Word frequency of the CHILDES corpus: Another perspective of child language features

Hanhong LI and Alex C. Fang
Department of Chinese, Translation and Linguistics
City University of Hong Kong

Abstract

Based on the corpus of the Child Language Data Exchange System (CHILDES), the current research explores the grammatical composition of child language in terms of word classes. Unlike past studies, the research reported in this article examines not only nouns and verbs but also adjectives, adverbs, prepositions, pronouns, determiners, and interjections. We investigated the word frequency patterns of these word classes in both child and maternal language in order to explore the correlation between input word frequency on the mother's side and the output word frequency on the child's side. Our results show that the frequency patterns for word classes differ between child language and maternal language. Due to children's mental development, young children use many more words with concrete and imageable referents such as nouns and pronouns than less concrete or imageable words such as adjectives, adverbs, prepositions, interjections and conjunctions. Data prove that nouns are the most frequently used word class in child language. In addition, children acquire monosyllabic words more easily and use them much more frequently than multi-syllabic words except for those with extremely high frequency. Moreover, our study reveals a positive correlation between the token volume of maternal language output and that of child language output, suggesting the important role of maternal language in children's language acquisition. A principle of comprehensible input should be highlighted in adults' speech to children in order to make them achieve larger vocabulary. This discovery in first language acquisition will hopefully help further studies in second or foreign language acquisition, learning and teaching.

1 Introduction

Research into word frequency and child language has been interesting to many scholars. Much of the previous research focusses on the study of nouns and verbs in terms of word frequency in child language. Though Sandofer, Smith and Luo (2000) try to analyze English-speaking parents' speech to children via the Child Language Data Exchange System (CHILDES), a large corpus of child language and adult language (cf. MacWhinney 2000, 2007), their study is mainly directed to the use of nouns and verbs. Moreover, previous studies direct their attention to parental input (production of language by parents) without comparing the child output (production of language by children). They do not compare the word frequency patterns between child language and parental language, and consequently fail to establish any correlation between them.

Word frequency plays an important role in our mental access to lexical information. Word frequency refers to "how often the word occurs in normal use of the language" (Nation and Warning 1997: 8). It is natural that some words occur more often than other words in our daily conversation or in certain situations. Carroll (1999) regards word frequency as one of the major factors which influence the process of accessing or retrieving lexical information from memory. Studies have demonstrated that phoneme recognition is accelerated with the use of high frequency words as compared to low frequency words (Foss 1969), as are visual word recognition tasks (Rubenstein, Garfield and Milliken 1970; Whaley 1978). High frequency words tend to be recognized more accurately and accessed faster in our mental lexicon.

Word frequency plays an active role in lexicon acquisition. Even children with language impairment learn verbs more efficiently if they are presented frequently and in an appropriate time spacing (Riches, Tomasello and Conti-Ramsden 2005). Brown (1958) pointed out that words that are frequently used in speech to children tend to match the children's cognitive predilections. Moreover, studies have shown that, when people of different ages are asked to write definitions of words, word frequency shows a strong influence on their definitions of adjectives (Marinellie and Johnson 2003) and the definitions of nouns and verbs (Marinelllie and Chan 2006). This suggests that the mental lexicon progresses and organizes high-frequency and low-frequency words differently.

Since word frequency impacts children's word cognition, the question is whether the word frequency in input will influence children's word frequency in output. Goldfield (1993) found a significant correlation between the word types of nouns used by mothers and the word types of nouns in their children's first 50 words (as cited in Sandhofer, Smith and Luo 2000). Researchers have found that early child vocabulary contains more nouns than verbs (Nelson 1973; Fenson et al. 1994). This is because "children hear more nouns than other kinds of words" and "it reflects biases in the child, that is a propensity for learning names for things before more relational terms" (Sandhofer et al. 2000: 562). Moreover, nouns tend to be acquired earlier than verbs (Gentner 1982), while verbs tend to be of lower imageability than nouns (Jones 1985; Masterson and Druks 1998; Bird, Franklin and Howard 2001). Imageability is a vital variable of the mental lexicon. Data show that imageability heavily influences age of acquisition and suggest that "the most imageable, most concrete concepts are acquired before less imageable, less concrete concepts" (Morrison, Chappell and Ellis 1997: 546). However, most of the research on imageablity and age of acquisition is based on subjective rating by experimental subjects (Juhasz 2005) instead of large corpora.

2 Aims of the study

Taking advantage of large modern corpora, the current study explores the word frequency patterns of child language in terms of different word classes including nouns, verbs, adjectives, adverbs, prepositions, pronouns, determiners, and interjections. The word frequency patterns of the counterparts in maternal language are investigated at the same time so as to study the correlation between the input and output word frequency of child language.

