

# Early Lexical Development: A Corpus-Based Study of Three Languages

Shuxia Liu (shuxia.liu@richmond.edu)

Xiaowei Zhao (xzhao2@richmond.edu)

Ping Li (pli@richmond.edu)

Department of Psychology, University of Richmond

Richmond, VA 23173 USA

## Abstract

In this study, we present a crosslinguistic analysis of early lexical development in English, Mandarin and Cantonese based on large-scale child-adult interaction transcripts in the CHILDES database. We examined patterns of lexical composition for several lexical categories (nouns, verbs, and adjectives) in children and their caregivers' vocabularies across eight different age groups from 13 to 60 months. A series of statistical methods were applied to analyze the developmental trajectories of lexical diversity of children's speech. Our study overcomes several major methodological limitations of previous studies, and provides a clear and detailed picture of the similarities and differences in lexical development patterns: (1) in all three languages, children's early language development shows increasing diversity as a function of time, and the child lexicon becomes more similar to that of their parents over time; and (2) language-specific differences in the linguistic input have strong influence on children's language output, which is reflected in the varying percentages of nouns, verbs, and adjectives in the child's productive lexicon at different developmental stages.

**Keywords:** Lexical Development, CHILDES, Corpus Study

## Introduction

It has been observed that English-speaking children's early words often include common nouns referring to objects with solid functional and physical properties like *ball*, *box*, *bubble*, and words that describe people or things in their immediate environment, such as *daddy*, and *mommy* (Bates, Dale & Thal, 1995). These "referential style" words are often learned first by children, while other categories such as verbs and adjectives join their vocabulary later. Closed-class words are acquired even later. It is widely accepted that the early vocabularies of English-speaking children have proportionally more nouns than words in other lexical categories (Gentner, 1982; Caselli, et al. 1995). Some researchers suggest that this "Noun Bias" in early language acquisition is a universal phenomenon that can be observed across all languages. Specifically, Gentner (1982) proposed the "*natural partitions hypothesis*", which states that there is a natural conceptual distinction between "concrete concepts" (nouns) and "predicative concepts" (verbs); and nouns are conceptually and perceptually more basic and simpler than other words, thus making nouns easier for children to grasp at a young age. However, some investigators disagree with this argument, especially when looking at languages that are typologically different from English. Crosslinguistic studies of Chinese and Korean have provided counter-evidence on the "Noun Bias" argument,

suggesting that nouns are not always learned first, and that in these languages the early predominant lexical category might be verbs (Choi, 1997, 2000; Tardif, 1996, 2006; Tardif, Gelman & Xu, 1999).

However, there were certain important methodological issues in these studies that might limit the reliability and validity of their results. First, most of the previous studies of early lexical composition patterns were based on relative small sample sizes (few participants or small-scale corpus). For example, Genter's study (1982) included only 16 children in total for six languages (thus, an average of two to four children in each language). Some later studies had a few more subjects, but the size was still not large enough -- usually around 20 subjects in each language (Choi, 2000; Tardif, 1996; Tardif et al, 1997). Generally speaking, individual children's lexical developmental trajectories can be different, and may exhibit diverse strategies in their vocabulary learning (e.g. referential style vs. expressive style as discussed in Nelson, 1973). Thus, variations with too few samples may limit the generalizability of the findings, especially in crosslinguistic contexts.

Second, previous studies often paid attention to children in specific age groups, or only investigated children's speech under specific contexts (e.g. toy-play in the laboratory). These factors, along with the small sample size, often led to a small lexicon available for analysis, thus again limiting the generalizability of findings. For example, Caselli et al (1995) studied the lexical composition according to parental reports of 659 English and 195 Italian children, but the subjects were limited to infants between 8 to 16 months of age, thus the lexicon that they could analyze was small, ranging from 50 to 100 words. In this case, researchers could not systemically compare the lexical composition patterns at different developmental times as children acquire an increasing number of words.

Third, the variations in methodologies used in previous studies often make it difficult to compare across languages and investigations. For example, it is quite common that different criteria were applied by investigators to judge the appropriate lexical category of a word. As discussed by Tardif et al. (1999), such variations will influence the extent to which the investigators discover a "Noun Bias", especially across different languages.

