

**The Development of Grammatical Competence
in Cantonese-speaking Children**

-Report of a project funded by RGC earmarked grant CUHK 2/91

by

**Thomas Hun-tak Lee
Colleen H Wong
Samuel Leung
Patricia Man
Alice Cheung
Kitty Szeto
Cathy S P Wong**

January 1996

Summary of Findings

We take it as axiomatic that children come to the language learning task with a rich set of linguistic universals. This innate endowment enables children to acquire the language they are exposed to on the basis of very limited data under highly diverse conditions. The past three decades of research have shown the early grammars of two- and three-year-olds to be exceedingly complex. Various syntactic categories (lexical and functional) are evidenced at around two years old or earlier, and by three years old, a wide range of noun phrases and verb phrases, as well as complex clauses, have appeared. The early grammars of children contain abundant evidence of parametric variation, reflecting the characteristics of their target language.

This project examines the grammar of young Cantonese-speaking children, with particular reference to typological characteristics of Cantonese, viz. classifiers and their role in the noun (or determiner) phrase, pro-drop, the interrogative and non-interrogative uses of *wh*-words, functional categories (including modal auxiliaries, aspect markers, and sentence final particles), focus adverbs, as well as the syntagmatic encoding of the definite/ indefinite distinction. The data are based on the longitudinal corpus of 8 children aged between one and a half and three year olds, as well as experimental studies carried out with children up to five and a half years old.

The project has created a tagged computer corpus of the language of 8 children coded in an internationally accepted format, 14 megabytes in size, the largest child language corpus to date in Chinese-speaking communities. This corpus, which we call the Hong Kong Cantonese Child Language Corpus, will be deposited both at the child language archive of Carnegie Mellon University and at the Arts Faculty server of the Chinese University. It will serve as a useful research tool for anyone interested to investigate early language, be they linguists, psychologists, philosophers, sociologists, or educationists.

We have gained a precise understanding of certain aspects of early grammar in Cantonese. The categories of classifiers, aspect markers, modal auxiliaries and sentence final particles are evidenced between 1 year 9 months and 1 year 11 months old, confirming the early onset of functional categories. Classifiers are used productively before age two; a core set of 11 nominal classifiers *go3*, *zek3*, *zi1*, *tiu4*, *zoeng1*, *lap1*, *gaan1*, *bun2*, *dou6*, *di1*, *deoi3* (個, 隻, 枝, 條, 張, 粒, 間, 本, 度, D, 對), including both sortal and mensural classifiers, are used by all children. Verbal classifiers are rare at this stage. The earliest aspect markers are the perfective marker *zo2* 咗, followed by the durative marker *zyu6* 住, and the progressive marker *gan2* 緊. Both the postverbal deontic modal auxiliary *dak1* 得 and the preverbal modal auxiliaries *wui3* 會 and *ho2ji3* 可以 are established at around two years old. Sentence final particles surface early as a prominent clausal constituent before two years old; at this stage they are found to be primarily non-interrogative in nature, and combinations of final particles are rare.

The articulation of the internal structure of the noun (or determiner) phrase follows a regular pattern. First bare nouns appear, then determiner-noun combinations, followed by numeral-classifier combinations. Classifier-noun combinations have a marked status. With respect to their external distribution, noun phrases of the afore-mentioned types invariably appear first in object position before they occupy subject positions. Pronouns are the exception, preferring subject to object positions.

Our study of pro-drop demonstrates convincingly that as in other languages, the rate of subject drop far exceeds that of object drop in child Cantonese. On the other hand, Cantonese-speaking children drop subjects at a much higher rate than their American counterparts (approx. 80% vs 30% at age two). Similarly, they drop objects much more than American children (approx. 47% vs 7% at age two). Thus, the early grammars of young children already reflect the typological features of their target language. The dropping of subjects and objects cannot be explained by processing limitations, since subjects of complex sentences are not necessarily dropped to a greater extent than those of simple sentences.

In a language in which wh-movement does not occur overtly, we have found asymmetries between subject and object, and between argument and adjunct in the acquisition of wh-words, echoing findings from other languages. Both the naturalistic and experimental data show that what-object (*matje* 乜嘢) questions are mastered before what-subject questions, and who-subject (*bingo* 邊個) questions before who-object questions. Where-argument (*bindou* 邊度) questions are acquired before where-adjunct-questions. Verb type is found to be an important factor in interrogative acquisition. With respect to non-interrogative uses of wh-words, a phenomenon present to a much greater extent in Chinese than in languages such as English, children show such uses by three years of age. The earliest non-interrogative uses of wh-words (e.g. *gei2* 幾 'how-(many/much)'; *dim* 點 'how'; *bin* 邊 'where'; *me1* 咩 'what') are restricted to intensification, rhetorical usage, and labeling of objects without names. The uses of wh-words in polarity contexts are not attested.

A salient feature of Cantonese is its ability to signal the definite/ indefinite contrast exclusively in terms of preverbal vs postverbal word order. Our experimental data suggest that children do not develop a reliable grasp of this syntagmatic distinction until very late, around five and a half years old. The findings on existential and ergative verbs show unequivocally that the cue to the definiteness distinction lies in word order and not in the presence of the existential verb *jau5* 有 'be/have'.

The syntactic distinction between prepositions and verbs in Cantonese is not as clearcut as in languages such as English. Our study of words which can function both as verbs and prepositions in the language, such as *bei2* 俾, indicates that the verb use appears earlier than the prepositional use.

In expressing quantificational meanings, Cantonese relies heavily on nonmovable adverbs located between the subject and the predicate. A prominent feature of child Cantonese as opposed to child English is that these quantifying adverbs occur early. We explored how young Cantonese-speaking children understand adverbs that signal the meanings of 'and', 'even', and reiteration, viz the adverbs *zoi* 再, *jau* 又, *zung* 仲, *dou* 都. Focus structures are not found before two years old, but emerge between two and three years of age. Children early on are sensitive to quantificational properties of particular adverbs such as the direction of quantification. Pragmatic scales are not yet formed.

Introduction

Thomas Hun-tak Lee, Colleen H Wong, and Samuel Leung

1. The report

This report contains the results of several years of research on the early grammar of Cantonese-speaking children. We consider a major outcome of the project the establishment of a longitudinal tagged corpus of 8 children covering the age range of approximately one and a half years to three years. This corpus, which we call the Hong Kong Cantonese Child Language Corpus has allowed us to gain some precise understanding of the early language of Cantonese-speaking children in a number of areas reported below. In addition, we have done cross-sectional experiments with older children on issues that are of interest from a crosslinguistic perspective: the syntagmatic signalling of the definite/indefinite distinction, and the implications of lack of overt movement of wh-phrases.

This report is divided into four parts. Part A consists of papers on selected areas of child Cantonese by various members of the team. These papers, based on the corpus we have built, are concerned with the development of classifiers, the internal structure of noun phrases and its relation to their external distribution, prepositions, focus adverbs, null arguments, the uses of wh-words, and the general issue of availability of functional categories before the age of two. Part B consists of two cross-sectional studies carried out on three and a half to five year old children. Part C contains commentaries on the corpus intended for a general academic readership. In Part D are found the diskettes of the corpus, a database encoded in the internationally accepted Chiles format of a size of around 14 megabytes. Three other diskettes contain three programs we have developed: a Chinese keyword in context search software (CHKWIC) specifically for the use of our corpus; a romanization program (ROMANIZE) which will automatically convert Chinese text files of the corpus to corresponding files with the Chinese characters changed to their romanized forms; and a lexicon program (DICTION) that will generate a listing of types and tokens of items of different word classes based on a set of tagged files. In addition there is a program (MLUMLT) for computing the mean length of utterance based on any section of an input file (defined by beginning and ending line numbers) of the corpus. The Chinese and romanized versions of the tagged corpus will be deposited in `humanum.arts.cuhk.hk`, while the romanized version will be deposited at the Chiles archive at the Carnegie Mellon University.

The papers in this report have been presented at seminars, workshops and conferences or published:

An earlier version of Ch. 2 was presented at the 1992 Linguistic Society of Hong Kong Annual Research Forum, and at a seminar at the Department of Speech and Hearing Sciences, University of Hong Kong.

An earlier version of Ch. 3 was read at a seminar at the Department of Linguistics, University of Hawaii.

Chs. 4 and 5 were presented at the Sixth North American Chinese Linguistics Conference held at the University of Southern California in 1994, and were subsequently published in the proceedings of the conference (1995) edited by J. Camacho and L. Choueiri.

An earlier version of Ch. 6 was given at the 1991 Linguistic Society of Hong Kong Annual Research Forum, and at a workshop of the *Linguistics Abroad* journal held at Beijing University in 1992.

An earlier version of Ch. 7 was presented at the 1992 Linguistic Society of Hong Kong Annual Research Forum, and also at the Chinese Linguistics Workshop of the 1993 Linguistics Society of America, held at Ohio State University.

A preliminary version of Ch. 8 was read at the Hong Kong Educational Research Association Annual Meeting in 1992, and also at the LSHK Annual Research Forum of the same year. The current version was presented at the Conference on the Processing of Chinese and other Asian Languages held at the Chinese University of Hong Kong in December of 1995.

Ch. 9 was presented at the Chinese Linguistics Workshop at the 1993 LSA Summer Institute held at the Ohio State University, and at the Fourth International Conference on Cantonese and other Yue Dialects held in December of 1994 at the City University of Hong Kong. It will appear in a volume on Cantonese Linguistics edited by Stephen Matthews.

Ch. 10 was presented at the 25th Annual Child Language Research Forum at Stanford University, in 1993, and appeared in the proceedings of the conference edited by Eve Clark.

Ch. 11 appeared in the *The Humanities Bulletin*, Faculty of Arts, The Chinese University of Hong Kong (1994), vol. 3.

Ch. 12 was published in the *Working Papers of the Department of English*, Hong Kong Polytechnic University, vol. 1 (1994).

2. The Hong Kong Cantonese Child Language Corpus

2.1 The datafiles

This database contains longitudinal data on the language of 8 Cantonese-speaking children, each recorded for approximately one year. The corpus contains 171 files coded in Childes format and tagged with a set of 33 word class labels.

These children were observed in their interactions with the caretakers, the investigator, and occasionally other adults who chatted with the children during the visits. Three research students carried out the observations and the recording. Patricia Man recorded Bohuen and Gakie; Alice Cheung recorded Bernard, Tsuntsun and Tinfan; and Kitty Szeto recorded Johnny, Jenny and Chunyat. The names of the children and the ages during which they were recorded are as follows:

NAME	Sex	Age at which recording began and ended	No. of files
Bohuen (wbh)	F	2;03;23 - 3;04;08	27
Gakei (cgk)	F	1;11;01 - 2;09;09	19
Bernard (mhz)	M	1;07;00 - 2;08;06	26
Tsuntsun (ckt)	M	1;05;22 - 2;07;22	25
Tinfan (ltf)	F	2;02;10 - 3;02;18	16
Johnny (hhc)	M	2;04;08 - 3;04;14	16
Jenny (LLy)	F	2;08;10 - 3;08;09	20
Chunyat (ccc)	M	1;10;08 - 2;10;27	22
			total 171

The file name indicates the name of the child and his/her age at the time of the recording. Each filename is made up of the initials of the child (the first three characters) and his/her age at the time of recording, in terms of Year (1 character), Month (2 characters), and Day (2 characters). All files have the suffix 'tag'. For instance, the file WBH20322.tag contains tagged utterances of Bohuen (wbh) when she was 2 years 3 months and 22 days old.

The files of the 8 children and their respective sizes are listed below:

Bohuen (27 files)

WBH20323 TAG	48,310	WBH20926 TAG	58,567
WBH20330 TAG	36,851	WBH21002 TAG	24,547
WBH20402 TAG	20,993	WBH21016 TAG	30,688
WBH20405 TAG	23,270	WBH21023 TAG	32,193
WBH20406 TAG	33,430	WBH21106 TAG	18,256
WBH20413 TAG	22,811	WBH21114 TAG	29,467
WBH20414 TAG	11,570	WBH21128 TAG	33,004
WBH20415 TAG	9,842	WBH30010 TAG	36,238
WBH20506 TAG	19,630	WBH30101 TAG	39,535
WBH20609 TAG	20,679	WBH30220 TAG	46,243
WBH20619 TAG	11,922	WBH30303 TAG	31,646

WBH20703 TAG	21,648	WBH30312 TAG	36,715
WBH20714 TAG	52,696	WBH30408 TAG	60,664
WBH20919 TAG	49,106		

Gakei (19 files)

CGK11101 TAG	88,455	CGK20318 TAG	32,461
CGK11108 TAG	38,193	CGK20325 TAG	31,182
CGK11122 TAG	21,503	CGK20408 TAG	45,931
CGK11129 TAG	55,229	CGK20430 TAG	68,227
CGK20008 TAG	35,575	CGK20503 TAG	36,682
CGK20207 TAG	42,112	CGK20711 TAG	59,831
CGK20221 TAG	68,347	CGK20808 TAG	65,528
CGK20228 TAG	31,564	CGK20818 TAG	79,915
CGK20304 TAG	56,490	CGK20909 TAG	95,250
CGK20311 TAG	37,122		

Bernard (26 files)

MHZ10700 TAG	43,806	MHZ20115 TAG	108,119
MHZ10800 TAG	82,730	MHZ20129 TAG	83,027
MHZ10814 TAG	30,205	MHZ20212 TAG	80,311
MHZ10828 TAG	29,063	MHZ20226 TAG	104,314
MHZ10904 TAG	109,848	MHZ20309 TAG	92,704
MHZ10918 TAG	77,601	MHZ20328 TAG	115,580
MHZ10925 TAG	110,265	MHZ20407 TAG	94,178
MHZ11010 TAG	99,495	MHZ20421 TAG	116,225
MHZ11023 TAG	109,863	MHZ20504 TAG	91,050
MHZ11106 TAG	95,845	MHZ20519 TAG	91,565
MHZ20003 TAG	99,218	MHZ20604 TAG	88,006
MHZ20016 TAG	105,301	MHZ20618 TAG	135,700
MHZ20101 TAG	127,255	MHZ20806 TAG	114,159

Tsuntsun (25 files)

CKT10522 TAG	18,620	CKT20016 TAG	100,568
CKT10703 TAG	10,142	CKT20108 TAG	127,686
CKT10710 TAG	12,637	CKT20205 TAG	146,398
CKT10800 TAG	104,711	CKT20215 TAG	159,272
CKT10807 TAG	97,010	CKT20303 TAG	132,050
CKT10821 TAG	94,224	CKT20317 TAG	109,179
CKT10907 TAG	110,450	CKT20400 TAG	123,376
CKT10914 TAG	51,177	CKT20414 TAG	111,579
CKT10929 TAG	126,734	CKT20500 TAG	128,133
CKT11030 TAG	117,009	CKT20514 TAG	107,821
CKT11113 TAG	112,869	CKT20618 TAG	125,791
CKT11127 TAG	102,393	CKT20702 TAG	130,499
CKT20009 TAG	142,026		

Tinfaan (16 files)

LTF20210 TAG	81,557	LTF20802 TAG	120,856
LTF20302 TAG	79,287	LTF20824 TAG	112,592
LTF20330 TAG	83,173	LTF20907 TAG	99,085
LTF20427 TAG	93,000	LTF21018 TAG	103,031
LTF20518 TAG	67,563	LTF21116 TAG	124,369
LTF20601 TAG	91,859	LTF30020 TAG	119,667
LTF20705 TAG	115,502	LTF30121 TAG	122,662
LTF20720 TAG	108,327	LTF30218 TAG	112,754

Johnny (16 files)

HHC20408 TAG	27,537	HHC20930 TAG	91,674
HHC20503 TAG	108,963	HHC21013 TAG	95,724
HHC20513 TAG	145,320	HHC21108 TAG	123,220
HHC20519 TAG	142,874	HHC30008 TAG	90,986
HHC20610 TAG	109,595	HHC30116 TAG	103,856
HHC20624 TAG	109,195	HHC30216 TAG	94,360
HHC20721 TAG	113,419	HHC30311 TAG	117,899
HHC20808 TAG	89,101	HHC30414 TAG	105,328

Jenny (20 files)

LLY20810 TAG	49,414	LLY30113 TAG	108,249
LLY20822 TAG	62,064	LLY30130 TAG	111,824
LLY20909 TAG	20,464	LLY30213 TAG	109,727
LLY20914 TAG	91,461	LLY30315 TAG	94,778
LLY20928 TAG	108,323	LLY30326 TAG	91,557
LLY21101 TAG	79,233	LLY30422 TAG	103,201
LLY21108 TAG	102,046	LLY30520 TAG	86,685
LLY21129 TAG	116,200	LLY30616 TAG	103,543
LLY30011 TAG	92,309	LLY30725 TAG	104,530
LLY30022 TAG	95,716	LLY30809 TAG	97,213

Chunyat (22 files)

CCC11008 TAG	44,585	CCC20523 TAG	120,525
CCC11100 TAG	19,170	CCC20608 TAG	111,275
CCC11121 TAG	28,616	CCC20624 TAG	98,160
CCC20110 TAG	98,190	CCC20706 TAG	117,815
CCC20117 TAG	57,474	CCC20713 TAG	111,903
CCC20206 TAG	83,178	CCC20800 TAG	114,384
CCC20213 TAG	109,816	CCC20817 TAG	87,806
CCC20307 TAG	105,992	CCC20907 TAG	115,663
CCC20323 TAG	97,494	CCC20923 TAG	102,006
CCC20410 TAG	100,141	CCC21013 TAG	111,722
CCC20507 TAG	83,663	CCC21027 TAG	123,784

2.2. The background of the 8 Cantonese-speaking children

a) Bohuen and Gakei

Both children are brought up in monolingual Cantonese-speaking working class families. Bohuen's father works in the warehouse of a mass transport company and her mother is a part-time piano teacher. The child has a younger brother who is about two years younger than her. They live with the child's grandmother and uncle. The child has already started attending a nursery school when data collection started. After school, she is taken care of by her parents and grandmother.

Gakei's father is a technician in a electronic company and her mother is a housewife. They live with the child's grandmother. Gakei's parents were both born in Hong Kong. The child is not yet in a nursery during the whole period of data collection. She is entirely taken care of by her mother.

b) Bernard, Tsuntsun and Tinfaan

All three are Cantonese-speaking children living in Hong Kong. Tsuntsun and Tinfaan were born in Hong Kong, while Bernard was born in Kent, United Kingdom and was brought back to Hong Kong at the age of 8 1/2 months old.

Tsuntsun is the only son of the family. His father is a Census & Survey Officer working in the government and his mother a secondary school teacher teaching Chinese and Religious Studies. Since his birth, he have been living in his maternal grandparents' house during weekdays and is taken care of by his grandmother. His parents visit him occasionally during the weekday evenings and take him back home on Friday nights to stay over the weekend. They communicate in Cantonese. When Tsuntsun was 1 year 10 months old, his mother went to study for a year in the United Kingdom. He started to attend a nursery at the age of 2 years 1 month.

Bernard is the only son of the family. His father is a lecturer in the Division of Construction and Land Use of the Hong Kong Polytechnic. His mother is a lecturer of the English Language Teaching Unit of the Chinese University of Hong Kong. Bernard's mother brought him back from the United Kingdom at the age of 8 1/2 months. He was then taken care of by his maternal grandmother at her house until the age of about 1 year 1 month. From that time to the age of 2 years 6 months, he was taken care of by a caretaker during the weekdays. He communicates in Cantonese, though his parents occasionally introduce to him some English terms. He started to attend the nursery play-groups at the age of 2 years 6 months.

Tinfaan is the youngest child in the family. She has a sister who is four years older. Her father is an engineer working in the government and her mother is a piano teacher teaching at home. During the first one-and-a-half years from her birth, she was taken care of mostly by a Filipino helper while her mother worked as a school music teacher. After her mother stopped working in school, Tinfaan is mostly taken care of by her mother, except at times when her mother has to give piano lessons or has to go out, when LTF is looked after by her Filipino helper. She communicates in Cantonese except when speaking to her Filipino helper, for which she uses 'something English-like' (as described by her mother). She started to attend kindergarten at the age of 2 years 9 months.

c) Johnny, Jenny and Chunyat

All three children were born in Hong Kong and were brought up in monolingual Cantonese-speaking families. They had not started going to a nursery during the period of data collection. Jenny is the youngest child in the family. She has an elder brother who is ten years older and an elder sister who is four years older. Jenny's

father is a businessman and her mother is a housewife. The family employs a Filipino helper, who speaks some Cantonese and English to the children.

Johnny is the youngest child in the family. He has an elder sister who is seven years older. His father is an engineer and her mother is a typist. The family employs a Thai helper and she speaks Cantonese to the children.

Chunyat is the only son in the family. His father is a merchant and his mother teaches English in a secondary school. They live with the child's maternal grandparents.

2.3. Tags

Below is a summary list of the syntactic categories used in coding the corpus. The romanizations are based on the Cantonese romanization scheme of the Linguistic Society of Hong Kong (LSHK) (see Matthews and Yip 1994: 400-401).