3 Methods of the study

Subject

The Manchester child language corpus was downloaded from the website of CHILDES database.[1] This corpus consists of the transcripts of audio recordings from a longitudinal study of 12 English-speaking children between the age of 1;8.22 (1 year 8 months and 22 days) to 2;0.25 (2 years and 25 days) with MLU (mean length of utterance) ranging between 1.06 and 2.27 in morphemes (cf. British English Manual[2]). The children were recruited through newspaper advertisements and local nurseries. All the children were first born, monolingual and cared for primarily by their mothers. Although socioeconomic status was not taken into account with respect to recruitment, the children were from predominantly middle-class families. There were six boys and six girls, three from Manchester and three from Nottingham. The transcripts for each child were numbered from 1 to 34 corresponding to the tape number and labeled (a) and (b) to correspond to the two 30-minute sessions within each recording.

The corpus was chosen on the following bases:

1. The corpus contains longitudinal data for a whole year. The children were audiotaped in their homes for an hour on two separate occasions in every 3-week period for one year.
2. In terms of gender control in the corpus, the project recruited six boys and six girls.
3. The corpus contains mainly dialogues between mothers and children. The participants were engaged in normal play activities with their mothers. For the first 30 minutes of each hour they played with their own toys whilst for the second 30 minutes toys provided by the experimenter were available to the child. For the duration of the recordings, the experimenter attempted as far as possible to remain in the background to allow contextual notes to be taken.
4. With around 2 million words, the corpus is large enough for a general analysis of word frequency. There are in total 6,725 different word types and 2,164,626 tokens in the corpus.

Procedure

There are two versions of CHILDES data: the CLAN version with its specialized CLAN software to search data, and the xml version.[3] Our current research is based on the xml version of the Manchester child language corpus, one component of the British English child language corpora.[4] Since the CHILDES corpora are mainly composed of dialogues between children and mothers, we had to set up two sub-corpora in the following steps.

First, with the help of the software Powergrep4.0[5] the sub-corpus of child language was extracted by using the regular expression <u who="CHI" uID="[^"<>]*?">(.*?)</u>, where CHI stands for children. Second, the sub-corpus of maternal language was extracted by using the regular expression <u who="MOT" uID="[^"<>]*?">(.*?)</u>, where MOT stands for mothers. Third, in order to research the grammatical composition of the major parts of speech, we use the regular expressions in Table 1 to extract the tokens and types in child and maternal language. The major parts of speech are nouns, verbs, adjectives, adverbs, pronouns, prepositions, conjunctions, determiners, and interjections, based on Table 45 of the CHAT manual.[6]

Table 1: The regular expressions to extract major parts of speech in child language

Word class	**Regular expression**
noun	<pos><c>n</c>(<s>.*?</s>)*</pos><stem>(.*?)</stem>
verb	<pos><c>v</c>(<s>.*?</s>)*</pos><stem>(.*?)</stem>
adj	<pos><c>adj</c>(<s>.*?</s>)*</pos><stem>(.*?)</stem>
adv	<pos><c>adv</c>(<s>.*?</s>)*</pos><stem>(.*?)</stem>
pronoun	<pos><c>pro</c>(<s>.*?</s>)*</pos><stem>(.*?)</stem>
preposition	<pos><c>prep</c>(<s>.*?</s>)*</pos><stem>(.*?)</stem>
conjunction	<pos><c>conj</c>(<s>.*?</s>)*</pos><stem>(.*?)</stem>
determiner	<pos><c>det</c>(<s>.*?</s>)*</pos><stem>(.*?)</stem>
interjection	<pos><c>int</c>(<s>.*?</s>)*</pos><stem>(.*?)</stem>

Other parts of speech have been adopted in this corpus besides the conventional parts of speech. Since these minor parts of speech have very few tokens, they are treated together as an individual group, named OTHERS for our data analysis (see Table 2).

Table 2: Minor parts of speech

Abbreviation	**Part of speech**
aux	auxiliary verb, including modals
bab	babbling
chi	child-invented form
co	communicator
fam	family-specific form
inf	infinitive marker to
L2	second-language form
neg	the negative marker
neo	neologism

on	onomatopoeia
part	participle
phon	phonol. consistent form
poss	possessive
post	postposed quantifier
ptl	particle
qn	quantifier
rel	relative pronoun
sing	singing
tag	tag marker
test	test word
unk	unknown
wplay	word play

After this we compared the mothers' word frequency, which is the children's input word frequency, with the children's output word frequency. Finally, we apply the Statistical Package for the Social Science (SPSS) , a statistic software, to see if there is a correlation between the mothers' and the children's word frequency.

Tools

We used the software Powergrep 4.0, the Manchester child language corpus in the CHILDES Database, and SPSS in the research procedure.

