

Running Head: CANTONESE EXPRESSIVE VOCABULARY

Is the Expressive Vocabulary of Young Cantonese Speakers Noun or Verb

Dominated?

Shek Kam Tse

Carol Chan

The University of Hong Kong

Hui Li

The Hong Kong Institute of Education

Date of Submission: May 8, 2004.

Correspondence should be addressed to Dr. Hui Li, School of Early Childhood

Education, The Hong Kong Institute of Education, 10 Lo Ping Road, N.T., Hong Kong.

Electronic mail may be sent to huli@ied.edu.hk.

The authors would like to extend their gratitude for the grant support of Professor Amy

Tsui, Dr. Terry Dolan, Dr. S. O. Leung and for the assistance in data collection of Ms. S.

K. Leung, Ms. W. Y. Choi, Mr. W. J. Lau, Ms. Dorcas Ng. Thanks go also to Dr. Twila

Tardif for her insightful comments and constructive suggestions on earlier drafts of this

paper.

Abstract

The spontaneous utterances produced by 492 Cantonese-speaking children aged 36 to 60 months in 30-minute toy play contexts were collected. The incidence and percentage of different lexical classes, the ratio of noun and verb types and tokens were calculated. A statistically significant predominance of verb usage was found in the expressive utterances of the children. No similar predominance was found for nouns. Linguistic, cultural and contextual variables possibly underlying these outcomes are explored and discussed.

Keywords: Cantonese, Expressive Vocabulary, Verb-dominated, Early Childhood

Is the Expressive Vocabulary of Young Cantonese Speakers Noun or Verb Dominated?

Most studies of the vocabulary acquisition of English-speaking children have found that nouns are learned earlier than verbs and that it is nouns that dominate early utterances (Goldfield & Reznick, 1990; Halliday, 1975; Markman, 1989; McShane, 1980; Nelson, 1973). Evidence from studies of Mandarin-speaking children (Tardif, Shatz & Naigles, 1997; Tardif, Gelman & Xu, 1999) and young children in Korea (Choi & Gopnik, 1995; Kim, McGregor & Thompson, 2000) has cast doubt on the universality of this phenomenon, with the discovery that verbs feature much more than nouns in very young children's utterances. A range of linguistic, cultural and contextual factors might be advanced to explain the differing patterns of lexical acquisition between the young children in the East Asian language and the English language studies. The researchers in the present study were interested in investigating patterns of lexical development in the case of Cantonese speaking preschoolers to see whether the acquisition patterns of 3-6 year olds follow the noun or verb dominant route.

Verb bias versus noun bias

For many years, studies both of English-speaking and non-English speaking children (Nelson, 1973; Halliday, 1975; McShane, 1980; Gentner, 1982; Markman, 1989; Goldfield & Reznick, 1990) found a predominance of nouns in early utterances and

vocabulary acquisition. The phenomenon whereby nouns are learned earlier than verbs and dominate children's early utterances is widely referred to as 'noun bias' (Gentner, 1982).

Despite the dominance of the noun bias assertion, there is gathering evidence that it may not apply universally, even in the acquisition of English. For example, a longitudinal study by Gopnik (1981) found that children generally acquired non-nominal words before nominals, and they tended to use non-nominal expressions in their earliest utterances. In other words, although it might still be true that many English-speaking children learn more nouns than verbs as their first words, this may not apply inevitably to all children. In addition, studies of the early acquisition of East Asian languages have yielded evidence that opposes the generality of the noun bias claim. Choi and Gopnik (1995) investigated Korean-speaking children's lexical development and found that most of the children displayed a period of rapid vocabulary growth characterised by a 'verb spurt' rather than the 'noun spurt' described by Goldfield and Reznick (1990). Moreover, for children who displayed both a verb and a noun spurt, the rapid increase in the acquisition of verbs tended to precede the rapid increase in the acquisition of nouns. Kim, McGregor and Thompson (2000) studied on the early productive vocabulary of Korean- and English-learning children and found that the Korean children learned significantly more verbs than did English-learning

children. They associated this difference with those in the linguistic and socio-pragmatic characteristics of the input addressed to the children, for example, Korean-speaking caregivers produced more activity-oriented utterances, more verbs, and more salient cues to verbs than did English-speaking caregivers.