Category	Example
1. adj = adjective	hung4
2. advf = focus adverb	zung6, dou1, jau6, zoi3
3. advi = adverb of intensity	hou3, gei2, gam13, zan1
4. advm = adverb of manner	maan6maan6dei2, ma4ma4dei2
5. advs = sentential adverb	bat1jyu4, gam2(joeng2), jat1cai4
6. asp = aspectual marker	zo2, zyu6, gan2, gwo3, hoi1
7. aux = auxiliary / modal verb	jing1goi1, hang2, ho2ji5, wui, sai2
8. cl = classifier	go3, zek3, bun2, bui1, di1
9. com = comparative morpheme	gwo3 (as in <i>dai6 gwo3</i>), di1 (as in <i>hung4 di1</i>)
10. conj = connective	dan6hai6, tung4maai4, waak6ze2
11. corr = correlative	jut6...jut6, jau6...jau6, gam2...gam3, jat1...jat1
12. ctc = clitic	dak1, dou3
13. det = determiner	nei1, go2, dai6
14. dir = directional verb	lok6, soeng5, ceot1, jap6, lai4
15. ex = expressive utterance	baai1baai3, zou2san4
16. gen = genitive marker	ge3
17. ins=emphatic inserted marker	gwai2 (as in <i>hou3 gwai2 leng3</i>)
18. nn = noun	ping4gwo2, ba4ba1
19. nnloc = locative noun phrase	soeng6mien6, leoi3mien6
20. nnpr = pronoun	ngo5, nei3, keoi3
21. nnpp = proper name	tin1faan4, Ga1kei4
22. neg = negative morpheme	m4, mai6, mou5
23. prt = post-verbal particle	faan1, sai3, can1, maai4, gwo3, ha2
24. prep = preposition	tung4maai2, hai2, bei2
25. q = quantifier	jat1, saam1, sap6, gei2, mui5
26. rfl = reflexive pronoun	zi6gei2

27. sfp = sentence final particle	&la3, &ga1 &ma3, &ge3 &le1
28. vd = ditransitive verb	bai2, bei2
29. verg = ergative verb	dit3
30. vf = function verb	hai6, jau5, hai2
31. vi = intransitive verb	siu3
32. vt = transitive verb	teo1
33. wh = wh words	mat1, mat1je5, dim2, dim2gaai2, dim2jeong2

2. 4. Chinese characters

a) The corpus has two versions, both for use in the DOS environment. The Chinese version requires the use of Eten 3.5 or later versions. As the data contain Cantonese characters which are not found in the standard GB or Big-5 character set, we have created userfonts to represent these Cantonese characters which are in common use in Hong Kong, but not in China or Taiwan. Anyone using the Chinese version of the corpus will need to copy the following files (which come with the corpus) to their Eten subdirectory:

usrfont.15m
usrfont.24m

b) The romanized version is derived from the Chinese tagged corpus by means of a conversion program based on a dictionary. Since a character may have different pronunciations (due to language variation or context), the romanized data files sometimes give more than one romanized form for a single character, separated by '^', a convention suggested by Brian MacWhinney. Thus, for example, the Cantonese morpheme for 'you' can have an alveolar lateral initial or an alveolar nasal initial. The morpheme will be rendered as 'lei^nei' in the romanized data. The romanized corpus contains the categorial tags below each romanized utterance, but it does not contain English glosses. In time, we hope to seek resources to enable us to disambiguate the romanized forms, and provide English glosses.

Please cite the following if data from this corpus are used:

Lee, Thomas H.T., Colleen H Wong, Samuel Leung, Patricia Man, Alice Cheung, Kitty Szeto, and Cathy S P Wong, "The Development of Grammatical Competence in Cantonese-speaking Children", Report of RGC earmarked grant 1991-94.

Table of Contents

Acknowledgments	i
Members of the project	ii
1. Introduction Thomas Hun-tak Lee, Colleen H Wong and Samuel Leung	iii
Part A: Studies based on the Longitudinal Corpus	
2. Classifiers in Cantonese-speaking Children from 1 year 5 months to 3 years 8 months Kitty Szeto	1
3. The Development of Noun Phrase Structure in Cantonese-speaking Children around Two Cathy S P Wong	26
4. The Acquisition of <i>Bei2</i> as Verb and Preposition in a Cantonese-speaking Child Colleen H Wong	65
5. Focus Adverbs in Child Cantonese Thomas Hun-tak Lee	84
6. The Realization and Omission of Subject and Object in Early Cantonese Patricia Man	107
7. Cantonese-speaking Children's Acquisition of Wh-words Alice Shuk-yee Cheung	125
8. Functional Categories in Child Cantonese Thomas Hun-tak Lee, Colleen H Wong and Cathy S P Wong	155

Part B: Cross-sectional Experimental Studies

- | | | |
|-----|---|-----|
| 9. | The Acquisition of (In)definiteness by Cantonese-speaking Children
Thomas Hun-tak Lee and Kitty Szeto | 175 |
| 10. | Cantonese-speaking Children's Comprehension of Wh-questions
Alice Shuk-yee Cheung and Thomas Hun-tak Lee | 203 |

Part C: Commentaries on the Hong Kong Cantonese Child Language Corpus

- | | | |
|-----|---|--|
| 11. | A Cantonese Child Language Corpus
Thomas Hun-tak Lee | |
| 12. | The Challenge in Creating a Hong Kong Cantonese
Child Language Corpus
Colleen Wong and Patricia Man | |

Part D: The corpus on diskettes

Acknowledgment

We wish to acknowledge the support of the following funding agencies for this project: The Research Grants Committee for a three-year earmarked grant (CUHK 2/91); the Hong Kong Polytechnic University for two studentships awarded to Patricia Man and Alice Cheung respectively, and for a research assistantship; the University of Hong Kong for a studentship awarded to Kitty Szeto; the Research Committee of the Chinese University of Hong Kong for a direct grant; the Freemason's Fund for East Asian Studies for a grant in support of computer applications related to the development of the corpus, and the Lee Hysan Foundation for a staff development grant.

The project would not have gotten off the ground without the support of the parents of the children and educationalists who generously allowed us access to their schools. We wish to thank in particular Ms. Leung Lai Ying of Lap Yan Kindergarten, and the Construction Association Kindergarten, Ms Bessie S M Pang of Suen Mei Kindergarten, and Ms. Linda Mak for their warm support for the project. The assistance of a number of students of CUHK- Juliet Du, Christine Lau, Catherine Tang, Gina Tse, Elaine Wong, Eva Yuen, and Fanny Yuen is duly acknowledged.

Our project has benefited immensely from the early works on longitudinal observation of Chinese children. In particular we wish to express gratitude to Tse Kwok-ping, who generously made available to us the Taiwan corpus his research team created; to Tseng Chiuyu and Tang Szewing for making available to us various bibliographical items of language acquisition research in Taiwan.

The latter phase of the research focussed on the computerization of the corpus and the development of software. We wish to record our thanks to Lawrence Cheung Yum-leung and Joseph Kwok, the programmers of the text search, romanization and lexicon generation programs we use for our analysis; to Wing-shing Wong and Liu Chi Leung of the Information Engineering department of the Chinese University for their advice and help in the development of the tagging software; to Brian MacWhinney and Li Ping for advice related to the coding and romanization of the Chinese corpus; to Tom Lai, Caesar Lun and Peter Tung for work related to the Cantonese phrase box based on the Linguistic Society of Hong Kong Cantonese Romanization scheme; to Kwan Tse-wan of the Humanities Computing Program for recommending that the corpus be placed on the server of the Arts Faculty of CUHK; to Man Tse Yuen, Chan Kwo, and Kitty Szeto for essential technical support.

Many colleagues have supported our research on Cantonese acquisition in various ways, by commenting on our papers, by sharing with us their research experiences, or by expressing interest in our work. We here express our deepest appreciation of the feedback and encouragement they have given us: Marjorie Chan, Cheung Kwan-hin, Cheung Yat-shing, Peter Cole, Jill de Villiers, Mary Erbaugh, Gu Yang, Gabby Hermon, Hsieh Hsin-I, James Huang, Audrey Li, Luke Kang-kwong, Steve Matthews, Mineyaru Nakayama, Neil Smith, James Tai, Gladys Tang, Benjamin T'sou, William Wang, Ken Wexler, Xu Liejiong, Virginia Yip, and Eric Zee.

Members of the Project

Alice Shuk-yee Cheung has a B.A. in English from the Chinese University of Hong Kong. She completed her M.Phil. thesis entitled "Acquisition of Wh-words by Cantonese-speaking Children" in 1995 at the Hong Kong Polytechnic University. She is now teaching at the Diocesan Girls' School.

Thomas Hun-tak Lee, principal investigator of the project, teaches Linguistics at the Department of English, Chinese University of Hong Kong.

Samuel Leung, co-investigator of the project, is lecturer in Linguistics at the Department of Speech and Hearing Sciences at the University of Hong Kong.

Patricia Man obtained her B.A. in Communication from the Hong Kong Polytechnic University, and completed her M.Phil. thesis from the HKPU entitled "Subject-object related distinctions and empty categories in child Cantonese" in 1993. She is now a graduate student at the Department of Chinese, Translation and Linguistics.

Kitty Szeto has a B.A. in English from the Chinese University of Hong Kong. She is currently an M.Phil. student at the Department of Speech and Hearing Sciences, University of Hong Kong working on a thesis entitled "The Acquisition of Cantonese Classifiers".

Cathy S P Wong has a B.A. in English and an M.Phil. in Applied Linguistics from the Chinese University of Hong Kong. She is currently a doctoral student at the Department of Linguistics, University of Hawaii, doing a thesis entitled "The development of noun phrase structure in Cantonese-speaking children".

Colleen H Wong, co-investigator of the project, is an associate professor at the Department of English, Hong Kong Polytechnic University.

Part A: Studies based on the Longitudinal Corpus

Classifiers in Cantonese-speaking Children from 1;05 to 3;08¹

Kitty SZETO

0. Introduction

Cantonese is a numeral classifier language, i.e. classifiers are used in construction with numerals. When counting in Cantonese, cardinal numbers are generally not used alone as in non-classifier languages like English. Enumeration or quantification usually requires the use of a classifier which is bound to a numeral form and which designates certain semantic features of whatever is being enumerated. Classifiers are important elements in Cantonese noun phrases and they are grammatically obligatory in a wide range of NP types.

Acquiring classifiers is not a simple task. Previous studies showed that children before two years old are found to have been using classifiers in spontaneous speech (Li 1968; Erbaugh 1986). On the other hand, Mandarin-speaking and Cantonese-speaking children likewise show immature use of classifiers up to age eight (Poon 1980; Fang 1985; Loke & Harrison 1986; Ying et al. 1987; Loke 1991; Mak 1991). Cross-linguistically, children acquiring Thai, which is also a numeral classifier language, are also shown not to have grasped the entire classifier system as old as ten years old (Gandour et al. 1984).

This paper reports on the classifier development of eight Cantonese-speaking children between the age 1;05;22 to 3;08;09. Longitudinal data collected shows that by two years old, Cantonese-speaking children are able to use classifiers productively and the classifier system expands in considerable rate and the children's classifier use generally conform to the adult pattern.

1. Classifiers and Classifier Languages

1.1 Definitions of Classifiers

Classifiers can be characterized as follows (Allan 1977: 285):

Classifiers are defined on two criteria:

(a) they occur as morphemes in surface structures under specifiable conditions; and

(b) they have meaning, in the sense that a classifier denotes some salient perceived or imputed characteristics of the entity to which an associated noun refers (or may refer).

Morphologically, classifiers occur as words or morphemes. Their syntactic distributions are regulated by "specifiable conditions". In numeral classifier languages such as Chinese, classifiers are obligatory in numerical expressions. The semantic function of classifiers is usually to indicate some intrinsic characteristics of the referents.

1.2 Types of Classifiers and Classifier Languages

Allan (1977: 286-287) identifies four kinds of classifier languages. They are,

- (i) numeral classifier languages, e.g. Chinese, Japanese, Thai.
- (ii) concordial classifier languages, e.g. Bantu.
- (iii) predicate classifier languages, e.g. Navajo.
- (iv) intra-locative classifier languages, e.g. Toba.

The types of classifiers correspond to the types of classifier languages. Numeral classifier languages are the paradigm type and they outnumber the other three types. Studies of classifiers usually focused on numeral classifiers and the term "classifier" is generally used to mean numeral classifiers. Typologically speaking, oriental languages typically have many more classifiers than other classifier languages (Allan 1977: 307).

Further subdivision is possible within different types of classifiers. Sortal and mensural classifiers are two major subtypes under numeral classifiers. A sortal classifier "individuates whatever it refers to in terms of the kind of entity it is" while a mensural classifier "individuates in terms of quantity" (Lyons 1977: 463). This distinction is found in Cantonese and a more detailed description of Cantonese classifiers will be given in the following session.

2. Cantonese Classifiers

Cantonese is a numeral classifier language i.e. classifiers are used in construction with numerals. They are bound morphemes and they occur in the following types of NPs:

- | | | |
|--|---|--------------|
| a. [Det - Num - CL - (N)] ² | e.g. <i>ni1 jat1 zek3 gau2</i> ³ | |
| | this one CL dog | 'this dog' |
| b. [Det - CL - (N)] | e.g. <i>ni1 zek3 gau2</i> | |
| | this CL dog | 'this dog' |
| c. [Num - CL - (N)] | e.g. <i>jat1 zek3 gau2</i> | |
| | one CL dog | 'one dog' |
| d. [Q - CL - (N)] | e.g. <i>mui5 zek3 gau2</i> | |
| | each CL dog | 'each dog' |
| e. [Wh - CL - (N)] | e.g. <i>bin1 zek3 gau2</i> | |
| | which CL dog | 'which dog' |
| f. [CL - N] | e.g. <i>zek3 gau2</i> | |
| | CL dog | 'the dog' |
| g. [CL _i - CL _i - (N)] | e.g. <i>zek3 zek3 gau2</i> | |
| | CL CL dog | 'every dog' |
| h. [N _x - CL - N _y] | e.g. <i>maa5lei6 zek3 gau2</i> | |
| | Mary CL dog | 'Mary's dog' |

NPs with classifiers are always referential in Cantonese and the head nouns can be omitted as long as they are recoverable from the context.

Cantonese classifiers cannot occur alone in NPs, i.e. a classifier cannot be the sole element in the entire noun phrase. The only exception to this is when the preceding discourse is explicit enough to interpret a single post-verbal classifier:

- (1) *nei5 zek3 biu1 hou2 leng3, ngo5 jau6 soeng2 maai5 (jat1) zek3 (biu1)*
 you CL watch very beautiful I also want buy (one) CL (watch)
 'Your watch is beautiful, I also want to buy one too'.
- (2) *ni1 di1 beng2 hou2 hou2sik6, nei5 sik6 (jat1) daam6 (beng2) si3 haa2*
 this CL biscuit very tasty you eat (one) CL (biscuit) try PRT
 'These biscuits are tasty, try a mouthful'.

In the above examples, the implicit referents should be *biu1* 'watch' in (1) and *beng2* 'biscuit' in (2). Moreover, the numeral *jat1* 'one' is understood and seems to have been omitted, leaving the classifier to stand alone in the NP. In fact, the unstressed *jat1* 'one' in Cantonese can always be omitted even in other types of noun phrases.

In Cantonese, NPs in the form [CL - N]_{NP} can be used pre-verbally to denote definite reference as in (3):

- (3) *bun2 syu1 hou2 hou2tai2.*
 CL book very interesting
 'The book is very interesting'.

This kind of [CL - N]_{NP} NPs are also used post-verbally to denote either definite or indefinite reference:

- (4) *ngo5 tai2 zo2 bun2 syu1*
 I read ASP CL book
 'I read *a* book' or 'I read *the* book'.

Cantonese classifiers are used after the Wh-words. However, not all classifiers can be used after *bin1*. Classifiers for standard units of measurement like *bong6* 'pound' or *maa5* 'yard' are ruled out:

- (5) *?/*bin1 bong6 ping4gwo2*
 which CL-pound apple
 'Which pound of apple?'
- (6) *?/*bin1 maa5 bou3*
 which CL-yard cloth
 'Which yard of cloth?'

Standard measurements do not cooccur with *bin1* 'which'. Instead, they are usually questioned by *gei2(do1)* 'how many/much':

- (7) *gei2(do1) bong6 ping4gwo2*
 how many CL-pound apple
 'How many pounds of apples?'
 (8) *gei2(do1) maa5 bou3*
 how many CL-yard cloth
 'How many yards of cloth?'

Gei2(do1) 'how many/much' is less restricted than *bin1* 'which' and it can be used with almost all classifiers except *di1*:

- (9) *gei2(do1) gaa3 ce1*
 how many CL car
 'How many cars?'
 (10) **gei2(do1) di1 ce1*
 how many CL car
 'How many cars?'
 (11) *gei2(do1) bui1 seoi2*
 how many CL-cup water
 'How many cups of water?'
 (12) **gei2(do1) di1 seoi2*
 how much CL water
 'How much water?'

Classifiers sometimes can replace the genitive marker. Instead of the prescribed genitive marker *ge3*, classifiers can be used to denote possession as well:

- (13) *ngo5 ge3 syu1*
 I GEN book 'My book/books'
 (14) *ngo5 bun2 syu1*
 I CL book 'My book'
 (15) *ngo5 di1 syu1*
 I CL book 'My books'

It is interesting to see that the use of different classifiers in (14) and (15) influence the implied number of the head noun referent. With the genitive marker *ge3* in (13), the number of the referent is ambiguous between singular and plural. The use of *bun2* in (14), however, suggests singularity while *di1* in (15) suggests plurality.

In numeral classifier languages, it is common to have a 'special pluralizing classifier, which occurs with demonstratives (but not with numerals) and replaces the semantically appropriate classifier that would be used in non-plural constructions' (Lyons 1977: 461). In Cantonese, *di1* is used for non-singular quantities. The distribution of *di1* is slightly different from other classifiers:

Table 1: Distributions of *-di1* and other classifiers in Cantonese Noun Phrases

[Det-Num-CL-N]	<i>ni1</i>	<i>jat1</i>	<i>di1</i>	<i>syu1</i>	<i>ni1</i>	<i>jat1</i>	<i>bun2</i>	<i>syu1</i>	'this book'
	this	one	CL	book	this	one	CL	book	'this book'
[Det-CL-N]	* <i>ni1</i>	<i>saam1</i>	<i>di1</i>	<i>syu1</i>	<i>ni1</i>	<i>saam1</i>	<i>bun2</i>	<i>syu1</i>	'these three books'
	this	three	CL	book	this	three	CL	book	'these three books'
[Num-CL-N]	<i>ni1</i>	<i>di1</i>	<i>syu1</i>	<i>syu1</i>	<i>ni1</i>	<i>bun2</i>	<i>syu1</i>	<i>syu1</i>	'this book'
	this	CL	book	book	this	CL	book	book	'this book'
[Q-CL-N]	<i>jat1</i>	<i>di1</i>	<i>syu1</i>	<i>syu1</i>	<i>jat1</i>	<i>bun2</i>	<i>syu1</i>	<i>syu1</i>	'one book'
	one	CL	book	book	one	CL	book	book	'one book'
[Q-CL-N]	* <i>sap6</i>	<i>di1</i>	<i>syu1</i>	<i>syu1</i>	* <i>sap6</i>	<i>bun2</i>	<i>syu1</i>	<i>syu1</i>	'ten books'
	ten	CL	book	book	ten	CL	book	book	'ten books'
[Q-CL-N]	? <i>so2jau5</i>	<i>di1</i>	<i>syu1</i>	<i>syu1</i>	* <i>so2jau5</i>	<i>bun2</i>	<i>syu1</i>	<i>syu1</i>	'all the books'
	all	CL	book	book	all	CL	book	book	'all the books'
[Q-CL-N]	? <i>cyun4bou6</i>	<i>di1</i>	<i>syu1</i>	<i>syu1</i>	* <i>cyun4bou6</i>	<i>bun2</i>	<i>syu1</i>	<i>syu1</i>	'all the books'
	all	CL	book	book	all	CL	book	book	'all the books'
[Wh-CL-N]	* <i>cyun4</i>	<i>di1</i>	<i>syu1</i>	<i>syu1</i>	<i>cyun4</i>	<i>bun2</i>	<i>syu1</i>	<i>syu1</i>	'the whole book'
	whole	CL	book	book	whole	CL	book	book	'the whole book'
[Wh-CL-N]	* <i>seng4</i>	<i>di1</i>	<i>syu1</i>	<i>syu1</i>	<i>seng4</i>	<i>bun2</i>	<i>syu1</i>	<i>syu1</i>	'the whole book'
	whole	CL	book	book	whole	CL	book	book	'the whole book'
[Wh-CL-N]	* <i>mui5</i>	<i>di1</i>	<i>syu1</i>	<i>syu1</i>	<i>mui5</i>	<i>bun2</i>	<i>syu1</i>	<i>syu1</i>	'every book'
	every	CL	book	book	every	CL	book	book	'every book'
[Wh-CL-N]	* <i>gei2</i>	<i>di1</i>	<i>syu1</i>	<i>syu1</i>	<i>gei2</i>	<i>bun2</i>	<i>syu1</i>	<i>syu1</i>	'a few books'
	a few	CL	book	book	a few	CL	book	book	'a few books'
[Wh-CL-N]	* <i>hou2do1</i>	<i>di1</i>	<i>syu1</i>	<i>syu1</i>	<i>hou2do1</i>	<i>bun2</i>	<i>syu1</i>	<i>syu1</i>	'many books'
	many	CL	book	book	many	CL	book	book	'many books'
[Wh-CL-N]	* <i>jat1</i>	<i>di1</i>	<i>bun3</i>	<i>syu1</i>	<i>jat1</i>	<i>bun2</i>	<i>bun3</i>	<i>syu1</i>	'one and a half book'
	one	CL	half	book	one	CL	half	book	'one and a half book'
[Wh-CL-N]	<i>bin1</i>	<i>di1</i>	<i>syu1</i>	<i>syu1</i>	<i>bin1</i>	<i>bun2</i>	<i>syu1</i>	<i>syu1</i>	'which book'
	which	CL	book	book	which	CL	book	book	'which book'
[CL-N]	* <i>gei2(do1)</i>	<i>di1</i>	<i>syu1</i>	<i>syu1</i>	<i>gei2(do1)</i>	<i>bun2</i>	<i>syu1</i>	<i>syu1</i>	'how many books'
	how many	CL	book	book	how many	CL	book	book	'how many books'
[CL-N]	<i>di1</i>	<i>syu1</i>	<i>syu1</i>	<i>syu1</i>	<i>bun2</i>	<i>syu1</i>	<i>syu1</i>	<i>syu1</i>	'the book'
	CL	book	book	book	CL	book	book	book	'the book'
[CL _i -CL _i -N]	* <i>di1</i>	<i>di1</i>	<i>syu1</i>	<i>syu1</i>	<i>bun2</i>	<i>bun2</i>	<i>syu1</i>	<i>syu1</i>	'every book'
	CL	CL	book	book	CL	CL	book	book	'every book'
[N _x -CL-N _y]	<i>ngo5</i>	<i>di1</i>	<i>syu1</i>	<i>syu1</i>	<i>ngo5</i>	<i>bun2</i>	<i>syu1</i>	<i>syu1</i>	'my book'
	I	CL	book	book	I	CL	book	book	'my book'

From the table above, it can be seen that the distributions of *di1* contrast mainly with other classifiers in noun phrases involving quantification or enumeration. Other than denoting singularity, *di1* can be used to mean 'all', 'several' or 'some' (cf. Cheung 1961):

- (16) *keoi5dei6 sik6 saai3 di1 daan6goul*
 they eat ASP CL cake
 'They ate all the cake'.
 (17) *jau5 di1 gau2 zau2 zo2*
 have CL dog go ASP
 Some of the dogs have gone.'
 (18) *keoi5dei2 sik6 zo2 jat1 di1 daan6goul*
 they eat ASP one CL cake
 'They ate some of the cake'.