4 Results

4.1 Grammatical composition of child language

The type and token information for each word class in the children's language is shown in Table 3. In terms of vocabulary use, we can see that children use many more tokens of pronouns, nouns, verbs and adverbs than any other word class (see Figure 1). They acquire and use more nouns than verbs. This finding agrees with other researchers' results (Nelson 1973; Fenson et al. 1994). Nouns tend to be acquired earlier than verbs (Gentner 1982) and nouns tend to be of higher imageability than verbs (Jones 1985; Masternson and Drunks 1998; Bird, Franklin and Howard 2001). In addition, Figure 1 shows that children use more nouns than any other word class, since “children hear more nouns than other kinds of words” and “it reflects biases in the child, that is a propensity for learning names for things before more relational terms” (Sandhofer et al. 2000: 562).

Moreover, children use many more adverbs and determiners than adjectives, prepositions, conjunctions and interjections as shown in Figure 1. This result matches other researchers’ predictions or beliefs that English-speaking children begin learning nouns before other types of vocabulary, and that function words, which tend to attract relatively low imageability ratings, are generally produced only when the child begins to use multiword utterances (Gentner, 1982; Bates et al. 1994; Bird et al. 2001).

If we regard the word type total as children’s vocabulary size, we can research the composition of different word classes in children’s vocabulary. Figure 2 shows that nouns, verbs and adjectives occupy the major parts of their vocabulary. Is it influenced by their maternal language? For a further study, see section 4.3.

Table 3: Types and tokens of child language

Word class	**Type**	**Token**	**Token proportion (token/total token%)**	**Type proportion (type/total type%)**
noun	3,080	114,141	19.66%	56.61%
verb	682	80,843	13.93%	12.53%
adjective	441	16,455	2.83%	8.11%
adverb	147	55,947	9.64%	2.70%
interjection	46	2,152	0.37%	0.85%
pronoun	52	93,280	16.07%	0.96%
determiner	37	41,322	7.12%	0.68%
preposition	43	17,021	2.93%	0.79%
conjunction	15	8,680	1.50%	0.28%
others	898	150,631	25.95%	16.50%
Total	5,441	580,472	100.00%	100.00%

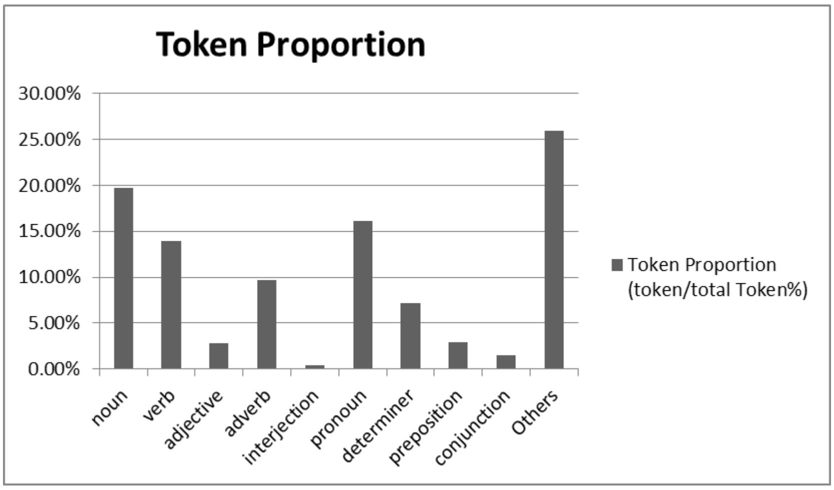

Figure 1: Token proportion of major word classes in child language

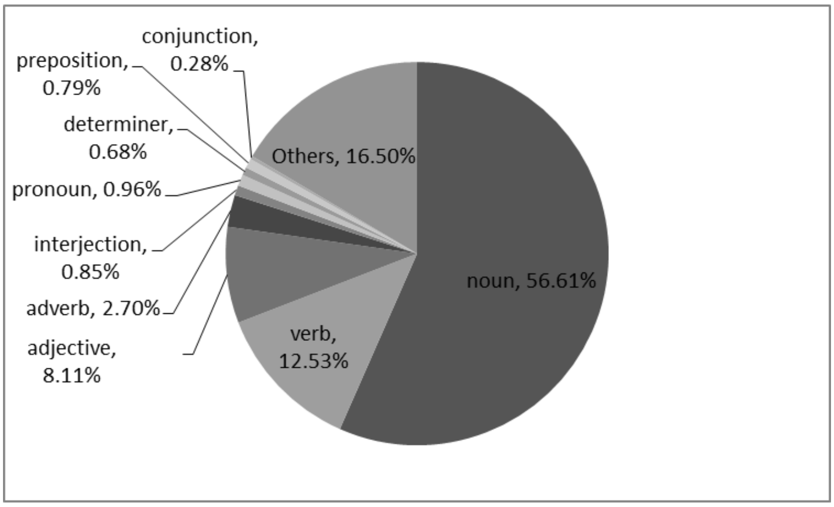

Figure 2: Type proportion of children's vocabulary

4.2 Grammatical composition of maternal language

Table 4 shows the totals and proportions of tokens and types of different word classes in the maternal language. Since each dialogue in our corpus is mainly between mothers and children, we can regard the maternal language as children's input language. In terms of tokens, i.e. vocabulary use, we find that mothers use many more pronouns than any other word class when they speak to children (see Figure 3). In mothers' speech to children, we notice many more pronouns, verbs and nouns than other word classes. Mothers use far fewer adjectives, adverbs, interjections, determiners, prepositions, and conjunctions as shown in Figure 3.