Fourth, previous studies often focused on measuring the ratio between two lexical categories: nouns and verbs. Inclusion of new measures for more lexical categories, such as the measure of lexical diversity in the current study,

would certainly provide us a more complete understanding of children's early lexical development in each language.

Our investigation here is designed to overcome these limitations by firstly examining a large-scale corpus that contains transcripts of child-parent interactions. It is based on the analysis of the Child Language Data Exchange System (CHILDES) (MacWhinney, 2000). The size of lexicon that we analyze is around 450 words, which is close to the size of vocabulary in the MacArthur-Bates Communicative Development Inventory (CDI, Dale & Ferson, 1996), a parental report measure of children's early vocabulary. Second, we focus on three languages: English, Mandarin and Cantonese,<sup>1</sup> and our study covers children across different age groups and examines their speeches under varying context situations. The age of children covered in these corpora includes the crucial periods of children's early language development, ranging from 13 to 60 months. Additionally, the materials are not restricted to a small number of parents-child dyads but include large amount of conversations under various context situations. Third, in our study, the obstacles posed by methodological variations are effectively overcome by a universal semantic criterion applied in the three languages for determining the lexical category of words. Fourth, our investigation covers a large number of common words in children's language across various grammatical categories, including not only nouns, verbs, but also adjectives; and within each category, we further divide words into different sub-categories according to their semantic features. Several lexical measures (e.g., the D-measure, cluster analysis, and noun-verb ratio) are further applied. These methodological improvements in our research are likely to provide better reliability, validity and generalizability than previous studies. Finally, unlike previous studies, our study focuses on the developmental pattern in children's lexical compositions at different ages, and how it is influenced by the speech of parents. This scenario provides a detailed and complete picture of children's lexical composition pattern in a crosslinguistic context.

## Methods

We investigated the similarities and differences of children's vocabulary development across languages and at different developmental stages. Our research is focused on child speech, but also includes the adult-to-child speech in the three target languages. We conducted two separate, but closely related sets of studies. The first study is based on a small but well balanced data set so that more rigorous statistical analyses (e.g., ANOVA and D-measure) could be run on the data. The second study involved a much larger sample size to ensure generalizability of our analyses in the first. In the second study, the percentages of the most

frequently occurring words in various lexical categories over total vocabulary were computed for all languages and at different developmental stages.

## Corpora for Each Language

Each of the three target languages has been widely investigated previously and there are many existing corpora for them in the CHILDES database. The English corpora were extracted from the American English database in CHILDES. In total, the data include speech from more than 700 children and their caregivers. The age of the children ranges from 13 months old to 60 months old. Files with children out of this age range were excluded from our analysis, as well as those files without age information. The Mandarin data that we used were extracted from the East Asian database. This database is smaller than the English data, including speech from around 300 Mandarin-speaking children and their caregivers. We chose the children with ages ranging from 14 to 60 months old. The Cantonese data involve the conversations of around 80 children and their caregivers. The size of utterances in the Cantonese data is comparable to that in the Mandarin data, and the ages of the children range from 15 to 60 months old.

## Developmental Age Groups

To investigate the lexical developmental trajectory of children, it is necessary to examine their lexical composition across time. In addition, we can also identify parental lexical composition across different ages of their children, to determine whether children's linguistic input and their productive speech show the same patterns. For this purpose, we partitioned all the transcripts into different groups according to the age of the children. In the first study, we included four age levels: 13-24 months, 25-36 months, 37-48 months, and 49-60 months. In the second study, we further split the age range into eight groups to give us a more detailed picture of the lexical development of the three languages. The eight age groups are: 13-18 months, 19-24 months, 25-30 months, 31-36 months, 37-42 months, 43-48 months, 49-54 months, and 55-60 months. Based on this scale with six months as a group, we obtained a rough picture of how the distributions of nouns, verbs, and adjectives change as a function of the children's linguistic experiences, along with the changes in the adult speech, if any.