When *di1* is used to mean 'all', an aspectual marker *saai3* is needed. Also, *di1* cannot be preceded by the numeral *jat1* when the meaning 'all' is intended. When *di1* is used to mean 'some', it can follow the existential *jau5* or it can be preceded by the numeral *jat1* 'one'. In other words, [*di1*-N]_{NP} can be used to mean either 'all' or 'some' as in examples (16) and (17) while [*jat1-di1*-N]_{NP} can only be used to mean 'some' as in (18). The only exception to [*jat1-di1*-N]_{NP} is when it is used pre-verbally, together with the aspectual marker *saai3* and the adverb *doul*:

- (19) *keoi5dei2 jat1 di1 daan6goul doul sik6 saai3*
 they one CL cake also eat ASP
 'They ate all the cake'.

2.1 Types of Cantonese Classifiers

The exact number of classifiers in Cantonese has not been agreed upon among linguists. Researchers working on Cantonese have different estimates. Mak (1991) estimates that Cantonese has more than thirty nominal classifiers; Poon (1980) estimates that the total number of classifiers is well under seventy-five while Matthews and Yip (1994) estimate there to be over sixty different classifiers. Data from this study shows that Cantonese-speaking children aged 1;05 to 3;08 used sixty-six different classifiers in spontaneous speech (see Table 7 in section 4 for a full list of classifiers produced by the children under study). The uncertainty in number may be due to dialectal differences and the non-orthographic nature of some classifiers. For example, the classifier for walls can be *dou6* or *bung6* and nasty smell can be classified as either *jat1 zam6 ceoi4* or *jat1 bung6 ceoi4* for which there is no standard written convention for both *bung6* and *zam6*. Furthermore, it is also uncertain whether the two phonemes *bung6* that are used for classifying walls and for classifying nasty smell are homophones. However fuzzy this kind of 'phonemic' classifiers may be, native speakers of Cantonese, including children as well, do use them in spontaneous speech.

The basic distinction of classifiers in Cantonese is shown in Figure 1 below:

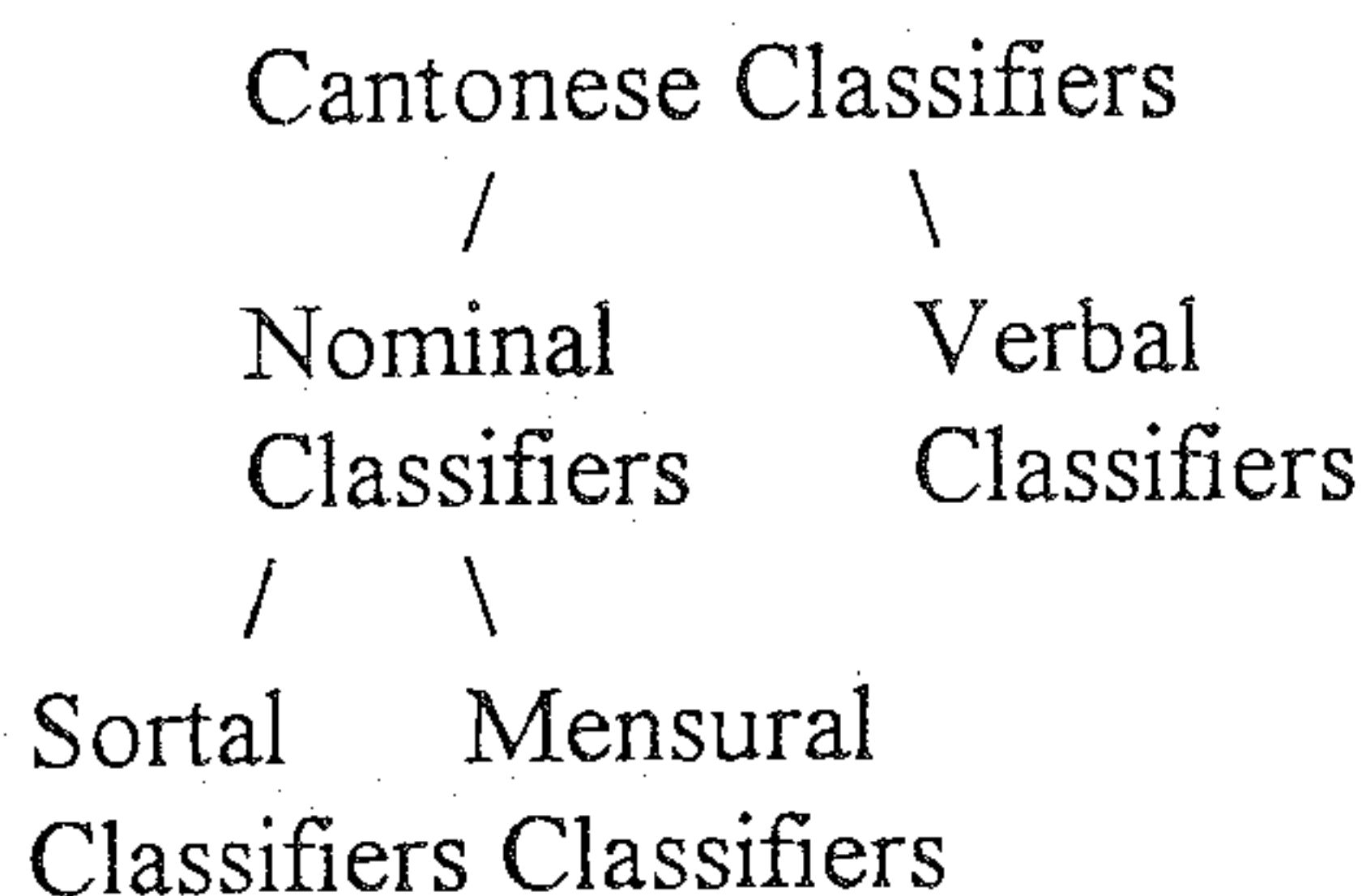


Figure 1: Types of Cantonese Classifiers

Nominal and verbal classifiers are differentiated according to whether they classify nouns or actions. Verbal classifiers are used to enumerate the number of times an action has taken place. They do not have explicit head nouns and they always occur post-verbally.

- (20) *keoi5 daa2 zo2 ngo5 loeng5 kyun4*
 s/he hit ASP I two CL
 'He punched me twice'.

Verbal classifiers are also present in other numeral classifier languages like Japanese. Japanese also has a set of verbal classifiers which modify verbs rather than nouns and they are part of expressions that specify the number of times an action is performed (cf. Sanches 1977)

Two major types of nominal classifiers are generally distinguished: sortal classifiers and mensural classifiers. These two differ with respect to their nature of classification. A sortal classifier individuates whatever it refers to in terms of the kind of entity that it is, whereas a mensural classifier individuates in terms of quantity (Lyons 1977: 463). In other words, sortal classifiers qualify while mensural classifiers quantify the head noun referent.

2.1.1 Sortal Classifiers

Most of the semantic categories used for sortal classification are based on perceptual attributes of the object being classified, such as shape, size, texture. Color, however, has been found to be one clear exception (Allan 1977, Lyons 1977). Sortal classifiers usually have some kind of features like shape or function encoded. They can be applied to a range of head nouns which agree with the features of the classifiers. Others do not have very well-defined features and they are sometimes called 'specific' classifiers. These are so-called as their cooccurrence with the head nouns is almost fixed. *Zaan2*, for instance, is only used to classify lamps. Sortal classifiers very often do not have exact English equivalent.

In many numeral classifier languages, there exists a general classifier which can be used with all nouns. *Ge3* in Mandarin Chinese is the general classifier (Chao 1968: 588, Mak 1991). Erbaugh (1986) also suggested that *ge3* in Mandarin can be used to replace most of the sortal classifiers. In Cantonese, *go3*, the allomorph of the Mandarin *ge3*, is often taken to be the general classifier (Loke & Harrison 1986, Loke 1991). However, it is still questionable

whether *go3* is the general classifier in Cantonese (cf. Mak 1991). Though it can be used with many nouns, there is still restriction for its use. For instance, *go3* is generally not used with substance denoting nouns. The acceptability of NPs like *jat1 go3 seoi2* 'one CL water' is marginal.

Shape classifiers are a prominent sub-group under sortal classifiers. It is often claimed that shape is an important basis for categorization. In Cantonese, examples of shape classifiers include *zi1* and *tiu4* for long objects, *zoeng1* and *faai3* for flat objects and *lap1*, *gau6* and *go3* for round objects.

Function classifiers are so called as they are said to classify objects according to their functions. For example, *baa2* classifies objects with handles, *gaa3* is used for vehicles, *bun2* for books and *gaan1* for buildings etc.

2.1.2 Mensural Classifiers

Unlike sortal classifiers, mensural classifiers do not have any qualitative functions. They are humanly determined units and they quantify the head nouns by some kind of measurement. They could in principle be used with almost all nouns.

A difference between mensural and sortal classifiers is their range of application. The choice of a sortal classifier depends on some characteristics or features of the head nouns. The potential head nouns that can go with any sortal classifier are selective and restrictive. Mensural classifiers, on the other hand, has a considerably wider range of application. There is less restriction governing the cooccurrence of mensural classifiers and the head nouns. Take *gau2* 'dog' as an example, the only sortal classifier that can be used with *gau2* 'dog' is *zek3* which suggests animacy. The mensural classifiers that can go with *gau2* 'dog' are *deoi1*, *kwan4*, *zaa6*, etc. Also, mensural classifiers can be questioned while sortal classifiers cannot:

- (21) *nei5 maai5 zo2 jat1 doi6 ding6 jat1 hap6 beng2gon1*
 you buy ASP one bag or one box biscuit
 'Did you buy a bag or a box of biscuits?'
 (22) **/nei5 maai5 zo2 jat1 zek3 ding6 jat1 tiu4 zoek3zai2*
 you buy ASP one CL-animacy or one CL-long bird
 'Did you buy an animate-thing bird or a long-thing bird?'

Mensural classifiers are found in most of the world's languages. For instance, they are also found in non-classifier languages like English:

- (23) two cups of water *loeng2 bui1 seoi2*
 one CL-cup water
 (24) two pounds of sugar *loeng5 bong6 tong4*
 two CL-pound sugar
 (25) a herd of cattle *jat1 kwan4 ngau4*
 one CL-herd cow

Sentences (23-25) above are examples of container classifiers; standard units of measurements and collective classifiers respectively. Many of the container classifiers are also

common nouns. For example, *bui1* 'cup' is used as a container classifier in *loeng5 bui1 seoi2* 'two cups of water'. On the other hand, *bui1* is a common noun in *loeng5 go3 bui1* 'two cups' in which case it takes a classifier *go3* as other common nouns do. Cantonese has all sorts of standard units of measurements for time like *fan1zung1* 'minute'; weight like *bong6* 'pound'; height like *cek3* 'foot' etc. Collective classifiers designate groups or collections of objects, for example, *deoi1* 'lump'; *kwan4* 'herd' etc.

The plural classifier *di1* is a special member under the group of mensural classifiers. As a mensural classifier, it quantifies the head noun referent but without any specific quantity. The quantity suggested varies between several, some or all (cf. examples (16) to (18)). Also as is characteristic of mensural classifiers, *di1* is compatible with almost all nouns.

2.2 Alternative Classifiers

Though the choice of classifiers is most often fixed, there is still some flexibility. Nouns like *coeng1* 'pistol' may be classified either by a shape classifier *zi1* (for long objects) or by a function classifier *baa2* (for objects with handles). In an experimental study, Mak (1991) found that Cantonese-speaking children tend more to use function rather than shape as the basis for categorization. Poon (1980) showed that in cases where there are two possible classifiers, Cantonese-speaking children preferred to use *go3* if it is one of the two choices.

Variation in classifier assignment may be due to dialectal difference. Older people speaking other Cantonese dialects may classify *caang2* 'orange' by *lap1* as in *jat1 lap1 caang2* 'one orange' meaning the same as *jat1 go3 caang2* 'one orange' (Matthews and Yip 1994). Sometimes, an alternative classifier is used intentionally instead of the prescribed ones to achieve certain purpose. A human being, besides being classified as *jat1 go3 jan4* 'one person' may be classified as *jat1 wai2 sin1saang1* 'a mister', having an honorific overtone; as *jat1 tiu4 jau2* 'one person' or *jat1 daau1 jau2* 'one person' which has some metaphoric extensions used to demean the person.

With the noun *je2* 'thing/stuff', which can possibly be classified by all classifiers, the choice of a classifier affects the interpretation of the head noun referent:

- (26) *ni1 zek3 je2*
this CL-animate thing/stuff
- (27) *sap6 zoeng1 je2*
ten CL-flat object thing/stuff
- (28) *go2 baa2 je2*
that CL-object with handle thing/stuff
- (29) *gei2do1 bong6 je2*
how many CL-pound thing/stuff
- (30) *bin1 di1 je2*
which CL-plural thing/stuff

Each of the classifier used in (26) to (30) is crucial in understanding the noun referent. Sortal classifiers classify according to the characteristic or features of the noun referent. *Zek3* in (26) suggests that the *je2* 'thing/stuff' so classified is probably animate; *zoeng1* in (27) suggests flat objects and *baa2* in (28) suggests the presence of a handle. On the other hand, mensural

classifiers classify according to the quantity of the noun referent. The thing/stuff classified in (29) is measured by *bong6* 'pound' and *di1* in (30) suggest that the stuff being referred to is non-singular.

2.3 Idiomatic Classifier Use

Cantonese classifiers can be used after the adjectives *daai6* 'big' and *sai3* 'small'. With these two adjectives, classifiers are used predicatively (Matthews and Yip 1994: 95):

- (31) *ni1 zek3 gau2 hou2 [daai6 zek3]*
 this CL dog very big CL
 'This dog is very big'.

This kind of idiomatic usage is possible with most classifiers except standard measures and *di1*:

- (32) **ni1 bong6 ping4gwo2 hou2 [daai6 bong6]*
 this CL-pound apple very big CL-pound
 'This pound of apple is very big'.
 (33) **ni1 di1 syu1 hou2 [daai6 di1]*
 this CL book very big CL
 'These books are very big'.

Another characteristic of [*daai6/sai3*- CL] construction is that the classifier used must be compatible with the implied referent. Thus it is ungrammatical to have:

- (34) **ni1 zek3 gau2 hou2 [daai6 faai3]*
 this CL-animate dog very big CL-flat object
 'This dog is very large flat object'.

Even though *daai6 faai3* is perfectly acceptable for flat objects like *geng3* 'mirror' or *min6baau1* 'bread', it cannot be used for *gau2* 'dogs' which takes the classifier *zek3*. Some of the [*daai6/sai3* - CL] constructions are fairly formulaic in Cantonese with a special meanings attached to them. For instance, *daai6 zek3*, *daai6 gau6* are used to mean the physical size of a man as being strong and sturdy-built; and *sai3 lap1* is used for persons who are physically small in size (Matthews & Yip 1994: 95).

3. Acquisition Studies

Children learning classifiers in different languages have been studied in the past few decades, e.g. Thai, Japanese, Mandarin Chinese and Cantonese etc. (cf. Gandour et al. 1984, Carpenter 1991 in Thai; Sanches 1977, Matusmoto 1987 in Japanese; Fang 1985, Erbaugh 1986, Loke & Harrison 1986, Ying et al 1987 in Mandarin Chinese; Poon 1980, Mak 1991 in Cantonese). Results from these studies show that acquiring classifiers is not a simple task. It is more easy for the children to grasp the classifier syntax in enumerating noun phrases while the correct matching between a classifier and a head noun is more difficult for the children.

The earliest reported classifiers are found to be produced by a girl speaking Malay Cantonese at the age of 1;05 in a longitudinal study conducted by Li (1968). Before 2 years old, the girl already knows the syntax of [Num-CL-N]_{NP} and [Dem-CL-N]_{NP}. Erbaugh (1986) also in a longitudinal study reported the use of classifiers in spontaneous speech by Mandarin-speaking children within the age two to three. The classifier use blossomed after 2;6. Both classifier types and tokens increased at a statistically significant rate. Before 2;6, children overwhelmingly used the general classifier. Classifier production is not frequent. In her two children under study, Erbaugh found a total of 13 classifier types and 60 classifier tokens used by one girl and 17 types and 68 tokens used by the other boy. Loke & Harrison (1986) noted a significant change from inappropriate use of classifiers to appropriate use between the ages 3 and 4.

Children rarely omit a classifier in obligatory context. In Erbaugh's (1986) study there are only six omissions of classifiers by children in over 44,000 utterances. Loke (1991) also shows that children seldom omit the classifier in obligatory context in enumerating objects. Children are correct with the classifier syntax in [Num-CL-N]_{NP} though the use of classifier may be incorrect. In experiments (cf. Ying et al. (1987)), children at 4 years old already conform to the [Num-CL-N]_{NP} pattern.

The first classifier that is being acquired is often said to be the general classifier. The children studied by Erbaugh in Mandarin Chinese overwhelmingly preferred to use the general classifier *ge3* especially before 2;6. Fang (1985) and Ying et al. (1987) both show that *ge3* is frequently used by 4 year olds Mandarin-speaking children. For Cantonese-speaking children, they also used *go3* excessively especially for the younger children (c.f. Poon 1980, Loke & Harrison 1986, Mak 1991) and the use decrease with age. Poon (1980) points out that in cases where two possible classifiers can be used, *go3* is always the preferred classifier choice.

Though classifiers appear in the early years, Mandarin-speaking children under 4 years old only have a limited knowledge of classifier semantics (Fang 1985) and children even up to 7 years old do not have a very good knowledge of classifier semantics (Ying et al. 1987). Similarly, Cantonese-speaking children at 6;6 have not yet reached adult proficiency (Poon (1980)).

4. Classifiers in Child Cantonese

4.1 The Data

Naturalistic speech data were recorded from eight Cantonese-speaking children who were brought up in monolingual families living in Hong Kong. The children were observed for a year and were aged between 1;05;22 and 3;08;09 (see Table 2 below). For the purpose of this paper, only the spontaneous utterances of the children were used. An utterance was considered spontaneous if it was not a repetition of either the whole or a part of an utterance found in the previous two conversational turns.

Table 2: List of Subjects under study

Child	Sex	Age of recording	No. of sessions
Tsuntsun	Male	1;05;22-2;07;02	25
Bernard	Male	1;07;00-2;08;06	26
Chunyat	Male	1;10;08-2;10;27	22
Gakei	Female	1;11;01-2;09;09	19
Tinfaan	Female	2;02;10-3;02;18	16
Bohuen	Female	2;03;22-3;04;08	27
Johnny	Male	2;04;08-3;04;14	16
Jenny	Female	2;08;10-3;08;09	20

4.2 Findings and Discussion

4.2.1 Age of Emergence

All the eight children use classifiers in their spontaneous speech. For the six older subjects Chunyat, Gakei, Tinfaan, Bohuen, Johnny and Jenny, classifiers are found at the beginning of recording. For the two youngest children Tsuntsun and Bernard, classifiers begin to appear in Tsuntsun's speech at 1;08;07 with three tokens and in Bernard's speech at 1;07;00 with one token. However, after the first occurrence of classifiers, there is then an absence of classifier use. The use of classifiers do not become productive until 1;11;13 for Tsuntsun and at 1;09;18 for Bernard and both types and tokens increase rapidly since then. Table 3 and 4 below list the number of classifier types and tokens used by Tsuntsun and Bernard before 2 years old.

