7% in 31-36 months, 1.8% in 37-42 months, and 1.8% in 43-48 months. The proportion of conjunctions increases gradually until 42 months, and it remains unchanged in 43-48 months. A similar trend is found in token proportion of conjunctions. Children produced only 0.2% conjunction tokens in 19-24 months, and became more frequently use conjunctions with increasing ages until 37-42 month, 1.4%. But the token proportion of conjunction decreases again in 43-48 months.

Prepositions: Children produced 0.7% preposition types in 19-24 months, 1.4% in 25-30 months, 1.6% in 31-36 months, 1.9% in 37-42 months, and 1.7% in 43-48 months. The proportion of prepositions is the smallest in the first stage, and it increases stage by stage until 37-42 months. After 42 months, the proportion is decreased again. As for token proportion of prepositions, a similar trend is found. Children produced only 0.5% of preposition tokens in 19-24 months, and became more frequently use prepositions with increasing ages until 37-42 month, 1.4%. But the token proportion of conjunction decreases again in 43-48 months.

WH-words: Children produced 1.3% WH-word types in 19-24 months, 1.8% in

25-30 months, 2.0% in 31-36 months, 1.8% in 37-42 months, and 2.1% in 43-48 months. The proportion of WH-words is the smallest in the first age stage, 19-24 months, and it increases in the next two age stage before 36 months. After 36 months, the proportion is reduced slightly in 34-42 months, but it increases again in 43-48 months, in which the proportion of WH-words is the highest. A similar trend of token proportion in WH-words is found. The proportion is lowest in 19-24 months, 1.3%, and it increases to 2.4 in 31-36 months. In the next age stages, the proportion of WH-tokens decreases in 37-42 months, and then increases again in 43-48 months.

Modals: Children produced 0.7% modal types in 19-24 months, 1.2% in 25-30 months, and 1.3% in 31-36 months, 37-42 months and 43-48 months. The proportion of modals is the smallest in the first age stage, 19-24 months, and it increases in the following three stages before 36 months. After 36 months, the proportion of modals keeps unchanged in the following stages. In terms of proportion of modal tokens, children produced fewer modals in 19-24 months, 0.3%. However, in later stages, the proportion increases to 1.3% and 1.9% in 37-42 months. It decreases slightly to 1.8% in 43-48 months.

Others: Other POS types accounts for 17.8% in 19-24 months, 14.1% in 25-30 months, 13.9% in 31-36 months, 13.1% in 37-42 months, and 15.5% in 43-48 months. The proportion of other POS types is the highest in the first age stage, 19-24 months, and it decreases in the next three stages before 42 months. The proportion increases again in 43-48 months, 15.5%. On the contrary, the proportion of tokens shows a different trend. The proportion is 25% in 19-24 months, and it falls to 23.1%. After 30 months, children increase the tokens gradually. The highest proportion is 27.4% in 43-48 months.

Based on the results above, children's vocabulary development trend was observed.

Therefore, proportions of word frequency can be used to present the development trend of children's early Mandarin vocabulary. In early stage, 19-24 months, nouns and verbs have the highest proportion of children's vocabulary. After children's 30 months, the proportion of nouns and verbs will decrease. The proportion of 8 POS (ADV, CL, QN, ADJ, CONJ, PREP, WH, and MOD) is a small proportion but it will increase after children's 2 years old. The proportion of other POS is another story. The proportion of types decreases after 24 months, but the proportion of tokens decreases in 25-30 months first and it increases after 30 months. The results of children's POS distribution indicates that children acquire content words first and frequently use them in early ages, whereas they acquire function words later, and increase the use of function words when they are older.

4.1.5. N/V ratio in all stages

Noun/verb ratio in each age stage is reported in this section. Table 4.9 provides the mean NVR of four types of noun/verb ratio in each age stage. As shown in Table 4.9, the mean NVR1 of 19-24 months showed the greatest value (M = 1.06, SD = 0.31), indicating a noun bias. NVR1 in the following stages implies a verb bias. The standard deviation shows that the biggest within group difference is 0.31 in 19-24 months. The later stages show a fluctuation in within group difference. It seems that children's individual difference is not steady.

NVR2 indicates a noun bias in all stages. The values are 1.4 in 19-24 months, 1.25 in 25-30 months, 1.02 in 31-36 months, 1.19 in 37-42 months, and 1.16 in 43-48 months. The lowest value reveals nearly no bias. The standard deviation shows the biggest within group difference is 0.45 in 19-24 months. The difference decreases to 0.27 gradually until 37-42 months. The difference increases to 0.35 again in 43-48 months. It

reveals that the children's individual difference gets smaller with increasing age.

NVR3 indicates a verb bias throughout five age stages. The values are 0.84 in 19-24 months, 0.68 in 25-30 months, 0.59 in 31-36 months, 0.67 in 37-42 months, and 0.64 in 43-48 months. The NVR3 in 19-24 months is the highest, indicating a weak verb bias, while the lowest value in 31-36 months indicates a strong verb bias. The standard deviation shows that the biggest within group difference is 0.30 in 19-24 months. In the later stages, it shows a fluctuant individual difference.

NVR4 has a similar tendency as NVR1. The NVR4 indicates a noun bias in 19-24 months and a verb bias in later age stages. The standard deviation shows that the biggest within group difference is 0.40 in 19-24 months and the difference decrease to 0.22 gradually until 37-42 months. The difference increases to 0.34 again in 43-48 months. It seems that the children's individual difference gets smaller as time goes by.

Table 4.9: Children's NVR.

		19-24 months	25-30 months	31-36 months	37-42 months	43-48 months
NVR1	Mean	1.06	0.92	0.82	0.91	0.89
	SD	0.31	0.22	0.27	0.21	0.26
NVR2	Mean	1.4	1.25	1.02	1.19	1.16
	SD	0.45	0.42	0.31	0.27	0.35
NVR3	Mean	0.84	0.68	0.59	0.67	0.64
	SD	0.30	0.19	0.23	0.16	0.25
NVR4	Mean	1.1	0.92	0.74	0.88	0.83
	SD	0.40	0.33	0.27	0.22	0.34

Figure 4.6 is the line graph of all NVR across five age stages. It depicts that four types of NVRs have a similar development trend. The NVRs are the highest in the early stages. 31-36 months is the turning point. NVRs decrease sharply before 31-36 months, and the ratios rise after then. As seen in the figure, NVR1 and NVR4 are extremely close which implies that including broad or strict of nouns and verbs do not have much difference. A weak noun bias is found in 19-24 months, and a verb bias is found in later

stages. NVR 2 which includes broad nouns and strict verbs shows a noun bias throughout all stages. On the contrary, NVR3 which includes strict nouns and broad verbs reveals a verb bias throughout all stages. All types of NVRs can provide a trajectory of children's development of using nouns and verb, but the inclusion of broad or strict nouns and verbs leads to different conclusions about noun bias or verb bias.

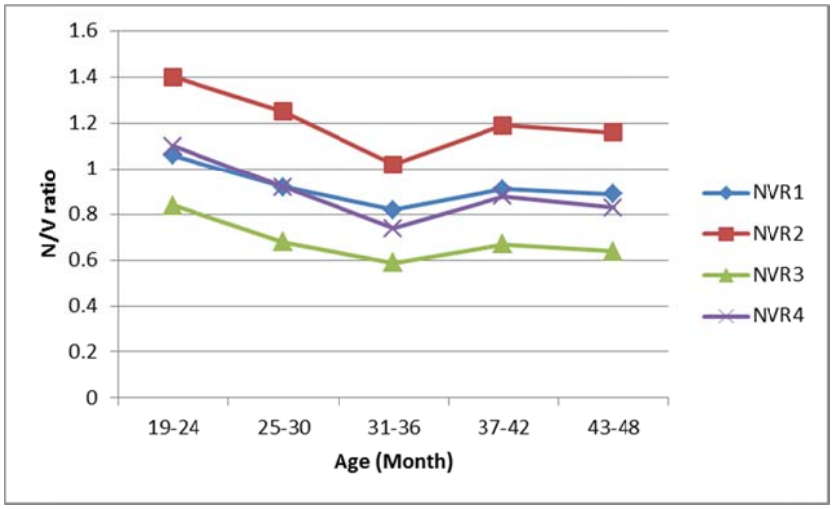


Figure 4.6: The development trend of Noun/Verb ratio.

4.1.6. D value in all stages

The following section will identify and discuss the trends of D values of children's all words and their nouns and verbs. Table 4.10 provides the result of children's D value. As seen in Table 4.10, D values of all words increase from 31.24 in 19-24 months gradually to the highest point 52.14 in 37-42 months, and it falls in 43-48 months. D value of nouns, broad verbs (D.verb.1), and strict verbs (D.verb.2) also reveal the same tendency, except for D value of nouns in 31-36 months shows a slightly decline. D values indicate that more lexical diversity of all words than diversity of nouns and verbs. Standard deviation shows that children's individual difference is fluctuated. Generally, the individual differences seem to be bigger in D value of all words than D value of nouns or verbs.

Table 4.10: Children's D values.

D		19-24 months	25-30 months	31-36 months	37-42 months	43-48 months
D.all	mean	31.24	44.85	46.16	52.14	45.4
	sd	13.25	13.56	14.18	10.85	10.18
D.noun	mean	12.31	16.73	16.5	19.04	15.51
	sd	7.5	7.42	9.2	7.88	8.82
D.verb.1	mean	9.93	17.69	22.56	25.14	21.06
	sd	5.41	9.4	9.29	9.93	4.77
D.verb.2	mean	7.35	12.66	17.15	18.93	16.19
	sd	4.34	6.21	6.77	7.43	4.27

Note: D.verb.1 is D of broad verbs and D.verb.2 is D of strict verbs.

Figure 4.7 depicts the development trend of D values. D.verb.1 is **D** of broad verbs and D.verb.2 is D of strict verbs. It reveals that all the D values have a rising tendency until 37-42 months and it declines after then. Besides, D values of nouns and D values of verbs are similar and strikingly close to one another. D values of broad verb and of strict verb are parallel; meanwhile D values of nouns overlaps D values of strict verb after 31-36 months. It indicates that lexical diversities of nouns and strict verbs are similar.

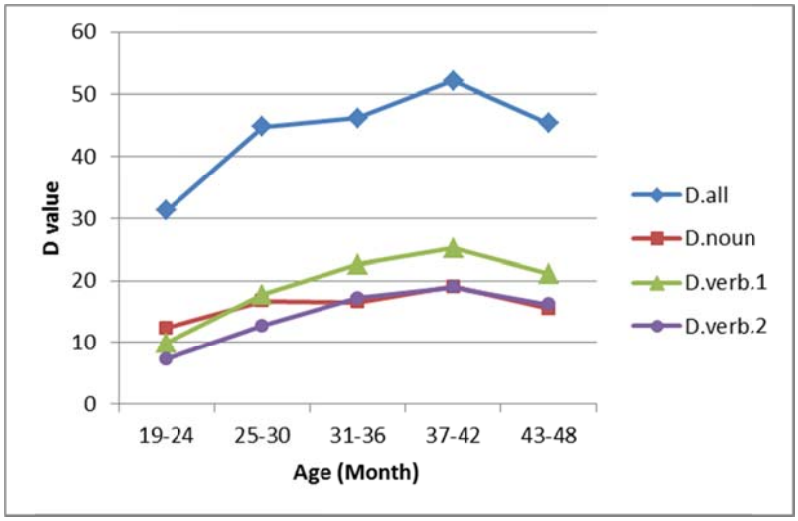


Figure 4.7: The development trend of D value.

In order to know whether age difference has effect on lexical diversity, a series of one-way ANOVA was conducted to D of all words, D of nouns, and D of verbs, with age group as the independent variable. The result of ANOVA of D of all lexicon shows that a significant difference between different age groups is found [$F(4,116) = 7.905$, $p < .05$]. The Scheffe post hoc test reveals that the mean D value of all lexicon in 19-24 months is significantly lower than that in other age groups ($p < .05$). There is no significant difference among the rest four age groups. As for D values of nouns, no significant difference between age groups was found.

When it comes to D values of broad verbs, the test of homogeneity of variances shows a significant difference [$F(4,116) = 3.115$, $p < .05$], which implies that equal variances were not assumed, so the Brown-Forsythe test was used. The result shows a significant difference between different age groups is found [$F(4, 110.736) = 12.115$, $p < .001$]. A Dunnett's T3 post hoc test implies that the mean D value of broad verbs in 19-24 months is significantly lower than that in other age groups ($p < .05$). The mean D value in 25-30 months is significantly lower than that in 37-42 months ($p < .05$). No significant difference of other groups was found.

The result of ANOVA of D of strict verb reveals that a significant difference between different age groups is found [$F(4,116) = 12.017$, $p < .05$]. The Scheffe post hoc test shows that the mean D values of strict verb in 19-24 months and 25-30 months are significantly lower than that in other groups. There is no significant difference between 19-24 months and 25-30 months, and no significant difference among the rest age groups either.

To sum up, D values of all words, of nouns, and of verbs all indicate the development trajectory of children's increasing lexical diversity as time goes by. A statistically significant age effect on lexical diversity is found in D values of all words

and of verbs. D values in 19-24 months and 25-30 months are significantly lower than D values in later stages. No significant difference is found between other stages.

4.2. Analysis of children's vocabulary organization

4.2.1. Distribution of semantic categories and conceptual levels

The distribution of semantic categories and conceptual levels of children's nouns will be illustrated in the following sections. The unidentifiable and incomplete words (64 types, 115 tokens) were excluded in the analyses. Table 4.11 provides the information of the total numbers and proportions of types and tokens in semantic categories.

As seen in Table 4.11, the top five main categories of noun types are people, tools, food, animals, and locations. These five categories account for 61.4% of children's all vocabulary in the samples. On the contrary, the numbers of types fewer than 20 are pronouns, furniture, numerals, and shapes, accounting for only 3.4% of children's vocabulary. The results may indicate that people, tools, food, animals, and locations appear the most frequently in children's early life, so children have more chance to acquire nouns of these categories. On the other hand, pronouns, furniture, numerals, and shapes are categories with fewer types themselves, so children acquire fewer words of these categories than categories with rich types.

The number of tokens was calculated to know how frequently children used the words in their speech. As shown in the Table 4.11, the top five frequently-used nouns are from categories of pronouns, people, animals, spatial words, and tools. These five categories account for 74.8% of all word tokens. Pronoun is the most frequently used category in children's speech, accounting for 29.5%. Although there are a few pronoun types, children use them the most frequently in their speech.

Table 4.11: The total numbers of nouns in semantic categories.

	People	Spatial	Pronouns	Tools	Vehicles
type	386	90	18	239	120
%	21.5%	5.0%	1.0%	13.3%	6.7%
token	4942	1367	6020	1086	863
%	24.2%	6.7%	29.5%	5.3%	4.2%
	Locations	**Natural**	**Clothing**	**Shapes**	**Body parts**
type	147	62	54	11	71
%	8.2%	3.5%	3.0%	0.6%	4.0%
token	644	351	285	24	663
%	3.2%	1.7%	1.4%	0.1%	3.2%
	Abstract nouns	**Toys**	**Food**	**Animals**	**Furniture**
type	129	82	172	158	17
%	7.2%	4.6%	9.6%	8.8%	0.9%
token	526	462	918	1871	106
%	2.6%	2.3%	4.5%	9.2%	0.5%
	Numerals	**Colors**	**SUM**		
type	14	25	1795		
%	0.8%	1.4%	100%		
token	19	281	20428		
%	0.1%	1.4%	100%		

Spatial words like 這裡 (zhè lǐ, here, 533 tokens) and 這邊 (zhè biān, here, 205 tokens) have a high frequency in children's speech, resulting in the high frequency of spatial words. 這裡 (zhè lǐ, here) and 這邊 (zhè biān, here) can be used to replace body parts or locations, as illustrated in the following examples. 這邊 (zhè biān, here) in example (4-1) can be 手 (shǒu, hand) or any body parts. In example (4-2) it can be 天空 (tiān kōng, sky). 這裡 (zhè lǐ, here) in example (4-3) may be 車庫 (chē kù, garage), and in example (4-4) it can be 牆壁 (qiáng bì, wall) or 桌子 (zhuō zi, table). The three categories with fewest tokens are furniture, shapes and numerals, the same trend as types, accounting for only 0.2%.

(4-1) From file "HTC01_CHENG303_03.cha"
*CHI: 我 這邊 癢 啦！

wǒ zhèbiān yang la
I have an itch here.

(4-2) From file "HTC01_Pan207_09.cha"

*CHI: 月亮 為什麼 在 這邊 呢？

yuèliàng wèishénme zài zhèbiān ne

Why is the moon here?

(4-3) From file "HTC01_CHOU204_04.cha"

*CHI: 在 這裡 洗 車 .

zài zhèlǐ xǐ chē

Wash cars here.

(4-4) From file "HTC01_CHOU210_10.cha"

*CHI: J 阿姨 可以 扶 這裡 .

J āyí kěyǐ fú zhèlǐ

Auntie J, you can put your hands here for support.

As for the numbers of words of conceptual levels, most of the noun types are basic-level words, accounting for 50.2 % of children's lexicon. Subordinate words account for 47.7%, and superordinate words is the fewest, only 2.1%. However, the number of noun tokens tells a different story. Children use basic-level words extremely frequently. Approximately 80.3% nouns are basic-level noun tokens, 17.4% of subordinate nouns, and 2.3% of superordinate words. It indicates that basic-level words are dominant in children's speech. Table 4.12 lists the proportions of types and tokens in conceptual levels.

Table 4.12: The total number of nouns in conceptual levels.

	basic	**subordinate**	**superordinate**	**SUM**
token	16406	3552	470	20428
%	80.3%	17.4%	2.3%	100%
type	901	857	37	1795
%	50.2%	47.7%	2.1%	100%

4.2.2. Semantic categories in different month stages

The following paragraph will present the results of semantic categories in each age stage. Table 4.13 is the top five semantic categories in each age stage, with the numbers

of word types and proportions provide. A complete table is Table 4 in the Appendix 3.

Table 4.13: Five semantic categories with the most noun types in each age stage.

19-24 months	25-30 months	31-36 months	37-42 months	43-48 months
People	People	People	People	People
81	146	128	154	111
23.1%	19.4%	19.2%	17.8%	23.6%
Tools	Tools	Tools	Tools	Animals
56	104	78	118	60
16.0%	13.8%	11.7%	13.6%	12.7%
Animals	Food	Food	Animals	Tools
45	91	77	80	45
12.8%	12.1%	11.5%	9.2%	9.6%
Food	Animals	Animals	Food	Vehicles
30	91	62	76	43
8.5%	12.1%	9.3%	8.8%	9.1%
Vehicles	Locations	Locations	Locations	Abstract Noun
20	52	56	67	32
5.7%	6.9%	8.4%	7.7%	6.8%

In general, people, tools, animals, and food are the dominate categories in each age. More vehicle nouns appear in children's lexicon in 19-24 months. Locations appear very often from 25 months to 42 months. What is more, the category of abstract nouns becomes a main category in 43-48 months. In 43-48 months, nouns of food are not produced very frequently anymore. When it comes to the fewest nouns, the categories of numerals, shapes, furniture, and colors appears to be fewer in children's speech. Children have few pronouns in 37-42 months and few natural phenomena nouns in 43-48 months.

The proportion of noun types of semantics categories in each age stage was examined. The trend of the proportion changes has 9 patterns, as shown in Table 4.14. Generally, a continued rising or falling may be a steady tendency of development. A decreasing proportion is found in people and animals until 42 months. The proportion of food rises in 25-30 months, but it keeps falling until 48 months. Meanwhile, an

increasing proportion is found in clothing and abstract nouns until 42 months, but it decreases in 43-48 months. The proportion of spatial words keeps rising across all age stages. Other categories have a fluctuated trend of rising and declining.

Table 4.14: The patterns of type proportion changes of semantic categories.

Semantic category	The patterns of noun type proportions			
	25-30 months	31-36 months	37-42 months	43-48 months
People, Animals	−	−	−	+
Tools, Toys	−	−	+	−
Natural, Furniture, Numerals	−	+	+	−
Pronouns, Vehicles	−	+	−	+
Food	+	−	−	−
Body parts, Colors	+	−	+	+
Locations, Shapes	+	+	−	−
Clothing, Abstract nouns	+	+	+	−
Spatial words	+	+	+	+

The number of nouns tokens in semantic categories was examined as well, and the five most frequently used semantic categories are provided in Table 4.15. A complete table is Table 5 in the Appendix 3. As shown in Table 4.15, people, pronouns, animals, spatial words, and tools are the most frequently used semantic categories in children's speech. Children did not use spatial words very often in 19-24 months; they produced vehicles nouns instead. As for the least frequently used categories, numerals, shapes, furniture, and colors have the lowest frequency in children's speech. The category of natural phenomena has a low frequency in 43-48 months, instead of colors.

The proportions of noun tokens in semantic categories were compared and patterns of increasing and decreasing proportions were found. Table 4.16 lists the 11 patterns of proportion changes. In terms of a falling trajectory, the use of people and food keeps declining across all stages, and the use of animals also keeps dropping until 42 months.

On the contrary, the use of pronouns keeps increasing in all stages, and the use of shapes and abstract nouns remains rising until 37-42 months. Other categories have an unsteady proportion changes.

Table 4.15: The five most frequently used semantic categories in each age stage.

19-24 months	25-30 months	31-36 months	37-42 months	43-48 months
People	**People**	**Pronouns**	**Pronouns**	**Pronouns**
629	1788	1235	2169	1158
36.3%	30.7%	28.7%	36.2%	44.7%
Pronouns	**Pronouns**	**People**	**People**	**People**
239	1219	1057	1101	367
13.8%	20.9%	24.6%	18.4%	.2%
Animals	**Animals**	**Animals**	**Animals**	**Animals**
196	611	382	435	247
11.3%	10.5%	8.9%	7.3%	9.5%
Tools	**Spatial Words**	**Spatial Words**	**Spatial Words**	**Spatial Words**
116	426	302	422	135
6.7%	7.3%	7.0%	7.0%	5.2%
Vehicles	**Tools**	**Food**	**Tools**	**Tools**
99	345	211	326	108
5.7%	5.9%	4.9%	5.4%	4.2%

Table 4.16: The patterns of token proportion changes of semantic categories.

Semantic category	The patterns of noun token proportions			
	25-30 months	31-36 months	37-42 months	43-48 months
People, Food	−	−	−	−
Animals	−	−	−	+
Tools, Naturals	−	−	+	−
Vehicles, Toys	−	+	−	−
Numerals	−	+	+	−
Furniture	−	+	+	+
Spatial words, Colors	+	−	+	−
Body parts	+	−	+	+
Locations, Clothing	+	+	−	−
Shapes, Abstract nouns	+	+	+	−
Pronouns	+	+	+	+

4.2.3. Conceptual levels in different month stages

The following section reports the number of nouns in conceptual levels in all stages. Table 4.17 presents the numbers of noun types in conceptual levels. As shown in Table 4.17, basic-level word dominates children's noun types in all stages. The number of basic-level words ranges from 239 to 522. The number of subordinate word ranges from 102 to 323. The number of superordinate word is the fewest, ranging from 10 to 22. The highest proportions of three levels are 68.1% of basic-level words in 19-24 months, 39.3% of subordinate words in 43-48 months, and 3.1% superordinate words in 31-36 months.

Table 4.17: The numbers of noun types of conceptual levels in each age stage.

		19-24 months	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
Superordinate	Type	10	13	21	22	14
	%	2.8%	1.7%	3.1%	2.5%	3.0%
Basic	Type	239	448	427	522	272
	%	68.1%	59.6%	63.9%	60.2%	57.7%
Subordinate	Type	102	291	220	323	185
	%	29.1%	38.7%	32.9%	37.3%	39.3%
SUM	Type	351	752	668	867	471

Generally, there are more basic-level words (from 57.7% to 68.1%) but fewer subordinate (from 29.1% to 39.3%) and superordinate words (from 1.7% to 3.1%) in earlier age stages, whereas the number of subordinate and superordinate words increases in later age stages. Besides, the proportion change indicates that children acquire more subordinate words than basic-level and superordinate words in 25-30 months. They acquire more basic-level and superordinate words than subordinate words in 31-36 months. After 36 months, children increase the proportion of subordinate words and reduce the proportion of basic-level words.

Figure 4.8 below depicts the proportion changes in all stages. It reveals that when the proportion of basic-level words decreases, the proportion of subordinate words increases, and vice versa. Besides, although the number of noun types of superordinate category increases in some stages, the proportion seems not to differ much.