In terms of word types, i.e. vocabulary size, it turns out that nouns, verbs, and adverbs occupy the major parts of the maternal vocabulary (see Figure 4). This tends to bear similarity with the grammatical composition of child language (see Figure 2). In order to investigate the relationship between maternal and child language, we have compared them in section 4.3.

Table 4: Types and tokens of maternal language

Word class	**Type**	**Token**	**Token proportion (token/total token%)**	**Type proportion (type/total type%)**
noun	5,354	199,533	13.24%	58.20%
verb	1,256	247,473	16.43%	13.65%
adjective	1,062	47,485	3.15%	11.54%
adverb	302	123,403	8.19%	3.28%
interjection	72	6,848	0.45%	0.78%
pronoun	60	321,999	21.37%	0.65%
determiner	40	96,899	6.43%	0.43%
preposition	62	67,904	4.51%	0.67%
conjunction	24	31,098	2.06%	0.26%
others	968	363,970	24.16%	10.52%
Total	9,200	1,506,612	100.00%	100.00%

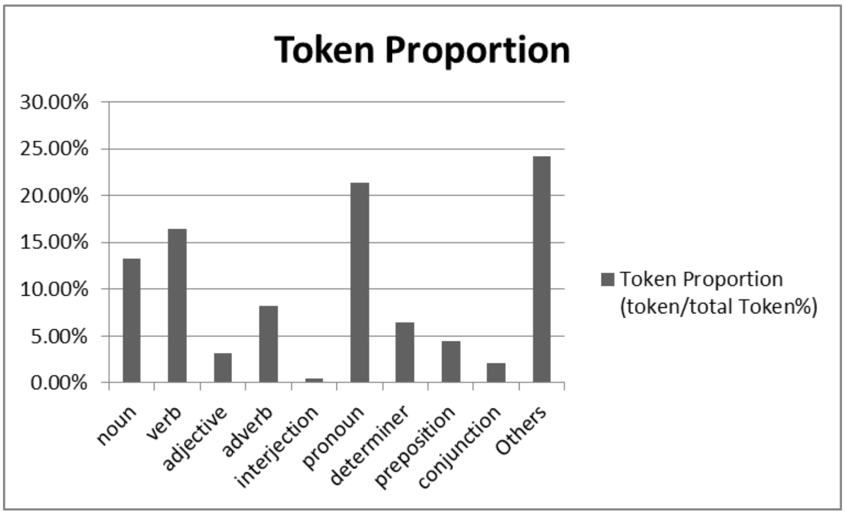

Figure 3: Token proportion of different word classes in maternal language

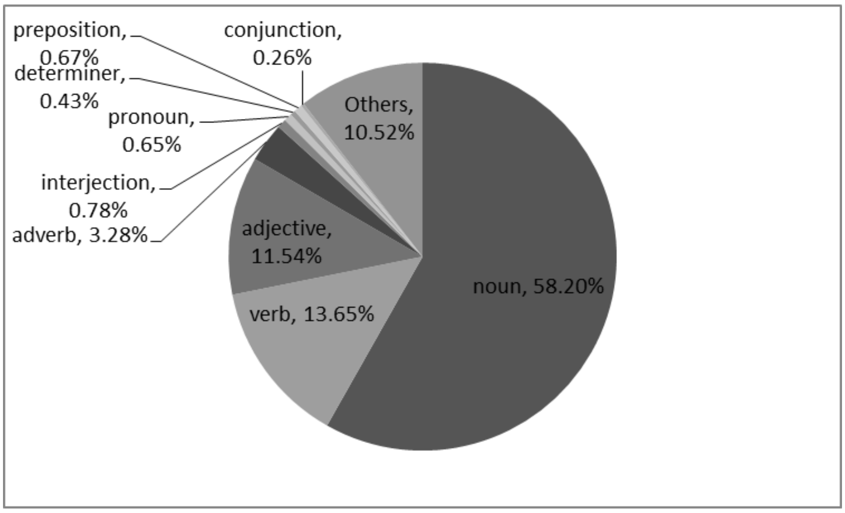

Figure 4: Type proportion of maternal vocabulary

4.3 Comparison between child and maternal language

4.3.1 Children's word tokens vs maternal word tokens

Figure 5 (based on Table 5) shows the difference of tokens between maternal language and child language in terms of different word classes. It turns out that the word tokens of each word class in maternal language are always more frequent than those in child language. Verbs and pronouns display the biggest gap between maternal and child language. Mothers use many more pronouns than their children.