Table 3: Number of Classifier Types and Tokens used by Tsuntsun before 2;0

Age	1;05;22	1;07;03	1;07;10	1;08;00	1;08;07	1;08;21	1;09;07	1;09;14	1;09;29	1;10;30	1;11;13	1;11;27
Tokens	0	0	0	0	3	0	0	2	0	1	19	8
Types	0	0	0	0	2	0	0	2	0	1	7	3

Table 4: Number of Classifier Types and Tokens used by Bernard before 2;0

Age	1;07;00	1;08;00	1;08;04	1;08;28	1;09;04	1;09;18	1;09;25	1;10;10	1;10;23	1;11;06
Tokens	1	0	0	0	1	9	11	28	18	18
Types	1	0	0	0	1	5	4	5	4	4

From 1;05;22 to 1;11;27, Tsuntsun is able to use nine different types of classifiers and they are (in order of first occurrence, items within semi-colons first occur in the same session): *go3*; *zek3*; *dou6*; *lap1*, *gau6*, *jip6*, *gin6*, *dil*; *bun2*. Bernard between the age 1;07;00 to 1;11;06 uses eight different types of classifiers: *go3*; *gaa3*, *dou6*, *deoi3*, *fanlzung1*; *bun2*; *zek3*, *seoi3*.

Consider also the data of Chunyat and Gakei before 2;0, the use of classifiers is already productive in the first recording session when Chunyat is at 1;10;08 and when Gakei is at 1;11;01. The number of types and tokens used by Chunyat and Gakei before 2 years old is listed in Table 5 and 6 respectively.

Table 5: Number of Classifier Types and Tokens used by Chunyat before 2;0

Age	1;10;08	1;11;00	1;11;21
Tokens	9	27	38
Types	2	5	2

Table 6: Number of Classifier Types and Tokens used by Gakei before 2;0

Age	1;11;01	1;11;08	1;11;22	1;11;29
Tokens	6	9	2	6
Types	2	6	2	3

Five different types of classifiers are found in Chunyat's speech from 1;10;08 to 1;11;21. The five types are: *dou6*, *dil*; *go3*, *zek3*, *bin1*. Gakei from 1;11;01 to 1;11;29 uses eight different classifier types: *go3*, *dou6*; *faai3*, *gaa3*, *dil*, *deoi3*; *zil*; *bin1*.

Looking at the data of Tsuntsun, Bernard, Chunyat and Gakei before 2;0, it is clear that classifiers start to emerge towards the end of the second year and by 2 years old, Cantonese-speaking children are able to use around 5 to 8 different types of classifiers.

4.2.2 Classifier Vocabulary

The total number of classifier types used by all eight children come up to 66 different types. Each child has a set of classifiers from 24 to 40 different types. Table 7 summaries the total number of classifier types and tokens and the percentage of use of each classifier.

Table 7: Percentage of Use of Classifiers and the Total Number of Classifier Types and Tokens used by Each Child

	Tsuntsun	Bernard	Chunyat	Gakei	Tinfaan	Bohuen	Johnny	Jenny
	1;05;22- 2;07;09	1;07;00- 2;08;06	1;10;08- 2;10;27	1;11;01- 2;09;09	2;02;10- 3;02;18	2;03;22- 3;04;08	2;04;08- 3;04;14	2;08;10- 3;08;09
Classifier Type								
1. <i>go3</i>	46.3	39.2	23.9	42.2	60	56.2	47.3	29.4
2. <i>zek3</i>	3.2	11.7	5.3	9.1	5.2	3.3	3.7	3.3
3. <i>zil</i>	0.2	0.4	0.9	1.7	0.05	2.6	1.4	0.6
4. <i>tiu4</i>	0.2	0.3	0.9	3	0.7	0.7	0.3	1.2
5. <i>zoengl</i>	1.6	1.6	1.2	0.9	0.9	1.3	0.5	1.4
6. <i>faai3</i>	/	1.2	/	0.7	0.4	/	0.2	0.1
7. <i>lap1</i>	1.8	0.1	0.6	0.4	1.1	2.4	0.8	1.4
8. <i>gau6</i>	1.2	0.3	0.6	0.2	/	0.7	0.7	0.1
9. <i>gaa3</i>	11.3	3.5	3.6	0.6	0.6	/	13.7	0.6
10. <i>baa2</i>	0.2	0.4	/	0.5	0.2	/	0.3	/
11. <i>gaan1</i>	1.2	1.2	0.2	0.2	1.36	1.52	0.9	0.04
12. <i>bun2</i>	1.4	1	0.9	3.8	0.9	0.2	0.6	1.1
13. <i>baan2</i>	0.2	1.2	/	/	0.05	/	/	0.4
14. <i>jip6</i>	0.2	0.7	/	/	/	/	/	/
15. <i>fo3</i>	/	/	/	/	/	/	0.1	/

16. <i>do2</i>	/	0.4	/	/	/	/	0.04	0.2
17. <i>po1</i>	0.6	0.3	/	0.6	0.3	0.6	0.4	0.04
18. <i>gin6</i>	1.2	/	0.3	0.1	0.6	/	0.2	0.4
19. <i>tou3</i>	/	/	0.1	/	/	/	0.04	0.1
20. <i>fuk1</i>	0.2	0.1	/	/	/	/	/	0.1
21. <i>sau5</i>	/	/	/	/	0.4	0.2	/	/
22. <i>wai2</i>	/	/	/	/	0.05	/	/	/
23. <i>fan6</i>	/	/	/	/	0.3	/	/	/
24. <i>fung1</i>	0.6	/	/	/	0.3	/	/	/
25. <i>beng2</i>	/	/	0.6	/	/	/	/	0.1
26. <i>dou6(mun4)</i>	/	/	/	0.1	0.1	/	/	0.04
27. <i>deng2</i>	/	/	/	/	0.2	/	/	0.2
28. <i>cang4</i>	/	/	/	/	0.05	/	/	/
29. <i>dou6</i>	24.9	27.6	50.2	16.4	17	13	17.7	30
30. <i>cyu3</i>	0.4	/	/	/	/	0.2	/	/
31. <i>bin1</i>	0.4	0.5	0.04	0.2	0.2	/	0.04	0.8
32. <i>bin6</i>	/	0.4	0.4*	0.3	0.2	/	0.08	/
33. <i>jat6</i>	/	/	0.1*	0.8	0.3	0.2	0.1	0.2
34. <i>dim2</i>	/	/	0.3	/	0.2	/	/	/
35. <i>seoi3</i>	/	0.4	/	/	0.05	0.4	/	0.3
36. <i>di1</i>	1.8	5.4	4.4	15.6	7.9	10.6	9.8	24
37. <i>deoi3</i>	0.2	0.5	0.5	0.1	0.05	0.4	0.3	0.3
38. <i>paai4</i>	/	/	0.04	/	/	/	/	/
39. <i>deoi1</i>	/	/	0.04	/	/	/	/	/
40. <i>daap6</i>	/	/	/	/	/	/	/	0.1
41. <i>baan1</i>	/	/	/	/	/	/	/	0.2
42. <i>zung2</i>	/	/	/	/	/	0.4	0.08	/
43. <i>joeng6</i>	0.2	/	/	0.2	/	/	0.04	/
44. <i>daam6</i>	0.2	/	0.1	0.2	0.2	/	/	0.2
45. <i>gang1</i>	/	/	/	0.4	/	/	/	/
46. <i>bong6</i>	/	/	0.3	0.1	/	/	/	0.3
47. <i>jjun4</i>	/	/	0.04	/	/	/	/	/
48. <i>man1</i>	/	0.3	0.4	0.3	0.1	0.4	/	0.3
49. <i>fan1zung1</i>	/	/	/	/	/	/	/	0.2
50. <i>buil</i>	/	0.1	0.4	0.1	0.1	2.2	0.04	0.04
51. <i>hap6</i>	/	0.1	/	/	0.1	0.2	/	/
52. <i>zeon1</i>	/	0.3	0.1	/	/	/	/	0.04
53. <i>baau1</i>	/	/	0.1	/	/	0.7	0.1	0.2
54. <i>gwun3</i>	/	/	/	/	/	/	0.08	/
55. <i>doi6</i>	/	/	/	/	/	/	0.04	0.04
56. <i>gaak3</i>	/	0.5	0.1	/	/	/	0.04	0.1
57. <i>seng1</i>	/	/	/	/	/	/	/	0.1
58. <i>hyun1</i>	/	/	/	0.2	/	/	/	/
59. <i>&gaak3</i>	/	/	0.2	/	/	/	/	/
60. <i>&daat3</i>	/	/	2.2	0.2	/	/	/	0.3
61. <i>&zaa6</i>	/	/	/	0.2	/	/	/	/
62. <i>&pet6</i>	/	/	/	/	/	/	0.04	/
63. <i>bou6</i>	/	/	/	/	/	0.4	/	/
64. <i>waak6</i>	/	/	/	/	/	/	0.08	/
65. <i>ci3</i>	/	/	1.1	0.5	0.1	1.1	0.3	1.5
66. <i>haa2</i>	0.2	/	/	/	0.05	/	/	/
Total no. of Types	25	28	33	31	35	24	33	40
Total no. of Tokens	497	735	2249	993	1835	461	2389	2335

As can be seen from table 7, Bohuen has the fewest types of classifiers used. This may be due to the fact that she also has the fewest number of classifier tokens. Compared with Tsuntsun, whose classifier tokens are similar to Bohuen, the number of classifier types of these two children are very similar. The number of types and tokens of classifiers do not seem to correlate with age.

With respect to the types of classifiers used, nominal classifiers are the majority. The use of verbal classifiers is extremely rare. Only four verbal classifiers: *ci3*, *haa2*, *bou6* and *waak6* are found. Table 8 below lists the total number of verbal classifiers found in each child.

Table 8: Total Number of Verbal Classifiers used by Each Child

	Tsuntsun	Bernard	Chunyat	Gakei	Tinfaan	Bohuen	Johnny	Jenny
<i>ci3</i>	0	0	10	2	1	0	6	10
<i>haa2</i>	1	0	0	0*	1	0	0	0
<i>bou6</i>	0	0	0	*0	0	2	0	0
<i>waak6</i>	0	0	0	0	0	0	2	0

Verbal classifiers are totally absent in Bernard's speech. The other children mostly used only one type of verbal classifiers. The use of verbal classifiers is by no means prominent or significant in child Cantonese.

However, the children do conform to the syntactic restrictions when they use these verbal classifiers. All of the verbal classifiers are used without the head noun and they all occur after a number or a quantifier in the form [Num/Q - CL]_{NP}. Only the oldest child Jenny uses quantifiers *hou2do1* 'many' and *gam3do1* 'so many' before the verbal classifier *ci3*. Besides, the classifier phrase either form an independent utterance or it is used after a verb. Only one exception is found with Chunyat. There is a case where Chunyat put the verbal classifier phrase before the verb:

35. (Chunyat at 2;06;24. Chunyat is drinking a bottle of milk and the investigator asks him how long does he need to finish the milk.)

- *INV: *gei2si4 sin1 jam2 jyun4 aa3?*
how long then drink finish SFP
'When can you finish drinking (the milk)?'
- *CHI: *saam1 ci3 jam2 jyun4 lo1*
three CL-times drink finish SFP
'It takes three times (for me) to finish.'

Among the nominal classifiers, the use of sortal classifiers outnumbers the use of mensural classifiers in all children. The proportion of sortal to mensural classifiers varies a great deal among individual child. Table 9 lists the number of sortal and mensural classifiers used by each child.

Table 9: Number of Types of Sortal and Mensural Classifiers used by Each Child

Child\ No. of types	Tsuntsun	Bernard	Chunyat	Gakei	Tinfaan	Bohuen	Johnny	Jenny
Sortal Classifiers	21	20	18	21	24	14	24	25
Mensural Classifiers	3	8	14	9	9	8	7	14

The inventory of the children's classifier system is substantial and the rate of expansion is considerable. Within a year, the two youngest children, Tsuntsun and Bernard, have developed a classifier system from zero to 25 and 28 different types respectively. The oldest child, Jenny, is able to use even up to 40 different classifier types from 2;08 to 3;08. This finding contrasts sharply with the previous longitudinal study on Mandarin Chinese which report a total of 17 classifiers used by 2 children aged 1;10 to 3;10 (Erbaugh 1986).

4.2.3 Core Set of Classifiers

Among the sixty-six classifiers found, eleven of them are used by all eight children: *go3*, *zek3*, *zi1*, *tiu4*, *zoeng1*, *lap1*, *gaan1*, *bun2*, *dou6*, *di1*, *deoi3*. This 'core' set of classifiers occupies more than 90% of classifier use of each child. The core set includes 4 sortal classifiers for shape (*zi1* for long rigid objects; *tiu4* for long flexible objects; *zoeng1* for flat objects; *lap1* for small round objects); 2 sortal classifiers for function (*gaan1* for buildings; *bun2* for books); 1 sortal classifier for animacy (*zek3*), 1 sortal classifier for place (*dou6*) and 1 sortal classifier for multiple classifications (*go3*). Only two mensural classifiers fall within the core set: *di1* and *deoi3*. *Di1* suggests fuzzy quantity and *deoi3* pertains to arrangement.

Go3 and *dou6* are by far the two most dominant sortal classifiers used by all children (see Table 7 above). Except Jenny, the 7 younger children used *go3* most while Jenny used *dou6* slightly more often than *go3* (29.94% for *dou6* and 29.42% for *go3*). *Di1*, on the other hand, is the most frequently used mensural classifier by all children. The other mensural classifiers occupied only from 0.04% to 2.17% of the total classifier use. Besides being the most frequently used sortal and mensural classifier, *go3*, *dou6* and *di1* are also among the first classifiers that the children used.

4.2.4 Early Classifier Errors

The majority of the children's classifiers are used without head nouns. The average percentage of NPs with head nouns used immediately after the classifier is 5.2% for Tsuntsun, 13.3% for Bernard, 11.3% for Chunyat, 31.2% for Gakei, 23.2% for Tinfaan, 30% for Bohuen, 13.2% for Johnny and 19.4% for Jenny⁴. [Det - CL]_{NP} is the most preferred NP types for the children's classifier use. It is also the first kind of NPs that Tsuntsun and Bernard use with classifiers.

Non-adult use of classifiers is not very frequent. In the following I will outline three types of classifier errors found in the children. The first is the use of an inappropriate classifier with a head noun; the second is the use of a single classifier alone and the third one is the double classifiers construction.

A. Inappropriate Classifier⁵

In learning to use classifiers, one of the major task of the child is to learn the correct matching of the classifier with the head noun. The data collected shows that children sometimes make mistakes in matching classifiers with the head nouns. Most of the inappropriate classifier choices occur with *go3*. There are a variety of head nouns that are used with the classifier *go3* incorrectly by the children, among which the more common ones include animals, vehicles, plants, body parts and food items etc.

Examples of *go3* used for animals:

- | | |
|--|--|
| 36. <i>jat1 m2 go3 jyu2</i>
one five CL fish | (Tsuntsun at 2;05;00)
'one five fish' |
| 37. <i>nil go3 maaulmaaul</i>
this CL cat | (Bernard at 2;06;04)
'this cat' |
| 38. <i>nil go3 jyu2</i>
this CL fish | (Chunyat at 2;09;23)
'this fish' |
| 39. <i>nil go3 maaulmaaul aa3</i>
this CL cat SFP | (Gakei at 1;11;29)
'this cat' |
| 40. <i>nil go3 aap3aap3</i>
this CL duck | (Tinfan at 2;07;20)
'this duck' |
| 41. <i>nil go3 lou5fu2 aak1</i>
this CL tiger SFP | (Bohuen at 2;10;23)
'this tiger' |
| 42. <i>loeng5 go3 jyu2</i>
two CL fish | (Johnny at 2;05;03)
'two fish' |
| 43. <i>nil go3 baak6tou3</i>
this CL rabbit | (Jenny at 2;01;30)
'this rabbit' |

Examples of *go3* used for vehicles:

- | | |
|--|---|
| 44. <i>tai2 nil go3 cel</i>
look this CL car | (Tsuntsun at 2;02;05)
'look at this car' |
| 45. <i>zeon3zeon3 jau5 nil go3 cel aa3</i>
Chunyat have this CL car SFP | (Chunyat at 2;06;08)
'Chunyat has this car' |
| 46. <i>nil go3 baalsi2</i>
this CL bus | (Gakei at 2;05;03)
'this bus' |
| 47. <i>daap3 gwo3 nil go3 syun4 lu3</i>
ride ASP this CL boat SFP | (Tinfan at 2;07;05)
'I have had a ride on this boat' |
| 48. <i>jau5 go3 baalsi2 aa3</i>
have CL bus SFP | (Bohuen at 2;04;02)
'there is a bus' |
| 49. <i>sei3 go3 cel</i>
four CL car | (Johnny at 2;05;03)
'four cars' |
| 50. <i>gam3 dol go3 celcel ge2</i>
so many CL car SFP | (Jenny at 3;01;13)
'so many cars' |

Examples of *go3* used for plants:

- | | |
|--|---|
| 51. <i>hai2 go3 faa1 lol</i>
at CL flower SFP | (Tsuntsun at 2;06;18)
'at the flower' |
| 52. <i>ni1 go3 faa1</i>
this CL flower | (Bernard at 2;02;26)
'this flower' |
| 53. <i>ni1 go3 sing3daan3 syu6</i>
this CL Christmas tree | (Chunyat at 2;09;07)
'this Christmas tree' |
| 54. <i>ni1 go3 faa1 lol</i>
this CL flower SFP | (Tinfaan at 2;06;01)
'this flower' |
| 55. <i>jau5 go3 syu6</i>
have CL tree | (Bohuen at 3;04;08)
'there is a tree' |
| 56. <i>loeng5 go3 faa1faa1</i>
two CL flower | (Jenny at 3;03;15)
'two flowers' |

Examples of *go3* used for body parts:

- | | |
|---|--|
| 57. <i>ni1 go3 hai6 sau2</i>
this CL is hand | (Chunyat at 2;10;27)
'this is a hand' |
| 58. <i>hou2 dai6 go3 ji5zai2</i>
very big CL ear | (Gakei at 2;02;28)
'a very big ear' |
| 59. <i>loeng5 go3 goek3</i>
two CL leg | (Tinfaan at 2;07;20)
'two legs' |
| 60. <i>ni1 go3 ngaan5zing1</i>
this CL eye | (Bohuen at 2;09;26)
'this eye' |
| 61. <i>loeng5 go3 sau2</i>
two CL hand | (Johnny at 2;10;13)
'two hands' |
| 62. <i>nei5 go3 ji5zai2</i>
you CL ear | (Jenny at 3;03;15)
'your ear' |

Examples of *go3* used for food items:

- | | |
|--|---|
| 63. <i>ni1 go3 syut3tiu2 laan6 zo2</i>
this CL popsicle broken ASP | (Bernard at 2;05;19)
'this popsicle is broken' |
| 64. <i>ni2 go3 daan2</i>
this CL egg | (Chunyat at 2;09;23)
'this egg' |
| 65. <i>ni1 go3 hoenglziul</i>
this CL banana | (Tinfaan at 2;07;20)
'this banana' |
| 66. <i>ni1 go3 hoenglziul hou2mei6 aa3</i>
this CL banana tasty ASP | (Bohuen at 3;01;01)
'this banana is tasty' |
| 67. <i>loeng5 go3 daan2</i>
two CL egg | (Johnny at 2;05;03)
'two eggs' |
| 68. <i>ni1 go3 naai1naai1 laa1</i>
this CL milk ASP | (Jenny at 3;00;11)
'this milk' |

Inappropriate classifier choice is less frequent in the two youngest children Tsuntsun and Bernard than in the older children. On the whole, classifier mismatch is not very frequent in the children's speech but it does occur throughout the period. All children occasionally used *go3* inappropriately. Besides *go3*, *zek3* is found to be used inappropriately by Gakei and once by Tinfaan and by Johnny:

Examples of classifier mismatch with *zek3*:

- | | |
|---------------------------------------|------------------------------|
| 69. <i>ngo5 zek3 doi2</i> | (Gakei at 2;03;18) |
| I CL bag | 'my bag' |
| 70. <i>keoi5 zek3 hau2</i> | (Gakei at 2;03;25) |
| s/she CL mouth | 'his/her mouth' |
| 71. <i>zek3 dan3</i> | (Gakei at 2;04;08) |
| CL chair | 'the chair' |
| 72. <i>ngo5 zek3 dan3</i> | (Gakei at 2;04;08) |
| I CL chair | 'my chair' |
| 73. <i>ngo5 zek3 mou4gan1</i> | (Gakei at 2;04;30) |
| I CL towel | 'my towel' |
| 74. <i>ngo5 zek3 dan3</i> | (Gakei at 2;04;30) |
| I CL chair | 'my chair' |
| 75. <i>go2 zek3 tou3</i> | (Gakei at 2;04;30) |
| that CL cover | 'that cover' |
| 76. <i>lo2 di1 ni1 zek3 bing1</i> | (Gakei at 2;05;03) |
| take CL this CL ice | 'take some ice of this kind' |
| 77. <i>ngo5 zek3 pet1pet1</i> | (Gakei at 2;05;03) |
| I CL hip | 'my hip' |
| 78. <i>ngo5 zek3 tou3</i> | (Gakei at 2;05;03) |
| I CL cover | 'my cover' |
| 79. <i>jau5 zek3 tau4</i> | (Gakei at 2;08;18) |
| have CL head | 'there is a head' |
| 80. <i>ni1 zek3 jyu2 aa3</i> | (Tinfaan at 3;01;21) |
| this CL fish SFP | 'this fish' |
| 81. <i>hou2 daai6 zek3 celcel aa3</i> | (Johnny at 2;08;08) |
| very big CL car SFP | 'a very big car' |

B. Single Classifier [CL]_{NP}

The children occasionally use a single classifier alone in a noun phrase. This is relatively more frequent in the youngest child Tsuntsun and is not found in Bernard's data. For Tsuntsun, the pattern persists over five months and finally disappears after he was 2;03;03. In Tsuntsun's data, some of the [CL]_{NP} are used as independent utterances and the others all follow the negator *m4*. In the other children, usually the single classifier is used in longer utterances. While some of them may be considered interpretable as referring to some specific referent mentioned in the previous discourse (cf. section 2 examples 1 to 2), the following list are the odd cases:

82. [di1]	(Tsuntsun at 1;11;13)
83. [go3 / bun2]	(Tsuntsun at 1;11;27)
84. [dou6]	(Tsuntsun at 2;00;09)
85. [zek3]	(Tsuntsun at 2;01;08)
86. [dou6]	(Tsuntsun at 2;02;15)
87. m4 [go3 / jip6 / lap1 / dou6 / gin6 / di1 / gau6]	(Tsuntsun at 1;11;13)
88. m4 [go3]	(Tsuntsun at 1;11;27)
89. m4 [gau6 / go3]	(Tsuntsun at 2;03;03)
90. [gaa3]	(Chunyat at 2;02;06)
91. mou5 [zoeng1] not-have CL	(Chunyat at 2;07;06) 'a sheet is missing'
92. m4 gin3 [zek3] no see CL	(Gakei at 2;02;21) 'one has disappeared'
93. [di1] m4 sik1 aa3 CL no know SFP	(Tinfan at 2;07;20) '(I) don't know these'
94. dim2gaai2 [go3] gam2 ge2 why CL like-this SFP	(Tinfan at 2;11;16) 'why is this like this?'
95. ni1 go3 e6, [zoeng1] le1, [zoeng1] le1 this CL SFP CL SFP CL SFP	(Bohuen at 3;03;03) 'this, where is the sheet, the sheet?'
96. [gaa3]	(Johnny at 2;05;13)
97. [pet6]	(Johnny at 2;06;10)
98. dou1 mou5 [gin6] also not-have CL	(Johnny at 3;01;16) 'there isn't a piece (of garment)'
99. hou2 ci5 [zoeng1] gam2joeng2 very like CL like-this	(Jenny at 3;01;30) 'like a sheet like this'

C. Double Classifiers Constructions [Det - CL - CL -N]

This type of double classifiers constructions are by no means frequent but they do appear in 7 out of 8 children. Except Bohuen, the double classifiers constructions are found in the speech of Tsuntsun, Bernard, Chunyat, Gakei, Tinfan, Johnny and Jenny. Usually, the two classifiers are not identical and they are sometimes used before or after a verb.