## Lexical Categories

In our study we paid particular attention to the percentages of three types of words – nouns, verbs and adjectives – over the total vocabulary size. For English, we referred to CDI (Dale & Ferson, 1996) to classify the lexical group to which each word belongs. For Mandarin and Cantonese, those words having exact translations in English were classified according to CDI, and the words unique to the two languages were classified according to dictionaries and grammar books (e.g. Lü, 2001; Institute of Linguistics of Chinese Academy of Social Sciences, 2002; Hao, Xing, Shu,

---

<sup>1</sup> Linguists continue to debate on whether Cantonese should be considered as a distinct language or as a dialect of the Chinese language. For convenience here we refer to Mandarin and Cantonese as two languages.

& Li, 2008). Also, we further split each lexical group into additional subcategories according to the semantic properties of these words. In this way, we were able to compare the crosslinguistic similarities and differences in depth.

### Technical Problems and Solutions

*Pinyin*, the Standard Mandarin Romanization system, was used to transcribe the speech in Mandarin-speaking children in CHILDES (some corpora also involve Chinese characters) database, Chinese corpora are transcribed either in Chinese characters or in the form of *Pinyin*. However, this method cannot differentiate the significant number of homophones in Chinese; for example, *zuo4* could mean do/make (做), sit (坐) and seat (座). Thus, for those transcripts that had only *Pinyin* codes, we simply recoded the transcripts using Chinese characters.

A second problem involved polysemantic words which belong to different lexical categories when presented in different contexts (e.g., *watch* in English, *xiang4* (象) in Chinese, meaning either “elephant,” or “resemble.”). In CHILDES, a tool has been used to disambiguate the meaning of these words under different contexts, involving the use of morphosyntactic analyzers (*MOR* and *POST*) in the CLAN program. In our study, we disambiguated such words according to the following 4 steps. (1) We extracted information from the morphosyntactic analyses and the already tagged words in the CHILDES data and used it as a basis of our analysis. (2) For a few untagged transcripts, we conducted the morphosyntactic analysis on them based upon the grammar database of the three languages (downloadable from the CHILDES website). (3) Although *MOR* and *POST* are very powerful tools (95% or more of words can be correctly tagged in English; MacWhinney, 2000), there were still some easily confused words that could not be correctly tagged. For this situation, we tagged them manually according to the context in which these words occurred. (4) For some other easily confused words that do not have enough context information, we simply classified these words according to their most frequently used denotations in dictionaries (Institute of Linguistics of Chinese Academy of Social Sciences, 2002; Miller, 1990).

### Study I

In the first study, we examined only the child’s speech. We chose a total of 72 files from the corpora; each file represented the conversation of one child and his/her caregivers. The ages of the 72 children ranged from 17 months to 59 months ( $M = 37.18$ ,  $SD = 12.94$ ). The children were split into four age groups as discussed earlier. In each age group, there were 18 children: six spoke English, six spoke Mandarin, and six spoke Cantonese. Each language included 24 subjects in total and we carefully matched the age distribution of the children in each language.

From the 72 files, the speech of each child was extracted. We surveyed the number of the noun and verb types in each

child’s speech, and computed the noun versus verb (N/V) ratio in types for each child. We also computed D-measure, a novel method to quantify vocabulary diversity of speech samples, by using a particular command from the CLAN program (*VOCD*). The calculation process of D-measure is based on the traditional method of measuring TTR (Type-Token ratio), and adequately considers the effects of varying sample sizes. It has been proven to be a valid and reliable measure of vocabulary diversity (McKee, Malvern, & Richards, 2000). High values of the D-measure reflect a high level of lexical diversity, which often indicates speakers’ good capacity of handling their language because they can diversify their vocabulary and avoid word repetitions in their speech. Thus, the N/V type ratio and the D-measure served as the two dependent measures in our statistical analyses. Two separate ANOVAs (3 languages x 4 age groups), one for each of the two dependent variables, were conducted.