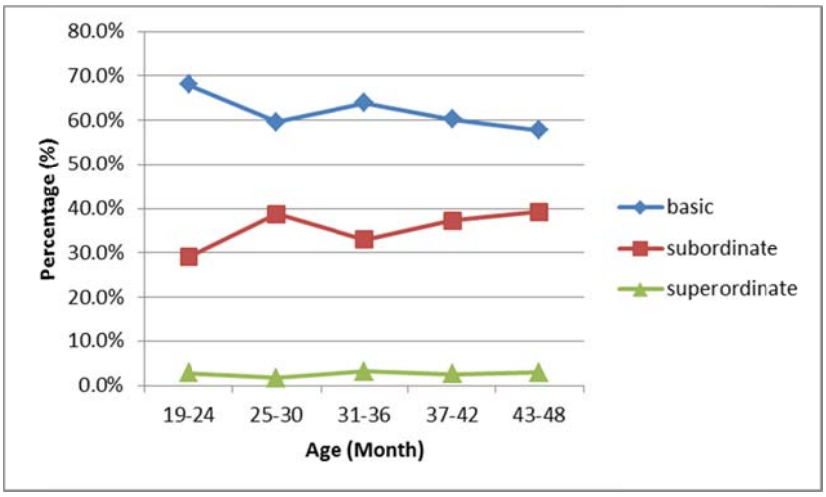


Figure 4.8: Noun type proportion changes of conceptual levels.

In addition to noun types in conceptual levels, noun tokens in conceptual levels were examined as well. Table 4.18 shows the numbers and proportions of noun tokens in conceptual levels. As shown in Table 4.18, basic-level word dominates children's noun tokens in all stages. The number of basic-level words accounts for 77.9% to 83%. The number of subordinate word accounts for 15.1% to 20.2% in the five stages. The number of superordinate word is the fewest, less than 150, accounting for only 1.8% to 3.3%.

In general, basic-level words account about 80% noun tokens in children's speech in all stages. Besides, the proportion reveals that children use more subordinate words in 25-30 months than other stages, but the proportion decreases in later stages. The number and proportion of superordinate words increase before 31-36 months, but the proportion fluctuates after 36 months.

Table 4.18: The numbers of noun tokens of conceptual levels in each age stage.

		19-24 months	25-30 months	31-36 months	37-42 months	43-48 months
Superordinate	Token	31	112	142	116	69
	%	1.8%	1.9%	3.3%	1.9%	2.7%
Basic	Token	1388	4532	3389	4969	2128
	%	80.0%	77.9%	78.9%	83.0%	82.2%
Subordinate	Token	316	1175	766	904	391
	%	18.2%	20.2%	17.8%	15.1%	15.1%
SUM	Token	1735	5819	4297	5989	2588

Figure 4.9 below depicts the proportion changes in all stages. It reveals that when the proportion of basic-level words decreases, the proportion of subordinate words increases, and vice versa. Besides, although the number of noun tokens of three levels changes in some stages, the figure shows that the changes seem not to differ much, except for subordinate words. Compared with the results of noun types on three conceptual levels, the development trend indicates that although the proportion of basic-level noun types decreases with increasing age, children keep using basic-level words frequently and even a little bit more frequently in later stages. Meanwhile, the proportion of subordinate noun types increases as time goes by, but children reduce the frequency of using subordinate words gradually.

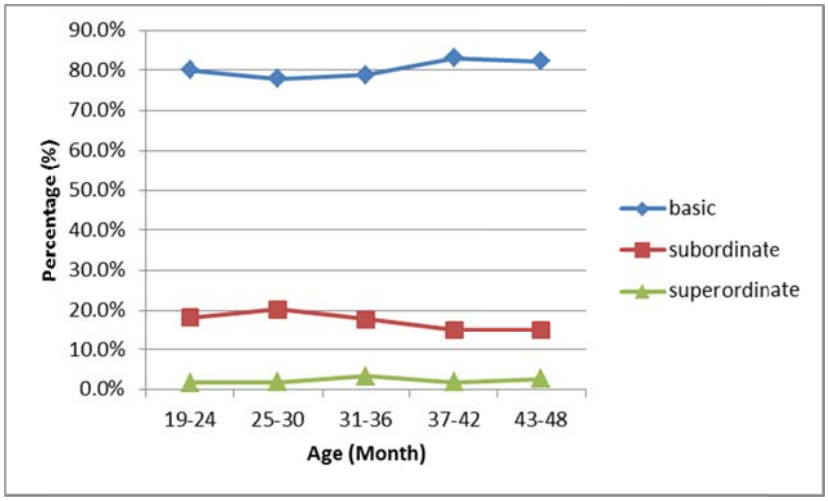


Figure 4.9: Noun token proportion changes of conceptual levels.

4.2.4. Development of conceptual levels of semantic categories

The development of conceptual levels in each semantic category in all stages is discussed in this section. The numbers of noun types and tokens of three conceptual levels in semantic categories are provided in Table 6 and Table 7 in Appendix 3. Since the sample sizes in stages are different, the raw frequency of types and tokens may not reflect the development trend well. Thus, raw frequency was transferred into percentage.

People:

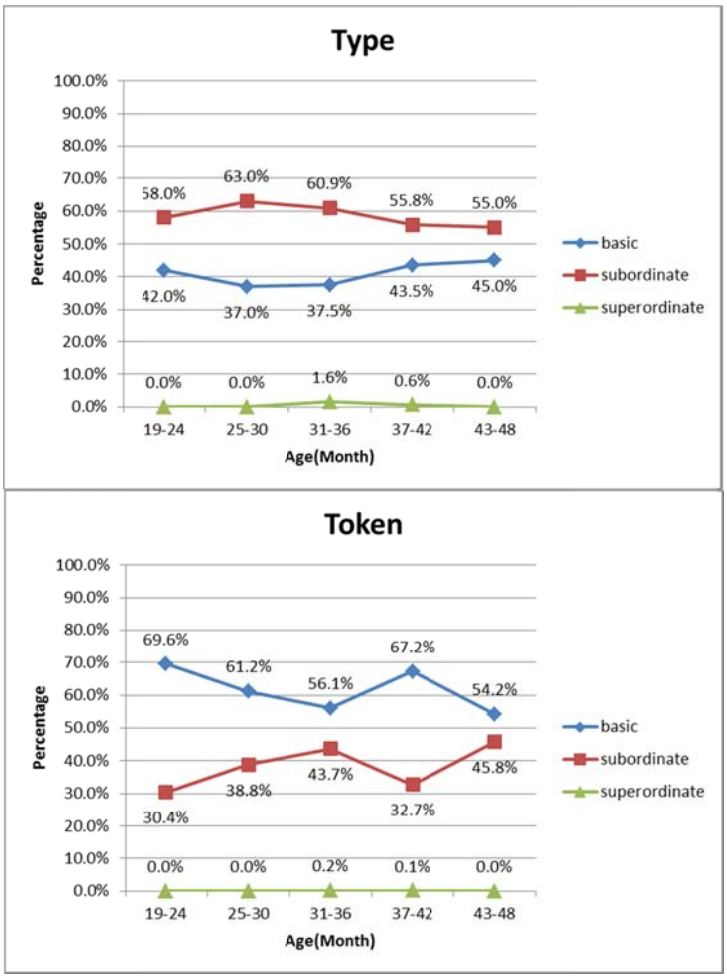


Figure 4.10: The proportion of conceptual levels in the category of people.

Figure 4.10 depicts the proportion change of three conceptual levels. Both

basic-level words and subordinate words appear in 19-24 months, and there are more subordinate words than basic-level words in all stages. The percentages show that more than 50% of nouns in the category of people are subordinate words. The figure shows that types of subordinate nouns of people decrease gradually, while basic-level nouns increases. Superordinate words in the category of people appear in 31-36 months and 37-42 months, and there are only 3 superordinate nouns.

Besides, the proportions of tokens of conceptual levels indicate that children use basic-level nouns of the category people much more frequently than subordinate nouns, although there are more subordinate nouns than basic-level nouns in the category of people. Before 36 months, the frequency of using basic-level nouns of people drops, while the frequency of using of subordinate nouns increases.

Pronouns:

All the pronouns are basic-level words, except for a subordinate noun 在下 which appeared only one time in 25-30 months. Both types and tokens show a rising trend of using different pronouns and using them frequently.

Body parts:

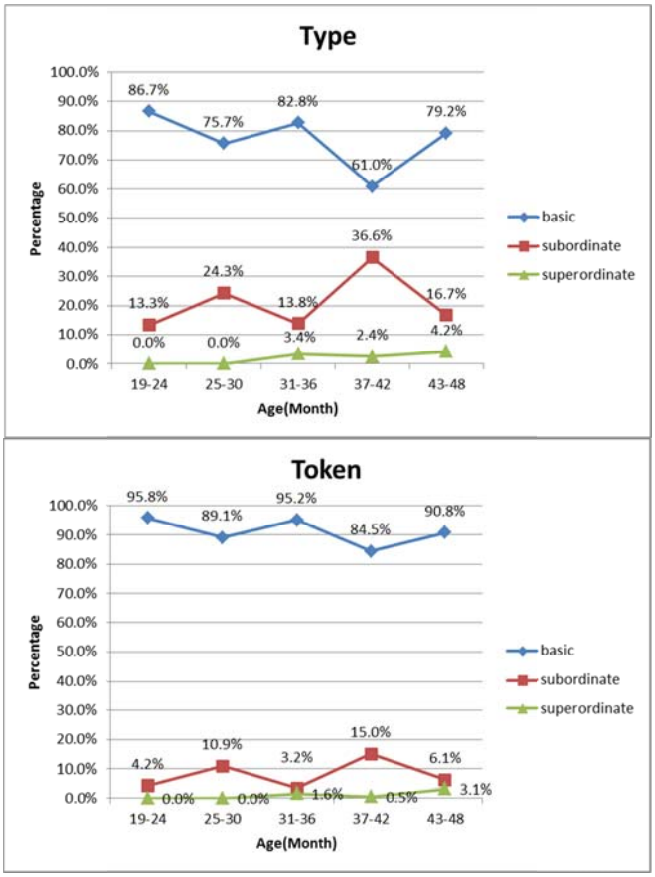


Figure 4.11: The proportion of conceptual levels in the category of body parts.

Figure 4.11 presents the percentage changes of conceptual levels of body parts. It shows that basic-level words are dominant in body parts, ranging from 61% to 86.7%. In general, although the trajectories are fluctuant, the proportion of basic-level words decreases, whereas the proportion of subordinate words increases. There are 36.6% of subordinate words in 37-42 months which is more than the proportions in other stages. Superordinate words of body parts appear in children's speech after 30 months, two tokens of 身體 (shēntǐ, body) in 31-36 months, one token of 全身 (quánshēn, whole body) in 37-42 months, and three tokens of 屍體 (shītǐ, corpse) in 43-48 months.

In terms of the tokens of body parts, the figure shows that a mean of 90% of body parts are basic-level words, ranging from 84.5% to 95.8%. It indicates children use

basic-level words of body parts with a greater frequency than subordinate and superordinate words.

Clothing:

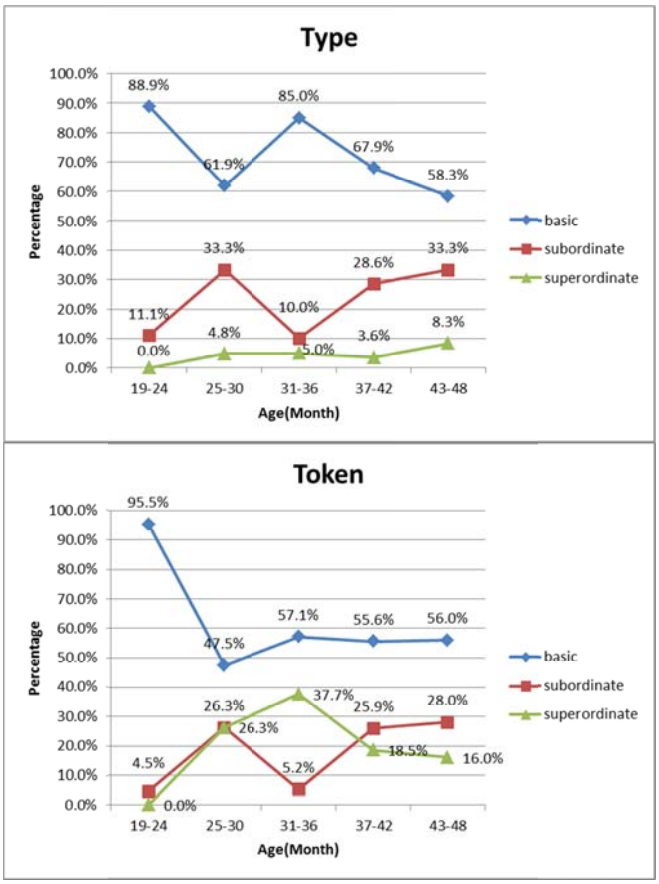


Figure 4.12: The proportion of conceptual levels in the category of clothing.

As shown in Figure 4.12, although the raw frequency of basic-level words in the category of clothing rises from 8 to 19, the proportion shows a decreasing trend from 88.9% to 58.3%. There is a tremendous drop of basic-level words in 25-30 months, and it rises again in 31-36 months. A gradual decreasing trend is found after 36 months. The proportion of subordinate words increases as time goes by, but a tremendous drop is found in 31-36 months. Superordinate words of clothing is 衣服 (yīfú, clothing) which appears from 25-48 months.

As for the frequency of using nouns about clothing, the highest proportion of

basic-level words is found in 19-24 months, and it falls to 47.5% in 25-30 months. After 30 months, the use of basic-level words is a steady trend, accounting for approximately 56% of all clothing words. Meanwhile, the frequency of subordinate and superordinate words rises in 25-30 months. The use of superordinate words in clothing words even accounts for 37.7% in 31-36 months. After 36 months, the frequency of using superordinate words drops whereas that of subordinate words rises.

Vehicles:

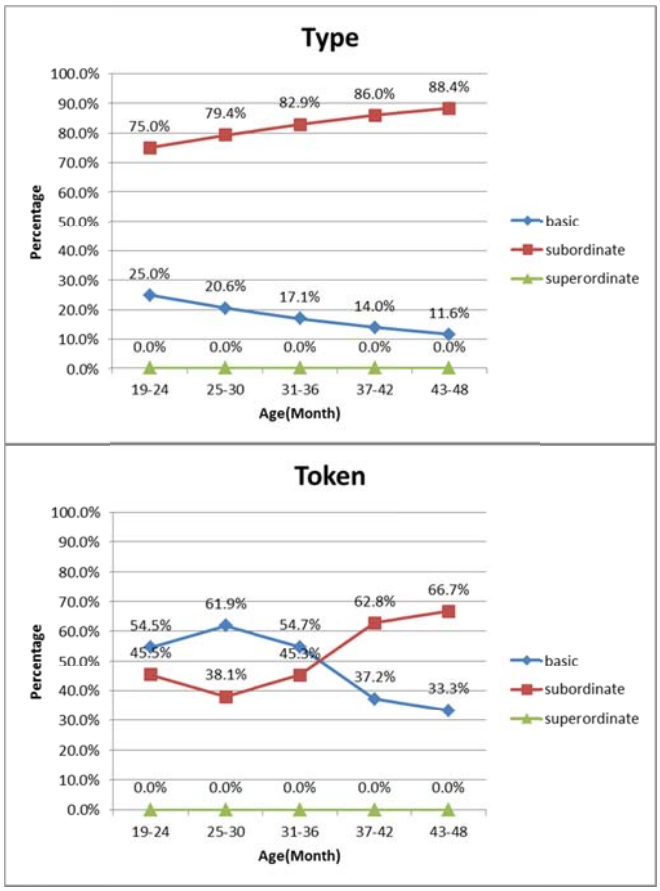


Figure 4.13: The proportion of conceptual levels in the category of vehicles.

Figure 4.13 presents the development trend of conceptual levels of vehicles. Two very noticeable trends are the steady decrease in basic-level words from 25% to 11.6%, and the steady increase in subordinate words from 75% to 88.4%. Besides, subordinate words outnumber basic-level words. There is no superordinate word found in these

stages. On the other hand, the proportion of tokens of vehicles shows a different trend. The frequency of using basic-level words is higher than subordinate words before 36 months, whereas the trend is reversed after 36 months, a higher frequency of using subordinate words than basic-level words.

Tools:

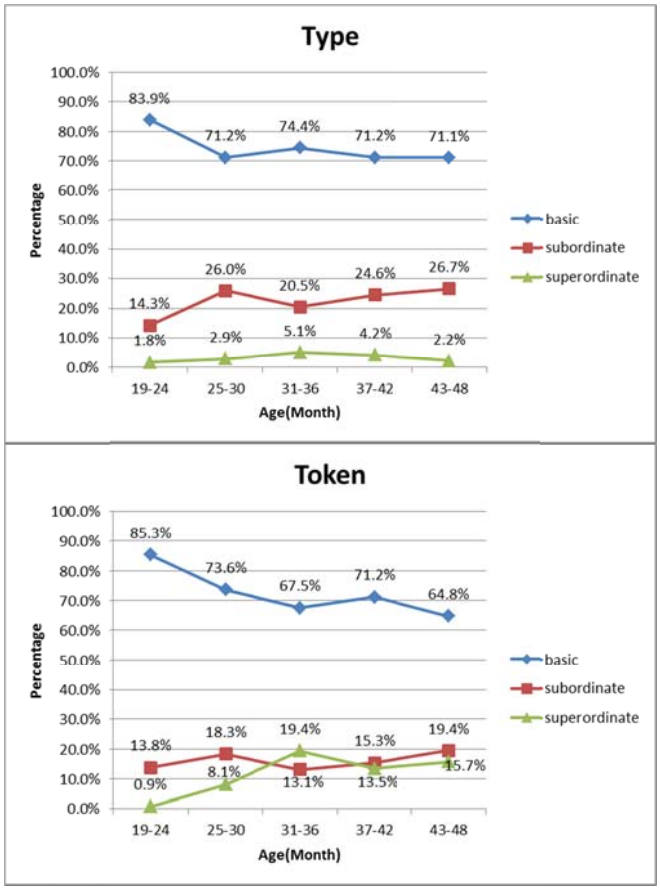


Figure 4.14: The proportion of conceptual levels in the category of tools.

The proportion of basic-level words of tools is 83.9% in 19-24 months, which is the highest proportion. In later stages, the proportion drops to 71.2%, and have a steady trend from then on. The proportion of subordinate words of tools increases from 14.3% to 26.7%. The number of types of superordinate words increases from 1 type to 5 types. The proportion of superordinate words shows a slight rise.

In terms of the frequency of using tool nouns, basic-level words are more

frequently used than subordinate and superordinate words. The falling tendency is found in the frequency of using basic-level words. The frequency of using subordinate word of tools seems not to change much. The use of superordinate words of tools is with increasing frequency until 36 months.

Furniture:

There are a few noun types of furniture. The number of basic-level words of furniture ranges from 3 to 7 types. The number of subordinate words is fewer than basic-level words, only 6 types are found in all stages. There are no superordinate words acquired in all stages. The token frequency also indicates that basic-level words of furniture are used with a higher frequency than subordinate words.

Food:

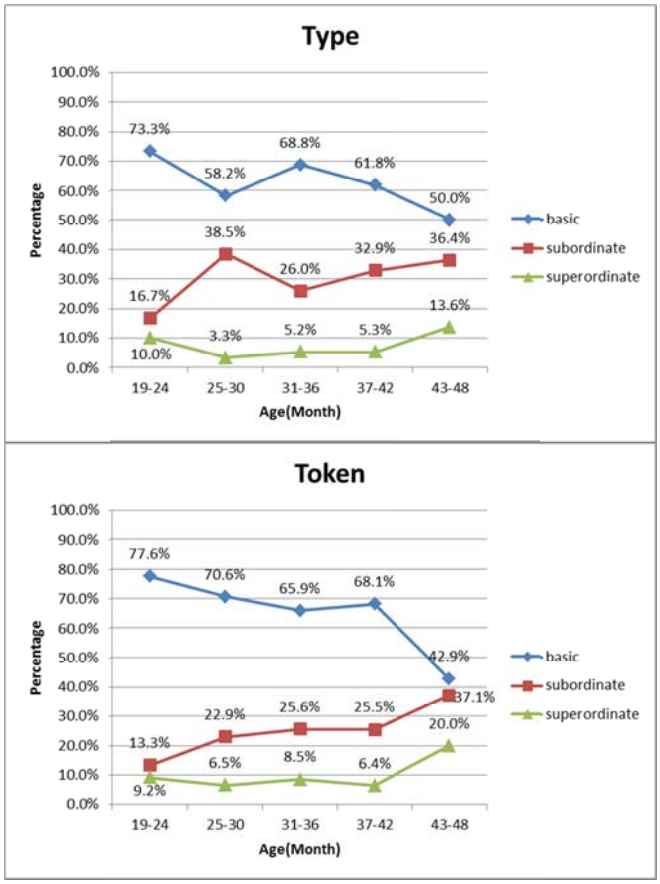


Figure 4.15: The proportion of conceptual levels in the category of food.

The proportion of basic-level word types of food outnumbers that of subordinate words and superordinate words. A declining trend is found in basic-level words from 73.3% to 50%, while rising trends are found in subordinate words from 16.7% to 36.4%, and in superordinate words from 3.3% to 13.6%. However, the raw frequency shows there are only 3 to 4 types of superordinate words. There is a remarkable increase of subordinate words in 25-30 months; meanwhile, the proportion of basic-level words and superordinate word decreases. The proportion of tokens shows that basic-level words are used more frequently. The falling trend of basic-level words indicates a decreasing frequency of using it. Meanwhile, the frequency of using subordinate words increases. The frequency of using superordinate words is low and steady, except for an increasing percentage in 43-48 months.

Animals:

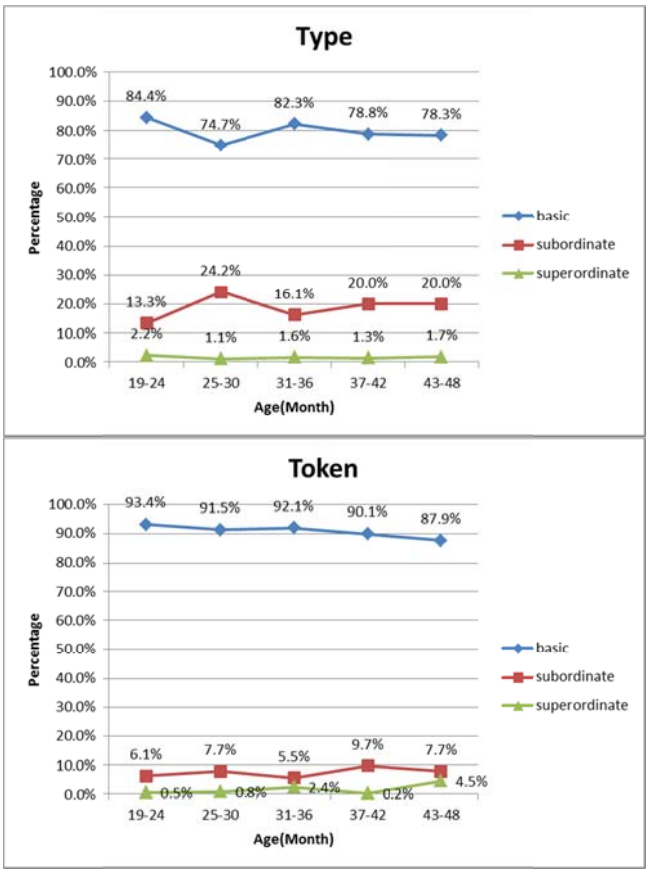


Figure 4.16: The proportion of conceptual levels in the category of animals.

As shown in Figure 4.16, most noun types of animals are basic-level words, and the figure shows a steady trend, except for a decrease in 25-30 months. The proportion of subordinate words accounts for 13.3% to 20%, and an increasing percentage to 24% is found in 25-30 months. There is only one superordinate word of animals, 動物 (dòngwù, animal). The token frequency also reveals that basic-level words are used more frequently than subordinate and superordinate words.