100. [ni1 go3 zek3 goek3] this CL CL foot	(Tsuntsun at 2;04;00)
101. [ni1 go3 zek3 ngaa5] this CL CL eye	(Bernard at 2;02;26)
102. [loeng5 daat3 tiu4 jyu2zai2] two CL CL fish	(Chunyat at 2;07;13)
103. [ni1 dou6 bun2 syu1] gam2 ge3 this CL CL book like-this SFP	(Chunyat at 2;10;13)
104. mit1 laan6 zo2 [go2 di1 bun2 syu1] gam2 gaa3 tear broken ASP that CL CL book like-this SFP	(Chunyat at 2;10;13)
105. [ni1 dou6 zi1 bat1] this CL CL pen	(Chunyat at 2;10;27)
106. [go2 go3 baa2 dou1] that CL CL knife	(Gakei at 2;04;30)

107. [ni1 go3 zoeng1 dan3] (Gakei at 2;04;30)
 this CL CL chair
108. [jat1 zek3 tiu4 cung4] (Gakei at 2;08;18)
 one CL CL worm
109. [jat1 go3 zek3 mo4gwu1] (Gakei at 2;08;18)
 one CL CL mushroom
110. ngo3 cai3 gaan1 [ni1 go3 gaan1 uk1] aa3 (Tinfan at 2;04;27)
 I build CL this CL CL house SFP
111. [ni1 dou6 dou6 mun4] (Tinfan at 2;04;27)
 this CL CL door
112. [ni1 go3 gaan1 uk1] (Tinfan at 2;07;20)
 this CL CL house
113. [ni1 go3 gaan1 uk1] (Tinfan at 2;09;07)
 this CL CL house
114. [ni1 go3 baa2 sin3] lo1 (Tinfan at 2;11;26)
 this CL CL fan SFP
115. maai5 [ni1 di1 di1 ce1] (Johnny at 2;05;19)
 buy this CL CL car
116. ngo5 maai5 [ni1 go3 gaa3 ce1] (Johnny at 2;09;30)
 I buy this CL CL car
117. [ni1 go3 baa2 gim3] (Johnny at 2;09;30)
 this CL CL sword
118. [ni1 go3 gaan1 uk1] gaa3 (Johnny at 2;10;13)
 this CL CL house SFP
119. [ni1 po1 po1 syu6] aa3 (Johnny at 2;10;13)
 this CL CL tree SFP
120. [ni1 go3 gaa3 ce1] mei5 tau4 ce1 (Johnny at 3;00;08)
 this CL CL car tail head car
121. jau6 zong6 [go2 gaa3 gaa3 ce1] (Johnny at 3;00;08)
 again crash that CL CL car
122. [loeng5 zek3 baa2 gim3] (Johnny at 3;01;16)
 two CL CL sword
123. [ni1 go3 zek3 ngaan5] aa3 (Johnny at 3;04;14)
 this CL CL eye SFP
124. [ni1 go3 zek1 sau2] li4li2 sau2 aa3 (Jenny at 2;09;14)
 this CL CL hand Jenny hand SFP
125. [ni1 di1 di1 haa4] lei4 gaa4 (Jenny at 3;00;22)
 this CL CL shoe SFP SFP
126. taam3 haa2 [ni1 di1 zek3 sau2] lo1 (Jenny at 3;01;30)
 visit PRT this CL CL hand SFP
127. ngo5 mou5 go2 e6 [go2 go3 zoeng1 soeng2] aa3 (Jenny at 3;01;30)
 I not-have that SFP that CL CL photo SFP
128. [ni1 dou6 zek3 goek3] aa3 (Jenny at 3;02;13)
 this CL CL foot SFP

From the above examples, it can be seen that the two classifiers may or may not be identical. If the two classifiers are not identical, the first classifier tends to be the more general classifier *go3*. The second classifier preceding the head noun, interesting enough, is often the

appropriate classifier for the head noun. Some of the double classifiers constructions may be understood as two separate NPs. For instance in (125), Jenny may mean [*ni1 di1*] (hai6) [*di1 haai4*] lei4 gaa4 'Are these shoes?' But in cases where the double classifiers construction is placed in a single argument position, then the whole phrase should be treated as one NP.

4.2.5 Other Classifier Knowledge

Besides the formal properties of classifiers, the children also used the classifiers idiomatically. Idiomatic use of *daai6 / sai3 go3* 'old/young' and *daai6 baa2* 'a lot' are found:

129. *zeon3zeon3 soeng5 tong4 aa3 daai6 go3 laa3* (Chunyat at 2;07;13)
 Chunyat up class SFP big CL SFP
 'Chunyat goes to school, he's old now'.
130. *keoi5 daai6 go3 neoi2 m4 sik1 waan2 tam4tam4zyun3* (Gakei at 2;05;03)
 s/he big CL girl no know play merry-go-round
 'S/he is old and s/he does not know how to play merry go round'.
131. *ngo5 daai6 go2 m4 sik6 gaa3* (Tinfan at 2;04;21)
 I big CL no eat SFP
 'I'm old and I don't eat (this)'.
132. *nei5 sai3 go3 dou1 jiu3 faan1 hok3 gaa3* (Tinfan at 2;07;20)
 you small CL also need go school SFP
 'You also need to go to school when you're small'.
133. *daai6 go3 jam2 naai1naai1 gaa3 laak3* (Bohuen at 2;10;02)
 big CL drink milk SFP SFP
 '(I) drink milk when I'm old'.
134. *ngo5 sai3 go3 gaa3* (Jenny at 3;00;11)
 I small CL SFP
 'I'm small'.
135. *nei5 hou2 daai6 go3 gaa3 laa3* (Jenny at 3;02;13)
 you very big CL SFP SFP
 'You're very old'.
136. *ngo5 le1, sai3 go3, ngo5 saam1 seoi3 aal maa3* (Jenny at 3;03;26)
 I SFP small CL I three year-old SFP SFP
 'I'm small. I'm three years old'.
137. *daai6 baa2 syu1* (Gakei at 1;11;29)
 big CL book
 'a lot of books'
138. *daai6 baa2 je2 aa3* (Johnny at 2;06;24)
 big CL thing SFP
 'a lot of things'
139. *ngo5 jau5 daai6 baa2* (Johnny at 3;02;16)
 I have big CL
 'I have a lot'.

Idiomatic classifier use is not found in the speech of the two youngest children Tsuntsun and Bernard which may suggest that children learn this kind of idiomatic classifier use in a later stage.

5. Conclusions

The investigation of classifier development in Cantonese-speaking children between the age 1;05 to 3;08 shows that classifiers emerge early. From the two youngest subjects, Tsuntsun and Bernard, classifiers first appear at 1;08 and 1;07 respectively and classifiers are then used continuously starting from 1;10 and 1;09 respectively. Cantonese-speaking children before two years old are able to use classifiers in their spontaneous speech and by two years old, they can use around 5 to 9 different types of classifiers. Classifier development during the first two years is rapid and the classifier system expands considerably within a year. There is a core set of 11 classifiers that are found to be used by all children and the three most frequent classifiers are *go2*, *dou6* and *di1*. Children's use of classifiers generally conform to the adult pattern though some errors like classifier mismatch, a single classifier NP or the double classifiers constructions are found occasionally.

References

- Allan, K. (1977) 'Classifier', in *Language*, 53 ii: 285-311.
- Carpenter, Kathie (1991) 'Later rather than sooner: extralinguistic categories in the acquisition of Thai classifiers', in *Journal of Child Language* 18: 93-113.
- Chao, Yuen-ren (1968) *A Grammar of Spoken Chinese*. Berkeley, University of California Press.
- Cheung, K. (1961) "Gong2 zau1 waa2 loeng3 ci4 dik1 jyu5 faat3 dak6 dim2" in *Zung1gwok3 Jyu5man4* Jan. 30-32.
- Erbaugh, Mary (1986) 'Taking Stock: The development of Chinese noun classifiers historically and in young children', in C. Craig (ed.) *Noun Classes and Categorization*. Philadelphia: John Benjamins, 399-435.
- Fang Fuxi (1985) "Si zhi liu sui er tong zhang wo han yu liang ci shui ping de shi yan yan jiu" in *Xinlixuebao*, vol. 4. 384-391.
- Gandour, J, S. Petty, R. Dardarananda, S. Dechongkit, and S. Mukngoen. (1984) 'The acquisition of numeral classifiers in Thai', in *Linguistics* 22: 455-479.
- Li, Lai-ha (1968) *The Development of the Syntax of a Cantonese-speaking child from 1 1/2 to 2 1/2 years of age*. M.A. dissertation. University of Malaya.
- Loke, Kit Ken & Godfrey Harrison (1986) 'Young Children's use of Chinese (Cantonese

and Mandarin) sortal classifiers', in Henry S.R. Kao and Rumjahn Hoosain (eds.), *Linguistics, Psychology, and the Chinese Language*. University of Hong Kong. 125-146.

Loke, Kit Ken (1991) 'A semantic analysis of young children's use of Mandarin shape classifiers', in *Child Language Development in Singapore and Malaysia*. Anna Kwan-Terry (ed.) Singapore University Press. 98-116.

Lyons, John (1977) *Semantics* Vol. 2. Cambridge: Cambridge University Press.

Mak, D. (1991) *The Acquisition of Classifiers in Cantonese*. Doctoral dissertation. University of Reading.

Matsumoto, Y. (1987) 'Order of acquisition in the Lexicon: Implications from Japanese numeral classifiers', in *Children's Language* vol.6, Hillsdale, NJ: Lawrence Erlbaum 229-259.

Matthews, S. & V. Yip (1994) *Cantonese: A Comprehensive Grammar*. London, NY: Routledge.

Poon, E. (1980) *Some aspects of the ontological development of nominal classifiers in Cantonese*. M.A. Dissertation, University of Hong Kong.

Sanches, M. (1977) 'Language acquisition and language change: Japanese numeral classifiers' in M. Sanches & B. Blount (eds.), *Sociocultural dimensions of language change*. New York: Academic Press. 51-62.

Ying H, G. Chen, Z. Song, W. Shao, Y. Guo (1987) "Si xhi qi sui er tong zhang wo liang ci de te dian", in Zhu M. (ed.) *Ertong Yuyan Fazhan Yanjiu*, East China Normal University. 91-103.

Notes

¹ The age is indicated as year;month;day.

² Abbreviations: Det - determiner; Num - numeral; CL - classifier; N - noun; NP - noun phrase; Q - quantifier; Wh - Wh-word; PRT - post-verbal particle; ASP - aspectual marker; GEN - genitive marker; SFP - sentence final particle

³ The romanization follows the Cantonese Romanization Scheme of the Linguistic Society of Hong Kong 1994.

⁴ This calculation may be rather loose in the sense that all the ambiguous cases of [Det - CL - N]_{NP} are included in the calculation. An NP in the form [Det - CL - N] could be interpreted as either [Det - CL - N]_{NP} or [Det - CL]_{NP} [N]_{NP} with a missing coupla in between. For example, an independent utterance of *ni1 zek3 gau2* 'this CL dog' can mean *ni1 zek3 (hai6) gau2* as a response to a previous question *ni1 zek3 hai6 mat1 je2* 'what's this'. If it is a response to the question *bin1 zek3 gau2 ngaau5 nei5* 'which dog bit you?', it should be analyzed as [Det - CL - N]_{NP}.

⁵ In considering inappropriate classifier choices, only utterances with explicit head nouns are considered.

The Development of Noun Phrase Structure in Cantonese-speaking Children Around Age Two

Cathy S.P. Wong

I. Introduction

The acquisition of a first language appears so natural and effortless that sometimes we take it for granted. Not until when we scrutinize each and every step that a small child takes to develop his language skill do we start to understand the complexity involved in the process of language learning and thus appreciate the endowment of our species-specific ability of to learn a native tongue.

Brown (1973) is one of the major works which provide us with a better understanding of the syntactic development of very young English-speaking children. In the following decades, the various aspects of English syntax have become the focus of first language acquisition. At the same time, the need to study languages other than English prevails as the idea that human beings are born with a mechanism which makes learning a first language possible gains support from both linguistic theories and first language acquisition findings.

It is undeniable that in spite of cross-linguistic differences, all human beings possess the ability to acquire a first language. However, linguistic variations do exist and not every individual learns the same language or in the same way. Only when we see repeated patterns in more than one individual and across languages can we claim that a universal mechanism is at work. In this sense, the findings of the acquisition of Chinese may provide crucial information which is not found in English or other European languages. Studies on the acquisition of Chinese are not scarce (Erbaugh 1982, 1986; Hu 1993; Lee 1981, 1986; Li 1989; Tse et al 1991; just to name a few). However, most of these studies (except a few) focus on a particular aspect of Chinese syntax and there has not been as much established findings about the general development of Chinese-speaking children. Furthermore, the studies on dialects like Cantonese are even fewer.

Compared with the findings of English-speaking children, relatively less is known about the acquisition of Chinese, a language so very different from English or most Indo-European languages. Chinese is an isolating and monosyllabic language (Cantonese being more so than Mandarin Chinese in both respects). It is a language which has lexical tones; it is basically monomorphemic; it consists of very few inflections: there is no case, gender, person or number inflection for nouns, nor is there any tense marking or person/number agreement on the verbs; the *wh*-words in questions remain in the same positions as they are in declarative sentences; it is a *pro*-drop language and so on. In view of the differences between the English and Chinese grammar, the issues being investigated and the results from studying children language in Chinese will definitely be very unlike those in English. For

example, instead of exploring when the past tense morphemes or subject-auxiliary inversion are acquired, researchers studying the acquisition of Chinese ponder over the issue when classifiers or sentence final particles emerge. Or while researchers are trying very hard to explain why English-speaking children omit subject noun phrases (which is not permitted in adult English) in the early stage of development, they should also be concerned why Chinese-speaking children should use subject noun phrases when they can be omitted without violating any adult grammar rules.

In short, the typological differences which Chinese manifests result in a different set of questions asked by child language researchers. At the same time, however, patterns found repeatedly across languages will help investigate the nature of the universal language acquisition mechanism. The present paper intends to examine the development of noun phrases of Cantonese-speaking children in the light of the above seemingly opposing but actually complementary forces.

In the sections to follow, I will first give a general description of Cantonese noun phrase structure in Section II and then point out the major characteristics of Cantonese noun phrases in Section III. Then a brief summary of previous studies on Chinese will be presented in Section IV, which is followed by the methodology described in Section V. The results and findings of the present study are given in Section VI. Section VII consists of the discussion of some of the findings. A conclusion is given in the final section, Section VIII.

II. Noun phrases in adult Cantonese

Cantonese noun phrases are head-final. In other words, the head noun always occurs at the right-most position in the phrase while all other elements occur in pre-nominal positions in a very strict order (cf. Matthews & Yip 1994, Chapter 6) as shown below:

[determiner -- numeral -- classifier -- head noun]

The adjectival phrase and relative clause, however, can occur either immediately before the head noun or right at the beginning of the noun phrase, that is. Usually, the nominalizer "ge3" is used when the head noun is modified by an adjectival phrase or relative clause.

(A) Bare Nouns or Proper Names As NPs

In its simplest form, a noun phrase consists only of a common noun or a proper name:¹

1

All romanizations of Cantonese in this paper follow the proposed Cantonese romanization scheme of the Linguistic Society of Hong Kong (1994).

- (1) cel lei4-gan2
 car come-ASP
 "The car is coming."
- (2) Ma3lei6 maai3-zo2 cel
 Mary buy-ASP car
 "Mary bought a car."

One interesting fact about Cantonese (and Mandarin as well) is that a bare noun occurring in subject position usually has definite reference (as in (1)) while definiteness is not always encoded in the NP when it is in the direct object position (as in (2)) (cf. Chen 1987: 88 on Mandarin).

(B) Head nouns modified by determiners, numerals and classifiers

When a determiner or a numeral is used, a classifier is obligatory. The possible types of structure including the head noun, classifier, numeral and determiner are as follows:

- (i) [classifier -- head noun]
 (3) gaa3 cel
 CL car "a/the car"
- (ii) [numeral -- classifier -- head noun]
 (4) jat1/saam1 gaa3 cel (cf. * jat1/saam1 cel)
 one/three CL car "a/three car/s"
- (iii) [determiner -- classifier -- head noun]
 (5) nei1 gaa3 cel (cf. * nei1 cel)
 this CL car "this car"
- (iv) [determiner -- numeral -- classifier -- head noun]
 (6) nei1 jat1 gaa3 cel (cf. * nei1 jat1 cel)
 this one CL car "this car"

The examples (3) through (6) clearly illustrate that the classifier is obligatory when a determiner or a numeral is used. The examples in (7) below show that the order of the determiner, numeral, classifier and head noun is very rigid.

(7) Word order in a Cantonese noun phrase: [determiner--numeral--classifier--noun]

- (a) nei1 jat1 gaa3 cel
 this one CL car "this car"
- (b) * nei1 gaa3 jat1 cel [determiner-classifier-numeral-noun]
 this CL one car
- (c) * jat1 nei1 gaa3 cel [numeral-determiner-classifier-noun]
 one this CL car
- (d) * jat1 gaa3 nei1 cel [numeral-classifier-determiner-noun]
 one CL this car

- (e) * gaa3 neil jat1 cel [classifier-determiner-numeral-noun]
 CL this one car
- (f) * gaa3 jat1 neil cel [classifier-numeral-determiner-noun]
 CL one this car

(C) Head nouns modified by adjectival phrases or relative clauses

A third type of structure involves a head noun modified by an adjectival phrase or a relative clause. As mentioned earlier, when an adjective phrase or relative clause modifies a head noun, the nominalizer "ge3" marker is normally used.

- (8) [hou2 leng3]_{AP} ge3 cel
 very beautiful POSS car
 "very beautiful cars"
- (9) [ngo5 kam4jat6⁶ maai3]_S ge3 syul
 I yesterday buy GEN book
 "books (which) I bought yesterday"

III. Four major characteristics of Cantonese noun phrases

Having introduced the major elements and basic structure of Cantonese noun phrases, I would like to discuss four major syntactic features of Cantonese noun phrases in the following paragraphs, which include: (a) the use of classifiers, (b) the count / mass non-distinction, (c) the pro-drop phenomenon, and (d) the possessive marker / nominalizer "ge3".

(A) Classifiers

One very noticeable feature in Cantonese, of course, is the obligatory use of the classifier with the determiner or numeral in Chinese. In English, it is grammatical to say "*that book*" or "*three books*". However, when that is translated into Chinese, each of those phrases have to contain a classifier denoting "printed matter":

- (10) go2 / saam1 bun2 syul
 that / three CL book "that / three book(s)"

In effect, all mentioning of the head nouns involves the use of a classifier one way or another, except when generic reference is intended. To complicate matters, classifiers fall into sub-types. The specific ones tend to be more arbitrary than the mensural ones in the sense that the specific classifiers are really specific. For example, the classifier "zoeng1" may seem to denote "flatness", such as tables, beds, pieces of paper and so on. However, not all flat objects can be used with "zoeng1". Books, for instance, are flat objects but they have a different classifier "bun2", indicating probably printed matter; hence, magazines can co-occur with "bun2". Newspaper, on the other hand, does not take "bun2" but "fan6" or

"zoengl". In any case, studies have shown that Mandarin-speaking children start off with the default or general classifier "ge3" (Erbaugh 1982, 1986; Tse et al 1991). Furthermore, there are some conceptual preferences, such as shape, involved in the acquisition of classifiers in Mandarin (Hu 1993) and Cantonese (Mak 1993).