Tardif (1996) found that 90% of 22-month-old Mandarin-speakers produced more verbs or action words than noun or object labels in their natural speech. Tardif and her colleagues (1999) later compared the proportions of nouns and verbs in the vocabulary of 24 English and 24 Mandarin-speaking infants (mean age 20 months) and their mothers. They found that Mandarin-speaking children used relatively fewer nouns and more verbs than their English-speaking counterparts did, and they suggest that the context itself may play an important role in determining the proportions of nouns in the children's early vocabulary. They argue that whether or not young language acquirers demonstrate a noun bias seems to depend on a variety of factors, including the methods by which their vocabulary is sampled and the contexts in which their speech is uttered.

A major problem with the reliability and validity of the data in the above studies concerns the way the vocabulary counts have been obtained. For example, in studies of Korean children, Au et al. (1994) based their evidence on mothers' reports or recollections of their children's vocabulary, whereas Choi & Gopnik (1995) based their

evidence on mothers' diaries or on-going recordings of all the words their children uttered, presumably in naturalistic settings. Whereas Au et al. reported more nouns in the children's early vocabulary, Choi and Gopik reported more verbs. This raises questions about whether parental reporting and naturalistic recording may yield different and possibly conflicting outcomes. Pine (1992) and Tardif (1996) propose that mothers are not completely unbiased observers and that parental reports appear to be biased toward an over-reporting of common nouns. It is well known that even adults may have mental concepts of nouns that they know but cannot articulate, and that neither adults nor children demonstrate perfect consistency of language use across all situations (Goldfield, 1993; Tardif et al., 1999). Nevertheless, studies of utterances in naturalistic settings may more validly reflect vocabulary acquisition and, for this reason, this was the chosen strategy in the present study.

The acquisition of Cantonese vocabulary

Regardless of whether children's vocabulary is noun-biased or verb-biased in English-speaking countries, there is no reason to assume *a priori* that the same will also be found in Chinese-speaking children. The overwhelming majority of studies of children's acquisition of Chinese feature Mandarin, rather than Cantonese. Mandarin is the name used by Westerners for the national language of China, which is officially named 'Putonghua' (Common Language). Mandarin is the spoken form of the Modern

Standard Chinese, which is widely accepted and used by about 77% of the 1.5 billion Chinese in the world, whereas about 130 million people speak Cantonese in Southern China, Hong Kong, Australia, Britain, Canada, and the United States. Cantonese is a dialect of Chinese and is mildly different from Mandarin as English is from Dutch in terms of phonology, vocabulary and pragmatics (Zheng, 1995).

The writers were interested in examining whether Tardif's findings about Mandarin-speaking children in Beijing also apply to Cantonese-speaking children in Hong Kong, and especially whether the patterns apply consistently as children grow older. In particular, would Cantonese speakers display a predominance of nouns in their vocabulary acquisition during the early childhood? To throw light on these questions, the spontaneous and natural utterances of 492 Cantonese children aged 36 to 60 months were collected and particular attention was paid to the incidence of nouns and verbs. The study endeavoured to investigate:

1. whether there is a statistically significant predominance of any particular lexical class in Cantonese-speaking children's vocabulary;
2. whether any pattern of predominance found applies consistently across ages and languages; and
3. Which socio-contextual factors might be accounting for any predominance found in the present study.

Method

Subjects

The young children in the present study were randomly selected from the participants of the International Educational Achievement (IEA) Pre-primary Project, which gathered data from Hong Kong pre-schoolers in order to establish a normative framework of developmental milestones for children aged 36 to 71 months. The sampling pool of the IEA Project consisted of all the Cantonese-speaking pre-schoolers in the 36 to 71 months age range in Hong Kong.