Natural phenomena and materials:

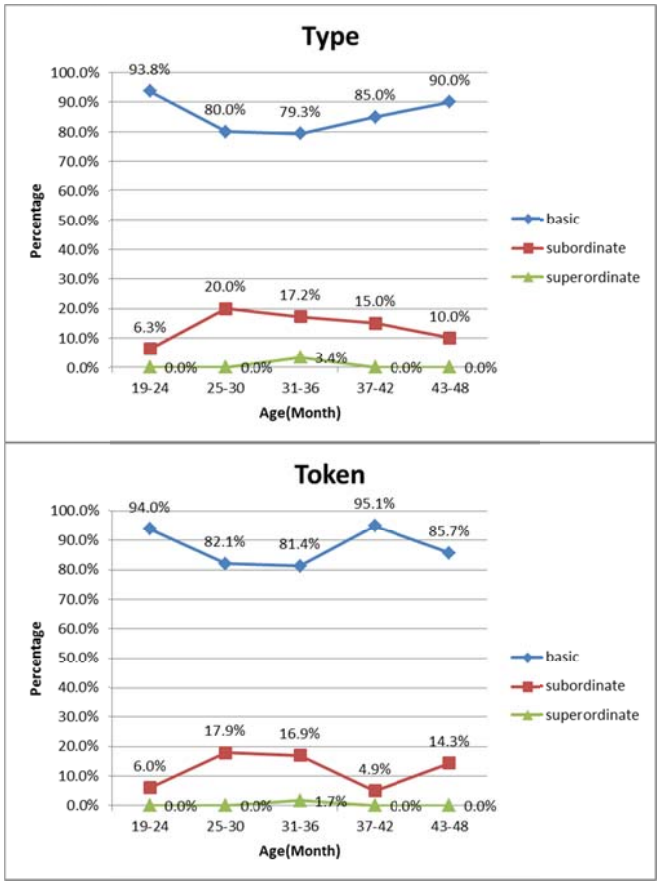


Figure 4.17: The proportion of conceptual levels in the category of natural phenomena and materials.

Figure 4.17 depicts the proportion of conceptual levels in natural phenomena and materials. There are more basic-level word than superordinate and subordinate words in natural phenomena and materials. Both raw frequency and proportion of basic-level

words show an increasing trend after 36 months. Meanwhile, the falling trend is found in subordinate words. There is only one superordinate word found in all stages, 天地 (tiāndì, the world). As for the frequency of using these nouns, basic-level words are used with a higher frequency than superordinate and subordinate words. The tendency of using basic-level words and subordinate is fluctuant.

Toys:

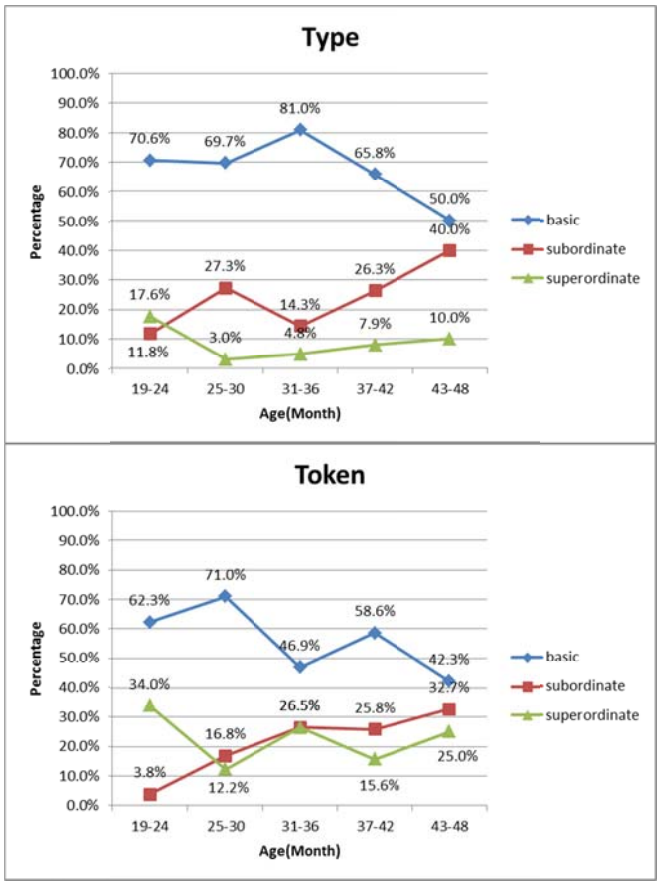


Figure 4.18: The proportion of conceptual levels in the category of toys.

As Figure 4.18 depicts, there are more basic-level noun types in toys than subordinate and superordinate words. A steady growth of subordinate words is found from 31 to 48 months, and meanwhile, the proportion of basic-level words decreases gradually. The proportion of superordinate word outnumbers than subordinate words in 19-24 months. The superordinate nouns of toys in 19-24 months are 具 (jù, toy), 具具

(jùjù, toy), and 玩具 (wánjù, toy). These three forms all mean toys. 具 (jù, toy) and 具具 (jùjù, toy) appeared only in this stage. The increasing proportion of superordinate words comes from new types 禮物 (lǐwù, gift) and 遊戲 (yóuxì, game) in later stages.

The frequency of using nouns of toys shows that basic-level words are used more frequently than subordinate and superordinate words. In general, a steady increasing trend of using subordinate words is found, while the trend of using basic-level words and superordinate words varies drastically. But the frequency of using nouns of three levels gets closer in 31-36 months and 43-48 months, especially the use of subordinate and superordinate words. Besides, the figure also shows that children use more superordinate words than subordinate words in 19-24 months.

Colors:

There are a small number of color words. Generally, basic-level words outnumber superordinate words and subordinate words, ranging from 3 types to 13 types. The numbers of basic-level words and subordinate words rise with increasing ages. There are only two superordinate words, 顏色 (yánsè, color) and 色 (sè, color). As for the frequency of using color words, basic-level words are the dominant level, ranging from 18 to 59 tokens. The frequency of using subordinate words is higher than that of superordinate words in 19-24 months and 43-48 months. However, superordinate words are used more frequently than subordinate words in 25-30 months, 31-36 months, and 37-42 months.

Shapes:

There are a few types in shapes, 7 basic-level words, one subordinate word, and 3 superordinate words. Shape words are acquired after children's 24 months. Subordinate and superordinate words of shape appear in 31-36 months.

The frequency of using shape nouns is low. Generally, basic-level words are used with a higher frequency than superordinate and subordinate words.

Locations and buildings:

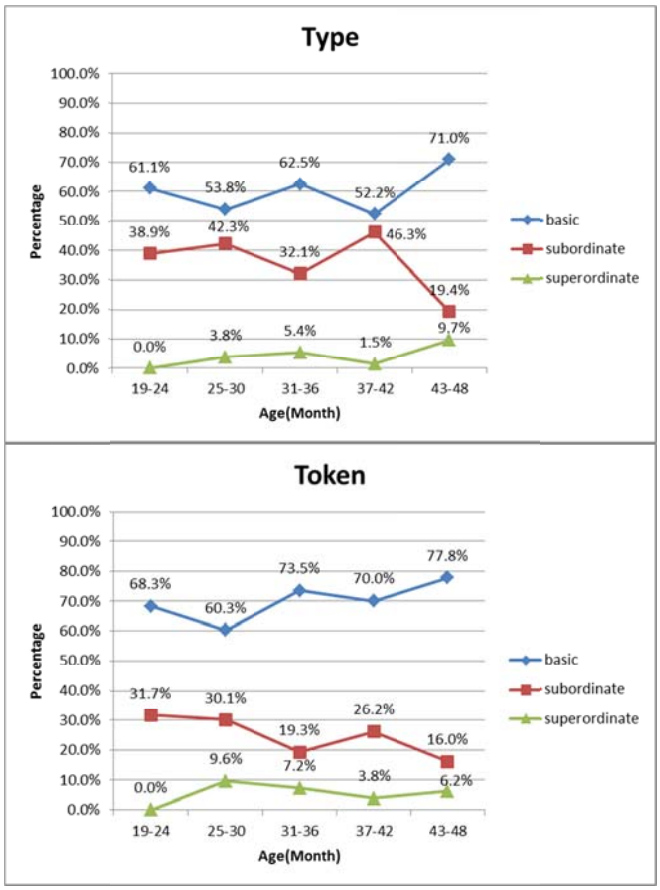


Figure 4.19: The proportion of conceptual levels in the category of locations and buildings.

Generally, there are more basic-level noun types of locations and buildings than subordinate and superordinate words. The proportions of basic-level words and subordinate words are very close to each other. A drastic increase of basic-level words and a sharp fall of subordinate words are found in 43-48 months. Only 4 types of superordinate location nouns are found in all stages, 位子 (wèizi, position), 地方 (dìfāng, place), 空間 (kōngjiān, space), and 世界 (shìjiè, world). As for the frequency of using location nouns, it is obvious that basic-level words are used with a higher

frequency than subordinate and superordinate words. A general increasing trend is shown in basic-level words, whereas the frequency of using subordinate and superordinate words declines.

Spatial words:

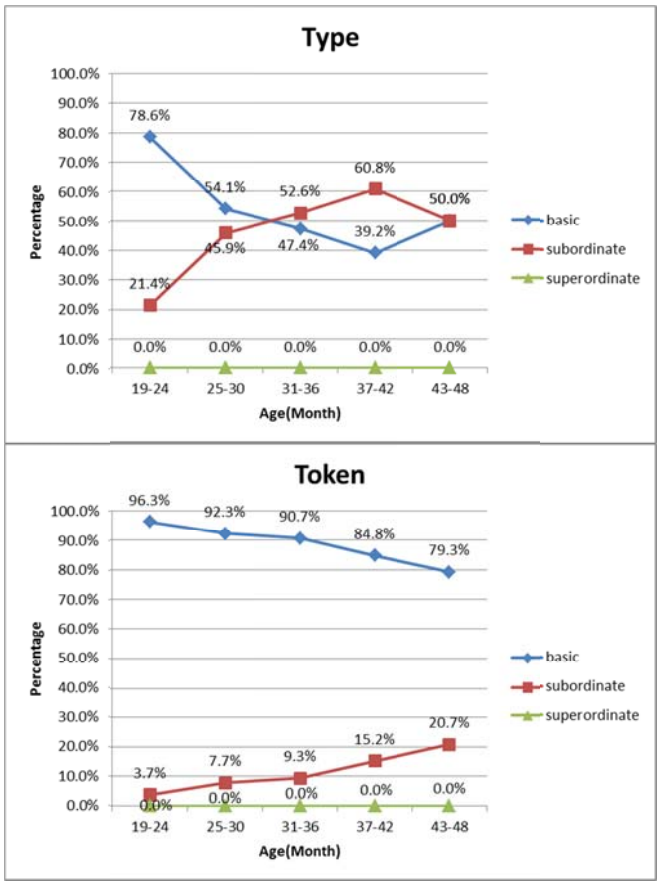


Figure 4.20: The proportion of conceptual levels in the category of spatial words.

The type proportion of spatial words reveals that basic-level words and subordinate words have a close proportion, except for a tremendous difference between them in 19-24 months. Basic-level words outnumber subordinates before 30 months, and the situation reverses after then. Basic-level words have a decreasing proportion while subordinate words have a rising proportion. The token proportion of using spatial words shows that basic-level words are used more frequently than subordinate words. In addition, two noticeable trends are a steady falling frequency of using basic-level words

and a steady rising frequency of using basic-level words.

Numerals:

There are few numeral nouns found in 19-24 months, 31-36 months and 37-42 months and most of the numeral nouns are used one time in the observed stage. Only one superordinate word appears in 19-24 months, and 2 in 37-42 months. Five basic-level numeral noun types appear in 31-36 months, and 7 basic-level numerals in 37-42 months.

Abstract nouns:

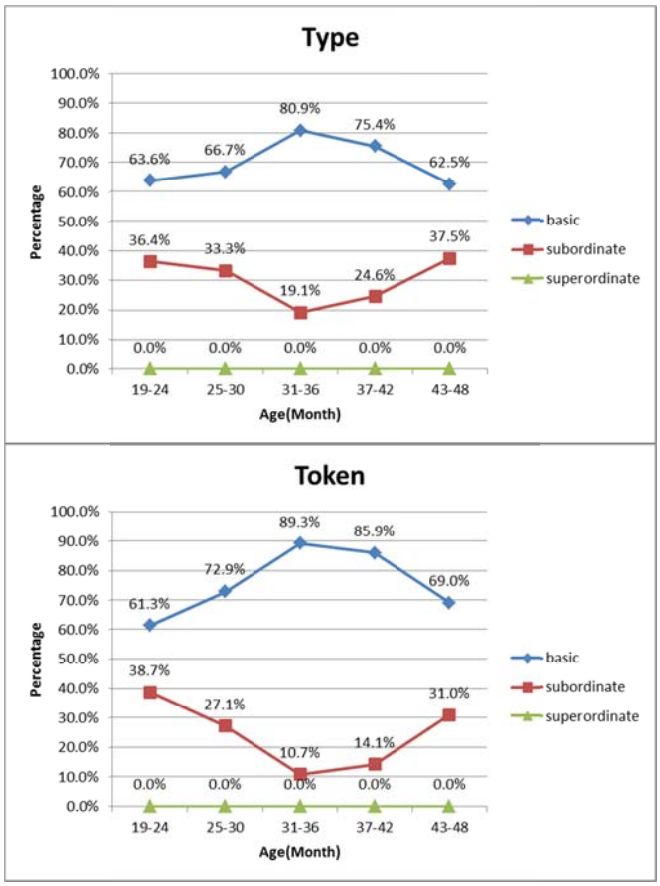


Figure 4.21: The proportion of conceptual levels in the category of abstract nouns.

Figure 4.21 reveals that basic-level words of abstract nouns outnumber subordinate words and basic-level words are used with higher frequency than subordinate words. In addition, it is obvious that the proportion of basic-level words keeps rising until 36

months, and keeps falling after 36 months. Meanwhile, the trend of subordinate words is a reversed trajectory. The proportion of subordinate words keeps falling until 36 months, and keeps rising after 36 months.

To sum up, it seems clear that there are more basic-level word types than superordinate and subordinate words in most semantic categories. There are 11 semantic categories showing a tendency that basic-level words outnumber subordinate words, and subordinate words outnumber superordinate words. The categories are tools, location, natural, clothing, body parts, toys, food, animals, pronouns, furniture, spatial words before 30 months, and abstract nouns. Two semantic categories, people and vehicles, show a tendency of more subordinate words than basic-level and superordinate words. Spatial words after 30 months also contain more subordinate words than basic-level words. Three semantic categories, shapes, colors, and numerals, show a trend of more basic-level words than superordinate words, and more superordinate words than subordinate words. The above paragraph is summarized below.

Table 4.19: The word frequency among conceptual levels.

(a) Type

Pattern	**Semantic Category**
Basic > subordinate > superordinate	Pronoun, body part, clothing, tool, furniture, food, animal, natural, toy, color, location, spatial (before 30 months), abstract nouns
Basic > superordinate > subordinate	Shape, numeral
Subordinate > basic > superordinate	People, vehicle, spatial (after 30 months)

(b) Token

Pattern	**Semantic Category**
Basic > subordinate > superordinate	People, pronoun, body part, tool, vehicle (before 36 months), furniture, food, toy (after 24 months), Clothing, animal, natural, location, spatial, abstract, color (24 months, 48 months)
Basic > superordinate > subordinate	Clothing (36 months), toy (before 24 months), tool (36 months), color, shape, numeral
Subordinate > basic > superordinate	Vehicle (after 36 months)

Note: “>” means outnumber.

Moreover, basic-level words are used with a higher frequency than superordinate and subordinate words in most semantic categories, except for vehicles after 36 months in which subordinate words is more frequently used than basic-level words. Subordinate words are used with a higher frequency than superordinate words in most semantic categories, except for clothing, tools and toys. Superordinate nouns of clothing, tools and toys are used more often than subordinate words in some stages. In addition to the development tendency of conceptual levels in semantic categories, the age stage of acquiring conceptual levels is compared, and six acquisition order patterns are found, as shown in Table 4.20.

(1) Basic-level words are acquired first, and then subordinate words are learned. There are no superordinate words observed. Categories in this pattern are pronouns and furniture.

(2) Basic-level words are acquired first, and in later stages subordinate and superordinate words are learned and appear at the same stage. The category of shapes develops in this pattern.

(3) Basic-level words and subordinate words appear at the same time, but no superordinate words exist in the five stages. Vehicles, spatial words, and abstract nouns are in this pattern.

(4) Basic-level words and subordinate words are acquired at the same time, and then superordinate words are acquired in later stages. The categories of people, location, natural, clothing, and body parts are in this pattern.

(5) Basic-level words, subordinate words, and superordinate words appear at the same time. Since nouns of three conceptual levels appear at the same stage, it is hard to identify the acquisition order of conceptual levels in this case. Tools, colors, toys, food, and animals are in this pattern.

(6) Superordinate words are acquired first, and then basic-level words are acquired. There are no subordinate words observed. Numerals are in this development pattern.

Table 4.20: Acquisition sequence of conceptual levels.

Pattern	**Semantic Category**
basic → subordinate	Pronouns (30), Furniture (30)
basic → subordinate = superordinate	Shapes (36)
basic = subordinate	Vehicles, Spatial words, Abstract nouns
basic = subordinate → superordinate	People (36), Location (30), Natural (36), Clothing (30), Body parts (36)
basic = subordinate = superordinate	Tools, Colors, Toys, Food, Animals
superordinate → basic	Numerals (36)

Chapter 5

Discussion and Conclusion

5.1. General discussion

This study aims to examine vocabulary development of Mandarin-speaking children in terms of growth patterns and organization, based on corpus data collected in Taiwan. Vocabulary growth was measured by computing the cumulative vocabulary, the frequency of POS, the proportion of POS, noun/verb ratio, the type/token ratio of POS, and D measure. Vocabulary organization was measured by computing the number of nouns in different semantic categories and conceptual levels.

5.1.1. Children's vocabulary growth

Children's vocabulary size

Cumulative vocabulary of each child and the mean cumulative vocabulary size in each age stage reveal a vocabulary growth, from 290 word types to 882 types. The vocabulary growth before 42 months is a steep distribution while the growth after 42 months is a flat distribution It indicates that the rate of vocabulary growth is higher in stages before 42 months. However, the rate of vocabulary growth slows down in 43-48 months. In terms of mean frequency in each age stage, the mean type frequency reveals that children produce more word types as they grow up, from a mean of 89.3 to 209.6 types. Mean token frequencies are 297.7 in 19-24 months, 636.3 in 25-30 months, 595.3 in 31-36 months, 808.0 in 37-42 months, and 732.6 in 43-48 months. The M-shaped distribution of mean token frequency seems to imply that children do produce more words in 25-30 months and 37-42 months, due to language development rather than sample sizes. Children may get familiar with new words acquired in previous stages,

and begin to use them frequently. The mean cumulative type frequency in this study is similar to but higher than that in Hsu's results (1996). Children in this study had 290 words before 2 years old, 741 words before 3 years old, and 882 words before 4 years old. Children in Hsu's study had 260 words before 2 years old, 634 words before 3 years old, and 771 words before 4 years old. The two study reach the same result in general.

POS frequency and proportion

In early stage, 19-24 months, nouns and verbs account for approximately 70% in children's vocabulary. After children's 30 months, the proportion of nouns and verbs will decrease. The proportion of 8 POS (ADV, CL, QN, ADJ, CONJ, PREP, WH, and MOD) will increase after 24 months. The distribution indicates that children acquire content words first and frequently use them in early ages, whereas they acquire function words later, and increase the use of function words in later stages. In terms of type frequency, children produced averagely more than 30 nouns and 30 verbs in 19-24 months, and in later stages, about 50 to 60 nouns and 65 to 75 verbs. The frequency distributions of nouns, verbs, and adjectives are an M-shaped distribution. The frequency rises in 25-30 months and 37-42 months, whereas it falls in 31-36 months and 43-48 months. On the contrary, the distributions of other parts-of-speech keep rising before 42 months and it drops after 42 months. As for token frequency, children produced more nouns and verbs than other POS in all stages. They produced about a mean of 93 nouns and 97 verbs in 19-24 months, but they produced about more than 100 nouns and verbs in later stages. The frequency distributions of nouns, verbs, adjectives, and modals are also an M-shaped distribution. The declined token frequency in 31-36 months and 43-48 months may imply that children learn to use other methods

to express their ideas, such as using more vocabulary of other lexical categories. This seems to be the case that the number of word tokens keeps increasing in ADV, CL, QN, CONJ, PREP, WH, and others until at the 37-42 months. Moreover, since word classes can be further grouped into open class and closed class, the result of POS in this study has shown different performance in open class and closed class. Open class or content words are produced earlier and have a higher proportion in children's speech than closed class or function words. The findings in this study support the idea that children's first words tend to be content words, while grammatical function words appear in children's speech later (Fernald & Marchman, 2006).

Examining the average number of types a child may produce, two quantitative changes and two qualitative changes were observed. A quantitative change occurred in 25-30 months when children have acquired and produced more different words in all word classes than in 19-24 months old. The mean total type frequency in 31-36 months (168.8) does not differ much from that in 25-30 months (166.4), so there may be nothing changed or may be a qualitative change. However, the type frequency of each POS in 31-36 months reveals that there does something changed. Children acquired and tended to use more words of all most lexical categories, expect for nouns and adjectives. There seems to be a qualitative change in this stage; children may learn new nouns of different categories and use a different strategy to construct their utterances, such as using more function words in their speech. In 37-42 months, a quantitative change occurred again, which seems to imply that children get familiar with what they learned in 31-36 months and use with a higher frequency. In 43-48 months, children produced fewer different word types in almost all word classes. This change may be caused by a more precise word use, such as using fewer words to express their ideas, or using other methods of constructing their speech beyond lexicon levels.

TTR

The type-token ratio of POS in each age stage provides some findings. Nouns, verbs, and POS classified as others have a flat distribution over five ages, which represents that TTR do not change much in different age stages. Previous section reports that children acquire more types of nouns and verbs, and children frequently use them as time goes by. It is also proven that the TTRs of nouns and verbs have a flat trajectory. The declining TTR of CL and MOD imply that children use some classifiers and modals more frequently in later stages. It supports that children acquire function words later and increase the use of function words in later stages.

Noun bias or Verb bias

Investigating the development of nouns and verbs leads to a question that whether there is a universal bias or not. Is there a nouns bias or verb bias in vocabulary of Mandarin-speaking children? This study tries to measure NVR by including broad and strict nouns and verbs in order to figure out the bias of Mandarin-speaking children in Taiwan. The results show that four types of NVRs have a similar development trend. NVR1 and NVR4 including broad or strict of nouns and verbs do not have much difference, a weak noun bias in 19-24 months and a verb bias in later stages. NVR2 which includes broad nouns and strict verbs shows a noun bias throughout all stages. On the contrary, NVR3 which includes strict nouns and broad verbs reveals a verb bias throughout all stages.

All types of NVRs can provide a trajectory of children's development of using nouns and verb. The NVRs are the highest in the early stages. The stage of 31-36 months is the turning point. NVRs decrease sharply before 31-36 months, and the ratios rise in later stages. However, the inclusion of broad or strict nouns and verbs leads to

different conclusion about noun bias or verb bias. Thus, researchers studying noun bias or verb bias should provide clear definitions of nouns and verbs.

In general, noun bias was observed in children's early vocabulary in this study. Nouns are first introduced into children's mental lexicon and children produce more nouns in age before 2 years old. The result seems to support Haryu and colleagues' explanation (2005) that the lack of verb morphology and the argument-dropping property of Mandarin makes verb learning even more difficult than nouns, and more difficult in Mandarin than other languages. On the contrary, children's vocabulary after 24 months show a verb bias tendency, and this may reflect the linguistic characteristic of Mandarin and also imply that children know arguments can be dropped and begin to drop arguments as in adult's speech.

Lexical diversity calculated from D values

Lexical diversity was measured by D. A higher value of D represents a higher diversity. The results show that D values of all vocabulary, of nouns, of broad verbs, and of strict verbs all have a similar developmental trend: D value keeps increasing until 37-42 months and it decline in 43-48 months. A statistically significant age effect on the D values, except for D of nouns, is found in 19-24 months and 25-30 months. The developmental trend indicates that children's speech become more and more diverse as they grow up, which supports the finding in Liu's study (Liu et al., 2008) that an increasing diversity and complexity was found in children's lexicon as time goes by. A similar lexical diversity development of Mandarin-speaking children is found in this study (age range = 19-48 months, M = 44.8, SD = 14.1) and in Liu's study (age range = 13-60 months, M = 44.21, SD = 19.11). The developmental trends observed in this study and Liu's study using different corpora may represent the lexical diversity of

typically-developing Mandarin-speaking children age before 4 years old. In addition, a higher D value of nouns than strict verbs before 30 months seems to reveal that children's lexical diversity in early stages comes from the diversity of nouns. Meanwhile, D values of nouns overlaps D values of strict verb after 31-36 months, indicating a similar lexical diversity.