Nevertheless, none of the studies have reported any findings on the syntactic development concerning classifiers. In other words, we still do not know whether Chinese children use the classifier with the determiner or the numeral first, with or without the head noun, in subject or object position first etc. Hopefully, this paper is going to provide some information to the syntactic development of classifiers in Cantonese-speaking children.

(B) No grammatical distinction between count / mass nouns

Another major difference between English and Cantonese is that the conceptual distinction of count / mass nouns is marked grammatically in English but not in Cantonese. This does not mean that Cantonese does not make such a distinction conceptually; it only refers to the fact that grammatically, one cannot distinguish count nouns from mass in Cantonese merely on the basis of morphological or syntactic cues.

In English, the count / mass distinction is explicitly encoded both morphologically and syntactically in the grammar. First, only count nouns can take the indefinite article "a" or "an" (e.g. *a book*, *an orange*; but *some bread*). Second, count nouns in English have two forms: singular and plural (e.g. *book* / *books*, *child* / *children*) while mass nouns have only one (*bread*). Third, there is agreement requirement in the English grammar which relies heavily on the count / mass distinction of the head noun. If the subject noun phrase of a sentence is a mass noun or singular count noun, the main verb has to be in its third person singular form. If the head noun of a noun phrase is a plural count noun, it has to take a plural determiner "*these* / *those*". In short, the count / mass distinction is an important part of the noun phrase in English grammar.

On the contrary, there is nothing to mark such a distinction in Chinese grammatically. First, there is no morphological marking for number: there is only one form for every noun and determiner. Second, both count and mass nouns occur in exactly the same syntactic environment in Cantonese. They can occur as bare nouns (while in English, singular count nouns cannot occur as bare nouns; it has to take an indefinite article or be in the plural form). Both count and mass nouns have to take a classifier when co-occur with a determiner or a numeral or both. In short, the grammatical structure of Cantonese noun phrases does not reflect a count / mass distinction at all.

Based on findings that English-speaking children are sensitive to the syntactic differences of count nouns as opposed to mass nouns and proper names, Bloom (1994) discusses the possible ways that English syntax helps children learn the meaning of novel nouns. Bloom & Keleman (1995: 6) proposes a set of syntax-semantics mappings which involves the count / mass distinction. According to their proposal, count nouns are supposed to map onto kinds of individuals, mass nouns refer to kinds of stuff and pronouns as well as proper names correspond to individuals. The question then is: What if a language does not have a count / mass distinction?

As pointed out in the previous paragraphs, Cantonese is a language which does not make a count / mass distinction grammatically (as noted by Bloom (1994: 322), too), and Cantonese children are able to learn new nouns and concepts of count, mass and everything. Therefore, it will be quite important to understand how the first nouns of Cantonese-speaking children are learned and developed. Are they sensitive to any syntactic cues? What kind of syntax-semantics mapping, if any at all, can we infer from the acquisition data from Cantonese?

(C) Pro-drop

A third characteristic is that Chinese is a "pro-drop" language. The noun phrases in Chinese can freely be omitted if enough discourse information is available, something which is not allowed in adult English. In first language acquisition, subject-drop is a widespread phenomenon across languages, even in non-pro-drop languages like English. The "null subject" issue in early child language has raised enormous interest in the field of first language acquisition research for linguists employing the Government and Binding approach (see Hyams 1994 and Rizzi 1994 for recent development) and those who do not (see O'Grady, Peters & Masterson 1989; Valian 1991, for example). The detail of the theories and discussions on the pro-drop issue are far beyond the scope of the present paper. One thing for sure is that regardless of the target first language, children all go through a period which they omit the subject NP. Moreover, a subject / object asymmetry exists in the pro-drop phenomenon (see for example, Wang, Lillo-Martin, Best & Levitt 1992).

Though a lot is known about dropping subjects in English-speaking children, relatively little is explored in Chinese from longitudinal naturalistic data. Hopefully, the findings from this paper will provide more comparable data, if not more insights, for the issue of omitted subjects in early children speech.

(D) The possessive marker "ge3" and possessive constructions in Cantonese

The possessive marker "ge3" in Cantonese is very similar to the English possessive marker "-s" in some ways. First, in both languages, the possessive marker is attached to the possessor and occurs in almost exactly the same syntactic

environment: [N-'s/-ge3 N]. For example, "Mary's car" in English will be equivalent to "Ma3lei6-ge3 ce1" (Mary-POSS car) in Cantonese. Second, the possessive marker is a grammatical morpheme in both languages. In other words, it is a morpheme which carries less transparent meaning and which are found to be acquired later than content words such as nouns and verbs in English (Brown 1973). Therefore, it will be interesting to see if such a grammatical morpheme also emerges late in the acquisition of languages other than English.

The major difference between the English and Cantonese possessive marker is that "ge3" in Cantonese has a much wider distribution: besides noun or proper names, it can be attached to pronouns, adjectival phrases and relative clauses as well (in which case it is sometimes called a nominalizer), which is not possible in English. In English, pronouns have the genitive form such as "my", "your", "their" etc. and the forming of relative clauses usually requires the use of an appropriate relative pronoun. The possessive marker "-s" has a very restricted distribution in English, probably only if the possessor is animated.

In Cantonese, using the possessive marker "ge3" is only one of the many ways of expressing possession. The following four are the options available (cf. Matthews & Yip 1994: 107-8):

- (i) with the possessive marker "ge3"
 [N / Pron -- ge3 -- N]
 (12) hok6saang1 / ngo5 ge3 syul
 student / I POSS book
 "(the) student's' / my books"
- (ii) with a classifier (and an optional determiner)
 [N / Pron -- (Det) -- CL -- N]
 (13) hok6saang1 / ngo5 (go2) bun2 syul
 student / I (that) CL book
 "that book of the student / mine"
- (iv) without "ge3" or classifier (used exclusively when the possessor is a pronoun and there is an inalienable relation between the pronoun and the head noun)
 [Pron -- N]
 (14) nog5 pang4jau5
 I teacher
 "my teacher"

From the findings in English, children tend to use the [N -- N] construction to indicate possession in the earliest stage of development (Brown 1973: 195-7). Therefore, given the similarity between the possessive marker in English and Cantonese, it is not surprising to find that the same phenomenon appears in children learning Cantonese. On the other hand, it will be interesting to see how the two groups of children diverge as their first language develops towards the adult form.

IV . What we know about the acquisition of Chinese

Mary Erbaugh's (1982) doctoral dissertation is the first and so far the most detailed longitudinal study on learning Chinese as a first language. She studied four Mandarin-speaking children aged 1;09 to 3:09 in Taiwan. As far as the noun phrase is concerned, Erbaugh (1982) makes the following generalizations about her Mandarin-speaking subjects.

First, most nouns produced are names for concrete objects. Second, personal pronouns are produced around aged 2;0, the first two being "I" and "you". The plural forms of these pronouns emerge after two and a half years old. The reflexive pronouns do not appear until after 2;0 for one child and 3;0 for another. Third, classifiers begin to emerge around the age of two, with the general classifier "ge" as the earliest to emerge. Fourth, the nominalizer *-de* (comparable to "ge3" in Cantonese) which is also used for marking possession, emerges as early as 1;11.

Tse et al (1991) is a more recent longitudinal study of 21 young children growing up in Taiwan aged from 0;07 to 6;00. Since the project covers all areas of phonology, morphology, syntax and pragmatic-discoursal functions, the part on syntax is brief but concise. Findings about noun phrases include the following points.

Nouns begin to emerge around 11 months, followed by a spurt at 13 months. Early nouns are typically monosyllabic or reduplicated forms denoting either kinship terms or concrete objects. Personal pronouns start to emerge at age 1;07. The pronoun for "you" and "I" emerge before "he/she". However, pronouns are found to be used more frequently only after age 2;01. Second, the demonstrative *zhe4ge* "this one" is found used productively around 1;03. Third, children start to use classifiers at around 1;07, mainly the general classifier *ge*. The [numeral-classifier-noun] pattern is found in 1;08. A wider variety of classifiers is found only after 1;10. However, children tend to retain the use of the generic classifier *ge* even though they have acquired other more specific ones. Finally, the use of possessive marker *"de"* is found to emerge around 1;10. It is used to mark possession, adjectival modification, nominalization, relative clause, in addition to being used as a sentence final particle.

There are no parallel findings reported on Cantonese yet since most studies on Cantonese focus on a specific area of syntax and are mainly experimental (Cheung 1991 on locative construction, Cheung & Lee 1993 on wh-words, Mak 1993 on classifiers, Man 1993 on null arguments and Wong 1990 on anaphora). Systematic longitudinal data on Cantonese is almost non-existent. One of the purposes of the present project is to fill this gap so that more comparable longitudinal data will be available to the field of first language acquisition.

V. Methodology

(A) Subject

The longitudinal data on which the analyses of this paper is based was from four youngest of the eight Cantonese-speaking children involved in this project, whose speech was tape recorded over a period of one year. All the children were born and raised in Hong Kong, speaking only Cantonese as their first language. The following table sums up the details of the four children.

Table 1 Details of Subjects

SUBJECT	AGE	SEX	SESSIONS	MLU RANGE
Tsuntsun (CKT)	1;05;22 -- 2;07;02	Male	25	1.05 -- 2.23
Bernard (MHZ)	1;07;00 -- 2;08;06	Male	26	0.24 -- 2.22
Chunyat (CCC)	1;10;08 -- 2;10;27	Male	22	1.54 -- 2.95
Gakei (CGK)	1;11;01 -- 2;09;09	Male	19	1.86 -- 3.36

(B) Procedure

Naturalistic data were collected from all four children by three different researcher assistants respectively. Alice worked with Tsuntsun and Bernard, Kitty with Chunyat and Patricia with Gakei. Hour-long audio tape recordings were made approximately bi-weekly by each of the research assistants interacting with the child in naturalistic settings, usually at the child's home. One of the adults of the family was always present when the recording was made. After each session, the audio tape was transcribed in handwritten Cantonese orthography by the research assistant who had conducted the recording as soon as possible. The handwritten transcriptions were then typed and computerized in CHAT format (MacWhinney 1991). Each of these files were later automatically tagged (i.e. given syntactic category labels) by a tagging program (Lee & Liu 1994).

(C) Data

Data for this paper was taken from transcribed sessions of all four children. In other words, from a total of 92 files. In order to examine the children's grammatical development, a more conservative approach was taken. Only spontaneous utterances of the children were used for analysis. Both exact imitations of adult speech and self repetitions were eliminated for this study. However, partial repetitions consisting of novel noun phrases were included. Any utterance of the child which had appeared in the previous two turns was regarded as repetition. Also, since this paper focuses on noun phrase structure, only those utterances containing nominal elements were extracted for analysis. In other words, utterances involving the following nouns, proper names, pronouns (including reflexive pronouns), determiners, classifiers, numerals, the possessive marker or nominalizer

"ge3" and wh-words were being analyzed for the development of noun phrase structure.

(D) Analysis

Since each lexical item in the speech was tagged with a syntactic label, each of the spontaneous utterances extracted from the 92 sessions was then grouped according to the four major nominal elements this paper intends to examine: (1) bare nouns, (2) pronouns, (3) classifiers and (4) the possessive marker / nominalizer "ge3". In addition to this, each of these four groups was further analyzed in terms of their relative position with respect to the main verb and the internal structure of the noun phrases itself. The sub-divisions were found to be quite useful in tracing the developmental pattern of each child especially when the children's language skills became more sophisticated. For example, when the child was at the one word stage, each of the utterances could simply be a bare noun, or a determiner plus a classifier. Once a child began to enter the multi-word stage, all sorts of nominal combination flourished. We found noun phrases containing a combination of [determiner -- classifier -- noun], or [pronoun -- classifier -- noun], or just [numeral -- classifier] or [determiner -- classifier]. Therefore, such grouping and sub-division help to plot the development of various nominal elements not just in terms of a certain category but also how a particular category interacts with another. A few utterances are taken from the corpus to illustrate how the utterances were analyzed, as shown below:

- (15) from Chunyat, session 1, aged 1;10;08
 *CHI: ne1 di1 ce1 (main line)²
 det cl nn (tagged line)
 this CL car
 "these cars"
 NP type: [Det-Cl-N] (analysis)
- (16) from Chunyat, session 22, aged 2;10;27
 *CHI: dou3 nei5 ne1 zi1 bat1 (main line)
 v pron det cl nn (tagged line)
 reach you this CL pen
 "now it's your pen's turn"
 NP types: [V-[Pron-Det-Cl-N]]

After all the utterances had been analyzed in terms of their NP types, the number of each type was tallied (identical utterances were counted as one), tabulated and calculated in percentage.

2

The main line was originally in Cantonese characters. For the sake of readers who are not able to read Cantonese, it was romanized.

Some errors were found in these children. These ungrammatical utterances were noted and put in a separate list for later analysis.

VI. Results and findings

(A) Overall development of children's noun phrases

The graphs in Figures 1 through 4 give a summary of the development of the four children of this study in terms of the total occurrences of noun phrases involving bare nouns, classifiers, pronouns and the possessive marker *ge3*. It is clearly shown that the two younger children use predominantly NPs involving bare nouns (an average of over 70% for both Tsuntsun and Bernard but only around 50% for Chunyat and Gakei). Also evident is the fact that, as the children grow older, the use of bare nouns decreases but the use of classifiers and pronouns is on the rise. Another interesting point is the individual differences demonstrated by Chunyat and Gakei in the use of pronouns and classifiers. While both of them are about the same age, Chunyat tends to use more classifiers than pronouns across all sessions (Figure 3). On the other hand, Gakei's use of pronouns increases gradually until it becomes dominant in the last session of her recording (Figure 4).

In addition to the diversity of the types of elements used as noun phrases, the types of structure pertaining to each of these different elements also increase. Figure 5 and Figure 6 illustrate this point. Since the possessive marker is very infrequent, it is excluded from this tabulation. These two figures show that the total number of types of NP structures used by Tsuntsun increases from 2 in the first session to 24 in the last one, from 3 to 23 for Bernard (Figure 5), from 7 to 31 for Chunyat and from 14 to 32 for Gakei (Figure 6).

The highest number of NP structure types in each child is 12 (last session of Tsuntsun), 14 (Session 22 of Bernard), 17 (Sessions 16, 19 and 29 of Chunyat) and 20 (Session 18 of Gakei) respectively. These numbers all point to the classifier-related structures. In all four children, the diversity of structure is most prominently shown in those related to the classifiers, which is not unlike adult speech. For Tsuntsun, no classifiers are used in the first three sessions but 12 different types of classifier-related structures are observed (Figure 5). A similar picture is found in Bernard, too. No classifiers are found in the second to the fifth sessions but a total of 14 types are being used in session 22, at the age of 2;05 (Figure 5). The older children Chunyat and Gakei doubtlessly show the same trend. Classifier-related structures grow from 3 to 17 for Chunyat and from 5 to 20 for Gakei (Figure 6).

Also shown in these two figures is the approximate time when each of these categories emerges. Let us assume that if a syntactic category, say the classifier, is found in a child's corpus in at least three different environment in the same session and this continues for at least three consecutive sessions, that syntactic category is

"acquired". Table 2 below sums up the developmental patterns of the three major syntactic categories in these four children.

Table 2 The Emergence of Nominal Categories			
	Bare Noun	Classifier	Pronoun
Tsuntsun	1;08	1;11	2;03
Bernard	1;08	1;09	--
Chunyat	1;10** 1;10**	2;08	
Gakei	1;11;01**	1;11;29	2;03

Note: ** indicates that the emergence may be earlier than the age specified but there is no way of verifying because that is the age when the child was first being recorded

Let us now proceed to take a look at the composition of the noun phrases of these children.

(B) The internal structure of children's noun phrase utterances

There are a maximum of 8 different types of structures related to nouns, 20 to classifiers and another 8 to pronouns found in the corpus of all four children. The simplest NP type consists of only one element, such as a proper name, while the most complex type involves a maximum of four elements, for example, a structure of [determiner--numeral--classifier--noun]. In this section, we will examine all the NP utterances only, that is, utterances which contain an NP only. In (C), NPs occurring with verbs will be investigated in greater detail.

B.1 Two-element NP utterances

I will start with the two-element NP, which consists of eight different types. For the purpose of identifying the emergence of a particular structure, I will assume that if a certain type of structure is used by any child continuously across three sessions, it is considered to be used productively by the child and the first session of that series will be regarded as the onset (this is indicated by the underlined numbers in these figures). Using this criterion, the order of emergence of all eight types of two-element NPs can be summarized in the following table, based on the data presented in Figures 7, 8, 9 and 10.

Table 3 The Emergence of Two-element NP Utterances

Subject	Order of Emergence of NP Types	Not Yet Emerged
Tsuntsun	[det-cl] > [adj-n] > [num-cl] / [wh-cl] > [n-n]	[pron-n], [cl-n], [cl-cl]
Bernard	[det-cl] > [n-n] > [adj-n] > [num-cl]	[wh-cl], [pron-n], [cl-n], [cl-cl]
Chunyat	[det-cl] / [n-n] > [num-cl] > [adj-n]	[wh-cl], [pron-n], [cl-n], [cl-cl]
Gakei	[n-n] > [det-cl] > [num-cl] > [pron-n]	[adj-n], [wh-cl], [cl-n], [cl-cl]

The [determiner--classifier] structure is found to be the first to emerge, at least for the majority of the children. Tsuntsun's first two-element NP emerges at around 1;11;27 (Session 12 in Figure 7) and Bernard's at around 1;09;04 (Session 5 in Figure 8). As for Chunyat, the [determiner--classifier] and [noun--noun] are used productively right from the start of recording, at the age of 1;10 (Figure 9). Similarly, Gakei shows a parallel pattern (Figure 10). However, her [determiner--classifier] structure does not satisfy our criterion of onset because even though it appears in the first session, it does not in the following one. Therefore, the onset of Gakei's [determiner--classifier] type has to be identified as two sessions after its first appearance, at the age of 1;11. Still, it is earlier than all other structures except the [noun--noun] NP.

On the other hand, the structures that none of the four children seem to have acquired are the [classifier--noun] and [classifier--classifier] NPs. Though there are some sporadic use of the [classifier--noun] structure by all four children, none of the occurrences satisfy our criterion. The [classifier--classifier] structure only appears twice in the one of the older children, Gakei's, corpus but never in the others'. It seems that the [classifier--classifier] is a very marked structure as compared to the rest.

B.2 Three-element NP utterances

There are only two types of NP structure that fall into the three-element category: the [determiner--classifier--noun] and [numeral--classifier--noun]. The fact that all tabulations (Figures 11, 12, 13 and 14) include these NP types occurring with verbs is due to the ambiguity of the [determiner--classifier--noun] structure.

On first glance, the [determiner--classifier--noun] sequence is a three-element NP. However, when children produce these NPs, they can be interpreted in two different ways. One is a single NP while the other is two NPs with an omitted linking verb, which is quite common even in adult speech. For example:

- (17) neil zek3 (hai6) maol
 this CL (be) cat
 "this cat / this is a cat"

In adult speech, the ambiguity can be resolved both by the length of pause between the classifier "go3" and the noun "maol" and the discorsal context. However, these cues are not reliably found in children speech. Therefore, in order not to impose too much grammatical knowledge on the part of the children, the [determiner--classifier--noun] structures are not treated as full three-element NPs until they are used with verbs.. When the [determiner--classifier--noun] NP is used

as either the subject or direct object of a verb. or an adjective, the ambiguity disappears:

- (18) neil zek3 mao1 hou2 leng3
 this CL cat very beautiful
 "this cat is very beautiful"
- (19) gin3-dou2 neil zek3 mao1
 see-RES this CL cat
 "(I) can see this cat"

In both of the above sentences, the [determiner--classifier--noun] NPs clearly cannot be interpreted otherwise. This is the reason why the tables containing three-element NPs include data involving the NP types occurring with verbs.

Now let us turn back to our data. Figure 11 shows that the youngest child Tsuntsun has not yet acquired any three-element NPs because none of the occurrences of three-element NPs satisfy our onset criterion. On the other hand, Bernard, who is just a little older than Tsuntsun, clearly produces some three-element NPs in the later sessions. Figure 12 illustrates that Bernard started to use [determiner--classifier--noun] utterances at around 2;01 (Session 15). However, as discussed in the previous paragraph, we cannot take these to be genuine three-element NPs. If we take a look at the columns where the [determiner--classifier--noun] NPs occur with a predicate, they appear at a much later stage, at 2;05 (Figure 12, Session 22). Similarly, the two older children also show the occurrence of [determiner--classifier--noun] without any predication before it is used with a predicate. The age of onset of Chunyat is 1;11 (Figure 13, Session 2) for non-predicated [determiner--classifier--noun] structure but 2;08 (Figure 13, Session 18) for predicated ones. Gakei shows a much earlier emergence of the predicated structure. The [determiner--classifier--noun] NP appears in both non-predicated and predicated forms at the same time, at age 2;02 (Figure 14, Session 6).