Table 5: Word tokens in child language vs in maternal language

Word class	**Child language**	**Maternal language**
noun	114,141	199,533
verb	80,843	247,473
adjective	16,455	47,485
adverb	55,947	123,403
interjection	2,152	6,848
pronoun	93,280	321,999
determiner	41,322	96,899
preposition	17,021	67,904
conjunction	8,680	31,098
others	150,631	363,970
Total	580,472	1,506,612

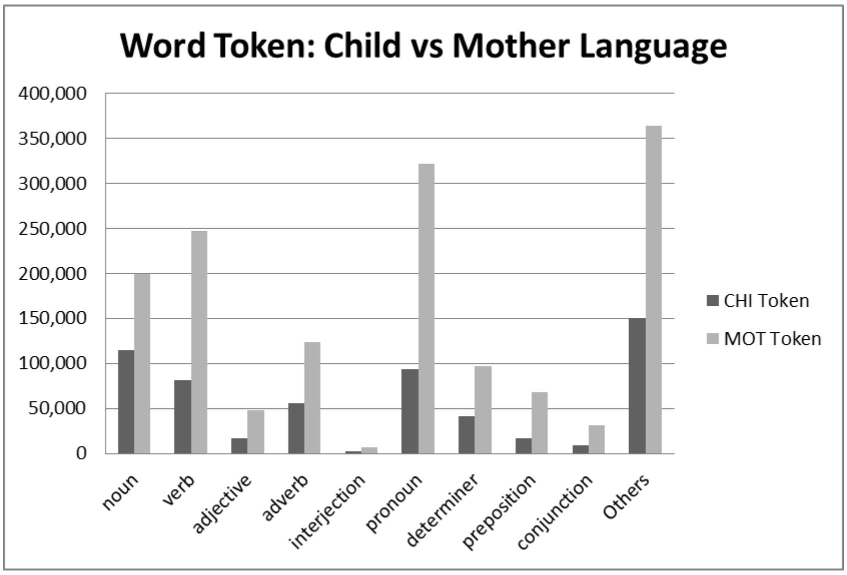

Figure 5: Tokens in child language vs in maternal language
CHI Token = token in child language
MOT Token = token in maternal language

In order to determine if there is a correlation between the word tokens in mothers' speech and the word tokens in children's speech, we used SPSS to analyze their Pearson correlation coefficient. The result in Table 6 is P<.001, r=0.936. That indicates that the number of words spoken by the children is significantly correlated with the number of words spoken to them by their mothers. It also demonstrates a close relationship between language input from mothers and language output from their children.

Table 6: The correlation between word tokens in child and maternal language

Pearson correlation	**CHI token**	**MOT token**
CHI token	1	.936(**)
Sig. (2-tailed)		.000
Group number	10	10

** Correlation is significant at the 0.01 level (2-tailed).
CHI token = token in child language
MOT token = token in maternal language

4.3.2 Children's word types vs maternal word types

Since word tokens are the frequency counts of the occurrences of different words while word types refer to the number of unique words, word types can better mirror a speaker's vocabulary size. In this case, we can compare children's vocabulary size with mothers' vocabulary size. Figure 6 (based on Table 7) shows that nouns, verbs and adjectives are the major components of both children's and mothers' vocabulary size. Naturally, due to language development, mothers know many more nouns, verbs, adjectives and adverbs than children. However, when it comes to interjections, pronouns, determiners, propositions and conjunctions, the gap between children and mothers is not large.

Table 7: Word types in child language vs in maternal language

Word class	**Child language**	**Maternal language**
noun	3,080	5,354
verb	682	1,256
adjective	441	1,062
adverb	147	302
interjection	46	72
pronoun	52	60
determiner	37	40
preposition	43	62
conjunction	15	24
others	898	968
Total	5,441	9,200