The sample in the present study consisted of 492 children selected from 68 pre-schools (58 kindergartens and 10 nurseries) located in Hong Kong Island, Kowloon and the New Territories. The subjects were randomly selected from each class at the participating pre-schools, representing three age groups (36, 48 and 60 months), with 82 boys and 82 girls in each age group. In order to minimise the influence of variability of home background and teaching approach, the number of children in each pre-school was limited to 10.

Communication task

A toy play context was set up in the selected children's classroom and furnished with a set of regular toys, including cooking materials, food and fruit, furniture and electrical appliances, hospital materials and vehicles. Each randomly arranged pair of

participants (boy/girl, boy/boy, or girl/girl) was left in the play corner to play for 30 minutes. They were encouraged to talk while they were playing and the 30-minute conversations were recorded using a cassette recorder. During the free play sessions, the researcher observed but did not intervene, and there were no other children in the classroom.

Transcription and coding

All conversations were audio-taped using high-fidelity equipment. Each conversation was first transcribed by one of two conversation researchers to a level of detail that captured all words and word fragments audible to the ear, as well as overlapping speech. Also transcribed were non-lexical fillers (such as “uh”) and other vocalisations (such as laughter). Transcriptions were made using a tape player that afforded automatic rewinding for repeated playing at slow and normal speeds. After each conversation was transcribed, other conversation researchers checked the transcript while listening to the tape to guarantee the transcripts achieved accuracy.

Coding of the lexical classes

Two specially trained research assistants entered the transcribed utterances into a computer using Microsoft Chinese Windows 98. Next, the Chinese script of each child’s oral language was segmented into utterances, then words (ci). Although no unchallenged definitions of a ‘word’ (ci) exist in Chinese linguistics, the writers believe

in the widely-accepted Wang's (1998) notion that a word (ci) is the smallest linguistic unit that can express meaning independently. Next, words were grouped as different classes according to their properties and the ways they were functioning in the sentences, by following a classification system modified from Zheng's (1995) and Ko's (1980) investigations (see Appendix 1). In the event, due to absences, the final database consisted of a total of 90,908 words from the 492 young children.

Reliability

The text transcript of the corpus was divided into halves, and two trained coders coded each half. Before starting the formal coding, the two coders coded six trials selected from the two halves, and the correlation between the two coding results was calculated, $r = .89$, $p < .001$, indicating excellent inter-rater consistency.

Results

The sample produced a mean of 48.65 (s.d. 37.35) utterances that yielded a mean of 184.77 (s.d. 52.38) vocabulary words for each child. The data included exclamations and onomatopoeia. Although the children's output varied both in the size of their overall vocabulary and in their syntactic development, the pattern of findings was consistent across the entire range of language abilities within the sample. Our previous analysis on the data of this corpus indicated that there was no obvious evidence of any systematic bias as a function of overall vocabulary size or the mean length of

utterances (for more details, see Tse, Chan, Kwang & Li, 2002).

Insert Table 1 and Figure 1 here

The means and standard deviations were calculated for each lexical class and the results are shown in Table 1. The percentages of each lexical class across the three age groups are shown in Figure 1. It can be seen that the majority of words produced by the sample are verbs, auxiliaries, pronouns, nouns and adverbs, which account for 80% of the total number of the words produced by the different age groups. Verbs, not nouns, were the most-frequently-used lexical class, and the proportion of verb tokens was consistent across the three age groups (23.26% for age 3, 22.91% for age 4, and 23.79% for age 5). By contrast, the proportion of noun tokens was consistently lower than that of verbs (13.72% for age 3, 12.11% for age 4, and 12.5% for age 5). A two-way log linear analysis with 3 (age group) × 12 (lexical class) between-subject factors was applied to the data for the different lexical classes. Pearson's chi-square analysis revealed a highly statistically significant main effect for lexical classes [$\chi^2_{(35)} = 59.80, p < 0.001$] across the different ages groups. No statistically significant effects for age and type-by-age interaction were found.