5.1.2. Children's vocabulary organization

Semantic category

The type frequency of sematic categories provides some findings. In general, the category of people, tools, animals, and food are the dominant categories in each age. Besides, vehicle nouns and locations are major nouns as well. What is more, the category of abstract nouns becomes a main category in 43-48 months. On the contrary, the categories of numerals, shapes, furniture, and colors appear less in children's speech. Children have few pronouns in 37-42 months and few natural phenomena nouns in 43-48 months. A continued rising or falling proportion may be viewed as a steady tendency of development. Therefore, a decreasing proportion is found in people and animals until 42 months. The proportion of food rises in 25-30 months, but it keeps falling until 48 months. Meanwhile, an increasing proportion is found in clothing and abstract nouns until 42 months, but it decreases in 43-48 months. The proportion of spatial words keeps rising across all age stages. Other categories have a fluctuated trend of rising and declining.

As for how often children use nouns of different semantic categories, the categories of people, pronouns, animals, spatial words, and tools are the most frequently used. Children produced vehicles nouns very often in 19-24 months. As for the least frequently used categories, numerals, shapes, furniture, and colors have the lowest

frequency in children's speech. The category of natural phenomena also has a low frequency in 43-48 months. In terms of a falling trajectory, the use of people and food keeps declining across all stages, and the use of animals also keeps dropping until 42 months. On the contrary, the use of pronouns keeps increasing in all stages, and the use of shapes and abstract nouns remains rising until 37-42 months. Other categories have an unsteady proportion changes.

In summary, the most concrete nouns and nouns which are the closest to children's life are acquired earlier, such as people, tools, animals, and food, and vehicles. Nouns in these categories are also used frequently. On the other hand, nouns which are abstract and far away from children's daily life are acquired later and used with a lower frequency, such as numerals, shapes, colors, and nouns of natural phenomena.

Conceptual levels

The results have shown that there are more distinct basic-level words but fewer subordinate and superordinate words in earlier age stages, whereas the numbers of subordinate and superordinate words increase in later age stages. The highest proportions of three levels are 68.1% of basic-level words in 19-24 months, 39.3% of subordinate words in 43-48 months, and 3.1% superordinate words in 31-36 months. Besides, the proportion change indicates that children acquire relatively more subordinate words in 25-30 months. They increased the proportion of basic-level words and superordinate words in 31-36 months. After 36 months, children increase the proportion of subordinate words and reduce the proportion of basic-level words. General speaking, when the proportion of basic-level words decreases, the proportion of subordinate words increases, and vice versa. In addition, although the number of noun types of superordinate category increases in some stages, the proportion seems not to

differ much.

As for the frequency of using nouns, basic-level words account about 80% noun tokens in children's speech in all stages. The proportion also reveals that children use more subordinate words in 25-30 months than other stages, but the proportion decreases in later stages. The number and proportion of superordinate words increase before 31-36 months, but the proportion fluctuates after 36 months. In general, although the number of noun tokens of three levels changes in some stages, the changes seem not to differ much, except for subordinate words. Compared with the results of noun types on three conceptual levels, the development trend indicates that although the proportion of basic-level noun types decreases with increasing age, children keep using basic-level words frequently and even a little bit more frequently in later stages. Meanwhile, the proportion of subordinate noun types increases as time goes by, but children reduce the frequency of using subordinate words gradually.

With regard to conceptual levels in semantic categories, three patterns of the word frequency in three conceptual levels were found. It seems that most categories have more basic-level word types than subordinate words, and more subordinate words than superordinate words. The categories are tools, location, natural, clothing, body parts, toys, food, animals, pronouns, furniture, spatial words before 30 months, and abstract nouns. One pattern is that subordinate words outnumber basic-level words, and basic-level words outnumber superordinate words. The categories of people, vehicles, and spatial words after 30 months developed in this way. Three semantic categories, shapes, colors, and numerals, show a trend of more basic-level words than superordinate words, and more superordinate words than subordinate words

Additionally, basic-level words are used with a higher frequency than superordinate and subordinate words in most semantic categories, except for vehicles

after 36 months in which subordinate words is more frequently used than basic-level words. Subordinate words are used with a higher frequency than superordinate words in most semantic categories, except for clothing, tools and toys. Superordinate nouns of clothing, tools and toys are used more often than subordinate words in some stages.

5.2. Conclusion

5.2.1. Answers to research questions

The purpose of this study is to examine the developmental trends of children's early Mandarin lexicon with a corpus-based method, and then to find out items which can be used to measure young children's early vocabulary development. Thus, two research questions of this study are:

(1) What are the major changes in early vocabulary development of Mandarin-speaking children?

(2) How can early vocabulary development of Mandarin-speaking children be indexed?

As the above results and discussion have shown, children's vocabulary increases as they grow up, and the cumulative vocabulary size rise from 290 to 882 words. The frequency and proportion of lexical categories supports again that content words are acquired before function words. Nouns and verbs are acquired before other lexical categories. The proportion of nouns and verbs decreases with increasing age, while the proportion of other lexical categories increases. The results of type-token ratio indicate that children acquire more nouns and verbs and also use them very often in all stages leading to a flat TTR distribution. The declining TTRs of classifiers and modals indicate that children produce them more frequently in later stages. Noun bias and verb bias debates may be resulted from the different inclusion of nouns and verbs. However, a

weak noun bias is still found in 19-24 months, and a verb bias is found in later stages. This is the same conclusion as the noun-verb ratios in Liu's study (2008) that Mandarin-speaking children produced more nouns in early stages and more verbs in later stages. Lexical diversity calculated from D values also supports what was reported in Liu's study (2008) that children's lexicon becomes increasingly diverse as they grow up.

Investigation of the development of semantic category reveals that the most concrete nouns and nouns which are the closest to children's life are acquired earlier, such as people, tools, animals, and food, and vehicles. Nouns in these categories are also used frequently. On the other hand, nouns which are abstract are acquired later and used with a lower frequency, such as numerals, shapes, colors, and nouns of natural phenomena. Investigation of the development of conceptual levels has shown that basic-level words were acquired first, followed by subordinate words and superordinate words. The timing of acquiring the superordinate nouns and subordinate nouns varied in different semantic categories. Basic-level words were used with a higher frequency than the other two levels. These results support the findings in studies of Jiang (2000) and Lee (2014).

This study has use a corpus-based method to explore vocabulary development of young Mandarin-speaking children in Taiwan. The language samples were collected from 10 typically-developing children before 4 years old. The results of the analyses of vocabulary size, POS frequency and proportion, TTR, noun bias, lexical complexity, semantic category, and conceptual levels may provide an estimated result for other studies of the same scale.

5.2.2. Limitation and future study

All of the measured items about vocabulary growth and vocabulary organization in this study have revealed general developmental trends of children's early Mandarin vocabulary. Although this study has tried to provide evidence to support the findings of children's early vocabulary acquisition of past studies, there are still some limitations in this study and some suggestions for future studies.

The limitations are about the frequency calculation and the analysis of polysemous words. Most of the results, except for NVR and D values, are obtained by counting frequency and computing the percentage. The calculation may be effected by different sample sizes. For instance, some categories have relatively fewer types intrinsically, such as conjunction words and modals in lexical categories, and shapes, colors and numerals in semantic categories. Type frequency calculation would lead to strange numbers in these categories. A better calculation needs to be considered in future works. Furthermore, the total frequency in an age group was computed when analyzing semantic categories and conceptual levels. It did not represent the real performance of every single child. Instead, it is more like a collection of all vocabulary from all children. If a study aims to examine individual performance, the calculation needs to be revised. Another limitation is that this study did not differentiate polysemous words in the analysis. A polysemous word has more than one meaning. For example, 菜 (cài) means vegetable which is a specific, or it can be ingredients of meals which is more general. It needs further consideration whether to count 菜 (cài) as one word or as more than one word which depends on the meanings it has.

Besides, suggestions for future studies are to compare with the MCDI wordlist and another corpus like CHILDE. Some vocabulary in TCCM would overlap with the vocabulary in MCDI wordlist and CHILDES. Overlapped vocabulary may be the core

lexicon or basic vocabulary in children's early language. With core lexicon or basic vocabulary, researcher can develop measurements or teaching materials for young children. In addition, the issue of language development of young children is worth including gender difference into analysis. Boys and girls may show different performance in acquiring and producing words of different lexical categories, semantic categories, and conceptual levels. Likewise, it is also worthy of comparing typically-developing children with children of specific language impairment (SLI) to examine whether there is a different development trend for SLI children.

In conclusion, this study has tried to provide evidence for the findings of past studies of language acquisition with a corpus-based method, and to provide a complete scope of early vocabulary development of Mandarin-speaking children before 4 years old in Taiwan. The methodology and results in this study may provide some inspiration to those who are interested in language acquisition of Mandarin or other languages.

References

Bancroft, D. (1995). Categorization, concepts and reasoning. In V. J. Lee & P. Das Gupta (Eds.), *Children's cognitive and language development*. Cambridge, Mass: Blackwell Publishers in association with the Open University.

Chang, C. (2003). Talking about the past: How do Chinese mothers elicit narratives from their young children across time. *Narrative Inquiry*, *13*, 99–126.

Cheung, H. (1998). The Application of MLU in Chinese. *The Journal of Speech-Language-Hearing Association*, *13*, 36–48.

Chinese Knowledge Information Processing Group (CKIP). (1998a). Introduction to sinica corpus: A tagged balance corpus for Mandarin Chinese. (CKIP Technical Report no.95-02/98-04). Nangang,Taipei: Academia Sinica Institute of Information Science.

Chinese Knowledge Information Processing Group (CKIP). (1998b). Segmentation principle for chinese language processing (CNS14366). Taipei: Chinese National Standards, Ministry of Economic Affairs, R.O.C.

Chinese Knowledge Information Processing Group (CKIP). (2011). E-HowNet Ontology Online. Retrieved May 20, 2015, from http://ehownet.iis.sinica.edu.tw/

Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in Korean: a cross-linguistic study. *Journal of Child Language*, *22*(03), 497–529.

Dale, P., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, *28*(1), 125–127.

Dhillon, R. (2010). Examining the "Noun Bias": A structural approach. *University of Pennsylvania Working Papers in Linguistics*, *16*(1), 51–60.

Fenson, L., Dale, P., Reznick, J., & Bates, E. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5).

Fernald, A., & Marchman, V. (2006). Language Learning in Infancy. In M. J. Traxler, & M.A. Grensbacher (Eds.), *Handbook of Psycholinguistics* (pp. 1027-1071). London: Elsevier.

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

Hamilton, A., Plunkett, K. I. M., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language*, *27*(03), 689–705.

Hao, M., Shu, H., Xing, a., & Li, P. (2008). Early vocabulary inventory for Mandarin Chinese. Behavior Research Methods, 40(3), 728–733.

Haryu, E., Imai, M., Okada, H., Li, L., Meyer, M., Hirsh-Pasek, K., & Golinkoff, R. M. (2005). Noun bias in Chinese children: Novel noun and verb learning in Chinese, Japanese, and English preschoolers. In A. Brugos, M. Clark-Cotton, & S. Ha (Eds.), *Proceedings of the 29th Annual Boston University Conference on Language Development* (Vol. 1, pp. 272–283). Somerville, MA: Cascadilla Press.

Heaps, H. S. (1978). *Information retrieval: Computational and theoretical aspects*. New York: Academic Press.

Hsu, J. H. (1996). *A study of the stages of development and acquisition of mandarin Chinese by children in Taiwan*. Taipei: Crane Publishing.

Huang, J.-Y. (2009). *Interactions and experiences between mothers and 2-year-old children reading picture books altogether*. National Taiwan Normal University, Taiwan.

International Organization for Standardization (ISO). (1982). Romanization of Chinese. ISO 7098.

Jackson-Maldonado, D., Thal, D., Marchman, V., Bates, E., & Gutierrez-Clellen, V. (1993). Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Language*, *20*(03), 523–549.

Jiang, Y.-W. (2000). *Basic-level effects in Chinese lexicon*. National Tsing Hua University, Taiwan.

Kuo, J., Tsay, J., & Peng, J. (2005). Basis Level Effect of Taiwanese Noun Acquisition. In *The Proceedings of 2005 Conference on Taiwan Culture: Linguistics, Literature, Culture and Education* (pp. 43–54). Chiayi: National Chiayi University.

Liu, S., Zhao, X., & Li, P. (2008). Early lexical development: A corpus-based study of three languages. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1343–1348). Austin, TX: Cognitive Science Society.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing*, *24*(4), 459–488.

McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, *15*(3), 323–338.

Mervis, C. B., & Crisafi, M. a. (1982). Order of acquisition of subordinate-, basic-, and superordinate-level categories. *Child Development*, *53*(1), 258–266.

Peng, X., & Chong, H. (2010). Mandarin-speaking children's early noun acquisition from the perspective of categorization. *Journal of Anhui University of Science and Technology (Social Science)*, *1*, 68–71.

Rosch, E. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350.

Rosch, E. (1978). Principles of categorization. In E. R. B. B. Lloyd (Ed.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439.

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology*, *32*(3), 492–504.

Tardif, T. (2006). But are they really verbs? Chinese words for action. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *When action meets word: How children learn verbs.* Oxford University Press, USA.

Tardif, T., Gelman, S. A., & Xu, F. (1999). Putting the "Noun Bias" in Context: A Comparison of English and Mandarin. *Child Development*, *70*(3), 620–635.

Tweedie, F. J., & Baayen, H. R. (1998). How variable may a constant be? Measures of lexical richness in perspective. *Computers and the Humanities*, *32*(5), 323–352.

Xia, F. (2000a). The part-of-speech tagging guidelines for the Penn Chinese Treebank (3.0). Technical Report No. IRCS-00-07. University of Pennsylvania.

Xia, F. (2000b). The segmentation guidelines for the Penn Chinese Treebank (3.0). Technical Report No. IRCS-00-06. University of Pennsylvania.

Zeng, T., & Zou, W. (2012). The development of category level words in Mandarin speaking children before 6 years old. *Journal of Psychological Science*, *35*(6), 1404–1409.

李佳樺 (Lee, Chia Hua) (2014)。苗栗四縣客語兒童早期名詞詞彙習得（未出版之碩士論文）。國立聯合大學，新竹。

張顯達 (Cheung, H.) (2003)。特定型語言障礙兒童之語言能力發展─發展轉變與類型之間的探索。行政院國家科學委員會專題研究計畫成果報告(NSC92-2411-H-002-055)。

張顯達、張鑑如、柯華葳、蔡素娟 (Cheung, H., Chang, C., Ko, H.-W., & Tsay, J.)（2011）。台灣兒童語言語料庫之建置。行政院國家科學委員會專題研究計畫(NSC96-2420-H-002-030)。

教育部國語推行委員會 (National Languages Committee Ministry of Education).（2000）。國小學童常用字詞調查報告書。台北市：教育部。檢自：http://www.edu.tw/files/site_content/M0001/primary/shindex.htm

梁衛蘭、郝波、王爽、楊豔玲、張致祥、左啟華 (Liang, W.-L., Hao, B., Wang, S., Yang, Y.-L., Zhang, Z.-X., Zuo, Q.-H.)、Tardif, T.、Fletcher, P. （2001）。中文早期

語言與溝通發展量表——普通話版的再標準化。中國兒童保健雜誌，9(5)，295-297。

盛青、鄧慧華、張光珍、梁宗保、陸祖宏 (Sheng, Q., Deng, H.-H., Zhang, G.-Z., Liang, Z.-B., & Lu, Z.-H.) (2012)。中國學步期兒童詞匯能力(普通話)發展特點研究。東南大學學報（醫學版）, *31*(4), 400–405.

葉家利 (Yeh, Chia-Li) (2009)。臺灣華語幼兒名詞和動詞之習得研究（未出版之碩士論文）。輔仁大學，新北市。

劉惠美、曹峰銘 (Liu, H.-M., & Tsao, F.-M.) (2010)。華語嬰幼兒溝通發展量表(臺灣版）。台北市：心理出版社。

Appendices

Appendix 1: The POS tagset in TCCM corpus (TCCM, 2011)

1	**ADV**	Adverb 副詞	很、一起、就
2	**ADJ**	Adjective 形容詞	小、新、大
3	**ASP**	Aspect marker 體貌詞	過、著、在
4	**CONJ**	Conjunction連接詞	和、跟、或
5	**DE**	Nominal marker, Adverbial marker, Complement marker	的、得、地
6	**DT**	Determiner 定詞	這、那
7	**INT**	Interjection 感歎詞	唉、噢、嗯
8	**CS**	Foreign words 外來語 (包括閩南語)	Mary、radio
9	**IDM**	Idiom 成語	一心二用、對不起、沒關係
10	**CL**	Classifier/measure word 量詞	公尺、一「雙」筷子、這「個」
11	**MOD**	Modal 法相詞	果然、也許、可以
12	**Nn**	Noun 一般名詞	桌子、水、公園
13	**Nloc**	Localizer 地方名詞	中間、地上、水裡
14	**Nppn**	Proper noun 專有名詞	台北市、小明、新光三越
15	**Npro**	Pronominal 代詞	你們、他、自己
16	**Nt**	Time noun 時間名詞	春天、現在、早上
17	**NEG**	Adverb of negation 否定副詞	不、沒、未
18	**ONM**	Onomatopoeia 擬聲詞	砰、嘩啦、咕
19	**PREP**	Preposition 介詞	到、把、在
20	**QN**	Quantifier 數詞	數字“一”,“二”,“三”...、許多
21	**SFP**	Sentence final particle 語尾詞	啊、呀、啦
22	**CHM**	Unknown word 未知詞	拇拇
23	**Va**	Stative verb 狀態動詞	高、大、辛苦
24	**Vc**	Shi-word 「是」字句	是、像
25	**Vi**	Intransitive verb 不及物動詞	跑、下雨、逛街
26	**Vt**	Transitive verb 及物動詞	吃、打、「有」字句
27	**Vr**	Reduplicated verb 疊詞	看看、敲「一敲」、走來走去、「越來越」胖
28	**WH**	Interrogatives “WH-”疑問詞	誰、為什麼、幹嘛

Retrieved from the TCCM website: http://taiccm.org/

Appendix 2: Examples of output from *vocd*.

Command line: vocd +t*CHI

Output:

From file <HTC01_WU107_01.cha>

好

媽媽 釣魚

好

阿姨

阿姨

嘟

阿姨

飲料

飲料

還 要

還 要 飲料

還 要 飲料

不 要

不 要

柳丁

要 布丁

吃

冰淇淋

冰淇淋 吃光光 啊啊

好

不 要

不 要

我 要 吧吧吧

吧吧吧

會 破 掉

要 布丁

鞋鞋

那 拖鞋

外面 有 飛飛

那 鳥 飛飛

這 小 西瓜

電扇

電

這 關 掉

關 掉

叔

啊

電視

啊

象棋

阿開姊姊

阿開姊姊

涼快

要 大 的

筆

媽媽

媽媽

要 尿尿

多多

請 兩 個

抓

抓 蝌蚪

要 買 蝌蚪

嗯

不 要

要 抱抱

嗯

要 抱抱

媽媽

三

嗯

我 不 要

不 要

一 條 狗狗

那 白 草

那 狗狗

不 要
尿尿
嗯
他 跑 掉
他 跑 掉
狗 大便
狗 大便
沒有
撞 到 頭 了
媽媽
菜菜
菜
菜
轉轉
有
在 水裡
雞
叮噹
嗯
不 要 啦
我 叮
好
青蛙
跑 掉
那 青蛙 跑 掉
青蛙 跑 掉
電池
石頭
要 丟 水溝
啊
有 泡沫
那 跑 掉
啊
泡沫
泡沫

是
玩
這 是 泡沫
泡沫
在 睡覺
泡沫
螞蟻
大 螞蟻
大 螞蟻
大 螞蟻
吃 菜菜
那 泡沫
一 隻 螞蟻
一 隻
兩 隻 螞蟻
不 要 了
泡沫 拜拜
看 泡沫
阿姨
看 泡沫
嗯
l 阿姨
不 要
嗯
嗯
不 要
不 要
阿姨 謝謝
廖芳芳
我 要 尿尿
嗯
烏龜 在 爬
媽媽

tokens	samples	ttr	st.dev	D
35	100	0.7514	0.068	39.752

36	100	0.7364	0.056	37.027
37	100	0.7370	0.070	38.214
38	100	0.7353	0.066	38.799
39	100	0.7390	0.057	40.795
40	100	0.7220	0.054	37.502
41	100	0.7232	0.059	38.728
42	100	0.7131	0.063	37.220
43	100	0.7209	0.051	40.042
44	100	0.7145	0.057	39.350
45	100	0.7093	0.055	38.948
46	100	0.7022	0.059	38.076
47	100	0.7043	0.059	39.410
48	100	0.6850	0.052	35.750
49	100	0.6978	0.054	39.465
50	100	0.6896	0.058	38.301

D: average = 38.586; std dev. = 1.240

D_optimum <38.56; min least sq val = 0.000>

tokens	samples	ttr	st.dev	D
35	100	0.7489	0.064	39.076
36	100	0.7450	0.059	39.178
37	100	0.7335	0.059	37.352
38	100	0.7242	0.060	36.133
39	100	0.7215	0.062	36.458
40	100	0.7248	0.066	38.166
41	100	0.7327	0.060	41.168
42	100	0.7136	0.056	37.332
43	100	0.7063	0.058	36.514
44	100	0.7075	0.057	37.649
45	100	0.7111	0.054	39.385
46	100	0.6985	0.055	37.215
47	100	0.7036	0.056	39.254
48	100	0.6944	0.056	37.863
49	100	0.6876	0.065	37.068
50	100	0.6950	0.049	39.592

D: average = 38.088; std dev. = 1.348

D_optimum <38.06; min least sq val = 0.000>

tokens	samples	ttr	st.dev	D
35	100	0.7571	0.063	41.309
36	100	0.7511	0.058	40.801
37	100	0.7465	0.064	40.664
38	100	0.7408	0.057	40.224
39	100	0.7290	0.062	38.234
40	100	0.7220	0.058	37.502
41	100	0.7234	0.060	38.788
42	100	0.7143	0.058	37.500
43	100	0.7088	0.062	37.102
44	100	0.7116	0.057	38.626
45	100	0.7060	0.054	38.146
46	100	0.7015	0.052	37.923
47	100	0.6977	0.065	37.832
48	100	0.7058	0.051	40.646
49	100	0.6935	0.050	38.437
50	100	0.6946	0.054	39.495

D: average = 38.952; std dev. = 1.328

D_optimum <38.89; min least sq val = 0.000>

VOCD RESULTS SUMMARY

==========

Types,Tokens,TTR: <94,232,0.405172>

D_optimum values: <38.56, 38.06, 38.89>

D_optimum average: 38.50

Appendix 3: The raw frequency tables.

Table 1: The total frequency of each POS in the whole corpus (including CS and CHM).

POS	Word Type	Percentage	Word Token	Percentage
ADV	109	2.26%	3098	3.86%
ADJ	49	1.02%	554	0.69%
ASP	4	0.08%	1706	2.12%
CHM	68	1.41%	1833	2.28%
CL	64	1.33%	3616	4.50%
CONJ	28	0.58%	811	1.01%
CS	720	14.96%	2718	3.39%
DE	4	0.08%	1744	2.17%
DT	9	0.19%	3467	4.32%
IDM	97	2.01%	252	0.31%
INT	73	1.52%	3351	4.17%
MOD	14	0.29%	1121	1.40%
NEG	3	0.06%	3544	4.41%
Nloc	90	1.87%	541	0.67%
Nn	1445	30.02%	11081	13.80%
Nppn	301	6.25%	1868	2.33%
Npro	22	0.46%	6847	8.53%
Nt	34	0.71%	206	0.26%
ONM	105	2.18%	506	0.63%
PREP	21	0.44%	790	0.98%
QN	60	1.25%	1344	1.67%
SFP	33	0.69%	4326	5.39%
Va	356	7.40%	5054	6.29%
Vc	3	0.06%	1723	2.15%
Vi	352	7.31%	3327	4.14%
Vr	72	1.50%	186	0.23%
Vt	661	13.73%	13223	16.47%
WH	17	0.35%	1453	1.81%
SUM	**4814**	**100.00%**	**80290**	**100.00%**

Table 2: The total number of POS types in each age stage.