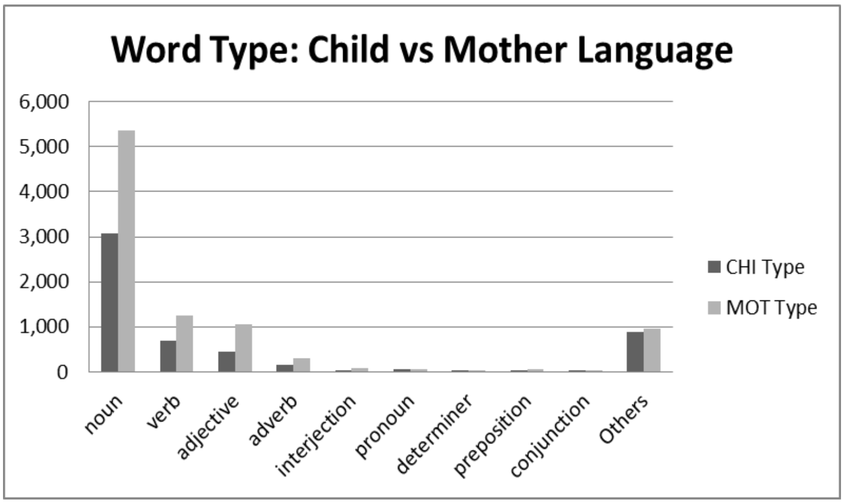

Figure 6: Types in child language vs in maternal language
CHI Type = type in child language
MOT Type = type in maternal language

Moreover, when we use SPSS to analyze the correlation between the types in child language and maternal language for different word classes, it turns out that there is a significant correlation between mothers' vocabulary size and children's vocabulary size with P<.001, r=0.991 (see Table 8). This result indicates that the variety of mothers' vocabulary in speech will influence the variety of their children's vocabulary.

Table 8: The correlation between word types in child language and those in maternal language

Pearson correlation	**CHI type**	**MOT type**
CHI type	1	.991(**)
Sig. (2-tailed)		.000
Group number	10	10

** Correlation is significant at the 0.01 level (2-tailed).
CHI type = type in child language
MOT type = type in maternal language

4.3.3 Children's word types vs maternal word tokens

Word tokens in child language cannot be equated exactly to the number of words children have acquired in their vocabulary since children may just repeat whatever they hear without understanding it. When it comes to whether children have acquired those words or not, the total of word types in child language can be more representative. In this case, the total of children's word types represents their vocabulary size in the children's language output while mothers' word tokens become the children's language input in the current corpus.

In order to explore the correlation between the children's language input and output, we carried out the Pearson's correlation coefficient between the word types in child language and word tokens in maternal language. The result in Table 9 is r=0.330, P<0.5. This means that the number of words acquired by children is not significantly related with the number of words spoken to them by their mothers. Obviously, language acquisition is not only a matter of input amount but also a matter of comprehension. Krashen (1985: 21) claims that "humans acquire language in only one way – by understanding messages, or by receiving 'comprehensible input'". Therefore, if mothers want to help their children to learn to speak, they need to not only talk to their children more but also make themselves understood to their children. Comprehensible input plays a significant role in language acquisition.

Table 9: The correlation between the number of words acquired by children and the number of words spoken to them by their mothers

Pearson correlation	**CHI type**	**MOT token**
CHI type	1	.330
Sig. (2-tailed)		.351
Group number	10	10

CHI type = type in child language
MOT token = token in maternal language

5 Discussion

The above data have shown interesting word frequency patterns of child language and maternal language. There are a few distinctive features for child language.

First, due to imageability and concreteness, pronouns and nouns are used very frequently by mothers and children as displayed in Figure 1 and Figure 3 (or in the combined Figure 5). Children acquire more nouns than verbs in terms of types (see Figure 2), and they use more nouns and pronouns than verbs in terms of tokens (see Figure 1).

If we look at the 20 nouns most frequently used by children (see Table 10), we find that most of them are concrete words which have immediate imageable referents near the children. For example, the word *mummy* is used most frequently by children and mothers since mummy is the main person who talks to the children. The synonyms of *mummy*, *mum* and *mama* are also among the ten most frequent nouns in child language. Words such as *car, train, door, house,* and *doll* could be the toys which were at their daily disposal.

Another interesting finding is that all the top 20 nouns are monosyllabic words which are easier for children to acquire, with the exception of *mummy*, *mama*, *baby* and *daddy* which are disyllabic and rank among the ten most frequently used nouns in child language. Word frequency thus plays a significant role in the children's word acquisition, as is evidenced by the number of words in the top 20 that are used by both mothers and their children, such as *car* and *baby* (see Table 10). This agrees with our previous finding in Table 6 which shows a significant correlation between word tokens in child language and those in maternal language.

Furthermore, pronouns such as *I, me* and *you* rank high in the word frequency list of child language. Obviously, the referents of these pronouns are easily found by the children. Therefore, with concrete and imageable referents, nouns and pronouns are spoken more frequently and learned faster than any other word class.