A multivariate analysis of variance (MANOVA) was applied, with age (3) and sex

(2) as the independent variables and the counts of each lexical class as the dependent measures. The results indicated non-significant effects for age, sex and age-by-sex interaction for any of the lexical classes, except for verbs, adverbs, exclamations, numerals and onomatopoeia. Unlike other lexical classes, children's production of verb tokens did differ significantly across age [$F(1,3) = 3.80, p < .05$], and there was a statistically significant age-by-sex interaction [$F(1,6) = 3.18, p < .05$]. Girls outperformed boys in verb tokens in all the age groups, except for that of 48 months. Follow-up post hoc analysis revealed a statistically significant difference between the groups of 48 and 60 months ($p < .05$). Since exclamations, numerals and onomatopoeia accounted for a tiny proportion of the entire vocabulary produced by the sample, and these aspects were not the main concern of the present study, no further analyses were carried out.

To further explore whether there is a statistically significant predominance of noun or verb types in Cantonese-speaking children's vocabulary, we took the effective way developed by Tardif, Gelman and Xu (1999) to calculate the ratio of nouns to verbs such as the following: $\text{Nouns} / (\text{Nouns} + \text{Verbs})$. This ratio of $N/(N + V)$ helps to realize our intention of conducting an explicit comparison of our data with that of other studies. If the ratio is above .50, it means that more nouns were produced, whereas ratio below .50 means more verbs were produced. As shown in Table 2, the ratio of $N/(N + V)$

is around .37 for the 3 age groups, $p_s < .005$, which means that all the Cantonese-speaking preschoolers in the present study produced significantly more verbs than nouns in the regular toys context. Noticeably, this ratio (.37) is very close to those (.38) of Mandarin-speaking children reported by Tardif and her colleagues (Tardif, Shatz & Naigles, 1997; Tardif, Gelman & Xu, 1999). Similarly, these ratios for Cantonese- (.37, $V > N$) and Mandarin-speaking children (.38, $V > N$) are significantly different from those of English-speaking children reported in Tardif, Gelman and Xu (1999) (.50, $N = V$). Unlike their English-speaking counterparts, the Cantonese- and Mandarin-speaking children showed a clear preponderance of verbs over nouns in their spontaneous vocabularies in these toy play contexts.

Insert Table 2 here

A further comparison between noun and verb usage was conducted using the top five nouns and verbs in the frequency lists of vocabulary items used by the subjects. As shown in Table 3, the frequency of the 5 most-used verbs sik6 (eat), waan4 (play), zou6 (do), tai2 (look) and bei2 (give)¹ ranged between 55 and 109, whereas that of the 5 most-frequent nouns jan4 (people), faan6 (rice), seoi2 (water), sin1saang1 (teacher),

¹ The numerals in Romanised forms represent tone (highest = 1, lowest = 6); 1 = high level, 2 = high rising, 3 = mid level, 4 = low falling, 5 = low rising, 6 = low level.

and caa4 (tea) varied from 22 to 59. This difference between the frequency of verb and noun tokens indicated that verbs feature more frequently in the children's utterances than nouns.

Insert Table 3 here

Discussion

As the largest investigation ever made of lexical development in Cantonese-speaking children, the present study was not deliberately designed to just ascertain whether verbs or nouns dominate Hong Kong children's expressive language. However, in light of the findings of Tardif and her colleagues (Tardif, 1996; Tardif, Shatz & Naigles, 1997; Tardif, Gelman & Xu, 1999), the present study found that the Cantonese-speaking children seem to use more verbs in their utterances than do English speakers (see Table 2). Nouns ranked low among the lexical classes in terms of incidence and percentages (see Figure 1). Given the fact that the nouns/verbs ratio is very close to that of Mandarin-speaking children reported by Tardif and colleagues, it can be claimed that verbs rather than nouns dominate the expressive vocabulary of Cantonese-speaking pre-schoolers in toy play contexts.