The other type of three-element NP structure seems to emerge at around the same time as the predicated [numeral--classifier--noun] NP. Bernard, began to use it in Session 24, at the age of 2;06 (Figure 12); Chunyat at 2;06 (Figure 13, Session 14) and Gakei at 2;08 (Figure 14, Session 17). However, none of them seem to be using the [numeral--classifier--noun] structure with any predicates productively yet, which they are for the [determiner--classifier--noun] NP.

The following table sums up the development of three-element NPs:

Table 4 The Emergence of 3-element NP Utterances

Subject	Types of 3-element Structures			
	[det-cl-n]	[det-cl-n]+V	V+[det-cl-n]	[num-cl-n]
Tsuntsun	--	--	--	--
Bernard	2;01	2;05	2;06	2;06
Chunyat	1;11	2;08	2;08	2;06
Gakei	2;02	2;02	2;03	2;08

To summarize this section on the internal structure of children's noun phrase, we find that some two-element NPs begin to emerge before 2;0 while all three-elements are used productively after age two. In addition, two-element NPs are more varied than the three-element ones.

(C) Noun phrase type with verb

In the paragraphs to follow, I will examine the development of three major nominal categories with respect to their relative position occurring with the verb. In other words, whether an NP is used as a subject or direct object of a verb is the issue of the section. The three nominal categories are those involving (1) bare nouns, (2) classifiers and (3) pronouns. Since any NP can occur in either the subject or object position or both, there are three different sub-types for each nominal category.

C.1 Bare noun

Figure 15 and Figure 16 present the data involving bare nouns occurring with verbs. It is found that all four children use bare nouns in both subject and object position productively. Gakei probably started using all three types of verb and bare noun combinations before she was first being recorded at the age of 1;11 (Figure 16, the last three columns). Therefore, there is no way of tracking the emergence of these three types of structures. The other three children, on the other hand, do give us a clear picture of the development of the use of bare nouns with verbs. The data from Tsuntsun, Bernard and Chunyat in Figures 15 and 16 illustrate that all three children follow the same path of using bare nouns first only as the direct object, then only as the subject and then both the subject and object simultaneously.

The data concerning Tsuntsun in Figure 15 shows that he started to produce [verb -- bare noun] structure at the age of 1;09 (the underlined number in column two in Figure 15). About two months later, at the age of almost two (1;11;27), bare nouns are found in the subject position and not long after that, at 2;0, a full-fledged sentence with bare nouns used as both the subject and object are produced.

Bernard shows an almost identical pattern (Figure 15, last three columns). The first bare noun direct objects show up at the age of 1;08, while the first bare

nouns subjects appear about a month later (1;09) and the [bare noun -- verb -- bare noun] complex structure emerges at 1;10. The same pattern is found in the older child Chunyat, too. His first bare noun direct objects appear at 1;10, first bare noun subjects at 1;11 and a verb with two bare noun arguments at around 2;02.

Not only does the developmental pattern of the children show a direct object preference, the frequency of the use of bare nouns in the two argument positions also demonstrates such a phenomenon, too. If we take a look at the percentage presented in Figures 15 and 16, we find that all four children use bare nouns more often as direct objects than subjects across all sessions. The average use of bare nouns (the bottom rows in Figures 15 and 16) in the object position is far greater than that in the subject position.

The following two tables summarize the two major findings discussed in the preceding paragraphs regarding the use of bare nouns in the subject and direct object position

Table 5a Bare Noun Related NPs

The Order of Development (Age of Onset)			
	V+BN	> BN+V	> BN+V+BN
Tsuntsun	1;09	1;11	2;00
Bernard	1;08	1;09	1;10
Chunyat	1;10	1;11	2;02
Gakei	1;11	1;11	1;11

Table 5b Bare Noun Related NPs

Frequency Count Average (% in terms of all Nominals)			
	V+BN	> BN+V	> BN+V+BN
Tsuntsun	17.99	6.67	2.51
Bernard	18.72	4.41	3.14
Chunyat	13.67	12.99	2.55
Gakei	19.45	7.69	3.67

In short, it is strongly evident that the preferred argument position for Cantonese-speaking young children to use bare nouns is the direct object position.

C.2 Classifiers

The next area we are going to examine concerns all NP types consisting of classifiers and whether they are used as subjects or objects. There are five different types of classifier-related NPs found in the corpus, namely, [determiner--classifier], [numeral--classifier], [classifier--noun], [determiner--classifier--noun] and [numeral--classifier--noun]. These NPs are then examined in terms of whether they occur

preverbally or post-verbally.³ Figures 17 through 20 are the detail of the results of Tsuntsun, Bernard, Chunyat and Gakei respectively.

Figure 17 illustrates that, among the five NP types mentioned above, Tsuntsun only uses three of them productively in argument positions: the [determiner-classifier] and [numeral--classifier] and [classifier--noun]. Among these three, only one occurs in the subject position. In other words, in addition to bare nouns, the only other NP structure which Tsuntsun uses as subjects is the [determiner--classifier] NP while a total of four (3 classifier NPs and the bare nouns) are used as direct objects. The frequency of occurrence of these NP types in Tsuntsun's corpus also points to the same phenomenon -- across all sessions, the use of classifier-related structures in the direct object position is more frequent than subject as demonstrated by the percentage in each session and the overall average in Figure 17.

Bernard shows a greater variety in the usage of classifier-related NP types (Figure 18). In addition to the three types used by Tsuntsun, Bernard also uses the three-element classifier NP [determiner--classifier--noun] in both the subject and object positions. Similar to the case in the use of bare nouns, the classifier-related NPs by Bernard emerge first and are used more frequently in the object position than the subject.

Figure 19 gives the detail of all classifier-related NPs used by Chunyat. The [classifier--noun] NP type is not used productively by Chunyat. The object / subject asymmetry is also observed in Chunyat for the classifier-related NPs. In the three NP types that he uses, the direct object position is the preferred one, just like Tsuntsun and Bernard. The three NP types, [determiner-classifier], [numeral--classifier], and [determiner--classifier--noun] are acquired first and are used more often as direct objects than subjects.

Gakei is no exception in preferring the object position. In addition, her NP types are more diverse than the rest. In fact, she is the only child who uses all five types of classifier-related NPs productively in either the subject or object position. Again, object NPs emerge earlier (with the exception of [determiner--classifier]) and are more frequently used than subject NPs.

Let me now summarize the findings on the development of classifier-related NPs in the following two tables.

³ These types should also be examined in terms of whether they occur in both the subject and object position at the same time, too, like the bare nouns. However, there are very few such cases and they are only sporadic, so they are not included in the tabulation here.

Table 6a Classifier-related NPs

Order of Development (Age of Onset)

NP Type	Tsuntsun	Bernard	Chunyat	Gakei
	Age	Age	Age	Age
[D+Cl] object	1;11	1;09	1;11;00	2;03
subject	2;01	1;10	1;11;21	2;02
[Num+Cl] object	2;04	2;05	2;05	2;08
subject	----	----	2;06	----
[Cl+N] object	2;03	2;03	----	2;03
subject	---	----	----	2;07
[D+Cl+N] object	----	2;06	2;06	2;03
subject	----	2;05	2;08	2;02
[Num+Cl+N] object	----	----	----	2;05
subject	----	----	----	----

Table 6b Classifier-related NPs

Frequency Count Average (% in terms of all Nominals)

NP Type	Tsuntsun	Bernard	Chunyat	Gakei
	Avg	Avg	Avg	Avg
[D+Cl] object	2.30	3.19	7.72	5.98
subject	1.23	2.09	8.90	4.38
[Num+Cl] object	0.36	0.35	0.74	0.53
subject	0.03	0.09	0.52	0.36
[Cl+N] object	0.38	0.62	0.10	3.64
subject	--	0.13	0.07	0.66
[D+Cl+N] object	0.17	0.37	0.46	0.84
subject	--	0.22	0.45	0.80
[Num+Cl+N] object	0.03	0.06	0.14	0.20
subject	--	--	--	0.02

In short, among all the NP types regarding classifiers, the [determiner--classifier] NP is the first to emerge, then followed by [numeral--classifier], [classifier--noun], [determiner--classifier--noun] and finally, [numeral--classifier--noun]. In most cases, the object position in each of these NP types is produced earlier and more frequently than the subject.

C.3 Pronoun

Contrary to the prevalent use of the bare noun and classifier NPs in the direct object position, the preferred argument position for pronoun use of all four children is the subject. This is illustrated in Figure 21 and Figure 22. The data

presented in these two figures show that the youngest child Tsuntsun started using pronouns as subjects first at the age of 1;11 and as objects at 2;01. Bernard started both at almost the same time -- 1;11 (subject use) and 2;01 (object use). The onset of Chunyat's pronoun use is a little later than the other three. For him, pronouns are used as subjects at 2;02 and as objects at 2;05. Gakei's case is a little difficult to decide because she already started using pronouns in both positions at the beginning of recording. However, her use of pronoun as object does not satisfy our onset criterion (i.e. a continuous production of that NP type over at least three consecutive recording sessions). Therefore, simply judging from the data we have collected, Gakei used pronouns as subjects first in the beginning of 1;11 while as objects later at the end of that same month.

Only the two older children (Chunyat and Gakei) show the productive use of pronouns in both positions within the same utterance, Chunyat at the age of 2;09 and Gakei at the age of 2;03.

The preference of using pronouns as subject is also illustrated by the frequency of use. The average of frequency (the last row in Figure 21 and Figure 22) of all four children clearly supports this observation. Except for a few sessions, the percentage of pronoun use in the subject position is higher than that in the object position. Moreover, the difference increases more drastically as the child grows older. As shown in Figure 22, Gakei, at the age of 2;09 (the last session), shows a contrast between 32.2 % (subject) and 5.8% (object). Chunyat's contrast at the age of 2;10 (the last session) is 15.9% (subject) as opposed to 2.7% (object).

The following is a summary of the pronoun development with regard to subject and object use.

Table 7a Pronoun Related NPs

The Order of Development (Age of Onset)			
	Pr+V >	V+Pr >	Pr+V+Pr
Tsuntsun	1;11	2;01	--
Bernard	1;11	2;01	--
Chunyat	2;02	2;05	2;09
Gakei	1;11;01	1;11;29	2;03

Table 7b Pronoun Related NPs

Frequency Count Average (% in terms of all Nominals)			
	Pr+V >	V+Pr >	Pr+V+Pr
Tsuntsun	2.87	1.85	0.18
Bernard	4.92	0.75	0.20
Chunyat	4.04	1.51	0.11
Gakei	19.13	4.88	1.26

From the findings presented above, it is clear that the use of pronouns in argument positions is greatly different from that of the other types of NPs.

(D) NPs expressing possession

As described in the section on adult Cantonese noun phrase structure, possession can be expressed in various ways, with or without the possessive marker. Six different types of NPs resembling adult possession construction are found in our data, namely, the [noun--noun], [pronoun--noun], [noun--determiner--classifier], [noun--classifier--noun], [pronoun--classifier--noun] NPs and those containing the possessive marker "-ge3". Nevertheless, the percentage of each of these NP types is quite low and they are used productively by only the two elder children. Furthermore, among these six, two of them can be considered as errors (the [noun--noun] and [pronoun--noun] types) because they are not used by adult speakers of Cantonese while two other involve ambiguity (the [noun--determiner--classifier] and [noun--classifier--noun] NPs, in which the first noun can be interpreted as addressing to a certain person instead of referring to the possessor). In effect, only two types among the six found in our corpus can be considered doubtlessly as expressing possession. They are the [pronoun--classifier--noun] and the one involving the possessive marker "-ge3". Because of all the factors mentioned above, it seems unwise to generalize too much based on the data we have so far.

The best we can say for sure now is that our data show that each of the two children who use any possessive expressions at all has chosen a different type: Chunyat started to use the possessive marker at around 2;05 (Figure 23, third column) while Gakei began to produce [pronoun--classifier--noun] as early as 2;03 (Figure 24, second column).

VII. Discussion

There are several issues to be addressed in this section. The first one has to do with the comparison between Mandarin and Cantonese children language findings. The second one is related to the preferred argument position of children's use of NPs. The last one concerns the development of NP types and the syntactic-semantic mapping in early children language acquisition.

(A) How different are Cantonese-speaking children from children learning other languages?

There are a lot similarities shared between English-speaking children and Cantonese-speaking children in the acquisition of their first language. One phenomenon repeatedly found across language is that content words are always learned before function words and Cantonese is no exception. In the case of Cantonese, grammatical morphemes such as the classifier and the possessive marker are found to emerge later than nouns. Another shared feature is that children of a

pro-drop language like Cantonese also prefer to use overt NPs often than in the object position, just like English-speaking children. A third similarity is that although Cantonese-speaking children have several options of expressing possession, the [N--N] structure is the one which they first rely on in spite of the fact that it is ungrammatical in adult speech. In this respect, Cantonese-speaking children resemble their English-speaking friends.

When compared with Mandarin-speaking children, Cantonese-speaking children tend to show a parallel developmental pattern in terms of the categories: nouns first, then classifiers, and later pronouns. Nevertheless, these categories seem to emerge later in Cantonese-speaking children than their Mandarin-speaking counterparts. This may be due to the more conservative method of analyses this paper has adopted. Or it may just be co-incidence that the four younger children of the present project develop these categories later. We need more data to confirm or reject any of the above speculations.

One noticeable difference lies in the emergence of the possessive marker ("ge3" in Cantonese and "de" in Mandarin). Both Erbaugh (1982) and Tse et al (1991) report that "-de" is used quite early to mark possession: 1;11 and 1;10 respectively. In our Cantonese data, however, the possessive marker "-ge3" does not emerge until after 2;03 (Gakei) or 2;05 (Chunyat).

Are there any major features in the Cantonese possessive marker which make it more difficult to learn? Among function words, why are some easier than others?

(B) Null subjects, too, in a pro-drop language?

While many researchers refer to the null subject phenomenon as dropping the subject, one can view the same fact as the subject being a more "marked" position than the object. What this implies is that, from a developmental point of view, we are claiming that if we find a child uses a certain construction, say relative clauses, as the subject, which is a "marked" position, it is then predicted that the same child should be using the same construction productively in the object position at the same time, if not earlier. From the data we presented in Section VI, this assumption that the subject is a marked position seems to hold pretty well (for bare nouns and classifiers) until we encounter the findings in the use of pronouns. This is really a puzzling fact to all linguists, from those who approach the pro-drop issue from a GB parametric point of view, or those who think subject use is related to tensed verbs, to those who attribute null subjects to performance constraints, or those who are purely looking at the developmental patterns.

In short, the findings on Cantonese-speaking children discussed in this paper now begs the question of why and how pronouns behave differently from other NPs. Is it some specific properties of Cantonese? It seems no studies, as far as I

know, have reported on similar findings and it is still an open question as to how universal or language specific this phenomenon is.

(C) What kind of syntactic-semantic mapping can possibly exist in cantonese?

The syntactic-semantic mapping proposal made by Bloom & Kelemen (1995) is based on a grammatical distinction between count and mass nouns in English and experimental studies showing that English-speaking children are sensitive to the syntax of these two groups of nouns. From our longitudinal data on Cantonese, there seems to be no evidence to either support or reject such a hypothesis. However, theoretically speaking, since Cantonese makes no distinction between count and mass nouns grammatically, if the mapped concepts of kinds of individual (count nouns) and kinds of stuff (mass nouns) are to be learned at all in Cantonese, they should not be based on a syntax-semantic mapping at all.

On the other hand, different semantic information is indeed encoded in different NP structures. For example, proper names and pronouns cannot have generic reference while most other structures can depending on the context. It is conceivable that in both English and Cantonese, children start off learning NP structures that make a minimal distinction between generic and non-generic reference instead of a kind of individual versus kind of stuff distinction.

To sum up, the claims of Bloom (1994) and Bloom & Kelemen (1995) cannot be disconfirmed by the Cantonese longitudinal acquisition data presented in this paper yet. This is definitely a very interesting issue. However, not until comparable experiments are conducted on Cantonese-speaking children will we be able to gain further insight into this issue.

VIII. Conclusion

It is really difficult to write a good conclusion for a paper which has tried to deal with so many issues at one time. Let me try to summarize this paper by reviewing what findings we have presented in this paper and what issues remain to be investigated as far as Cantonese noun phrases are concerned.

First, we now know that the four younger children in our project show a similar developmental pattern of the emergence of nominal categories -- bare nouns are produced first, followed by NPs involving classifiers and then pronouns. We also find that content words emerge earlier than function words and that the possessive marker is the latest to emerge among the nominal grammatical morphemes. Also reported about Cantonese acquisition data is that the internal structure of NPs used by children around two years old are predominantly two-elements and mostly classifier-related. In terms of the subject / object asymmetry issue, Cantonese-speaking children behave like children learning other languages-- they prefer the object position, except for the use of pronouns.

The findings lead to a number of interesting questions which require further investigation. One of them is why the use of pronouns by Cantonese-speaking children show a preference for the subject position. Another is the issue of syntax-semantics mapping in first language acquisition. What exactly are being mapped?

References

- Bloom, Paul. 1994. Possible names: The role of syntax-semantics mappings in the acquisition of nominals. *Lingua* 92, 297-329.
- Bloom, Paul and Kelemen, Deborah. 1995. Syntactic cues in the acquisition of collective nouns. *Cognition* 56, 1-30.
- Brown, R. 1973. *A first language: the early stages*. Cambridge, Mass.: Harvard University Press.
- Chen, Ping. 1987. On referentiality vs. nonreferentiality, identifiability vs. nonidentifiability, specificity vs. nonspecificity and genericity vs. individuality in Chinese. *Zhongguo Yuwen* 2, 81-92.
- Cheung, Alice Shuk-Yee & Lee, Thomas, Hun-Tak. 1993. Cantonese-speaking Children's Comprehension of Wh-Questions. In Eve Clark (ed.), *The Proceedings of the 25th Stanford Child Language Research Forum*. Stanford: Center for the Study of Language and Information.
- Cheung, Sik-Lee. 1991. *The acquisition of locative constructions in Cantonese-speaking children*. Doctoral dissertation, Stanford University.
- Erbaugh, Mary. 1982. *Coming to order: Natural selection and the origin of syntax in the Mandarin speaking child*. Michigan: UMI Dissertation Services.
- Erbaugh, Mary. 1986. Taking stock: The development of Chinese noun classifiers historically and in young children, in C. G. Craig (ed.), *Noun Classes and Categorization*, Amsterdam and Philadelphia: John Benjamins, pp. 399-436.
- Hu, Qian. 1993. *The Acquisition of Classifiers by Young Mandarin-speaking Children*, PhD dissertation, Boston University.
- Hyams, Nina. 1994. V2, Null Arguments and COMP Projections. In Teun Hoekstra & Bonnie D. Schwartz (Eds.), *Language Acquisition : Studies in Generative Grammar. Papers in honor of Kenneth Wexler from the 1991 GLOW Workshops*. Amsterdam: John Benjamins Publishing Co.

- Lee, Thomas Hun-Tak. 1981. *Acquisition of negation in a Mandarin-speaking child*. M.A. dissertation, University of Hong Kong.
- Lee, Thomas Hun-Tak. 1986. Acquisition of quantificational scope in Mandarin Chinese. *Papers and Reports on Child Language Development* 25, 57-64.
- Lee, Thomas Hun-Tak & Liu, Chi-Leung. 1994. *A Tagging program for Cantonese language acquisition data*. Hong Kong: Chinese University of Hong Kong.
- Li, Ping. 1989. *Aspect and aktionsart in child Mandarin*. Doctoral dissertation, University of Leiden.
- Linguistic Society of Hong Kong. 1994. *The LSHK Cantonese romanization scheme*. Hong Kong: Linguistic Society of Hong Kong.
- MacWhinney, Brian. 1991. *The CHILDES project: Tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum.
- Mak, David Lai-Woon. 1993. *The acquisition of classifiers in Cantonese*. Ph.D. dissertation. University of Reading.
- Man, Patricia Yuk-Hing. 1993. *Subject-object distinctions and null NPs in child Cantonese*. M.Phil. dissertation, Hong Kong Polytechnic.
- Matthews, Stephen & Yip, Virginia. 1994. *Cantonese: A comprehensive grammar*. London: Routledge.
- O'Grady, William, Peters, Ann & Masterson, Deborah. 1989. The transition from optional to required subjects. *Journal of Child Language* 16, 513-29.
- Rizzi, Luigi. 1994. Early Null Subjects and Root Null Subjects. In Teun Hoekstra & Bonnie D. Schwartz (Eds.), *Language Acquisition : Studies in Generative Grammar. Papers in honor of Kenneth Wexler from the 1991 GLOW Workshops*. Amsterdam: John Benjamins Publishing Co.
- Tse, Kwok-Ping, Tang, Ting-Chi, Shie, Yu-Huei. & Li, Cherry. 1991. *Chinese children's language acquisition and development*. Taipei: Department of English, Taiwan Normal University.
- Valian, Virginia. 1991. Syntactic subjects in the early speech of American and Italian children. *Cognition* 40, 21-81.
- Wong, Colleen. 1990. *The development of anaphora in Chinese listening and reading*. Doctoral dissertation, State University of New York at Albany.

Appendices

A. Notes to figures:

All numbers in the following figures represent the percentage of a particular structure with respect to the occurrence of the total number of nominal elements found in that particular session.