POS	19-24	25-30	31-36	37-42	43-48	All stage
Noun	367	770	681	885	480	1892
Verb	288	673	661	763	399	1444
ADV	21	50	59	76	42	109
CL	10	26	33	33	22	64
QN	13	21	32	41	22	60
ADJ	9	23	16	20	15	49
CONJ	6	14	14	20	17	28
PREP	3	15	16	18	9	21
WH	8	11	11	13	11	17
MOD	4	10	8	10	7	14
Others	86	161	137	149	114	328
All POS	**815**	**1774**	**1668**	**2028**	**1138**	**4026**

Table 3: The total number of POS tokens in each age stage.

POS	19-24	25-30	31-36	37-42	43-48	All stage
Noun	1767	5846	4321	6011	2598	20543
Verb	1856	6729	5126	6882	2920	23513
ADV	101	661	718	1189	429	3098
CL	223	753	892	1158	590	3616
QN	102	238	390	447	167	1344
ADJ	35	184	110	167	58	554
CONJ	19	117	225	353	97	811
PREP	31	182	212	270	95	790
WH	83	337	374	416	243	1453
MOD	21	283	225	416	176	1121
Others	1419	4396	4076	6122	2883	18896
All POS	**5657**	**19726**	**16669**	**23431**	**10256**	**75739**

Table 4: The numbers of noun types of semantic categories in each age stage.

Category	19-24 months	25-30 months	31-36 months	37-42 months	43-48 months
People	81	146	128	154	111
%	23.1%	19.4%	19.2%	17.8%	23.6%
Spatial words	14	37	38	51	28
%	4.0%	4.9%	5.7%	5.9%	5.9%
Pronoun	9	14	16	17	13
%	2.6%	1.9%	2.4%	2.0%	2.8%
Tools	56	104	78	118	45
%	16.0%	13.8%	11.7%	13.6%	9.6%
Vehicles	20	34	41	50	43
%	5.7%	4.5%	6.1%	5.8%	9.1%
Location	18	52	56	67	31
%	5.1%	6.9%	8.4%	7.7%	6.6%
Natural	16	25	29	40	10
%	4.6%	3.3%	4.3%	4.6%	2.1%
Clothing	9	21	20	28	12
%	2.6%	2.8%	3.0%	3.2%	2.5%
Shapes	0	3	5	5	1
%	0%	0.4%	0.7%	0.6%	0.2%
Body parts	15	37	29	41	24
%	4.3%	4.9%	4.3%	4.7%	5.1%
Abstract nouns	11	45	47	65	32
%	3.1%	6.0%	7.0%	7.5%	6.8%
Toys	17	33	21	38	20
%	4.8%	4.4%	3.1%	4.4%	4.2%
Food	30	91	77	76	22
%	8.5%	12.1%	11.5%	8.8%	4.7%
Animals	45	91	62	80	60
%	12.8%	12.1%	9.3%	9.2%	12.7%
Furniture	4	6	7	10	4
%	1.1%	0.8%	1.0%	1.2%	0.8%
Numerals	1	0	5	9	0
%	0.3%	0.0%	0.7%	1.0%	0.0%
Colors	5	13	9	18	15
%	1.4%	1.7%	1.3%	2.1%	3.2%
SUM	**351**	**752**	**668**	**867**	**471**

Table 5: The numbers of noun tokens of semantic categories in each age stage.

Category	19-24 months	25-30 months	31-36 months	37-42 months	43-48 months
People	629	1788	1057	1101	367
%	36.3%	30.7%	24.6%	18.4%	14.2%
Spatial words	82	426	302	422	135
%	4.7%	7.3%	7.0%	7.0%	5.2%
Pronoun	239	1219	1235	2169	1158
%	13.8%	20.9%	28.7%	36.2%	44.7%
Tools	116	345	191	326	108
%	6.7%	5.9%	4.4%	5.4%	4.2%
Vehicles	99	247	190	231	96
%	5.7%	4.2%	4.4%	3.9%	3.7%
Location	41	146	166	210	81
%	2.4%	2.5%	3.9%	3.5%	3.1%
Natural	50	106	59	122	14
%	2.8%	1.8%	1.4%	2.0%	0.5%
Clothing	22	80	77	81	25
%	1.3%	1.4%	1.8%	1.4%	1.0%
Shapes	0	4	7	11	2
%	0%	0.1%	0.2%	0.2%	0.1%
Body parts	48	193	124	200	98
%	2.8%	3.3%	2.9%	3.3%	3.8%
Abstract nouns	31	118	121	185	71
%	1.8%	2.0%	2.8%	3.1%	2.7%
Toys	53	131	98	128	52
%	3.1%	2.3%	2.3%	2.1%	2.0%
Food	98	323	211	251	35
%	5.6%	5.6%	4.9%	4.2%	1.4%
Animals	196	611	382	435	247
%	11.3%	10.5%	8.9%	7.3%	9.5%
Furniture	8	18	23	34	23
%	0.5%	0.3%	0.5%	0.6%	0.9%
Numerals	1	0	5	13	0
%	0.1%	0.0%	0.1%	0.2%	0.0%
Colors	22	64	49	70	76
%	1.3%	1.1%	1.1%	1.2%	2.9%
SUM	**1735**	**5819**	**4297**	**5989**	**2588**

Table 6: The numbers of types in three conceptual levels of semantic categories.

People	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	34	54	48	67	50
subordinate	47	92	78	86	61
superordinate	0	0	2	1	0
Body parts	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	13	28	24	25	19
subordinate	2	9	4	15	4
superordinate	0	0	1	1	1
Clothing	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	8	13	17	19	7
subordinate	1	7	2	8	4
superordinate	0	1	1	1	1
Vehicles	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	5	7	7	7	5
subordinate	15	27	34	43	38
superordinate	0	0	0	0	0
Tools	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	47	74	58	84	32
subordinate	8	27	16	29	12
superordinate	1	3	4	5	1
Furniture	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	4	5	7	6	3
subordinate	0	1	0	4	1
superordinate	0	0	0	0	0
Food	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	22	53	53	47	11
subordinate	5	35	20	25	8
superordinate	3	3	4	4	3
Animals	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	38	68	51	63	47
subordinate	6	22	10	16	12
superordinate	1	1	1	1	1
Natural phenomena	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	15	20	23	34	9
subordinate	1	5	5	6	1
superordinate	0	0	1	0	0
Toys	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	12	23	17	25	10
subordinate	2	9	3	10	8

superordinate	3	1	1	3	2
Colors	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	3	9	7	13	10
subordinate	1	2	0	3	4
superordinate	1	2	2	2	1
Shapes	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	0	3	3	4	0
subordinate	0	0	1	0	0
superordinate	0	0	1	1	1
Location	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	11	28	35	35	22
subordinate	7	22	18	31	6
superordinate	0	2	3	1	3
Numerals	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	0	0	5	7	0
subordinate	0	0	0	0	0
superordinate	1	0	0	2	0
Abstract nouns	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	7	30	38	49	20
subordinate	4	15	9	16	12
superordinate	0	0	0	0	0
Spatial words	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	11	20	18	20	14
subordinate	3	17	20	31	14
superordinate	0	0	0	0	0
Pronouns	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	9	13	16	17	13
subordinate	0	1	0	0	0
superordinate	0	0	0	0	0

Table 7: The numbers of tokens in three conceptual levels of semantic categories.

People	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	438	1094	593	740	199
subordinate	191	694	462	360	168
superordinate	0	0	2	1	0
Body parts	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	46	172	118	169	89
subordinate	2	21	4	30	6
superordinate	0	0	2	1	3
Clothing	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	21	38	44	45	14
subordinate	1	21	4	21	7
superordinate	0	21	29	15	4
Vehicles	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	54	153	104	86	32
subordinate	45	94	86	145	64
superordinate	0	0	0	0	0
Tools	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	99	254	129	232	70
subordinate	16	63	25	50	21
superordinate	1	28	37	44	17
Furniture	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	8	13	23	30	22
subordinate	0	5	0	4	1
superordinate	0	0	0	0	0
Food	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	76	228	139	171	15
subordinate	13	74	54	64	13
superordinate	9	21	18	16	7
Animals	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	183	559	352	392	217
subordinate	12	47	21	42	19
superordinate	1	5	9	1	11
Natural phenomena	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	47	87	48	116	12
subordinate	3	19	10	6	2
superordinate	0	0	1	0	0
Toys	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	33	93	46	75	22
subordinate	2	22	26	33	17

superordinate	18	16	26	20	13
Colors	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	18	52	44	59	59
subordinate	3	5	0	4	10
superordinate	1	7	5	7	7
Shapes	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	0	4	5	10	0
subordinate	0	0	1	0	0
superordinate	0	0	1	1	2
Location	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	28	88	122	147	63
subordinate	13	44	32	55	13
superordinate	0	14	12	8	5
Numerals	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	0	0	5	11	0
subordinate	0	0	0	0	0
superordinate	1	0	0	2	0
Abstract nouns	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	19	86	108	159	49
subordinate	12	32	13	26	22
superordinate	0	0	0	0	0
Spatial words	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	79	393	274	358	107
subordinate	3	33	28	64	28
superordinate	0	0	0	0	0
Pronouns	**19-24 months**	**25-30 months**	**31-36 months**	**37-42 months**	**43-48 months**
basic	239	1218	1235	2169	1158
subordinate	0	1	0	0	0
superordinate	0	0	0	0	0

Appendix 4: Lists of parts of speech except for nouns and verbs

ADV									
19-24 months		**25-30 months**		**31-36 months**		**37-42 months**		**43-48 months**	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
好	27	好	139	就	103	就	199	就	83
就	12	還	113	也	100	好	155	好	54
還	12	就	70	還	86	也	147	還	46
又	10	也	68	好	84	還	107	也	45
也	6	都	30	都	56	都	84	很	31
太	5	太	28	很	35	很	45	都	30
已經	4	再	27	這麼	26	再	43	比較	16
這麼	4	很	22	才	21	又	39	才	12
都	4	還要	22	再	20	等一下	27	再	11
再	3	等一下	12	又	18	比較	25	又	11
等一下	3	這麼	11	太	18	先	21	先	6
然後	2	已經	10	一起	15	太	19	只	6
先	1	又	8	最	14	已經	19	本來	6
剛	1	最	8	等一下	14	這麼	19	等一下	6
好像	1	才	7	通通	9	才	16	一起	5
常	1	真	6	先	7	最	15	剛剛	5
很	1	一點	5	快	6	現在	14	已經	5
才	1	快	5	結果	6	一直	12	統統	4
最	1	快要	5	一直	5	剛剛	12	以前	3
蓋	1	通通	5	剛剛	5	還要	12	只有	3
還要	1	好像	4	只	5	只	11	太	3
		比較	4	現在	5	本來	9	更	3
		須要	4	已經	4	一起	7	最	3
		一直	3	一會兒	3	一點	7	現在	3
		偷偷	3	接下來	3	好像	6	真	3
		剛	3	比較	3	後來	6	趕快	3
		剛剛	3	真	3	快	6	一	2
		只	3	越	3	一	5	一點	2
		新	3	一點	2	必	5	亂七八糟	2
		還有	3	剛才	2	真	5	另外	2
		一起	2	只是	2	結果	5	後來	2
		以後	2	夠	2	還是	5	快	2
		怎	2	從前	2	上次	4	非常	2
		更	2	快要	2	以前	4	以後	1
		正	2	趕快	2	差一點	4	剛才	1
		蠻	2	還是	2	從前	4	原來	1
		那麼	2	那麼	2	快要	4	好像	1
		下次	1	都	2	通通	4	快要	1
		先	1	乾脆	1	下次	3	有一點	1
		剛剛好	1	以後	1	剛	3	還是	1
		剛好	1	其實	1	剛才	3	還要	1
		差一點	1	剛	1	天天	3	那麼	1
		已經	1	剛剛好	1	更	3		
		已	1	剛好	1	那麼	3		
		早	1	各自	1	原來	2		
		時	1	多	1	另外	2		
		突然	1	大	1	從來	2		
		趕快	1	好像	1	正	2		
		還好	1	小	1	每次	2		
		重新	1	差一點	1	趕快	2		
				後來	1	重新	2		

ADV									
19-24 months		25-30 months		31-36 months		37-42 months		43-48 months	
word	freq	word	freq	word	freq	word	freq	word	freq
				更	1	馬上	2		
				最後	1	一下子	1		
				當然	1	亂	1		
				老	1	倒底	1		
				還有	1	剛好	1		
				還要	1	卻	1		
				非常	1	同	1		
				馬上	1	多	1		
						將要	1		
						幸好	1		
						從小	1		
						忽然	1		
						慢慢	1		
						每天	1		
						永遠	1		
						直	1		
						真是	1		
						細細	1		
						絕對	1		
						至	1		
						起先	1		
						這次	1		
						還好	1		
						還有	1		
						非常	1		

ADJ									
19-24 months		25-30 months		31-36 months		37-42 months		43-48 months	
word	freq	word	freq	word	freq	word	freq	word	freq
大	16	小	90	小	43	小	63	小	19
小	10	大	32	大	20	大	29	大	15
多	3	多	23	多	18	壞	19	多	9
好	1	臭	7	白	8	多	19	漂亮	2
好吃	1	壞	5	漂亮	4	好	6	舊	2
死	1	小小	5	另外	3	新	6	長	2
白	1	電動	3	假	2	胖	4	中大	1
紅	1	好	2	小小	2	小小	3	亂七八糟	1
胖	1	長	2	臭	2	臭	3	圓	1
		髒	2	黑	2	假	2	壞壞	1
		乖乖	1	上上	1	另外	2	好	1
		個別	1	尖	1	真	2	小小	1
		大大	1	掉	1	短	2	新	1
		懶惰	1	肥	1	冰	1	白	1
		新	1	胖胖	1	同一	1	香香	1
		有小	1	電動	1	圓	1		
		甜	1			破	1		
		白	1			笨	1		
		紅紅	1			變	1		
		老	1			頑皮	1		
		醜	1						
		香香	1						
		高	1						

ASP									
19-24 months		**25-30 months**		**31-36 months**		**37-42 months**		**43-48 months**	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
了	89	了	370	了	314	了	444	了	195
在	12	在	58	在	55	在	63	在	21
著	11	著	9	著	10	過	18	著	5
		過	7	過	6	著	16	過	3

CONJ									
19-24 months		**25-30 months**		**31-36 months**		**37-42 months**		**43-48 months**	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
因為	8	還有	51	然後	51	然後	92	然後	26
還有	6	因為	21	還有	44	還有	83	因為	15
然後	2	跟	13	可是	41	因為	71	跟	13
就是	1	可是	10	跟	29	跟	37	還有	11
或	1	和	4	因為	26	可是	24	可是	9
跟	1	就是	4	就是	21	就是	11	如果	4
		然後	4	不然	3	還是	9	的話	4
		但是	3	所以	3	和	7	還是	3
		不然	2	結果	2	結果	6	但是	2
		免得	1	但是	1	所以	2	和	2
		只是	1	只有	1	雖然	2	就是	2
		如果	1	和	1	不然	1	不然	1
		而且	1	要是	1	且	1	另外	1
		還是	1	還是	1	但	1	只有	1
						但是	1	結果	1
						只有	1	要是	1
						如果	1	還要	1
						或是	1		
						與	1		
						除非	1		

DE									
19-24 months		**25-30 months**		**31-36 months**		**37-42 months**		**43-48 months**	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
的	64	的	290	的	402	的	631	的	291
		得	14	得	8	得	23	得	16
		地	2					之	2
								地	1

DT									
19-24 months		**25-30 months**		**31-36 months**		**37-42 months**		**43-48 months**	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
這	182	這	664	這	855	這	831	這	377
那	8	那	90	那	78	那	217	那	124
		第	8	每	5	哪	6	每	1
		哪	5	哪	3	每	4		
		每	2	上	1	上	2		
		下	1	其	1	別	1		
						第	1		

INT									
19-24 months		**25-30 months**		**31-36 months**		**37-42 months**		**43-48 months**	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
嗯	150	嗯	559	嗯	285	嗯	384	嗯	157

INT									
19-24 months		**25-30 months**		**31-36 months**		**37-42 months**		**43-48 months**	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
啊	105	啊	119	哇	55	啊	135	那	57
哇	28	哇	66	啊	54	那	130	啊	43
哦	26	喔	38	那	32	喔	41	欸	40
嗚	14	那	29	喔	29	嘿	34	嘿	32
哎	9	哦	19	哎呀	23	哇	33	喔	23
耶	9	咦	15	哼	9	哎	31	蛤	19
咦	8	唉喲	11	哦	7	欸	28	喂	14
蛤	7	喂	10	唉	7	唉喲	25	呃	11
喔	6	哼	9	喂	7	哎喲	13	哇	11
嘿	6	嗚	9	嘿	7	喂	11	呀	8
呀	5	唉	7	阿弟	5	哦	10	哎喲	6
喂	5	嘿	7	咦	4	阿	10	唉唷	6
唉唷	4	噎	6	哎	4	唉唷	7	哦	5
噫	3	耶	6	嗚	4	咦	6	唉呀	5
哇哇	2	哎喲	5	嗯嗯	4	呃	5	咦	4
哎呀	2	嗯嗯	5	唉呀	3	唉	4	喔喔	4
哎喲	2	嗯嗯嗯	5	唉喲	3	嗚	4	阿	4
哼	2	哎	4	蛤	3	蛤	4	唔	3
唉呀	1	哎呀	4	阿	3	呀	2	欸喲	3
噯	1	哼哼哼	4	哎呦	2	呼	2	哎	1
阿	1	唉呀	4	喂喂喂	2	喂喂	2	哎呀	1
		哇啊	3	嗨	2	喂喂喂	2	唉	1
		唉唷	3	嘻	2	幹	2	唉呦	1
		啊啊	3	噯	2	呵	1	喲	1
		噢	3	呀	1	呵呵	1	嗨	1
		天	3	咿阿	1	哎呀	1	噢	1
		呀	2	哎呦喂	1	哎呦	1	噫	1
		哼哼	2	哎喲	1	哎耶	1	疑	1
		嗨	2	唉唷	1	哦哦	1		
		蛤	2	喀	1	唉呀	1		
		阿	2	喂喂喂喂	1	唉呦喂	1		
		呃	1	喲	1	唔	1		
		呦	1	嘎嘎	1	喔喔喔	1		
		咯	1	嘿嘿	1	喝	1		
		哇哇	1			嗯嗯	1		
		哇哇哇	1			嘻	1		
		啊呀	1			噢	1		
		啊啊啊	1			噫	1		
		喔咿喔	1			耶	1		
		喔喔	1			這	1		
		喲	1			那那	1		
		嘿嘿	1			那那那	1		
		噓	1						
		噯	1						

IDM									
19-24 months		**25-30 months**		**31-36 months**		**37-42 months**		**43-48 months**	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
ABC	1	ABCD	1	謝謝	13	謝謝	13	沒關係	5
謝謝	3	謝謝	13	九九九	8	不好意思	9	一二三	3
造飛機	2	不客氣	4	剪刀石頭布	8	再見	9	謝謝	2

IDM									
19-24 months		**25-30 months**		**31-36 months**		**37-42 months**		**43-48 months**	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
七八九	1	一二三	3	一二三	7	沒關係	6	一二三一二三四	1
七八九十	1	一二三五六七八	3	水中游	4	一二三	5	一二三四	1
三	1	對不起	3	依比呀呀	2	一二三四五六	4	一隻兩隻三隻四隻五隻六隻	1
不客氣	1	一二三五六	2	早	2	亂七八糟	4	再見	1
依比呀呀	1	一二三四	2	沒關係	2	一二三四五	3	吉普吉普吉吉普	1
傷腦筋	1	再見	2	火車汽車各幾部	2	早	3	咚咚隆咚嗆咚咚隆咚嗆	1
哥哥爸爸真偉大	1	受不了	2	一一二一二一二一	1	剪刀石頭布	2	對不起	1
啦啦啦啦	1	祝你生日快樂	2	一二	1	造飛機	2	皮膚癢	1
娃娃也出來走走走	1	ㄅㄆㄇ	1	一二一	1	一二	1	皮膚癢皮膚癢又不癢	1
小狗汪汪開汽車嘟嘟嘟	1	一七八八	1	一二三四	1	一二一二一二一	1	皮膚癢皮膚癢癢又癢香港腳香港腳	1
拜拜	1	一二	1	一二三四五六七八九十	1	一二三六五七八	1	英勇無敵	1
滴嗒嗒滴嗒嗒我在吹喇吧	1	一二三四五六	1	一零三五一零九	1	一二三四五六七	1	足爽擦一擦就不癢足爽擦一擦就不癢	1
自言自語	1	七八九五四三五四三五四七八九	1	三四一二三四	1	一二三四五六七八九	1		
這樣走走那樣走走	1	七八五六	1	三四五九	1	九五八零九九	1		
		三四	1	不客氣	1	五節	1		
		三四五	1	你好	1	傷腦筋	1		
		亂七八糟	1	依比呀呀啦	1	八八八八	1		
		二三	1	十一十二十三十四十五十六十七十八十九二十	1	八四二八	1		
		依比呀呀	1	哥哥爸爸	1	受不了	1		
		兩頭尖尖這麼大個	1	嗚嗚嗚數一數	1	哥哥爸爸	1		
		冰燕弟弟冰水積聚裡滴	1	嘰哩咕嚕滾下來	1	對不起	1		
		叮叮噹叮叮噹	1	對不起	1	小氣鬼喝涼水	1		
		四五六七八	1	小汽車嘟嘟嘟	1	小蜜蜂小蜜蜂小蜜小蜜蜂	1		
		四五六七八九	1	小火車嗚嗚嗚	1	我變我變我變變	1		
		搖到外婆橋	1	我們來數一數	1	沒有關係	1		
		搖搖到橋	1	拔蘿蔔	1	造飛機造飛機來到青草地	1		
		沒關係	1	生日快樂	1	魔鏡魔鏡 xx 上誰最漂亮	1		

IDM									
19-24 months		25-30 months		31-36 months		37-42 months		43-48 months	
word	freq	word	freq	word	freq	word	freq	word	freq
		爬呀爬呀過山河	1						
		老師好	1						
		胡扯蛋	1						
		臉一擠	1						
		臉一擠脖一縮	1						
		螃蟹一爪八個	1						

CL									
19-24 months		25-30 months		31-36 months		37-42 months		43-48 months	
word	freq	word	freq	word	freq	word	freq	word	freq
個	209	個	651	個	750	個	976	個	535
條	3	本	26	隻	26	隻	41	隻	9
隻	3	隻	20	張	17	台	21	把	8
支	2	張	7	次	12	本	17	根	5
口	1	腳	7	台	11	種	12	色	5
天	1	件	5	本	10	次	11	台	3
張	1	棵	5	口	9	歲	9	張	3
種	1	次	5	片	6	件	8	捲	3
部	1	朵	4	堆	5	塊	8	次	3
面	1	顆	3	條	5	條	6	天	2
		包	2	種	5	輛	6	本	2
		口	2	些	4	些	5	種	2
		名	2	天	4	天	5	分鐘	1
		天	2	件	3	口	4	包	1
		卷	1	把	3	支	4	層	1
		台	1	年	2	頁	3	樣	1
		塊	1	輛	2	分鐘	2	歲	1
		家	1	鍋	2	張	2	段	1
		杯	1	頂	2	把	2	盒	1
		枝	1	公分	1	片	2	碗	1
		瓶	1	公斤	1	顆	2	輛	1
		盤	1	公里	1	下	1	遍	1
		碗	1	刀	1	位	1		
		間	1	塊	1	公里	1		
		雙	1	層	1	包	1		
		面	1	朵	1	對	1		
				杯	1	層	1		
				滴	1	手	1		
				瓶	1	排	1		
				筆	1	朵	1		
				罐	1	眼	1		
				艘	1	碗	1		
				顆	1	篇	1		

MOD									
19-24 months		25-30 months		31-36 months		37-42 months		43-48 months	
word	freq	word	freq	word	freq	word	freq	word	freq
會	14	會	219	會	141	會	263	會	101
能	3	可以	38	可以	46	可以	111	可以	64
一定	2	要	7	能	17	能	22	能	5

可以	2	真的	6	真的	13	可	5	真的	3
		好	4	可	3	應該	4	准	1
		能	3	用	3	敢	4	敢	1
		敢	2	一定	1	該	4	用	1
		用	2	敢	1	一定	1		
		必	1			用	1		
		該	1			真的	1		