To account for this finding, one needs to consider various cultural, linguistic and

contextual influences. Structurally, Cantonese, like Mandarin, is also pro-drop language, in which verbs are obligatory and rarely omitted (Tardif, Shatz, & Naigles, 1997). For example, Tse (1993) found that, among the 18 types of syntactic structures in Cantonese, there are 16 in which verbs are necessary. By contrast, English is a non-pro-drop language, in which more noun phrases, and presumably more common nouns, are required for communication. In more grammatically correct English, overt subject and object noun phrases are obligatory, and nouns may occupy a perceptually salient position at the end of a sentence. As a result, a young child's sentence in English like "Me apple, banana no nice" has no verbs but these are implicit and the meaning is quite clear to English speakers. Furthermore, even regular English verbs may change in sound during declension and irregular verbs can be an absolute nightmare for the learning of English by non-native speakers. Hence, the finding that nouns seem to be acquired earlier than verbs in English might be a reflection of the language-specific features.

In contrast, verbs are more constant and often occur alone as a complete sentence in Cantonese (Tse, 1993). For example, the word 'sik6' (eat), the most-frequently used verb in the study's sample, can serve as a sentence by itself. In the 30-minute spontaneous play situation, when a boy presented an apple to his playmate and said "sik6" (eat), the children could understand who should 'eat' and

what should be 'eaten' on the basis of such a verb sentence. In fact, during the data gathering sessions, it was common for the children to use verbs to replace nouns temporarily. They appeared to prefer to use phrases including verbs (such as 'the thing that we [use to] eat') rather than the noun involved ('bowl'). In other words, certain linguistic features of Cantonese might help account for the finding that verbs may be more conspicuous than nouns in young children's utterances.

Furthermore, socio-cultural distinctions between English and Chinese speakers may also have contributed to the apparent differences in linguistic patterns.

English-speaking mothers engage in naming activities with their children more than Korean-speaking mothers do (Choi & Gopnik, 1995), and they display a pattern of greater noun than verb usage. In contrast, Korean and Mandarin-speaking care-givers consistently produced a relatively higher proportion of verb types in their speech vocabularies than did English care-givers (Choi & Gopnik, 1993, 1995; Tardif, 1993, 1996; Tardif, Shatz, & Naigles, 1997; Tardif, Gelman & Xu, 1999). In addition, Tardif (1993) found that Mandarin-speaking mothers placed verbs at the beginning and end of utterances with much higher frequencies than they placed nouns. Whereas the syntactical and semantic form of Mandarin used by mothers of young children tends to emphasise verbs, mothers tend to highlight nouns more in the case of English.

Although there were no productive speech samples of the Cantonese-speaking

mothers in the present study, it is reasonable to assume that the type of language children regularly hear around them will influence their daily speech and eventually have impacts on the predominance of verbs in the early lexicon of Cantonese infants in Hong Kong. There is a complex bi-directional interaction between the child's pre-linguistic cognitive structures and the shape of the specific language the child hears in the early stages of language acquisition. The predominance of verbs in the utterances of adult Cantonese-speakers may have long-term influences on their children's lexical development. This might help explain why children who are encouraged to focus on verbs in the everyday Cantonese they hear around them acquire verbs early. Thus, it would be inappropriate to claim that the Cantonese-speaking children in the present study produced more verbs simply because of the nature of Cantonese grammar.

In line with Tardif et al's findings (1999), the present study also found that contextual features in the toy play context might have had a critical impact on the subjects' utterances. Each pair of participating children (boy/boy, girl/girl, or boy/girl) was placed in a free play corner containing cooking, eating, transport, hospital and general play artefacts. They were encouraged to talk while they were playing and, not surprisingly, the utterances produced were dominated by verbs associated with cooking, eating, transport, hospitals and play. For example, their most frequently used

verbs were sik6 (eat), waan4 (play), zou6 (do), tai2 (look) and bei2 (give) (see Table 3).