### Noun/Verb Ratio

Table 1. ANOVA for ratio of noun types versus verb types

	SS	DF	MS	F	Sig.
Age	11.89	3	3.96	1.55	.212
Language	19.61	2	9.08	<b>3.82</b>	<b>.027*</b>
Age x Language	13.07	6	2.12	.85	.54
Error	153.89	60	2.57		
Total	352.65	72			

\*  $P < .05$

As seen in Table 1, the analysis results showed a significant main effect of language,  $F(2, 60) = 3.82$ ,  $P = .027$ . However, there was no main effect for age,  $F(3, 60) = 1.55$ ,  $P = .212$ . Furthermore, no interaction effect between language and age was found,  $F(6, 60) = .849$ ,  $P = .537$ . Following a significant main effect of language, an LSD post-hoc test for language at the .05 alpha level was conducted, which yielded the following effects. English speaking children displayed higher mean ratio of nouns vs. verbs ( $M = 2.19$ ,  $SD = 2.71$ ) than Mandarin speaking children ( $M = .98$ ,  $SD = .34$ ) and Cantonese speaking children ( $M = 1.23$ ,  $SD = .55$ ). There was no significant difference in the ratio of nouns and verbs between Mandarin and Cantonese speaking children.

Table 2. N/V ratio changes across language and age groups

Age Group (months)	Mandarin		Cantonese		English	
	Mean	SD	Mean	SD	Mean	SD
13-24	1.15	.33	1.46	1.01	3.87	5.30
25-36	.81	.39	1.32	.30	1.86	.72
37-48	1.09	.28	1.17	.26	1.34	1.54
49-60	.86	.30	.96	.20	1.68	.74
<b>Mean</b>	.98	.34	1.23	.55	2.19	2.71

The mean ratios of noun types vs. verb types under different situations are shown in Table 2. It is clear from

these results that English speaking children use more types of nouns than verbs, a clear “Noun Bias” (2.19). But children in the other two language groups have relatively weak “Noun Bias” (1.23, Cantonese) or even no “Noun Bias” (0.98, Mandarin). Mandarin and Cantonese are more similar in terms of N/V ratios. The results found here clearly show the crosslinguistic differences in lexical composition. In addition, although the difference as a result of age groups is not significant, we can still see some developmental patterns from Table 2. For all three languages, when children are younger than 24 months old, they display a “Noun Bias” with more noun types than verb types. However, as the children age, the N/V ratio becomes smaller, even for English, which means that relatively more and more types of verbs have entered the children’s vocabulary.

### D-measure

Table 3. ANOVA for D-measure

	SS	DF	MS	F	Sig.
Age	14042.4	3	4680.8	<b>19.92</b>	<b>.000*</b>
Language	4490.8	2	2245.4	<b>9.55</b>	<b>.000*</b>
Age x Lang	1470.9	6	245.2	1.04	.407
Error	14100.8	60	235.0		
Total	223052.5	72			

\* P < .01

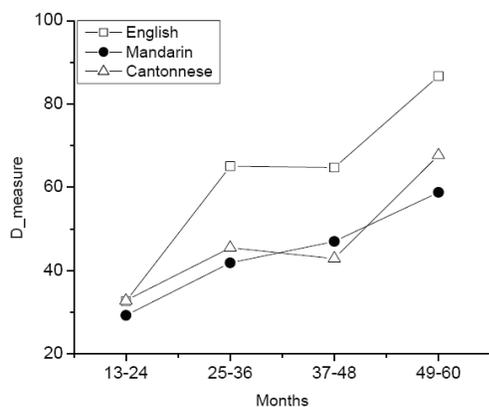


Figure 1. D-measures (lexical diversity) change with age.