NEG									
19-24 months		**25-30 months**		**31-36 months**		**37-42 months**		**43-48 months**	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
不	166	不	692	不	473	不	735	不	392
沒有	138	沒有	220	沒有	250	沒有	298	沒有	84
沒	5	沒	16	沒	26	沒	30	沒	19

ONM									
19-24 months		**25-30 months**		**31-36 months**		**37-42 months**		**43-48 months**	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
嗶	29	嘻嘻	34	哈	20	哈	17	哈	11
嘟嘟	8	嘻嘻嘻	10	嘻嘻	11	嘻	9	嘻	9
哈	7	嘟	9	喀擦	8	喵	7	嘿嘿	6
嘟	7	咻	7	咻	5	叩	5	碰	6
汪	7	叮噹	6	汪	5	咚	5	嘟	4
汪汪	7	呱呱	6	叩	4	咻	5	呃	3
咚咚	3	咚	6	嘻	4	哈哈哈	5	咻	3
咳	3	嘻	6	哈哈	3	哈哈	4	哈哈	3
呱	2	哈	5	哈哈哈	3	嗡嗡嗡	3	喀拉喀拉喀拉	3
咕咕咕	2	汪	5	吱吱	2	咪咪	2	咕	2
嘰	2	汪汪	5	咕咕	2	噓	2	咚咚	2
叮噹	1	碰	5	咚	2	噗	2	叱叱	1
吓	1	叩	4	咚咚咚	2	欸欸	2	呵呵呵	1
咻	1	呱	4	嘟	2	劈剖劈剖劈剖	1	咚咚咚	1
哈哈	1	嗚	4	嘰咕	2	咚咚	1	咪	1
喵	1	咻咻	3	angangangangang	1	哇哇哇	1	哎呀呀	1
嘻	1	咻咻咻	3	叩叩	1	啪	1	哎哎哎	1
碰碰	1	哈哈	3	咕嘰	1	啾啾啾	1	啪	1
郭郭郭	1	哈哈哈	3	咕嚕咕嚕	1	喀喀喀	1	啪呃	1
		砰砰砰	3	咚咚	1	嗡嗡嗡嗡	1	啪呃啊	1
		叮咚	2	哇哇	1	嘟嘟嘟	1	啾啾啾	1
		咚咚	2	嘟嘟	1	嘰	1	嘎嘎	1
		哈啾	2	嘟嘟嘟	1	嘰剖啦剖啦	1	嘎嘎 ki 嘎嘎	1
		啾啾	2	嘰哩咕嚕	1	嘻嘻	1	嘟嘟	1
		嗶	2	嘰嘰	1	嘻嘻嘻	1	嘟嘟嘟	1
		嘟嘟	2	嘰嘰瓜瓜	1	嘿嘿嘿嘿	1	嘟嘟嘟嘟	1
		砰	2	嘻嘻嘻	1	碰碰碰	1	嘰嘰嘰嘰	1
		砰砰	2	嘻嘻嘻嘻	1	蹦蹦	1	嘿嘿嘿	1
		叭	1	嘿嘿	1	蹦蹦蹦	1	戈戈戈戈	1
		叮	1	布穀	1	鈈呀鈈呀	1	挪	1
		呦鳴	1	汪汪	1	鈈鈈鈈鈈鈈鈈	1	澎澎澎	1
		呸	1	滴滴	1			登登	1
		呼	1	砰	1			碰碰	1
		咕咕咕	1	碰	1			空	1
		咚咚咚咚	1	碰碰碰	1				

ONM									
19-24 months		25-30 months		31-36 months		37-42 months		43-48 months	
word	freq	word	freq	word	freq	word	freq	word	freq
		咚							
		哞	1						
		哞哞	1						
		唧	1						
		喵喵	1						
		嗚嗚	1						
		嘰	1						
		嘿嘿嘿	1						
		恰	1						
		砰砰砰砰	1						
		砰砰砰砰砰砰	1						
		碰碰	1						

PREP									
19-24 months		25-30 months		31-36 months		37-42 months		43-48 months	
word	freq	word	freq	word	freq	word	freq	word	freq
在	21	在	78	在	72	把	80	跟	25
把	8	把	32	把	56	在	68	把	23
用	2	用	17	從	22	用	33	用	12
		幫	16	跟	17	跟	31	從	11
		跟	12	用	11	被	14	在	10
		從	8	被	10	幫	12	被	7
		被	8	幫	8	和	9	比	3
		和	2	和	6	從	5	像	2
		對	2	按	2	到	4	按	2
		比	2	比	2	對	3		
		像	1	到	1	比	3		
		到	1	對	1	往	2		
		將	1	把他	1	像	1		
		為	1	照	1	按	1		
		給	1	等	1	為	1		
				給	1	照	1		
						給	1		
						遭	1		

QN									
19-24 months		25-30 months		31-36 months		37-42 months		43-48 months	
word	freq	word	freq	word	freq	word	freq	word	freq
一	67	一	135	一	213	一	228	一	83
好多	10	兩	25	兩	28	兩	60	兩	41
五	6	三	15	三	22	三	26	五	7
兩	5	好多	14	二	17	四	16	三	6
二	3	五	8	六	16	六	12	四	4
四	3	一點	7	0	12	二	10	別的	3
多少	2	七	6	五	12	八	10	一下	2
一半	1	八	4	四	12	五	9	一半	2
一點	1	一點點	3	八	10	好多	7	一點點	2
七	1	九	3	十	7	一點	6	全部	2
三	1	六	3	七	6	七	5	好多	2
六	1	四	3	一點	4	一下	4	幾	2
十	1	幾	3	好多	4	全部	4	有的	2
		二	2	一下	3	十	4	七	1

QN									
19-24 months		**25-30 months**		**31-36 months**		**37-42 months**		**43-48 months**	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
		一下	1	九	3	幾	4	二	1
		一半	1	一零	2	第三	4	八	1
		十	1	一點點	2	第五	4	十二	1
		全部	1	三九一七八六四	2	一點點	3	半	1
		多	1	零	2	二十二	3	多少	1
		百百	1	一二三五五五	1	多少	3	整	1
		整	1	一二三四五	1	九	2	第一	1
				一片片	1	二十三	2	第三	1
				三十一	1	台	2		
				三百	1	第二	2		
				二零	1	一七四	1		
				別的	1	一些	1		
				半	1	一來	1		
				幾	1	三百	1		
				整個	1	三點	1		
				有	1	二十一	1		
				滿	1	二十四	1		
				點	1	元元	1		
						全	1		
						兩百	1		
						六百	1		
						別的	1		
						十二	1		
						壹	1		
						好幾	1		
						第一	1		
						頁頁	1		

SFP									
19-24 months		**25-30 months**		**31-36 months**		**37-42 months**		**43-48 months**	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
啊	69	啊	221	啊	245	啊	596	啊	236
啦	50	啦	129	啦	125	呀	229	啦	197
耶	29	喔	90	呀	103	啦	216	呢	60
哦	16	呀	74	嘛	81	喔	187	喔	52
勒	11	嘛	44	喔	70	呢	97	呀	50
喔	11	耶	44	呢	61	耶	96	耶	38
呀	9	呢	35	耶	57	嘛	44	嘛	31
哇	9	嗎	31	嗎	38	嗎	34	吧	23
呢	8	吧	19	哼	16	吧	33	哦	23
嗎	8	哦	18	哦	13	哼	32	嗎	15
嘛	7	勒	9	吧	11	哦	27	哪	12
囉	4	咧	7	哇	8	囉	20	囉	12
吶	3	囉	7	勒	6	噢	17	勒	11
了	2	喲	6	好了	6	哪	15	哼	10
吧	2	吶	2	吶	5	欸	10	欸	8
喂	2	哇	2	了	3	好了	6	噢	4
唷	1	唷	2	囉	3	哇	5	吶	2
啊啊	1	好了	2	咧	2	勒	4	咧	2
喲	1			唷	1	吶	4	喲	2
				啦啦啦	1	唷	4	哇	1
				而已	1	咧	3	唷	1
				阿	1	而已	3	阿	1

SFP									
19-24 months		**25-30 months**		**31-36 months**		**37-42 months**		**43-48 months**	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
						阿	3		
						哩	2		
						喲	2		
						唉	1		
						嗯	1		
						嘿	1		
						好啊	1		

WH									
19-24 months		**25-30 months**		**31-36 months**		**37-42 months**		**43-48 months**	
word	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**	**word**	**freq**
誰	41	什麼	104	為什麼	118	什麼	160	什麼	93
什麼	24	哪裡	90	怎麼	96	為什麼	77	怎麼	62
怎麼	7	誰	66	什麼	84	怎麼	74	為什麼	25
哪裡	4	怎麼	45	哪裡	28	怎麼樣	34	哪	20
那裡	3	為什麼	13	誰	22	誰	26	怎麼樣	14
幹嗎	2	那裡	8	幹嘛	11	哪裡	19	幹嘛	9
哪	1	幹嘛	4	那裡	6	哪	11	誰	9
幹嘛	1	怎	3	哪	5	幹嘛	7	哪裡	5
		幹嗎	2	那	2	哪裏	2	哪裏	3
		哪裏	1	怎	1	哪邊	2	怎樣	2
		那	1	那邊	1	怎樣	2	怎	1
						怎麼辦	1		
						那裡	1		

Appendix 5: Lists of nouns

Nouns from 19-24 months

People													
Superordinate		Basic-level						Subordinate					
Type/Token		34 Type / 438 Token						47 Type / 191 Token					
		媽媽	168	爹地	5	小朋友	1	巴比	40	米老鼠	3	P 阿姨	1
		阿姨	95	奶奶	4	老婆婆	1	阿拉丁	20	趙晶晶	3	傑克	1
		爸爸	49	妹妹	4	叔	1	嘉嘉	18	陳應之	3	利	1
		姊姊	14	阿伯	4	爸	1	斑比	16	唐老鴨	2	史努比	1
		媽	12	阿嬤	4	阿兵哥	1	阿丁	14	軒軒	2	吳小秀	1
		人	11	老婆	3	阿婆	1	二姊	5	阿比	2	廖芳芳	1
		寶寶	9	叔叔	3	姨	1	嘉	5	阿開姊姊	2	楊小風	1
		老公公	7	姊	3	獵人	1	阿卜	5	仙女	1	牛媽	1
		弟弟	7	新娘子	3	壞人	1	雞婆	4	堂叔	1	球球	1
		哥哥	6	爺爺	3	魔鬼	1	F 阿姨	4	堂叔叔	1	白雪公主	1
		姑姑	5	老師	2			阿琳	4	多多	1	聖誕老公公	1
		阿公	5	媽咪	2			T 阿姨	3	強棒	1	阿弟	1
								Y 姊姊	3	精靈	1	阿拉	1
								安安叔叔	3	貝比	1	馨馨	1
								小元	3	C 阿姨	1	龍貓	1
								小紅帽	3	L 阿姨	1		
Pronouns													
Superordinate		Basic-level						Subordinate					
Type/Token		9 Type / 239 Token						Type/Token					
		我	84	你	15	那	3						
		他	66	自己	9	我們	1						
		這	57	他們	3	這些	1						
Body Parts													
Superordinate		Basic-level						Subordinate					
Type/Token		13 Type/ 46 Token						2 Type/ 2 Token					
		手	13	鼻子	3	骨頭	2	腸子	1				
		頭	6	大便	2	手手	1	鼻涕	1				
		聲音	5	屁股	2	眼睛	1						
		嘴巴	4	耳朵	2								
		眉毛	3	腳	2								
Clothing													
Superordinate		Basic-level						Subordinate					
Type/Token		8 Type/ 21 Token						1 Type/ 1 Token					
		鞋	9	背心	1			拖鞋	1				
		鞋子	5	襪子	1								
		帽子	2	襪襪	1								
		帽帽	1	鞋鞋	1								
Vehicles													
Superordinate		Basic-level						Subordinate					
Type/Token		5 Type/ 54 Token						15 Type/ 45 Token					
		車子	27					火車	12	垃圾車	2	紅燈	1
		車車	11					公共汽車	8	火車頭	2	老虎土機	1
		飛機	8					卡車	5	計程車	2	貨車	1
		車	6					綠燈	3	飛狗巴士	2	車頭	1
		汽車	2					灰狗巴士	3	挖土機	1	自強號	1
Tools													
Superordinate		Basic-level						Subordinate					
1 Type/ 1 Token		47 Type/ 99 Token						8 Type/16 Token					
機器	1	錶	9	釘子	2	火柴	1	鉛筆	4				
		錶錶	9	錄音機	2	燈燈	1	拉雜書	3				
		報紙	6	電話	2	盒	1	手槍	2				
		書	6	傘	1	相片	1	金龜油	2				
		油漆	4	包	1	相簿	1	雨傘	2				
		筆	4	匙	1	網子	1	機關槍	1				
		被被	4	叉子	1	膠水	1	鋼筆	1				
		收音機	3	壺壺	1	蠟燭	1	電扇	1				
		燈	3	扇子	1	袋	1						

		錄音帶	3	手把	1	釘	1						
		錢	3	手電筒	1	鏡	1						
		錶錶錶	3	把把	1	鑰匙	1						
		時鐘	2	斧頭	1	電池	1						
		紅包	2	杯子	1	電燈	1						
		自來水	2	枕頭	1	電視	1						
		藥藥	2	橡皮	1								
Furniture													
Superordinate		Basic-level						Subordinate					
Type/Token		4 Type/ 8 Token						Type/Token					
		椅子	3	床	1								
		門	3	椅椅	1								
Food													
Superordinate		Basic-level						Subordinate					
3 Type/ 9 Token		21 Type/ 76 Token						5 Type/ 13 Token					
飲料	5	果汁	18	葡萄	3	柳丁	1	肝油	6				
菜	3	果凍	9	餅乾	3	湯	1	冰淇淋	2				
水果	1	水	9	布丁	2	牛奶	1	胡蘿蔔	2				
		蘋果	7	熱狗	2	芭樂	1	西瓜	2				
		糖糖	4	菜菜	2	茶	1	巧克力	1				
		油	3	蛋糕	2	茶茶	1						
		糖	3	咖啡	1	養樂多	2						
Animals													
Superordinate		Basic-level						Subordinate					
1 Type/1 Token		38 Type/ 183 Token						6 Type/12 Token					
動物	1	魚	30	狗狗	5	蟑螂	2	吸玻璃魚	4				
		豬豬	11	鳥鳥	5	阿狗	2	蝌蚪	3				
		鳥	11	猴子	4	鹿鹿	2	野狼	2				
		豬	10	蛇	4	公	1	大肚魚	1				
		馬	10	蟲	4	咪咪	1	鯊魚	1				
		狗	9	鹿	4	小狗	1	鳥兔	1				
		貓咪	8	海豹	3	海豚	1						
		貓	7	青蛙	3	牛牛	1						
		長頸鹿	7	兔	2	老虎	1						
		烏龜	6	大象	2	蜘蛛	1						
		螞蟻	6	河馬	2	蟲蟲	1						
		鴨鴨	6	牛	2	雞	1						
		小鳥	5	獅子	2								
Natural phenomena and natural materials													
Superordinate		Basic-level						Subordinate					
Type/Token		15 Type/ 47 Token						1 Type/ 3 Token					
		泡沫	10	太陽	3	洞洞	1	山洞	3				
		花	8	星星	2	湖	1						
		石頭	6	月亮	2	草	1						
		雨	5	土	1	電	1						
		洞	4	天空	1	星	1						
Toys													
Superordinate		Basic-level						Subordinate					
3 Type/ 18 Token		12 Type/ 33 Token						2 Type/ 2 Token					
玩具	14	球球	8	娃娃	3	滑梯	1	五木棋	1				
具具	3	球	6	泡泡	2	球球球	1	象棋	1				
具	1	小木偶	4	不倒翁	1	神燈	1						
		積木	4	嗶嗶	1	鞭炮	1						
Colors													
Superordinate		Basic-level						Subordinate					
1 Type/1 Token		3 Type/ 18 Token						1 Type/3 Token					
色	1	紅色	13					米色	3				
		藍色	3										
		黃色	2										
Shapes													
Superordinate		Basic-level						Subordinate					
Type/Token		Type/Token						Type/Token					
Locations													
Superordinate		Basic-level						Subordinate					
Type/Token		11 Type/ 28 Token						7 Type/ 13 Token					

		馬路	9	樓梯	2	水溝	1	中正紀念堂	5	跑道	1		
		動物園	4	美國	2	馬戲團	1	阿婆家	3	中紀堂	1		
		地板	3	家	1	屏東	1	二樓	1	東方百貨	1		
		醫院	3	房子	1			教室	1				
Spatial words													
Superordinate		Basic-level						Subordinate					
Type/Token		11 Type/ 79 Token						3 Type/3 Token					
		這裡	36	上面	2	那邊	2	圖片上	1				
		裡面	19	北	2	外面	1	水裡	1				
		這邊	7	後面	2	邊	1	山上	1				
		那裡	5	裡	2								
Numerals													
Superordinate		Basic-level						Subordinate					
1 Type/1 Token		Type/Token						Type/Token					
號碼	1												
Abstract nouns													
Superordinate		Basic-level						Subordinate					
Type/Token		7 Type/19 Token						4 Type/ 12 Token					
		從前	6	今天	2			交通音樂	6				
		音樂	5	名字	1			大家一起來	3				
		寶貝	2	巧妃	1			新年	2				
		話	2					強棒出擊	1				

Nouns from 25-30 months

Abstract nouns													
Superordinate		Basic-level						Subordinate					
Type/Token		54 Type/1094 Token						92 Type/ 694 Token					
		媽媽	321	阿婆	7	阿祖	3	乖乖	160	小燕姊姊	4	小妹妹	1
		阿姨	149	姐姐	6	人人	2	JC	76	假叔叔	3	小姐姐	1
		爸爸	137	小姐	6	壞人	2	軒軒	51	大姊	3	日本新娘	1
		媽	132	怪物	6	外婆	2	J 阿姨	47	天使	3	木頭閒人	1
		阿公	53	寶寶	5	巫婆	2	巴比	40	Y 姑姑	3	C 舅舅	1
		小丑	31	小朋友	5	醫生	2	小柔	30	元	3	F 阿姨	1
		姊姊	25	叔叔	4	鬼	2	嘉嘉	21	哈利	3	光光	1
		哥哥	24	婆	4	外公	1	米老鼠	16	嘟嘟	3	吳漢漢	1
		奶奶	21	媽咪	4	女	1	唐老鴨	13	多力	3	吳阿秀	1
		人	16	小偷	4	姊	1	乖	11	安可姊姊	3	周	1
		婆婆	16	新娘子	4	姑姑	1	馨馨姊姊	10	小乖	3	多多龍	1
		妹妹	13	爸	4	姨媽	1	J	9	小矮人	3	天龍特攻隊	1
		新娘	11	阿嬤	4	客人	1	聖誕老公公	9	珍珍	3	威威	1
		姨	9	伯伯	3	市長	1	二姊	8	仙女	2	小于	1
		阿媽	9	哥	3	王子	1	頑皮豹	8	大哥哥	2	小燕	1
		魔鬼	9	新郎	3	老公公	1	白雪公主	8	小阿姨	2	小珍	1
		公主	7	警察	3	舅舅	1	陳水扁	8	幼兒	2	小白	1
		爺爺	7	閒人	3	阿祖	1	大爸爸	7	美少女	2	小英姊姊	1
								咪咪小花貓	7	阿 zo	2	小阿姨	1
								老巫婆	6	LJC	2	小雅	1
								小乖乖	6	T 阿姨	2	巴	1
								湯姆	6	小咪姑姑	2	扁	1
								姨嬤	5	小珍珍	2	比比	1
								瑪雅	5	小紅帽	2	水扁	1
								阿琳	5	小達	2	潘	1
								H 老師	4	泰山	2	玲玲	1
								仔仔	4	咪咪大爸爸	1	皮皮	1
								元元	4	咪咪小花貓	1	軒	1
								唐	4	土地公	1	鍾甲甲	1
								小叮噹	4	大爸	1	阿丁	1
												頑皮豹	1
												黃黃洲	1
Pronouns													
Superordinate		Basic-level						Subordinate					
Type/Token		13 Type/ 1218 Token						1 Type/ 1 Token					
		我	581	自己	24	別	3	在下	1				
		你	250	我們	20	人家	1						
		他	195	人家	9	你們	1						
		這	90	別人	5								
		那	36	他們	3								
Body Parts													
Superordinate		Basic-level						Subordinate					
Type/Token		28 Type/ 172 Token						9 Type/ 21 Token					
		眼睛	26	頭	6	嘴	1	鼻涕	7				
		手	24	大便	5	尿	1	右腳	4				
		腳	21	肚子	3	屁屁	1	大嘴巴	2				
		腳腳	15	脖子	3	指甲	1	牙縫	2				
		嘴巴	13	鬍子	3	牙齒	1	香港腳	2				
		頭髮	11	屁股	2	痱子	1	腳趾	1				
		手手	8	皮	2	背背	1	膝	1				
		耳朵	8	眉毛	2	頭頭	1	長腳	1				
		聲音	8	背	2	髮髮	1	頭腦	1				
						鼻子	1						
Clothing													
Superordinate		Basic-level						Subordinate					
1 Type/ 21 Token		13 Type/ 38 Token						7 Type/ 21 Token					
衣服	21	鞋子	10	帶子	2	褲子	2	拖鞋	10	雨衣	2		
		鞋	6	帽子	2	髮夾	2	內褲	3	溜冰鞋	1		
		鞋鞋	5	手套	2	帽	1	托鞋	2	運動鞋	1		
		尿布	2	褲	2	戒指	1	涼鞋	2				

						新娘衣服	1						
Vehicles													
Superordinate		Basic-level						Subordinate					
Type/Token		7 Type/ 153 Token						27 Type/ 94 Token					
		車	49	紅綠燈	7			火車	16	跑車	3	輪子	2
		飛機	48	汽車	1			禮賓車	11	卡車	2	大卡車	1
		車車	26	波機	1			腳踏車	11	喇叭車	2	大大船	1
		車子	21					乖乖車	7	戰鬥機	2	摩托車	1
								吉普車	5	戰鬥飛機	2	直昇機	1
								三輪車	4	計程車	2	老爺車	1
								娃娃車	4	警察車	2	蒸汽火車	1
								救護車	4	賓車	2	象車子	1
								公共汽車	3	賽車	2	光陽汽車	1
Tools													
Superordinate		Basic-level						Subordinate					
3 Type/ 28 Token		74 Type/ 254 Token						27 Type/ 63 Token					
東西	25	剪刀	11	管子	4	刀	1	手提包	5	小牛頓	3	小玻璃	1
機器	2	書	11	蓋子	4	吹風機	1	手槍	5	表非鳴	3	抹布	1
工具	1	湯匙	11	麥克風	4	娃娃東西	1	蠟筆	5	垃圾桶	2	毛線	1
		錄音帶	11	口紅	3	扣扣	1	大哥大	4	故事書	2	水瓶	1
		玻璃	9	吸塵器	3	掃把	1	萬金油	4	白旗	2	煮飯匙	1
		藥	9	吸管	3	旗子	1	茶杯	3	皮包	2	瓦斯爐	1
		藥藥	8	手帕	3	板子	1	路燈	3	簽字筆	2	鉛筆	1
		子彈	7	被被	3	板擦	1	鉛筆盒	3	面速力達母	2	降落飛傘	1
		碗	7	包	2	棉被	1	電鍋	3	奶奶瓶	1	雨傘	1
		鍋子	7	拖把	2	照相機	1						
		電池	7	插頭	2	燈台	1						
		刀刀	6	放大鏡	2	相框	1						
		眼鏡	6	棍子	2	筷子	1						
		筆	6	燈	2	耳機	1						
		錄音機	6	畫	2	蠟燭	1						
		錢錢	6	痱子粉	2	錢	1						
		刀子	5	盤子	2	鍋	1						
		叉子	5	相片	2	鏟	1						
		報紙	5	票	2	鏡	1						
		杯	5	肥皂	2	鐘	1						
		槍	5	錄影帶	2	鑰匙	1						
		鏟子	5	電話	2	電燈	1						
		鏡子	5	電風扇	2	電腦	1						
		杯子	4	保麗龍	1	電視	1						
		毛巾	4	冰箱	1								
Furniture													
Superordinate		Basic-level						Subordinate					
Type/Token		5 Type/ 13 Token						1 Type/ 5 Token					
		椅子	6					壁櫥	5				
		門	3										
		櫃子	2										
		抽屜	1										
		桌子	1										
Food													
Superordinate		Basic-level						Subordinate					
3 Type/21 Token		53 Type/ 228 Token						35 Type/ 74 Token					
菜	19	糖糖	21	飯飯	4	麵包	2	巧克力	7	野糖	2	舒跑	1
水果	1	水	16	餅	4	包子	1	燒餅	6	X 子果汁	1		
飲料	1	漢堡	16	餅乾	4	大蒜	1	百百果汁	6	叉燒飯	1		
		果汁	12	香蕉	4	橘子	1	紅蘿蔔	5	大燒派	1		
		豆漿	12	冰冰	3	瓜	1	乖乖熊	4	棒棒糖	1		
		飯	12	油	3	米米	1	玉米湯	3	楊桃汁	1		
		咖啡	11	茶	3	肉	1	葡萄汁	3	油蛋	1		
		蛋	9	薯條	3	菜菜	1	雞肉	3	泡麵	1		
		麵	7	蘋果	3	蓮霧	1	冰冰淋	2	牛奶餅	1		
		湯	6	餅餅	3	蜂蜜	1	冰淇淋	2	芹菜	1		
		牛奶	5	饅頭	3	豆	1	哈蜜瓜	2	草莓糖	1		
		糖果	5	奶	2	豆豆	1	小魚煎蛋	2	蘑菇	1		