As there were just two participants in each session, nouns were usually unnecessary whereas verbs were essential in the simple, often incomplete and context-dependent speech. In effect, the toy play context allowed the children to omit nouns when communicating in exactly the same manner favoured by the adult speakers around them.

Moreover, the paired children tended to use pronouns, not nouns, to refer to concrete items. Because the conversations involved two playmates in a small corner, the children had no difficulty in understanding what the pronouns referred to, for example 'this', 'that', 'it', 'you' and 'he'. Thus pronouns ranked higher than nouns in terms of absolute numbers in the present study. Although the fact that context plays a significant role in the speech of Cantonese-speaking children is consistent with the findings of Mandarin-speaking children reported by Tardif et al. (1999), the children in the present study ranged in age from 36 to 60 months and could be expected to be able to use pronouns freely in daily speech. Tardif et al.'s (1999) 22-month-olds had just started to use pronouns. This age difference between the samples may account for the superior use of pronouns by the children in the present study.

Whilst it would seem reasonable to claim that there is evidence opposing the predominance of nouns in the Cantonese-speaking young children, this does not imply

that steps should be taken by educators to redress an imperfection. Nor can it be assumed that the findings are wholly representative of the overall speech used by the sample in their everyday communications. It would be necessary to gather data from neighbouring age groups, from care-givers and from a wide range of contexts for the resolution and clarification of this issue. What is certain is that cultural, contextual, linguistic influences and the interactions among them are accountable for the differences in children early speech.

The present study has its limitations. First, it was confined to a cross-sectional approach because there were insufficient resources to execute a longitudinal study. Second, the care-givers of the participating children were not involved in the present study. Further research is needed to investigate the impact of parental input on children's lexical development. Nevertheless, the findings may have some contribution to theory building and they may have practical implications for early language education. The finding that Cantonese-speakers display a predominance of verbs in their language in early childhood may indicate a clear direction for language teachers in kindergartens. It is suggested that Hong Kong pre-school teachers replace the extensive copying, dictation and rote memory exercises that predominate the kindergarten classroom (Ho & Bryant, 1997; Opper, 1992) with the active teaching approach, involving more activity, movement, role play and other "doing" related

activities in the teaching of Cantonese.

References

- Au, T. K., Dapretto, M., & Song, Y. K. (1994). Input vs. constraints: Early word acquisition in Korean and English. Journal of Memory and Language, *33*, 567-582.
- Chen, P., & Xu, Z. Y. (1991). Ertong zui chu cihui de houde children's earliest vocabulary words. Unpublished manuscript, Beijing University, Beijing, China.
- Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in Korean: A cross-linguistic study. Journal of Child Language, *22*(3), 497-529.
- Choi, S., & Gopnik, A. (1993, April). Nouns are not always learned before verbs: An early verb explosion in Korean. Paper presented at Stanford Child Language Research Forum, Stanford, CA.
- Erbaugh, M. S. (1992). The acquisition of Mandarin. In D. I. Slobin (Ed.), The cross-linguistic study of language acquisition (Vol. 3. pp. 373-455). Hillsdale, NJ: Erlbaum.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), Language development: Language, thought, and culture (Vol. 2, pp. 301-334). Hillsdale, NJ: Erlbaum.
- Goldfield, B. A. (1993). Noun bias in maternal speech to one-year-olds. Journal of Child Language, *20*, 85-99.
- Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content,

and the vocabulary spurt. Journal of Child Language, 17, 171-183.

Gopnik, A. (1981). Development of non-nominal expressions in 1-to 2-year-olds. In P. Dale & D. Ingram (Eds.), Child language: An international perspective (pp. 238-259). Baltimore: University Park Press.

Halliday, M. A. K. (1975). Learning how to mean: explorations in the development of language. London: Edward Arnold.

Ho, S.H.C., & Bryant, P. (1997). Learning to read Chinese beyond the logographic phase. Reading Research Quarterly, 32(3), 276-290.