Another separate 3 X 4 ANOVA was conducted to evaluate the effects of language and age differences on the D-measure. The results (Table 3) indicate both language and age have significant main effects on lexical diversity. For the main effect of age,  $F(3, 60) = 19.92, P < .01$ . For the main effect of language,  $F(2, 60) = 9.55, P < .01$ . However, no interaction effect between language and age was found,  $F(6, 60) = 1.04, P = .41$ . Following the significant main effects of age and language, an LSD post-hoc test at the .05 alpha level was conducted, which yielded the following effects. First, English speaking children reported higher D-measure ( $M = 62.26, SD = 18.45$ ) than Mandarin speaking children ( $M = 44.21, SD = 19.11$ ) and Cantonese speaking children ( $M = 47.21, SD = 18.45$ ). There was no significant

difference in the D-measure between Mandarin and Cantonese speaking children. Second, differences between any two age groups are significant, except for the difference between 25-36 months old children and 37-48 months old children. From Figure 1 we can see that the three languages have similar developmental patterns in terms of the D-measure. As time goes by, children’s speech becomes increasingly diverse, and it further reflects children’s development of their language competence compared to that in the early age groups. On the other hand, Mandarin and Cantonese’s developmental patterns are more similar to each other than to English, showing also clear crosslinguistic differences.

### Study II

The purpose of Study II is to provide us with a more complete picture of how languages develop across time in terms of speech complexity and lexical composition. To increase the generalizability of the findings, we increased the sample size of this study to include all the available and age-appropriate data from CHILDES as discussed in the method Section. We classified the data files within 13 and 60 months into eight age groups with six months as the scaling unit. We also obtained both child speech and adult speech samples in order to investigate the similarities and differences between the language input and children’s productive output in each language. As a consequence in this scenario, we dealt with 48 situations (3 languages x 8 development levels x 2 groups: adult/children) in total. Here, we first extracted the child speech and the adult speech into separate files for each of the 24 Languages x Age groups. The N/V ratio was again calculated. We also examined the lexical compositions in the vocabularies of the 48 situations. In particular, certain numbers<sup>2</sup> of the most frequently occurred word types in the vocabularies were recorded. We examined the vocabulary composition within three lexical categories of nouns, verbs, and adjectives by splitting each category into more detailed subgroups. The procedure led to 30 subcategories for each vocabulary of the 48 situations. The percentages of the subcategories over the total word numbers (in types) can be used to describe the lexical composition of each vocabulary. Based on this detail information (treating each subcategories as a variable), we conducted a cluster analysis of the vocabularies of the 48 situations to determine the similarity and difference across language, age, and people.

### Noun/Verb Ratio in Detail

Following the analyses in Study I, we again calculated the noun vs. verb type ratio in both children’s (Figure 2a) and adults’ (Figure 2b) vocabularies with half year as the scaling unit. From Figure 2.a, we can see that the three languages follow a similar developmental pattern. Irrespective of

<sup>2</sup> Around 450-500, but the number is much smaller for age groups of children younger than 36 months old, due to the fact that younger children are not able to produce high volume of words.

language, it is clear that there are more nouns than verbs in children’s vocabularies at the earliest stage – a clear “Noun Bias” for all languages (N/V ratio is larger than 1 under these situations). But as children age, the “Noun Biases” become weaker, approaching the level of adults vocabularies as shown on Figure 2.b. In addition, compared with English, the “Noun Bias” in Mandarin and Cantonese is much weaker; for certain age ranges, the number of nouns is quite close to, or even lower than verbs (e.g., for Mandarin in 19-24 months). This difference between languages is also seen in Figure 2.b, where English-speaking adults show a stronger N/V ratio than Chinese or Cantonese speaking adults.