Table 10: Comparing the 20 nouns most frequently used by children and mothers

	Children		**Mother**	
Rank	**Frequency**	**Noun**	**Frequency**	**Noun**
1	4963	*mummy*	5361	*mummy*
2	3241	*car*	3782	*bit*
3	2577	*baby*	2753	*car*
4	1911	*train*	2747	*baby*

5	1774	*daddy*	2471	*daddy*
6	1243	*mum*	2059	*train*
7	1211	*man*	1880	*look*
8	1125	*doll*	1635	*way*
9	1106	*bit*	1611	*color*
10	975	*mama*	1487	*doll*
11	957	*horse*	1392	*thing*
12	891	*cow*	1338	*box*
13	844	*egg*	1335	*man*
14	786	*bridge*	1247	*boy*
15	782	*house*	1213	*horse*
16	781	*way*	1206	*girl*
17	752	*truck*	1158	*book*
18	710	*cat*	1141	*cat*
19	707	*toy*	1110	*egg*
20	702	*dog*	1076	*dog*

Second, the verb is also of high frequency in the children's language. We find that verbs are closely related with the children's own sense of movement and their own desires. For example, among the 30 most frequent verbs, we find verbs associated with psychological desires such as *want, get, like, need, know* and *think*, and others associated with physical movements such as *go, come, put, sit, do, eat* and *play* (see Table 11). Moreover, mothers tend to describe what babies are doing or express what they want their babies to do. That is why the children can understand and learn these words faster. It is interesting that all the top 30 verbs in the children's language are monosyllabic. By contrast, although mothers used *remember*, which has three syllables, many times, the children still failed to acquire it. Generally speaking, children tend to acquire monosyllabic verbs faster and use them more frequently than multi-syllable verbs.

Table 11: Comparing the 30 verbs most frequently used by children and mothers

	Children		Mother	
Rank	**Frequency**	**Verb**	**Frequency**	**Verb**
1	19737	*be*	87018	*be*
2	7091	*go*	14517	*have*
3	6757	*want*	10939	*go*
4	6720	*get*	9291	*get*
5	2958	*have*	8542	*want*
6	2904	*put*	8226	*put*
7	2461	*do*	8146	*think*
8	1789	*come*	7421	*do*
9	1520	*make*	5992	*come*
10	1219	*like*	5399	*see*
11	1160	*sit*	4302	*know*
12	1148	*see*	3740	*look*
13	982	*need*	3706	*like*
14	979	*know*	3293	*say*
15	950	*find*	3076	*let*
16	911	*eat*	2989	*make*
17	834	*take*	2194	*need*
18	762	*look*	2165	*take*
19	711	*play*	1936	*find*
20	675	*let*	1786	*play*
21	573	*stick*	1647	*sit*
22	510	*fall*	1620	*eat*
23	509	*think*	1605	*tell*
24	481	*say*	1375	*give*
25	444	*draw*	1161	*will*
26	388	*open*	1119	*remember*

27	372	*fix*	789	*fall*
28	348	*read*	784	*draw*
29	341	*build*	772	*pull*
30	290	*buy*	754	*keep*

Third, as shown in Figure 1 and Table 3, children tend to use far fewer adjectives, prepositions and conjunctions due to the low imageability of these words and children's mental development. The reason is that adjectives are more abstract and less imageable than nouns and pronouns, which have concrete referents. Prepositions involve spatial analysis of different objects, which requires more mental intelligence. Conjunctions require logical analysis of different syntactic elements, which is not easy for a two-year-old child.

Fourth, word tokens in children's language input vastly outnumber those in the children's language output, and a comprehensible input helps to build up children's vocabulary size. Mothers produce many more word tokens than their children (see Figure 5). Moreover, as demonstrated in Table 6, there is a correlation between word tokens spoken by children and those spoken by mothers to children, which indicates that the quantity of language input influences children's language output. In this case, in order to make children produce more language output, mothers have to talk to them more.

However, a large quantity of language input does not ensure successful language output. Table 9 shows that the number of words *acquired* by children is not significantly correlated with the number of words spoken to them by their mothers. Obviously, comprehensibility of the input is a factor. In other words, if we want children to really acquire words instead of just murmuring and repeating sounds, we have to make them understand what we are saying to them. By doing so, we can help children build up their vocabulary faster and talk more properly.

6 Conclusion

By exploring the grammatical composition of child language in terms of word classes, our study shows that the frequency patterns for word classes in child language differ from those of maternal language. Children tend to use many more words with concrete and imageable referents such as nouns and pronouns than less concrete or imageable words such as adjectives, adverbs, prepositions, interjections and conjunctions. Nouns are the most frequently used word class in child language. Moreover, children tend to acquire monosyllabic words more easily and use them much more frequently than multi-syllabic words except for those with extremely high frequency such as *mummy* and *daddy*. Furthermore, more input becomes a requisite for output in child language. Our study reveals a positive correlation between the token total of maternal language output and the token total of child language output, suggesting the important role of maternal language in children's language acquisition. Though maternal word frequency has a great impact on child word frequency, a principle of comprehensible input should be highlighted in adults' speech to children so that children can achieve larger vocabulary size and reproduce concrete concepts comprehensibly. This discovery in first language acquisition will hopefully help further studies in second or foreign language acquisition, learning and teaching.