Kim, M., McGregor, K.K., & Thompson, C.K. (2000). Early lexical development in English- and Korean-speaking children: Language-general and language-specific patterns. Journal of Child Language, 27(2), 225-254.

Ko, W. L. (1980). The Research on Cantonese. Hong Kong: Commercial Press.

Markman, E. M. (1992). Constraints on word learning: Speculations about their nature, origins, and domain specificity. In M. R. Gunnar & M. P. Maratsos (Eds.), Modularity and constraints in language and cognition: Minnesota symposium on child psychology (Vol. 20, pp. 59-101). Hillsdale, NJ: Erlbaum.

Markman, E. M. (1989). Categorization and naming in children. Cambridge, MA: MIT Press.

McShane, J. (1980). Learning to talk. Cambridge: Cambridge University Press.

Nelson, K. (1973). Structure and strategy in learning to talk. Monographs of the Society for Research in Child Development, 38 (Serial Nos. 1-2), 1-136.

Opper, S. (1992). Hong Kong's Young Children: Their Pre-schools and Families (pp. 21-27). Hong Kong: Hong Kong University Press.

Pinker, S. (1984). Language learnability and language development. Cambridge, MA: Harvard University Press.

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

Tardif, T. (1993). Adult-to-child speech and language acquisition in Mandarin Chinese. Unpublished doctoral dissertation, Yale University, New Haven, CT.

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.

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

Tse, S. K. (1993). The Composing Process of Hong Kong Children in Primary Schools. Unpublished doctoral thesis, University of Nottingham, United Kingdom.

Tse, S. K., Kwong, S.M, Chan, C., & Li, H. (in press). Sex Differences in Syntactic Development: Evidence from Cantonese-Speaking Pre-Schoolers in Hong Kong.

Manuscript submitted for publication.

Wang, N. (1998). A Corpus of Ci. Hong Kong: Peaceful Book Publication.

Wu, T., & Xu, Z. Y. (1979/1980). A preliminary analysis of children during the first three years. Journal of Chinese Sociology and Anthropology, 12(3). (Originally published in Chinese as “Chusheng dao san sui ertong yanyu fazhan jilu de chubu fenxi,” Xinlixue Bao, 1979(2), pp. 153-165)

Yuan, B. (1977). You yige congchuan yanjiu dao kan liang sui youer de yuyan fazhan [A longitudinal study of the language development of a 2-year-old child]. Zhongguo Jiazheng (Chinese Home Economics), 6, 11-16.

Zheng, D.O. (1995). Jinri Yueyu [Cantonese Today]. Guangzhou: Ji'nan University Press.

Appendix 1

Lexical Classes and examples

Classes	Definition	Example
Noun	Semantically denoting object, thing, person, idea, or concept; and syntactically being the subject of a sentence or the object.	<u>lou5</u> <u>si1</u> (teacher), <u>fo2</u> <u>ce1</u> (train),
Verb	Semantically denoting action and activity; and syntactically forming the predicate of a sentence together with other elements.	<u>mei4</u> <u>siu3</u> (smile) , <u>cai3</u> (construct)
Adjective	Semantically being the word that describes qualities; and syntactically being used either to modify nouns, or to be combined with the verb to form the predicate of a sentence.	<u>dung3</u> (cold) , <u>mei5</u> (beautiful)
Adverb	Syntactically being used to modify a verb, adjective or other adverb.	<u>cyun4</u> <u>bou6</u> (all) , <u>ne1</u> <u>paai4</u> (recently)
Pronoun	Syntactically being used instead of a noun or noun phrase.	<u>ngo5</u> (I, me) <u>nei5</u> (you)
Auxiliary	Syntactically assisting a verb to express grammatical function such as tense, aspect, and mood.	<u>gwo3</u> (already), <u>iau6</u> (again)
Numeral	Chinese numbers being used to denote	<u>jat1</u> (one),