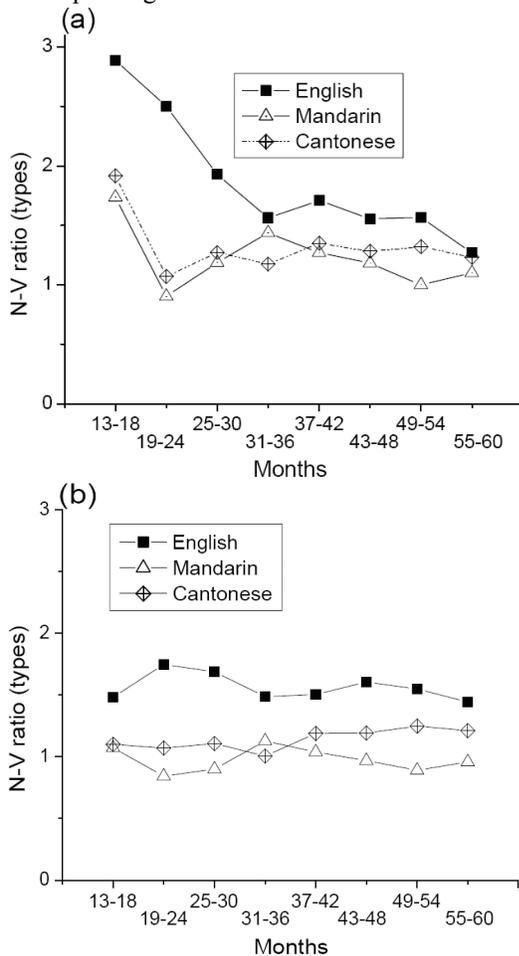


Figure 2. Noun/verb ratio (in types) as a function of age: (a) children, and (b) adults.

### Cluster Analysis

Based on the detailed information of lexical composition (with 30 subcategories), we applied a cluster analysis on the 48 situations as mentioned above. Cluster analysis is a type of interdependence multivariate statistical technique, which can group observations into smaller clusters, and the observations in same cluster are more similar to each other in structure than to observations in other clusters (Hair, Anderson, Tatham, & Black, 1998). The dendrogram of the analysis shows that the 48 situations are by and large

clustered into three large groups according to language (Figure 3). This means that the lexicons of people (children and adults) speaking the same language have similar lexical compositions. The language factor is the most important factor that distinguishes the 48 different situations under the child language context. In addition, the lexicons of Mandarin and Cantonese speaking people are more similar in lexical compositions, as the clusters of the two languages are closer and on the same branch of the cluster tree. This result is consistent with the earlier findings from the ANOVA analysis. A cluster analysis was also applied to analyze the similarity of the situations occurring under each language. The results show that, in the same language, adult lexicons often share a similar composition pattern, which differs from the lexical compositions of children. But as children grow up, their lexical composition patterns become increasingly similar to those of their parents.

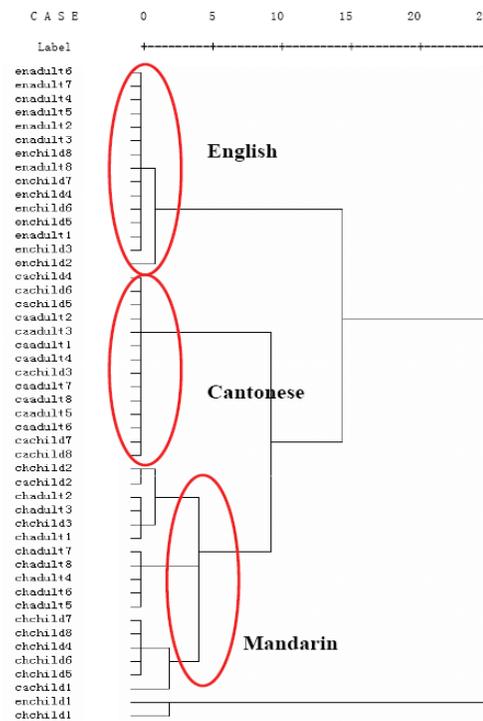


Figure 3. Cluster analysis of the similarity among the vocabularies of 48 situations. Ward method was used.

### Conclusion

We can summarize the major findings from our corpus-based studies as follows: (1) In all three languages, children’s early vocabularies show roughly similar patterns: an increasing diversity and complexity in lexicon as a function of time/age (as shown in D-measure of children’s speech) (Figure 1 and Tables 3); children’s vocabularies also become more similar to those of their parents over time (Figure 2). (2) Crosslinguistic variations in children’s linguistic input have strong influences on their language output, which is reflected in the varying percentages of nouns, verbs, and adjectives in the total words children are able to speak in the three languages (Figure 3).