Acknowledgement

The work reported in this article was supported in part by research grants received from City University of Hong Kong (Grant Nos 7008062, 7008002 and 7002387). The authors would like to acknowledge the assistance received from the members of the Dialogue Systems Group, in particular Ms Kara Warburton, for comments on an earlier draft.

Notes

1. http://childes.psy.cmu.edu/data/Eng-UK/
2. http://childes.psy.cmu.edu/manuals/
3. http://childes.psy.cmu.edu/data/local.html
4. http://childes.psy.cmu.edu/data-xml/Eng-UK/
5. http://www.powergrep.com/
6. http://childes.psy.cmu.edu/manuals/chat.pdf

References

Bates, Elizabeth, Virginia Marchman, Donna Thal, Larry Fenson, Philip Dale, J. Steven Reznick, Judy Reilly and Jeef Hartung. 1994. Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language* 21: 85–123.

Bird, Helen, Sue Franklin and David Howard. 2001. Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers* 33 (1): 73–79.

Brown, Roger. 1958. How shall a thing be called? *Psychological Review* 65: 14–21.

Carroll, David W. 1999. *Psychology of language*. Second edition. Pacific Grove, Calif.: Brooks/Cole.

Caselli, Maria Cristina, Elizabeth Bates, Paola Casadio, Judi Fenson, Larry Fenson, Lisa Sanderl and Judy Weir. 1995. A cross-linguistic study of early lexical development. *Cognitive Development* 10: 159–199.

Fenson, Larry, Philip S. Dale, J. Steven Reznick, Elizabeth Bates, Donna J. Thal and Stephen J. Pethick. 1994. Variability in early communicative development. *Monographs of the Society for Research in Child Development* 59(5) Serial No. 242.

Foss, Donal J. 1969. Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. *Journal of Verbal Learning and Verbal Behavior* 8 (4): 457–462.

Genter, Dedre. 1982. Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S.A. Kuczaj (ed*.). Language development.* Volume 2: *Language, thought and culture*, 301–334. Hillsdale, NJ: Lawrence Erlbaum.

Goldfeld, Beverly A. 1993. Noun bias in maternal speech to one-year-olds. *Journal of Child Language* 20: 85–100.

Jones, Gregory V. 1985. Deep dyslexia, imageability and ease of predication. *Brain & Language* 24: 1–19.

Juhasz, Barbara J. 2005. Age-of-acquisition effects in word and picture identification. *Psychological Bulletin* 131 (5): 684–712.

Krashen, Stephen D. 1985. *The input hypothesis: Issues and implications*. New York: Longman.

MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk*. Third edition. Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, Brian. 2007. The TalkBank Project. In J. C. Beal, K. P. Corrigan and H. L. Moisl (eds.). *Creating and digitizing language corpora.* Volume 1: *Synchronic databases*, 163–180. Houndmills: Palgrave-Macmillan.

Marinellie, Sally A. and Yen-Ling Chan. 2006. The effect of word frequency on noun and verb definitions: A developmental study. *Journal of Speech, Language, and Hearing Research* 49 (5): 1001–1021.

Marinellie, Sally A. and Cynthia J. Johnson. 2003. Adjective definitions and the influence of word frequency. *Journal of Speech, Language, and Hearing Research* 46 (5): 1061–1076.

Masterson, Jackie and Judit Druks. 1998. Description of a set of 164 nouns and 102 verbs matched for printed word frequency, familiarity and age-of-acquisition. *Journal of Neurolinguistics* 11: 331–354.

Morrison, Catriona M., Tameron D. Chappell and Andrew W. Ellis. 1997. Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology A* 50 (3): 528–559.

Nation, Paul and Robert Warning. 1977. Vocabulary size, text coverage and word lists. In N. Schmitt and M. McCarthy (eds.). *Vocabulary: Description, acquisition and pedagogy*, 6–19. Cambridge: Cambridge University Press.

Nelson, Katherine. 1973. Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development* Vol. 38, Nos 1–2, Serial No. 149:1–135.

Riches, Nick G., Tomasello, M. and Gina Conti-Ramsden. 2005. Verb learning in children with SLI: Frequency and spacing effect. *Journal of Speech, Language, and Hearing Research* 48 (6): 1397–1411.

Rubenstein, Herbert, Lonnie Garfield and Jane A. Millikan. 1970. Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior* 9: 487–494.

Sandhofer, Catherine M., Linda B. Smith and Jun Luo. 2000. Counting nouns and verbs in the input: Differential frequencies, different kinds of learning? *Journal of Child Language* 27: 561–585.

Whaley, C.P. 1978. Word–nonword classification time. *Journal of Verbal Learning and Verbal Behavior* 17 (2): 143–154.