	the quantity of objects, things, persons, ideas, or concepts.	<u>ji6</u> (two)
Classifier	Semantically measuring or classifying a noun; and syntactically appearing in between the noun and the numeral or between the noun and the demonstrative pronoun.	<u>faai3</u> (piece), <u>shuang</u> 1 (pair)
Conjunction	An uninflected word used to link together words or sentence parts.	<u>waak6 ze2</u> (or) , <u>tung4</u> (and)
Preposition	Part of speech, usually indeclinable in form, used together with a noun phrase to show the relationship between that phrase and other words in the sentence.	<u>ceoi4 zo2</u> (except) , <u>hoeng3 zyu6</u> (towards)
Exclamation	An utterance conveying intensity of emotion, often by means of particular emphasis, stress...etc.	<u>waa1</u> (gosh), <u>haa1 haa1</u> (haha)
Onomatopoeia	The formation of words imitating natural sounds.	<u>ding1 dong1</u> (ding dong) , <u>paang4</u> (bang)

Table 1

Means and Standard Deviations for the Production of Vocabulary Tokens across the 3Age Groups

	Group of 3-4 Years		Group of 4-5 Years		Group of 5-6 Years	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Verb	40.30	21.79	41.88	20.00	47.17	25.40
Auxiliary	29.54	15.99	30.43	16.26	33.13	18.54
Pronoun	28.70	15.99	28.87	14.94	29.58	18.13
Noun	23.77	15.46	22.14	12.60	24.79	15.44
Adverb	14.46	9.90	17.02	10.04	17.62	11.54
Classifier	11.07	7.42	12.08	7.41	13.07	9.85
Adjective	8.41	6.20	8.14	5.39	9.70	6.86
Exclamation	4.52	5.56	6.33	5.43	7.05	7.34
Preposition	2.27	3.11	2.12	3.13	1.99	2.48
Numeral	1.98	3.02	4.78	10.85	4.66	14.09
English words	.79	1.74	.65	1.42	.65	1.49
Conjunction	.64	1.18	1.25	5.46	.72	1.30
Onomatopoeia	.64	1.69	.38	1.07	.57	1.41

Table 2

Comparisons of the Noun and Verb Types and Ratios with Other Relevant Data

	Nouns	Verbs	Nouns	Summary
	Mean	Mean	Mean	
Cantonese				
36 months ($n=164$)	18.76	31.88	.37*	V > N
48 months ($n=164$)	20.07	33.79	.37*	V > N
60 months ($n=164$)	22.46	37.66	.37*	V > N
Mandarin				
22 months($n=10$)(Tardif, 1996)	11.9	18.7	.38	V > N
20 months ($n=24$) (Tardif et al,1999)#	4.0	6.5	.38	V > N
English				
20 months ($n=24$) (Tardif et al,1999)#	4.9	4.9	.50	N = V

- * Denotes that the Nouns/(Nouns + Verbs) ratio is significantly difference from .50, $p < .005$.
- # Denotes that the data were just driven from the “regular toys “context (Tardif, Gelman & Xu, 1999).

Table 3

The Top Five Nouns and Verbs in the Frequency List

		36 months	48 months	60 months
		(<u>n</u> = 164)	(<u>n</u> = 164)	(<u>n</u> = 164)
Verbs	1. <u>sik</u> 6 (eat)	102	101	97
	2. <u>waan</u> 4 (play)	81	64	74
	3. <u>zou</u> 6 (do)	60	71	66
	4. <u>tai</u> 2 (look)	55	74	64
	5. <u>bei</u> 2 (give)	68	61	61
Nouns	1. <u>jan</u> 4 (people)	59	58	56
	2. <u>faan</u> 6 (rice)	48	55	50
	3. <u>seoi</u> 2 (water)	34	46	49
	4. <u>sin</u> 1 <u>saang</u> 1 (teacher)	35	31	26
	5. <u>caa</u> 4 (tea)	22	41	24

Figure 1: Distribution (Percentage) of Lexical Classes across the Three Age Groups