What are the driving forces for these similarities and differences in children's early lexical development? Some cross-linguistic studies suggest that the presence or absence of the "Noun bias" in different languages might be due to multidimensional factors. First, children's language input might affect their early vocabulary composition. According to Sandhofer, Smith and Luo (2000), in both English- and Mandarin- speaking parents' speech, nouns often have a "flat distribution" in frequency, while verbs follow a "steep distribution", which causes nouns to be more easily learned regardless of language. It is also suggested that Chinese adults emphasize verbs over nouns when they speak to their children, while English parents use more nouns than verbs (Tardif, et al. 1999). Second, the characteristics of the to-be-learned language itself can also be important. For example, in a recent review, Tardif (2006) stated that usage of verbs tends to be highly specific in Chinese; whereas in English, general purpose verbs are commonly used. This characteristic can partly contribute to the privilege of verbs in Chinese children's early words.

In a recent computational study (Zhao & Li, 2008), we investigated the lexical development of English and Mandarin in a neural network model. The results are consistent with our corpus analyses here; and our simulation results suggest that phonemic lengths and the occurrence frequencies of words in children's language input also play important roles in the particular lexical compositions observed in different languages. More importantly, our study presents a dynamic developmental picture for early lexical acquisition, which is dependent on the joint contributions of mechanisms of learning and characteristics of the learning environment.

Large-scale corpus-based studies such as ours presented here, along with empirical and computational studies, should allow us to further elucidate the mechanisms of early lexical development, and of language acquisition in general (see Li, Zhao, & MacWhinney, 2007; Zhao & Li, 2008).

### Acknowledgments

This research is based on first author's Master Thesis completed at the University of Richmond. We thank the UR Graduate Research Committee for their kind support.

### References

- Bates, E., Dale, P.S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *Handbook of child language*. Oxford: Basil Blackwell.
- Caselli, M.C., Bates E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, 10, 159-199.
- Choi, S. (1997). Language-specific input and early semantic development: Evidence from children learning Korean. *The crosslinguistic study of language acquisition, Vol. 5: Expanding the contexts* (pp. 41-133). Lawrence Erlbaum Associates Publishers.
- Choi, S. (2000). Caregiver input in English and Korean: Use of nouns and verbs in book-reading and toy-play contexts. *Journal of Child Language*, 27(1), 69-96.
- Dale, P.S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, 28, 125-127.
- 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* (pp. 301-334). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hao, M., Xing, A., Shu, H., & Li, P. (2008). Early vocabulary inventory for Mandarin Chinese. *Behavior Research Methods*. (in press)
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W.C. (1998). *Multivariate data analysis* (5<sup>th</sup> ed.). Upper Saddle River, NJ: Prentice Hall.
- Institute of Linguistics of Chinese Academy of Social Sciences (CASS). (2002). *Modern Chinese dictionary*. Beijing, China: The Commercial Press.
- Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*, 31, 581-612.
- Lǚ. S. (2001). *Eight hundred words in modern Chinese*. Beijing, China: The Commercial Press.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum.
- McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, 15, 323-338.
- Miller, G. A. (1990). WordNet: An on-line lexical database. *International Journal of Lexicography*, 3, 235-312.
- Nelson, K. (1973). Structure and strategy in learning to talk. *Monograph of the Society for Research in Child Development*, 38, 1-2, Serial #149.
- Sandhofer, C., Smith, L., & Luo, J. (2000). Counting nouns and verbs in the input: Differential frequencies, different kinds of learning?. *Journal of Child Language*, 27(3), 561-585.
- Tardif, T. (1996). Nouns are not always learned before verbs. Evidence from Mandarin speaker's early vocabularies. *Developmental Psychology*, 32, 492-504.
- Tardif, T. (2006). But are they really Verbs? Chinese words for action. In K. Hirsh-Pasek, R.M. Golinkoff (Eds.), *Action Meets Word: How Children learn Verbs*. New York: Oxford University Press.
- Tardif, T., Gelman, S.A., & Xu, F. (1999) Putting the "noun bias" in context: a comparison of English and Mandarin. *Child Development*, 70, 620-635.
- Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language*, 24(3), 535-565.
- Zhao, X., & Li, P. (2008). Vocabulary development in English and Chinese: A comparative study with self-organizing neural networks. *This Volume*.