		葡萄	5	汽水	2	飼料	1	巧克力糖	2	豆乾	1		
		蛋糕	5	草莓	2	餅干	1	木瓜	2	豆甘	1		
		蛋蛋	5	蝦	2	養樂多	1	炸雞	2	雞皮	1		
		布丁	4	開水	2	鳳梨	1	狗狗菜	2	香菇	1		
		糖	4	香焦	2	麵麵	1	蛋炒飯	2	高麗菜	1		
		蜜	4	鮮奶	2								
Animals													
Superordinate		Basic-level						Subordinate					
1 Type/5 Token		68 Type/559 Token						22 Type/47 Token					
動物	5	熊熊	62	牛	8	鶴	2	鸚鵡	10				
		咪咪	53	青蛙	8	鹿	2	吸玻璃魚	5				
		魚	42	狐狸	7	鹿鹿	2	黨魚	5				
		河馬	30	貓	7	企鵝	1	斑馬	4				
		蛇	25	蝴蝶	6	孔雀	1	烏鴉	3				
		大象	19	蟲蟲	6	小熊	1	小白兔	2				
		狗狗	18	尾巴	5	小鳥	1	毛毛蟲	2				
		老鼠	18	獅子	5	犀牛	1	金魚	2				
		鴨鴨	15	兔寶寶	4	狼	1	六七蛇	1				
		恐龍	14	喵喵	4	老鷹	1	圈圈蛇	1				
		兔	13	蚊子	4	蚯蚓	1	大大魚	1				
		熊	13	蟲	4	蜈蚣	1	大魚蛇	1				
		兔兔	12	豬	4	蜜蜂	1	天鵝	1				
		馬	12	鴨子	4	蝦子	1	小丑魚	1				
		鴨	12	鵝	4	蝸牛	1	小鴨	1				
		狗	11	蜘蛛	3	螳螂	1	狗熊	1				
		猴子	11	小狗	2	蟑螂	1	白兔	1				
		烏龜	10	毛	2	袋鼠	1	白狗	1				
		老虎	10	猴猴	2	豹	1	綿羊	1				
		螞蟻	10	翅膀	2	雞	1	螢火蟲	1				
		長頸鹿	10	蛇蛇	2	鴕鳥	1	野狼	1				
		鳥	9	貓咪	2	龍龍	1	鯨魚	1				
		兔子	8	貓貓	2								
Natural phenomena and natural materials													
Superordinate		Basic-level						Subordinate					
Type/Token		20 Type/ 87 Token						5 Type/ 19 Token					
		花	35	池塘	2			山洞	13				
		太陽	6	石頭	2			聖誕樹	3				
		木頭	6	刺	1			大水	1				
		樹	6	地球	1			大海池	1				
		洞	5	月亮	1			哈雷彗星	1				
		草	5	樹葉	1								
		星星	4	河	1								
		樹木	3	泥土	1								
		花花	3	煙	1								
		天	2	電	1								
Toys													
Superordinate		Basic-level						Subordinate					
1 Type/ 16 Token		23 Type/ 93 Token						9 Type/ 22 Token					
玩具	16	娃娃	16	帶帶	2	電影	1	籃球	5				
		琴	14	泡泡	2	面具	1	象棋	5				
		球	13	吉他	1	鞭砲	1	氣球	3				
		小木偶	7	單棍	1			汽球	3				
		煙火	7	單槓	1			卡通影片	2				
		球球	7	嚕叮	1			國旗	1				
		圖	4	機器人	1			彈珠	1				
		鈴鐺	4	溜滑梯	1			新娘娃娃	1				
		喇叭	3	積木	1			鋼琴	1				
		鞭炮	3	跳繩	1								
Colors													
Superordinate		Basic-level						Subordinate					
2 Type/ 7 Token		9 Type/ 52 Token						2 Type/ 5 Token					
顏色	4	紅色	17	咖啡色	3			粉紅色	4				
色	3	綠色	12	白色	2			橘紅	1				
		黃色	7	白	1								

		橘色	5	黑色	1								
		紫色	4										
Shapes													
Superordinate		Basic-level						Subordinate					
Type/Token		3 Type/ 5 Token						Type/Token					
		勾勾	2										
		圈圈	1										
		手印	1										
Locations													
Superordinate		Basic-level						Subordinate					
2 Type/ 14 Token		28 Type/ 88 Token						22 Type/ 44 Token					
地方	12	家	19	大陸	3	購物中心	1	福利中心	8	土地公廟	1	健寶	1
空間	2	房間	14	美國	3	電梯	1	台大	5	媽媽家	1	固特異	1
		飛機場	5	房子	2	馬路	1	遠東購物中心	4	教室	1		
		動物園	4	教會	2	高樓大廈	1	阿里山	4	書房	1		
		路	4	樓梯	2	台北	1	虎頭山	3	爺爺家	1		
		德國	4	水溝	2	台北市	1	廚房	2	牆腳	1		
		停車位	3	加油站	1	日本	1	八德鄉	2	石頭路	1		
		地板	3	工廠	1	桃園縣	1	新民幼稚園	2	阿公家	1		
		學校	3	樓	1			主臥室	1	雅雅家	1		
		嘉義	3	牆壁	1			叔叔家	1	人人幼稚園	1		
Spatial words													
Superordinate		Basic-level						Subordinate					
Type/Token		20 Type/ 393 Token						17 Type/ 33 Token					
		這裡	222	下	3			地上	5	桌子下	1		
		這邊	54	上	2			家裡	5	樹上	1		
		裡面	34	裡	2			樓下	4	照片裡	1		
		那邊	15	邊	2			嘴巴裡	2	筒子裡	1		
		外面	13	一邊	1			床床上	2	腳上	1		
		上面	10	前面	1			手上	2	袋袋裡	1		
		後面	9	右	1			桌子上	2	頭上	1		
		旁邊	8	那邊	1			肚子裡	2				
		那裡	8	背面	1			地下	1				
		右邊	5	邊	1			山洞裡	1				
Numerals													
Superordinate		Basic-level						Subordinate					
0 Type/ 0 Token		0 Type/ 0 Token						0 Type/ 0 Token					
Abstract nouns													
Superordinate		Basic-level						Subordinate					
Type/Token		30 Type/ 86 Token						15 Type/ 32 Token					
		現在	24	運動	2	生日	1	兒童天地	7	讚美詩	1		
		故事	10	怪頭	2	辦法	1	光陽新生代	6	跟屁蟲	1		
		味道	7	以前	2	關係	1	英文	3	幼寶班	1		
		歲	4	以後	2	體操	1	大家一起來	2	無敵鐵金剛	1		
		音樂	4	昨天	2	今天	1	灰姑娘	2	青城派	1		
		事	3	個性	1	剛才	1	醜小鴨	2				
		一下	3	力氣	1	後來	1	小時候	2				
		呆子	2	字	1	明天	1	印尼話	1				
		時候	2	星媽	1	晚上	1	咪咪故事	1				
		話	2	決戰	1	白天	1	新娘故事	1				

Nouns from 31-36 months

People													
Superordinate		Basic-level						Subordinate					
2 Type/ 2 Token		48 Type/ 593 Token						78 Type/ 462 Token					
傢伙	1	媽媽	185	醫生	7	小丑	2	乖乖	153	阿弟	2	克地老師	1
兄弟	1	阿姨	70	妹妹	5	工人	2	乖	47	光小姐	1	加菲貓	1
		爸爸	56	新娘子	5	老婆婆	2	J 阿姨	46	光華	1	吳阿秀	1
		新娘	30	新郎	5	阿媽	2	JC	33	光華伯	1	大衛叔叔	1
		媽	25	舅舅	5	女兒	1	軒軒	17	天使	1	寇力	1
		人	23	國王	4	姐姐	1	美女	12	好朋友	1	寇力老師	1
		小朋友	23	伯伯	3	孩子	1	桃太郎	10	好貝	1	寧寧姐姐	1
		奶奶	16	司機	3	小孩	1	強強	9	媽媽老師	1	小仁	1
		阿公	16	姨	3	怪物	1	阿秀	9	小人	1	小揚哥哥	1
		老師	15	寶寶	3	朋友	1	妖怪	8	小妞妞	1	小東東	1
		哥哥	13	爺爺	3	警察	1	護士小姐	8	小姊姊	1	小涵	1
		小姐	10	同事	2	郵差	1	小飛俠	7	小子	1	小貞	1
		弟弟	10	哥	2	阿兵哥	1	小柔	5	小寶貝新娘	1	小馬叔叔	1
		叔叔	9	壞人	2	阿婆	1	小秀	5	強盜	1	小黃	1
		媽咪	9	大人	2	阿嬤	1	小美人魚	5	海盜船長	1	巴比	1
		老公公	7	姊姊	2	鬼	1	白雪公主	5	熊熊新娘	1	方芳	1
								T 阿姨	4	熊熊老師	1	東東	1
								一休	4	白雪	1	海方礦工	1
								大師傅	3	米老鼠老師	1	珍珍	1
								小阿姨	3	精靈	1	米老鼠	1
								皮媽	3	聖誕老公公老師	1	蝙蝠俠	1
								奇奇	3	軒	1	阿袖	1
								小叮噹	3	G 阿姨	1	雷東東	1
								青青阿姨	3	J	1	青青	1
								光頭阿公	2	Lily 阿姨	1	頑皮豹	1
								小飛俠	2	W	1	龍龍	1
Pronouns													
Superordinate		Basic-level						Subordinate					
Type/Token		16 Type/ 1235 Token						Type/Token					
		我	544	自己	14	其他	2						
		你	302	別人	6	這些	1						
		他	255	人家	5	那些	1						
		他們	42	那	4	別	1						
		我們	39	大家	3								
		這	14	你們	2								
Body Parts													
Superordinate		Basic-level						Subordinate					
1 Type/ 2 Token		24 Type/ 118 Token						Type/Token					
身體	2	手	19	屁屁	4	牙齒	2	口水	1				
		頭	12	屁股	4	腳腳	2	手指頭	1				
		頭髮	11	耳朵	4	骨頭	2	肚臍	1				
		眼睛	9	聲音	3	嘴	1	鼻孔	1				
		腳	9	臉	3	手手	1						
		鼻子	9	頭頭	3	痘子	1						
		肚子	8	大便	2	眼	1						
		嘴巴	5	牙	2	肚肚	1						
Clothing													
Superordinate		Basic-level						Subordinate					
1 Type/ 29 Token		17 Type/44 Token						2 Type/ 4 Token					
衣服	29	帽帽	14	蝴蝶結	2	帶子	1	項鍊	3				
		襪子	5	袖子	2	帽子	1	美女皮裙子	1				
		褲子	3	鞋	2	新娘衣服	1						
		鍊子	3	鞋鞋	2	裙子	1						
		帽	2	啾啾	1	褲褲	1						
		新衣	2	外套	1								
Vehicles													
Superordinate		Basic-level						Subordinate					
Type/Token		7 Type/104 Token						34 Type/ 86 Token					
		車	40	船	10			火車	19	直昇機	2	挖土機	1

		飛機	18	汽車	7			輪子	6	起重機	2	消防車	1
		車車	15	捷運	2			娃娃車	5	輪胎	2	爺爺車車	1
		車子	12					爺爺車	5	鐵軌	2	破車	1
								摩托車	4	馬車	2	禮賓車	1
								螺旋槳	3	子彈列車	2	腳踏車	1
								計程車	3	乖乖車	1	跑車	1
								輪船	3	公務車	1	轎車	1
								阿祥叔叔車	3	公車	1		
								公共汽車	2	卡車	1		
								單車	2	大車	1		
								大卡車	2	怪手	1		
								工機車	2	拉車	1		
Tools													
Superordinate		Basic-level						Subordinate					
4 Type/ 37 Token		58 Type/ 129 Token						Type/Token					
東西	33	書	9	槍	2	棉被	1	奶瓶	4				
機器	2	錢	8	毛巾	2	水壺	1	寶劍	3				
垃圾	1	剪刀	7	盒子	2	燈	1	衛生紙	3				
東東	1	湯匙	6	盤子	2	燈燈	1	黑板	3				
		刀子	5	筆	2	瓶子	1	ok 繃	1				
		洗衣機	5	紙	2	畫	1	咖啡杯	1				
		手錶	4	繩子	2	碗	1	垃圾桶	1				
		眼鏡	4	茶壺	2	票	1	布袋	1				
		錄音機	4	被子	2	罈	1	故事書	1				
		電話	4	電燈泡	2	蓋子	1	洗臉盆	1				
		刀	3	信	1	藥	1	瓦斯爐	1				
		棍子	3	卡片	1	螺絲	1	茶壺被	1				
		針	3	叉叉	1	蠟燭	1	茶壺被被	1				
		錄音帶	3	叉子	1	被	1	面紙	1				
		電視	3	吹風機	1	被被	1	鳥籠	1				
		刷子	2	咖啡機	1	鏡子	1	小牛頓	1				
		劍	2	大砲	1	電燈	1						
		包	2	子彈	1	麥克風	1						
		唱片	2	扇子	1								
		杯杯	2	板擦	1								
Furniture													
Superordinate		Basic-level						Subordinate					
Type/Token		7 Type/ 23 Token						Type/Token					
		門	9	抽屜	1								
		桌子	7	櫥子	1								
		窗門	3	窗簾	1								
		床	1										
Food													
Superordinate		Basic-level						Subordinate					
4 Type/ 18 Token		53 Type/ 139 Token						20 Type/ 54 Token					
水果	9	水	10	香腸	3	水餃	1	奶油	11	蘆筍汁	1		
菜	6	蛋糕	7	楊桃	2	湯	1	冬瓜茶	9	豆沙	1		
飲料	2	蘋果	6	水水	2	湯湯	1	水蜜桃	5				
晚飯	1	蛋	6	油	2	漢堡	1	魚丸	5				
		飯	6	玉米	2	瓜瓜	1	巧克力	3				
		飯飯	6	瓜	2	米	1	巧克力糖	3				
		糖糖	5	番茄	2	粽子	1	可樂	2				
		豆漿	5	糖果	2	糖	1	棒棒糖	2				
		香蕉	5	草莓	2	羊羹	1	西瓜	2				
		麵包	5	菜菜	2	肉肉	1	雞蛋	2				
		咖啡	4	餅乾	2	茶	1	冰淇淋	1				
		果汁	4	饅頭	2	茶茶	1	啤酒	1				
		桃子	4	麵	2	葡萄	1	滷蛋	1				
		肉	4	麵麵	2	蓮藕	1	烏龍麵	1				
		火腿	3	冰	1	蝦	1	玉米湯	1				
		牛奶	3	包子	1	豆豆	1	瓜瓜湯	1				
		芒果	3	桃	1	鮮奶	1	生日蛋糕	1				

		養樂多	3	椒	1			紅蘿蔔	1				
Animals													
Superordinate		Basic-level						Subordinate					
1 Type/ 9 Token		51 Type/ 352 Token						10 Type/ 21 Token					
動物	9	熊熊	83	蜘蛛	4	鵝	2	野豬	6				
		咪咪	56	蟲	4	小猴子	1	小白兔	5				
		魚	38	尾巴	3	汪汪	1	野狼	2				
		狗狗	22	烏龜	3	海獅	1	黑熊	2				
		蟲蟲	14	熊貓	3	犀牛	1	哈巴狗	1				
		大象	13	貓咪	3	狐狸	1	斑馬	1				
		馬	11	兔	2	狼	1	毛毛蟲	1				
		老鼠	8	公雞	2	老虎	1	金魚	1				
		小魚	6	小狗	2	蚊子	1	黨魚	1				
		河馬	6	小鳥	2	蝸牛	1	小丑魚	1				
		牛	6	烏賊	2	螞蟻	1						
		青蛙	6	熊	2	象	1						
		蛇	5	猩猩	2	豬豬	1						
		蝦子	5	螃蟹	2	馬馬	1						
		豬	5	貓	2	鴨	1						
		狗	4	鱷魚	2	鴨子	1						
		猴子	4	鳥	2	鴨鴨	1						
Natural phenomena and natural materials													
Superordinate		Basic-level						Subordinate					
1 Type/ 1 Token		23 Type/ 48 Token						5 Type/ 10 Token					
天地	1	洞	6	影子	2	毒	1	山洞	4				
		天	5	樹	2	沙	1	大洞	2				
		地球	3	鐵	2	洞洞	1	太陽光	2				
		月亮	3	雪	2	海	1	大海	1				
		花	3	風	2	石頭	1	陽光	1				
		葉子	3	山	1	空氣	1						
		夏天	3	星星	1	雨	1						
		太陽	2	梅花	1								
Toys													
Superordinate		Basic-level						Subordinate					
1 Type/ 26 Token		17 Type/ 46 Token						3 Type/ 26 Token					
玩具	26	黏土	10	泡泡	2	積木	1	汽球	14				
		娃娃	5	洋娃娃	2	蹺蹺板	1	國旗	10				
		呼啦圈	4	電影	2	陀螺	1	氣球	2				
		溜滑梯	4	魔鏡	2	電動玩具	1						
		球	4	圖畫	1	風箏	1						
		磁鐵	4	燈籠	1								
Colors													
Superordinate		Basic-level						Subordinate					
2 Type/ 5 Token		7 Type/ 44 Token						Type/Token					
顏色	3	紅色	12	白色	3								
色	2	綠色	9	紫色	3								
		黃色	9	白	1								
		藍色	7										
Shapes													
Superordinate		Basic-level						Subordinate					
1 Type/1 Token		3 Type/ 5 Token						1 Type/ 1 Token					
花紋	1	圓圈圈	3					橢圓形	1				
		圈圈	1										
		圓圈	1										
Locations													
Superordinate		Basic-level						Subordinate					
3 Type/ 12 Token		35 Type/ 122 Token						18 Type/ 32 Token					
位子	6	房子	18	房屋	2	樓	1	遠企購物中心	5	中研院			
地方	5	家	16	樓梯	2	機場	1	遠企購物中心	4	台大			
世界	1	房間	13	家	2	欄杆	1	電動扶手梯	3	台大醫院			
		停車場	12	新竹	2	洗車廠	1	書房	2	浴室			
		學校	9	飛機場	2	醫院	1	柴油路	2	白沙灣			

		動物園	7	餐廳	2	電話亭	1	游泳池	2	遠東百貨公司			
		電梯	4	公司	1	飛機場	1	高速公路	2	阿里山			
		教會	3	地板	1	高樓大廈	1	埔心牧場	2				
		牆壁	3	城堡	1	停車場	1	辦公室	1				
		路	3	平交道	1	動物園	1	高架橋	1				
		屏東	3	扶手梯	1	大溪	1	中正紀念堂	1				
		幼稚園	2	柱子	1								
Spatial words													
Superordinate		Basic-level						Subordinate					
Type/Token		18 Type/ 274 Token						20 Type/ 28 Token					
		這裡	117	那邊	8	一邊	2	地上	4	樓上	1	桌桌上	1
		這邊	52	後面	6	上	2	家裡	3	熊熊上	1	背上	1
		裡面	27	下面	5	底下	2	門口	3	電話上	1	肚子裡	1
		上面	20	前面	5	中間	1	手上	2	溝裡	1	床上	1
		旁邊	9	裡頭	4	裡	1	頭上	1	椅子上	1	地下	1
		那裡	9	對面	3	邊	1	輪船上	1	袋袋裡	1	天上	1
								樓下	1	海洋裡	1		
Numerals													
Superordinate		Basic-level						Subordinate					
Type/Token		5 Type/ 5 Token						Type/Token					
		七月	1										
		三十五年	1										
		三點	1										
		兩號	1										
		十五點半	1										
Abstract nouns													
Superordinate		**Basic-level**						**Subordinate**					
Type/Token		38 Type/ 108 Token						9 Type/ 13 Token					
		以前	14	名字	3	動作	1	媽媽歌	2				
		現在	9	之後	3	後來	1	小班	2				
		時候	9	晚上	2	前天	1	歌仔戲	2				
		故事	7	剛才	2	青天	1	英文	2				
		今天	7	事	2	知識	1	布袋戲	1				
		以後	5	以後	2	事情	1	賴皮鬼	1				
		歌	4	工班	2	年	1	唐詩	1				
		音樂	4	體操	1	字	1	給愛麗絲	1				
		樣子	3	寶貝	1	以前	1	有時候	1				
		進風	3	辦法	1	水準	1						
		笨蛋	3	歲	1	月	1						
		問題	3	意思	1	一下	1						
		昨天	3	晚	1								

37-42 months

People														
Superordinate		Basic-level						Subordinate						
1 Type/ 1 Token		67 Type/ 740 Token						86 Type/ 360 Token						
人類	1	媽媽	176	同事	5	阿媽	2	乖乖	60	小阿姨	2	警察伯伯	1	
		爸爸	83	女生	5	俠	1	J 阿姨	50	小飛俠	2	護士小姐	1	
		阿姨	71	姨	4	先生	1	聖戰士	38	米鬼	2	C 小姐	1	
		人	64	媽咪	4	公主	1	乖	34	賢丈	2	P	1	
		老師	43	小孩	4	女兒	1	白雪公主	15	Eva 阿姨	2	W 阿姨	1	
		媽	19	小朋友	4	女朋友	1	JC	7	小倫	2	eva 阿姨	1	
		警察	19	消防隊員	4	姊	1	月光俠	7	小倫阿姨	2	仙度麗娜	1	
		皇后	18	爺爺	4	姐姐	1	唐老鴨	6	小矮人們	2	元元哥哥	1	
		小偷	17	男生	4	姑姑	1	艾美兔	6	小重	2	大信	1	
		奶奶	14	老太婆	4	寶寶	1	哥哥俠	5	林阿姨	2	大信丸	1	
		姊姊	13	醫生	4	小賊	1	石老師	5	清清阿姨	2	安妮	1	
		小孩子	13	阿嬤	4	戰士	1	米奇	5	聖戰士	2	小信	1	
		弟弟	13	姨丈	3	新娘子	1	米妮	5	阿強	2	小場	1	
		妹妹	12	老大	3	新郎	1	太空戰士	4	中國人	1	小柔小姐	1	
		哥哥	11	兒子	2	水手	1	亂馬	4	嘟嘟	1	小矮人	1	
		王子	10	同學	2	老公公	1	布魯托	4	外國人	1	王哥哥	1	
		伯伯	8	學生	2	老板	1	軒軒	4	大舅舅	1	珊珊	1	
		魔鬼	8	小丑	2	舅媽	1	壞人俠	3	太空人	1	義路	1	
		叔叔	7	爸	2	郵差	1	太空超人	3	太陽公公	1	聖誕老公公	1	
		新娘	7	牧者	2	鎮長	1	胖胖人	3	女人	1	艾美	1	
		壞人	6	護士	2	阿兵哥	1	J	3	好朋友	1	蔡阿姨	1	
		小姐	6	阿伯	2			小毅	3	嬰兒	1	鄭先生	1	
		巫婆	6	阿公	2			強強	3	小人	1	鍾阿姨	1	
								耶和華	3	小叔	1	阿袖	1	
								壞姊姊	2	小哥哥	1	陳小廷	1	
								家婆	2	小子	1	電光俠	1	
								小媽媽	2	敵人	1	高飛	1	
								小弟弟	2	星星俠	1	黃老師	1	
								小草	2	月光俠	1			
Pronouns														
Superordinate		Basic-level						Subordinate						
Type/Token		17 Type/ 2169 Token						Type/Token						
		我	829	那	16	這些	2							
		你	583	自己	16	那些	1							
		他	441	別	8	其他	1							
		這	114	人家	5	大家	1							
		我們	113	別人	3	本身	1							
		他們	33	你們	2									
Body Parts														
Superordinate		Basic-level						Subordinate						
1 Type/ 1 Token		25 Type/ 169 Token						15 Type/ 30 Token						
全身	1	手	22	屁股	5	肚	2	頭腦	6	大嘴巴	1			
		頭髮	19	肚子	5	肚皮	2	臭腳丫	4	眼淚	1			
		耳朵	16	臉	5	手手	1	左腿	3	肚臍	1			
		腳	15	腿	4	眼珠子	1	果汁臉	3	腸子	1			
		頭	15	鼻子	4	腳腳	1	鼻涕	3	舌頭	1			
		嘴巴	12	大便	3	腿腿	1	手指頭	2	辮子	1			
		牙齒	11	屁屁	3	骨頭	1	下巴	1					
		眼睛	11	屁	2			口水	1					
Clothing														
Superordinate		Basic-level						Subordinate						
1 Type/ 15 Token		19 Type/ 45 Token						8 Type/ 21 Token						
衣服	15	耳環	7	上衣	1	皮帶	1	項鍊	6	長裙	1			
		鞋子	7	口袋	1	花環	1	絲襪	5					
		褲子	6	外套	1	裙	1	愛迪達	4					
		袖子	4	帽	1	褲	1	內褲	2					
		裙子	4	帽帽	1	鞋鞋	1	托鞋	1					
		襪子	3	手套	1			毛襪	1					
		帽子	2	扣子	1			滑鞋鞋	1					
Vehicles														

Superordinate		Basic-level						Subordinate					
Type/Token		7 Type/ 86 Token						43 Type/ 145 Token					
		車	30					火車	20	娃娃車	2	太空船	1
		飛機	25					推ㄟ車	17	挖土機	2	小包車	1
		車子	17					紅燈	14	摩托車	2	小鳥飛機	1
		船	6					綠燈	9	救火車	2	急救車子	1
		汽車	4					腳踏車	8	方向盤	2	挖路車	1
		車車	3					霹靂車	7	爺爺車車	2	消防車	1
		捷運	1					輪胎	6	輪船	2	王冠軍車	1
								拖車	5	釣竿車	2	貨櫃	1
								遊覽車	5	阿公車	2	貨車	1
								阿卡車	5	乖乖車	1	蹦蹦車	1
								摩拖車	3	乖車	1	輪子	1
								警車	3	公共搭車	1	金光車	1
								卡大車	2	剪刀車	1	電風扇車	1
								卡車	2	吊車	1		
								大卡車	2	太空汽車	1		
Tools													
Superordinate		Basic-level						Subordinate					
5 Type/ 44 Token		84 Type/ 232 Token						29 Type/ 50 Token					
東西	36	藥	30	筆	2	毛巾	1	雲梯	7	手機	1		
機器	5	書	18	線	2	洗手台	1	貼紙	4	擦手布	1		
垃圾	1	錢	16	膠布	2	火柴	1	奶瓶	2	故事書	1		
東東	1	剪刀	10	膠水	2	炸彈	1	彩色筆	2	水盆	1		
藥品	1	針	9	被子	2	照相機	1	書包	2	畫板	1		
		鏡子	9	被被	2	牙簽	1	機關槍	2	百寶箱	1		
		包包	7	鉤鉤	2	牙膏	1	水桶	2	皮包	1		
		眼鏡	6	錄影帶	2	盒子	1	煮飯鍋	2	福杯	1		
		杯子	4	鑰匙	2	碗	1	菜刀	2	筆記本	1		
		枕頭	4	電池	2	磚頭	1	藥丸	2	粉筆	1		
		棉被	4	電視	2	竿	1	藥瓶	2	腊筆	1		
		燈	4	髒東西	2	繩子	1	鉛筆	2	茶杯	1		
		玻璃	4	傘	1	蓋子	1	錢筒	2	西藥	1		
		籃子	4	冰箱	1	藥藥	1	雨傘	2	鉛筆盒	1		
		鐵馬武器	4	刀子	1	袋子	1	手槍	1				
		叉子	3	包	1	被	1						
		槍	3	圖片	1	錄音帶	1						
		湯匙	3	大頭針	1	錢錢	1						
		盤子	3	天秤	1	鐘	1						
		馬桶	3	扇	1	鐘鐘	1						
		刀	2	手錶	1	長劍	1						
		劍	2	收音機	1	電熱器	1						
		卡片	2	書書	1	電腦	1						
		吸塵器	2	書本	1	電話	1						
		扇子	2	望遠鏡	1	電風扇	1						
		梯子	2	杖	1	預防針	1						
		棍子	2	槓槓	1	飛彈	1						
		祕密武器	2	橡皮筋	1	麥克風	1						
Furniture													
Superordinate		Basic-level						Subordinate					
Type/Token		6 Type/ 30 Token						4 Type/ 4 Token					
		椅子	9					厚椅架	1				
		門	9					紅門	1				
		桌子	6					鐵門	1				
		窗戶	3					電動門	1				
		床	2										
		抽屜	1										
Food													
Superordinate		Basic-level						Subordinate					
4 Type/ 16 Token		47 Type/ 171 Token						25 Type/ 64 Token					
菜	9	水	16	飯飯	3	柿子	1	冰塊	18	綠豆湯	1		
點心	5	草莓	14	麵	3	殼點心	1	口香糖	6	綠豆球	1		
早餐	1	糖	13	咖啡	2	水水	1	脆笛酥	5	綠豆稀飯	1		

飲料	1	糖果	11	披薩	2	油膏	1	炒雞蛋	4	花生米	1		
		蘋果	11	水餃	2	漢堡	1	木瓜	3	花生肉	1		
		果汁	8	火鍋	2	火腿	1	牛肉乾	3	花蜜水	1		
		花蜜	8	牛奶	2	花生殼點心	1	冰淇淋	2	葡萄水	1		
		葡萄	7	肉	2	蕃茄	1	棒棒糖	2	蜜蕉	1		
		蛋糕	7	肉肉	2	蜂蜜	1	蘿蔔糕	2	雞腿	1		
		花生	5	茶	2	蜜	1	酸梅	2				
		餅干	5	蛋	2	蝦蝦	1	雞蛋	2				
		湯	4	蛋蛋	2	豆腐	1	脆笛酥	2				
		飯	4	豆豆	2	釋迦	1	冰水	1				
		李子	3	開水	2	鳳梨	1	巧克力	1				
		綠豆	3	餅乾	2	龍眼	1	稀飯	1				
		菜菜	3	香蕉	2			紅蘿蔔	1				
Animals													
Superordinate		Basic-level						Subordinate					
1 Type/ 1 Token		63 Type/ 392 Token						16 Type/ 42 Token					
動物	1	熊熊	42	恐龍	6	蟲	2	毛毛蟲	8				
		咪咪	28	翅膀	6	蟲蟲	2	大鼻象	7				
		魚	21	蝴蝶	6	蠶	2	白兔	7				
		熊	20	蝸牛	6	西瓜動物	2	大牛	3				
		豬	15	豬豬	6	鱷魚	2	白熊	3				
		小鳥	13	獅子	5	企鵝	1	小鳥狗狗	2				
		蜜蜂	13	貓	5	喵喵	1	瓢蟲	2				
		鴨	13	長頸鹿	4	小熊	1	野狼	2				
		大象	12	鴨鴨	4	小狗	1	小白兔	1				
		老鼠	12	兔	3	小貓咪	1	小象	1				
		猴猴	11	咕咕咕	3	尾	1	小鼻象	1				
		貓咪	11	河馬	3	海豚	1	母雞	1				
		青蛙	10	狐狸	3	牛牛	1	臭貓貓	1				
		馬	10	老虎	3	羊咩咩	1	鯊魚	1				
		狗狗	9	蚊子	3	蜘蛛	1	鰻魚	1				
		尾巴	8	蛇	3	蟲子	1	黑豬	1				
		烏龜	8	蜈蚣	3	象象	1						
		狗	8	螞蟻	3	雞	1						
		鳥	8	猴子	2	騾子	1						
		牛	7	羊	2	魚刺	1						
		兔子	6	螃蟹	2	鴨子	1						
Natural phenomena and natural materials													
Superordinate		Basic-level						Subordinate					
Type/Token		34 Type/ 116 Token						6 Type/ 6 Token					
		花	15	星星	3	大理花	1	幽谷	1				
		洞	14	沙子	3	天空	1	聖誕樹	1				
		太陽	11	洞洞	3	森林	1	茉莉花	1				
		月亮	5	籽	3	沙垢	1	雪人	1				
		樹	5	雲	3	河	1	雪球	1				
		電	5	天	2	泥土	1	電光	1				
		樹葉	4	山	2	海	1						
		沙	4	殼	2	花花	1						
		石頭	4	泡沫	2	鐵	1						
		空氣	4	煙	2	星	1						
		風	4	葉子	2								
		地球	3	山	1								
Toys													
Superordinate		Basic-level						Subordinate					
3 Type/ 20 Token		25 Type/ 75 Token						10 Type/ 33 Token					
玩具	16	球	13	面具	3	泡泡	1	氣球	9	鋼珠	1		
禮物	2	球球	8	風箏	3	球棒	1	鋼琴	9				
遊戲	2	魔鏡	8	機器人	2	積木	1	保齡球	5				
		溜滑梯	5	笛子	2	翹翹板	1	高爾夫球	3				
		黏土	5	麻將	2	金剛	1	棒球	2				
		娃娃	4	假人	1	風鈴	1	小喇叭	1				
		彈簧	3	呼啦圈	1	鼓	1	小提琴	1				
		撲克牌	3	圖畫	1			籃球	1				

		磁鐵	3	木馬	1			跳旗	1				
Colors													
Superordinate		Basic-level						Subordinate					
2 Type/ 7 Token		13 Type/ 59 Token						Type/Token					
色	3	紅色	15	橘色	2	綠	1	粉紅色	2				
顏色	4	綠色	11	紅	2	藍	1	橘子色	1				
		白色	9	黑色	2	黃	1	葡萄色	1				
		黃色	7	白	1								
		藍色	6	紫色	1								
Shapes													
Superordinate		Basic-level						Subordinate					
1 Type/ 1 Token		4 Type/ 10 Token						Type/Token					
形	1	圈圈	6										
		三角形	2										
		圓圈圈	1										
		圓形	1										
Locations													
Superordinate		Basic-level						Subordinate					
1 Type/ 8 Token		35 Type/ 147 Token						31 Type/ 55 Token					
地方	8	家	42	小學	3	百貨公司	1	西藥店	4	婆婆家	2	空房子	1
		房子	21	大學	2	皇宮	1	新民	4	客廳	2	糖果店	1
		學校	9	幼稚園	2	花園	1	西藥店	4	青年公園	2	美容院	1
		警察局	7	橋	2	路	1	保羅街	3	麥當勞	2	足球場	1
		美國	7	台灣	2	遂道	1	東園	3	兒童樂園	1	辦公室	1
		動物園	6	天國	1	隧道	1	五樓	2	六樓	1	土地公廟	1
		樓梯	6	學	1	電梯	1	客廳	2	地下街	1	東院	1
		公園	5	屋子	1	風車	1	屋頂	2	大樓梯	1		
		城堡	4	柱子	1	三峽	1	教室	2	姨家	1		
		房間	4	洗車場	1	德國	1	棒球場	2	廚房	1		
		醫院	4	牆壁	1	花蓮	1	中興街	2	無瑰城堡	1		
		地下道	3	王國	1			健寶園	2	玫瑰花城堡	1		
Spatial words													
Superordinate		Basic-level						Subordinate					
Type/Token		20 Type/358 Token						31 Type/ 64 Token					
		這裡	116	裏面	3			世界上	8	路上	2	殿中	1
		這邊	74	裏頭	3			地下	5	口裡	1	眼睛裡	1
		裡面	53	邊	2			家裡	5	嘴巴裏	1	背上	1
		那邊	23	上	1			床上	5	嘴巴裡	1	路邊	1
		上面	21	中	1			天上	4	地上	1	車上	1
		下面	16	中間	1			肚子裡	4	天空上	1	電視上	1
		那裡	16	前	1			水邊	3	屋子裡	1	面前	1
		外面	12	底下	1			青草地上	3	座位上	1	頭裡	1
		後面	8	裡	1			桌子上	2	樓上	1	鼻子上	1
		旁邊	4	裡頭	1			椅子上	2	樓下	1		
								海邊	2	樹葉上	1		
Numerals													
Superordinate		Basic-level						Subordinate					
2 Type/ 2 Token		7 Type/ 11 Token						Type/Token					
數字	1	八	3	五號	1								
號碼	1	十四號	3	三點鐘	1								
		二	1	五點	1								
		五	1										
Abstract nouns													
Superordinate		Basic-level						Subordinate					
Type/Token		49 Type/ 159 Token						16 Type/ 26 Token					
		現在	25	車禍	3	火氣	1	灌籃高手	4				
		故事	11	今天	3	祕密	1	英文	3				
		時候	11	事	2	筵席	1	大力水桶	3				
		氣魄	10	心腸	2	舞	1	大聰明	2				
		笨蛋	9	時間	2	詩篇	1	中國功夫	2				
		以前	7	禮拜	2	辦法	1	小時候	2				
		昨天	7	以後	2	關係	1	大笨蛋	1				
		一下	5	過去	2	靈魂	1	小氣鬼	1				
		明天	5	一生	1	音樂	1	小笨蛋	1				

		功課	4	世	1	魔術	1	猴猴故事	1				
		福	4	事情	1	打勤	1	神經病	1				
		晚上	4	名字	1	剛才	1	螞蟻舞	1				
		味道	3	恩惠	1	時候	1	討厭鬼	1				
		字	3	慈愛	1	現	1	仙履奇緣	1				
		寶貝	3	時後	1	點鐘	1	哥林前	1				
		布	3	歌	1			美少女戰士	1				
		樣子	3	歲	1								

43-48 months

People													
Superordinate		Basic-level						Subordinate					
Type/Token		50 Type/ 199 Token						61 Type/ 168 Token					
		媽媽	31	媽咪	3	大人	1	阿拉丁	44	強強	2	小希希	1
		人	21	朋友	3	好人	1	大彬彬	7	米老鼠	2	小琪姐姐	1
		爸爸	15	管理員	3	姑姑	1	太空戰士	6	阿袖	2	小舅舅	1
		壞人	11	叔叔	2	小偷	1	小彬彬	6	麗雅公主	2	小重	1
		阿姨	10	外婆	2	小朋友	1	楊國欽	6	仁德	1	小青阿姨	1
		舅舅	9	女兒	2	工程師	1	賈方	6	太空人	1	小鳴	1
		老師	8	女生	2	怪魔	1	聖戰士	5	太空戰士	1	彬彬	1
		弟弟	7	小孩子	2	戰士	1	小矮人	4	小弟弟	1	披風俠	1
		妹妹	6	巫婆	2	新娘子	1	月光俠	4	月光俠	1	昆昆	1
		小孩	6	王	2	新郎	1	白雪公主	4	森林之王	1	星星俠	1
		哥哥	5	醫生	2	爸	1	阿弟	4	男兒	1	機器戰警	1
		警察	5	乘客	1	皇后	1	陳裴娟	4	白白公主	1	江老師	1
		司機	4	人人	1	老板	1	小馬	3	美如	1	聖戰士	1
		媽	4	伯伯	1	郵差	1	戰鬥王	3	阿美人	1	聖誕老公公	1
		魔鬼	4	兒子	1	阿公	1	豬小弟	3	霹靂劍隊	1	蝙蝠俠	1
		公主	3	公公	1	阿媽	1	小毅	3	TS 阿姨	1	貞德	1
		婆婆	3	同學	1			梁尚賜	3	光仁	1	阿丁	1
								阿拉	3	唐老鴨	1	阿拉巴	1
								姨婆	2	國欽	1	阿昆	1
								小馬伯伯	2	大華	1		
								希希	2	小俊	1		
Pronouns													
Superordinate		Basic-level						Subordinate					
Type/Token		13 Type/ 1158 Token						Type/Token					
		我	483	他們	18	別人	2						
		你	279	那	15	你們	1						
		他	252	自己	10	那些	1						
		這	49	人家	6								
		我們	40	別	2								
Body Parts													
Superordinate		Basic-level						Subordinate					
1 Type/ 3 Token		19 Type/ 89 Token						4 Type/ 6 Token					
屍體	3	腳	21	眼睛	3	頭髮	2	香港腳	3				
		手	16	鼻子	3	屁屁	1	手指頭	1				
		頭	9	後背	2	眼珠	1	鼻水	1				
		耳朵	7	牙齒	2	臉部	1	鼻涕	1				
		嘴巴	6	聲音	2	鬍子	1						
		屁股	4	肚子	2								
		脖子	4	臉兒	2								
Clothing													
Superordinate		Basic-level						Subordinate					
1 Type/ 4 Token		7 Type/ 14 Token						4 Type/ 7 Token					
衣服	4	帽子	4	吊帶	1			外褲	3				
		紗巾	3	戒指	1			襪褲	2				
		圍兜兜	2	襪子	1			牛仔褲	1				
		帽帽	2					褲襪	1				
Vehicles													
Superordinate		Basic-level						Subordinate					
Type/Token		5 Type/ 32 Token						38 Type/ 64 Token					
		車子	13					跑車	6	老爺噴空車	2	碰碰船	1
		車	12					火車	4	警察金龜車	2	老爺福特車	1
		船	3					太空梭	3	飛機車	2	老爺車	1
		汽車	2					福特車	3	三輪船	1	聒聒車	1
		飛機	2					賽車	3	五號車	1	裝甲車	1
								輪子	3	大炮新車	1	賽跑車	1
								一輪船	2	大車	1	輪胎	1
								吉普車	2	娃娃車	1	遊艇	1
								尖尖車	2	戰鬥機	1	金龜車	1
								打鼓船	2	扣扣船	1	釣竿車	1
								直升機	2	摩托車	1	電車	1

								稻草車	2	汽艇船	1	香煙船	1
								老爺吉普車	2	眼睛船	1		
Tools													
Superordinate		Basic-level						Subordinate					
1 Type/ 17 Token		32 Type/ 70 Token						12 Type/ 21 Token					
東西	17	刀	9	筆	2	游泳圈	1	書包	5	魚缸	1		
		劍	7	線	2	燈	1	鉛筆	3				
		書	5	鏡片	2	碗	1	垃圾桶	2				
		子彈	4	電池	2	竹竿	1	藥膏	2				
		油漆	4	剪刀	1	籠子	1	衛生紙	2				
		錄影帶	4	包包	1	繩子	1	整潔衛生紙	1				
		錄音帶	4	卡片	1	藥	1	筆芯	1				
		刀子	2	夾子	1	針	1	貼紙	1				
		圖片	2	手錶	1	電燈	1	錢包	1				
		棍子	2	機械手	1	電視	1	降魔無敵劍	1				
		眼鏡	2	橡皮筋	1			霹靂劍	1				
Furniture													
Superordinate		Basic-level						Subordinate					
Type/Token		3 Type/ 22 Token						1 Type/ 1 Token					
		門	16					自動門	1				
		椅子	4										
		桌子	2										
Food													
Superordinate		Basic-level						Subordinate					
3 Type/ 7 Token		11 Type/ 15 Token						8 Type/ 13 Token					
水果	5	水	2	牛奶	1	餅乾	1	冰淇淋	3	月餅	1		
早餐	1	豆花	2	糖果	1	香蕉	1	冰塊	2	棒棒糖	1		
點心	1	青菜	2	蘋果	1	鳳梨	1	口香糖	2	聖代	1		
		飯	2	豆漿	1			楊桃汁	2	菊花茶	1		
Animals													
Superordinate		Basic-level						Subordinate					
1 Type/ 11 Token		47 Type/ 217 Token						12 Type/ 19 Token					
動物	11	獅子	30	兔子	3	喵喵	1	太空獅子	3				
		小鳥	17	怪獸	3	大猩猩	1	空間魚	3				
		馬	16	河馬	3	小魚	1	太空馬	2				
		犀牛	12	烏龜	3	海馬	1	小雞	2				
		長頸鹿	12	蟲	3	狗	1	毛毛蟲	2				
		蛇	10	貓頭鷹	3	虫	1	大蟒魚	1				
		牛	9	小咪咪	2	蜘蛛	1	太空河馬	1				
		猴子	9	小蛾	2	螃蟹	1	太空長頸鹿	1				
		老虎	9	恐龍	2	象	1	母雞	1				
		大象	8	狗狗	2	豬	1	老鷹鳥	1				
		角	8	羊	2	金龜	1	飛蛾	1				
		尾巴	6	老鷹	2	青蛙	1	鯊魚	1				
		魚	6	袋鼠	2	駱駝	1						
		老鼠	5	金龜子	2	鱷魚	1						
		小狗	4	雞	2	鴨子	1						
		鴕鳥	4	公雞	1								
Natural phenomena and natural materials													
Superordinate		Basic-level						Subordinate					
Type/Token		9 Type/ 12 Token						1 Type/ 2 Token					
		天	2	天空	1	花	1	聖誕樹	2				
		太陽	2	樹樹	1	葉子	1						
		樹	2	洞洞	1	風	1						
Toys													
Superordinate		Basic-level						Subordinate					
2 Type/ 13 Token		10 Type/ 22 Token						8 Type/ 17 Token					
玩具	9	面具	7	溜滑梯	2			保齡球	6	國旗	1		
遊戲	4	拼圖	2	飛毯	2			阿拉丁神燈	3	小球球	1		
		木馬	2	太空玩具	1			大當棍	2	阿拉丁阿拉神燈	1		
		模型	2	球	1			氣球	2				
		溜梯梯	2	球球	1			七七球	1				
Colors													
Superordinate		Basic-level						Subordinate					
1 Type/ 7 Token		10 Type/ 59 Token						4 Type/ 10 Token					

顏色	7	紅色	18	白色	3			紅黃色	4				
		藍色	14	黑色	3			白黃色	3				
		綠色	7	咖啡色	2			粉紅色	2				
		黃色	7	灰色	1			黑黃色	1				
		橘色	3	紫色	1								
Shapes													
Superordinate		Basic-level						Subordinate					
1 Type/ 2 Token		Type/Token						Type/Token					
形狀	2												
Locations													
Superordinate		Basic-level						Subordinate					
3 Type/ 5 Token		22 Type/ 63 Token						6 Type/ 13 Token					
地方	2	家	15	地板	2	醫院	1	天橋	6				
空間	2	學校	7	樓梯	2	馬路	1	麥當勞	3				
位子	1	台北	7	台	1	台北市	1	小路	1				
		西門	5	台北市	1	圓山	1	青年公園	1				
		房子	4	家家	1	新門	1	台北公園	1				
		動物園	3	房間	1	香港	1	新民	1				
		房屋	3	花市	1								
		學校	3	角落	1								
Spatial words													
Superordinate		Basic-level						Subordinate					
Type/Token		14 Type/ 107 Token						14 Type/ 28 Token					
		這裡	42	下面	3			頭上	8	天上	1		
		這邊	18	後面	3			背上	5	手上	1		
		裡頭	9	旁邊	3			房子裏	2	書上	1		
		上面	7	外面	2			河裏	2	沙裏	1		
		裏面	6	那裡	2			海裏	2	火車上	1		
		裡面	6	前面	1			五號車上	1	車子裏	1		
		那邊	4	裏	1			台上	1	馬路上	1		
Numerals													
Superordinate		Basic-level						Subordinate					
0 Type/ 0 Token		0 Type/ 0 Token						0 Type/ 0 Token					
Abstract nouns													
Superordinate		Basic-level						Subordinate					
Type/Token		20 Type/ 49 Token						12 Type/ 22 Token					
		樣子	8	時候	2	禮拜	1	賈方負仇記	4	臭討厭	1		
		今天	8	歌	2	笨蛋	1	大笨蛋	3	賈方復仇記	1		
		昨天	5	從前	2	點	1	大笨頭	3	雞雞覺	1		
		剛才	4	晚上	2	一下	1	王八蛋	3	霹靂神光	1		
		現在	3	名字	1	下午	1	大笨豬	2	一零一忠狗	1		
		功課	2	字	1	最後	1	中班	1				
		故事	2	神經	1			小班	1				