B. Keys

Age -- xxxxx: x (year) xx (month) xx (day)

e.g. 10522 should be read as "1;05;22), which means "one year five months and twenty-two days"

C. Abbreviations

A =	adjective
Adj =	adjective
AVG =	average
BareNm =	proper names
BareN =	bare noun
BN =	bare noun
Cl =	classifier
ClLoc =	classifier (locative)
D =	determiner
N =	bare noun
Num =	numeral
m =	numeral
Pr =	pronoun
Pron =	pronoun
Sess =	the nth session of recording
V =	verb
W =	wh-words
Wh =	wh-words

FIGURE 1
TSUNTSUN: SUMMARY OF NP TOTAL

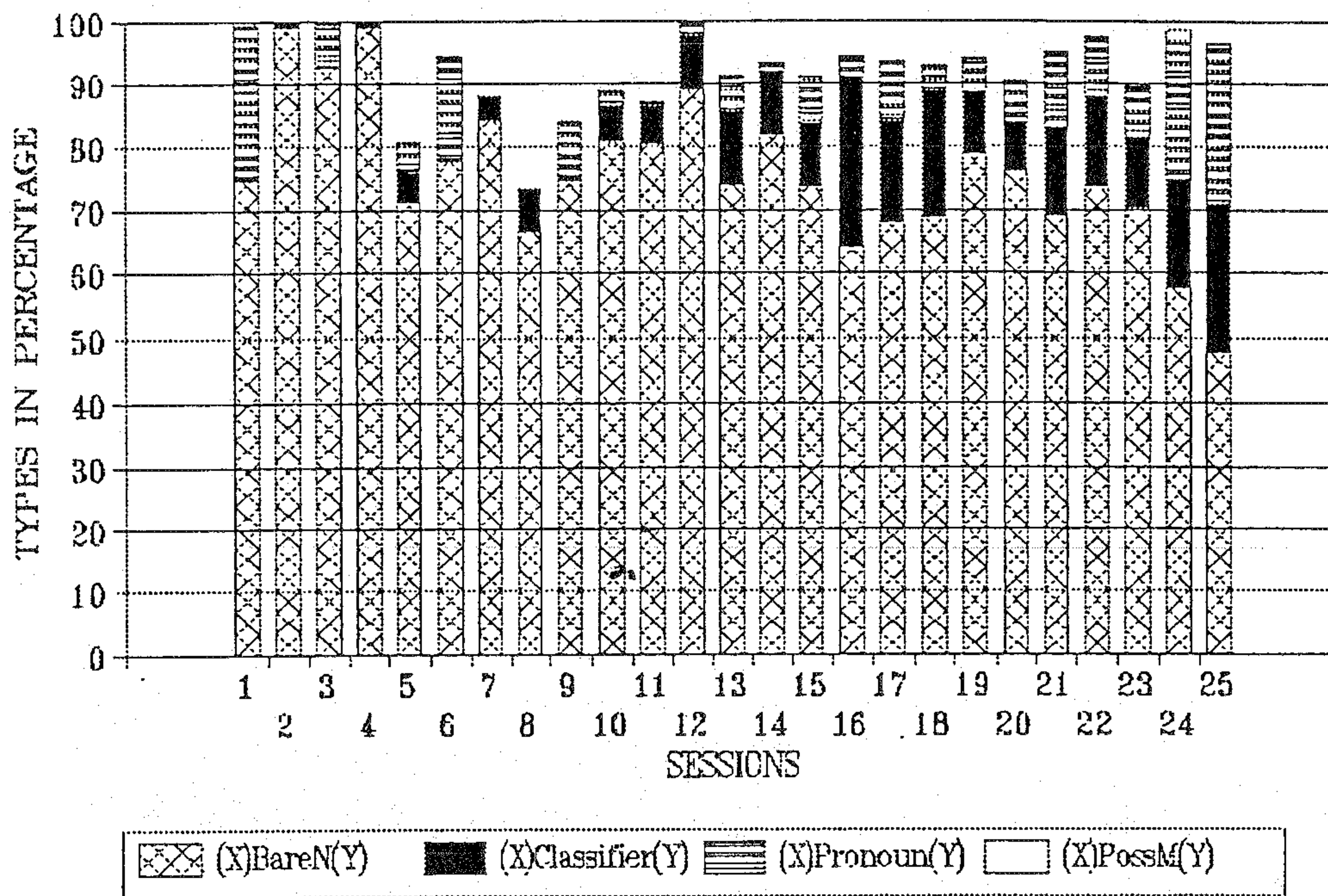


FIGURE 2
BERNARD: SUMMARY OF NP TOTAL

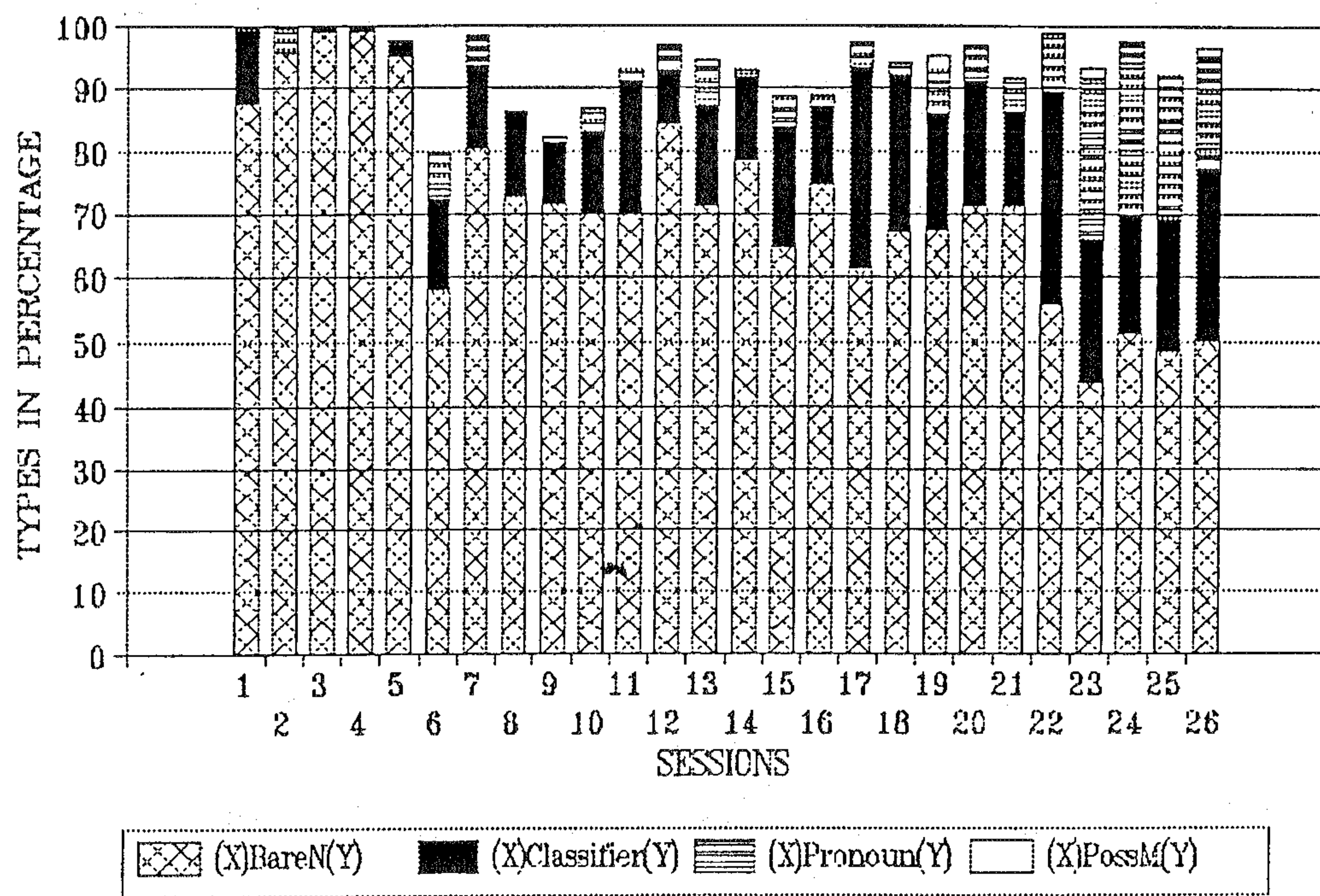


FIGURE 3
CHUNYAT: SUMMARY OF TOTAL

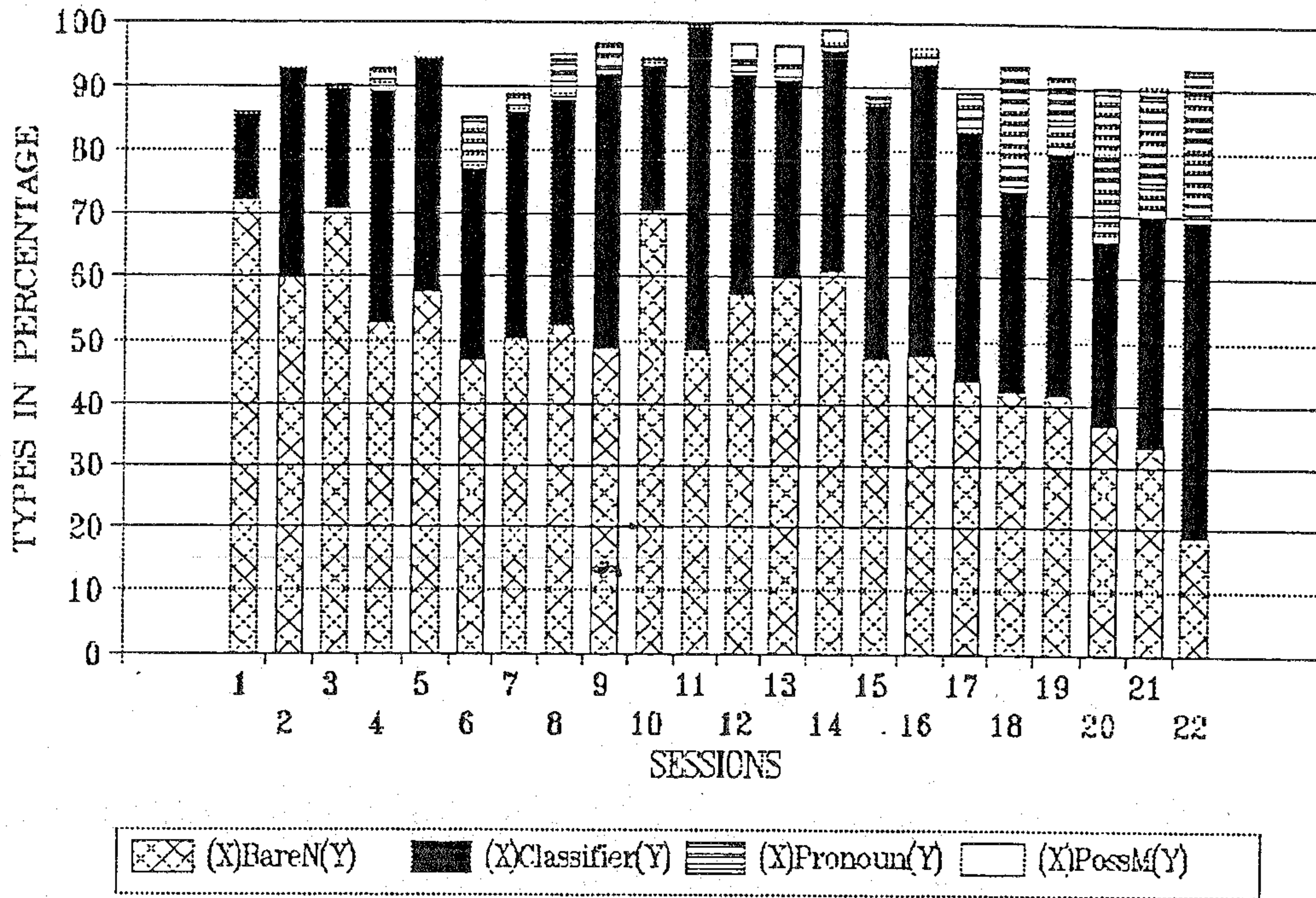


FIGURE 4
GAKEI: SUMMARY OF NP TOTAL

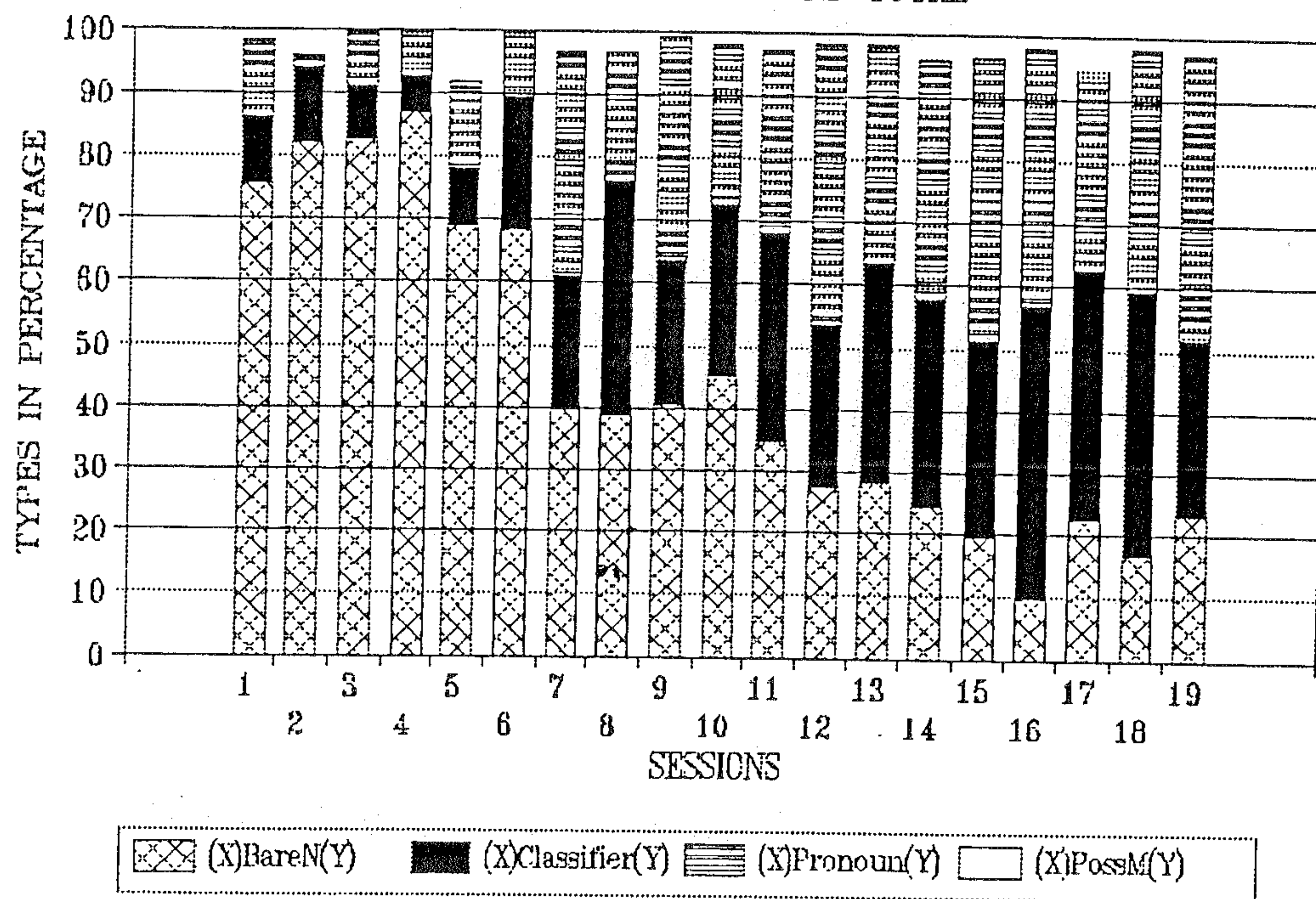


**Figure 5 Frequency Count of Structure Types
Tsuntsun & Bernard**

Tsuntsun						Bernard					
Age	Sess	BN	CI	Pron	Total	Age	Sess	BN	CI	Pron	Total
10522	1	1	—	1	2	10700	1	2	1	—	3
10703	2	2	—	—	2	10800	2	3	—	1	4
10710	3	2	—	1	3	10814	3	2	—	—	2
10800	4	3	—	—	3	10828	4	4	—	—	4
10807	5	3	1	1	5	10904	5	5	1	—	6
10821	6	3	—	2	5	10918	6	3	3	2	8
10907	7	3	1	—	4	10925	7	4	4	1	9
10914	8	3	1	—	4	11010	8	5	5	—	10
10929	9	3	—	2	5	11023	9	6	5	—	11
11030	10	3	2	1	6	11106	10	7	4	2	13
11113	11	4	2	—	6	20003	11	8	7	2	17
11127	12	5	3	1	9	20016	12	7	4	2	13
20009	13	4	5	2	11	20101	13	8	6	3	17
20016	14	5	3	1	9	20115	14	7	5	1	13
20108	15	6	6	3	15	20129	15	7	7	2	16
20205	16	6	7	2	15	20212	16	8	7	2	17
20215	17	6	9	6	21	20226	17	8	13	3	24
20303	18	6	7	2	15	20309	18	5	10	1	16
20317	19	6	6	3	15	20328	19	8	10	3	21
20400	20	7	6	4	17	20407	20	7	10	3	20
20414	21	6	8	6	20	20421	21	8	10	2	20
20500	22	7	8	4	19	20504	22	8	14	4	26
20514	23	6	8	3	17	20519	23	8	12	2	22
20618	24	6	9	4	19	20604	24	7	10	5	22
20702	25	6	12	6	24	20618	25	6	8	2	16
						20806	26	8	11	4	23

**Figure 6 Frequency Count of Structure Types
Chunyat & Gakei**

Chunyat						Gakei					
Age	Sess	BN	CI	Pron	Total	Age	Sess	BN	CI	Pron	Total
11008	1	4	3	—	7	11101	1	7	5	2	14
11100	2	6	5	—	11	11108	2	6	5	1	12
11121	3	5	5	—	10	11122	3	7	2	1	10
20110	4	8	10	2	20	11129	4	6	5	2	13
20117	5	7	7	—	14	20008	5	5	5	2	12
20206	6	7	9	2	18	20207	6	7	13	3	23
20213	7	5	9	3	17	20221	7	7	8	2	17
20307	8	7	8	3	18	20228	8	5	6	3	14
20323	9	5	6	1	12	20304	9	6	12	4	22
20410	10	7	8	2	17	20311	10	7	9	5	21
20507	11	7	8	1	16	20318	11	6	12	6	24
20523	12	7	13	3	23	20325	12	6	8	4	18
20608	13	8	12	3	23	20408	13	8	8	7	23
20624	14	8	14	2	24	20430	14	8	14	5	27
20706	15	8	13	1	22	20503	15	7	10	5	22
20713	16	8	17	2	27	20711	16	4	12	6	22
20800	17	7	9	2	18	20808	17	7	15	8	30
20817	18	8	9	5	22	20818	18	5	20	6	31
20907	19	7	17	5	29	20909	19	8	17	7	32
20923	20	7	17	5	29						
21013	21	8	13	7	28						
21027	22	8	16	7	31						

Figure 7 Tsuntsun's Two-element NPs

Age	Sess	A+N %	N+N %	Pr+N %	D+Cl %	m+Cl %	Cl+N %	Wh+Cl %
10522	1	--	--	--	--	--	--	--
10703	2	--	--	--	--	--	--	--
10710	3	--	--	--	--	--	--	--
10800	4	--	--	--	--	--	--	--
10807	5	--	--	--	--	--	--	4.8
10821	6	--	--	--	--	--	--	--
10907	7	--	--	--	--	--	--	--
10914	8	6.7	--	--	6.7	--	--	--
10929	9	--	--	--	--	--	--	--
11030	10	--	--	--	--	--	--	2.8
11113	11	--	--	--	--	--	--	--
11127	12	--	--	--	<u>4.4</u>	--	--	--
20009	13	--	2.9	--	4.3	--	--	1.4
20016	14	--	--	--	5.0	--	--	--
20108	15	<u>1.0</u>	4.9	--	3.9	--	1.0	--
20205	16	2.5	--	--	2.5	<u>1.7</u>	--	<u>0.8</u>
20215	17	0.6	<u>0.6</u>	--	2.6	0.6	0.6	0.6
20303	18	1.4	2.1	--	2.8	5.5	--	0.7
20317	19	0.6	1.3	--	2.5	--	--	0.6
20400	20	0.7	2.0	--	--	--	--	1.4
20414	21	0.6	3.0	--	2.4	2.4	0.6	--
20500	22	0.8	1.6	--	3.2	2.4	--	--
20514	23	0.7	--	--	2.1	0.7	--	--
20618	24	1.2	--	1.2	1.2	1.8	0.6	0.6
20702	25	0.8	0.8	0.4	0.8	0.4	0.4	--

Figure 8 Bernard's Two-element NPs

Age	Sess	A+N %	N+N %	Pr+N %	D+Cl %	m+Cl %	Cl+N %	W+Cl %
10700	1	--	--	--	12.5	--	--	--
10800	2	--	--	--	--	--	--	--
10814	3	--	--	--	--	--	--	--
10828	4	4.3	--	--	--	--	--	--
10904	5	--	2.4	--	<u>2.4</u>	--	--	--
10918	6	--	--	--	6.0	--	--	--
10925	7	--	--	--	3.3	3.3	--	--
11010	8	--	<u>1.5</u>	--	4.5	3.0	--	--
11023	9	--	2.9	--	1.0	--	--	--
11106	10	<u>1.3</u>	1.3	--	1.3	--	--	--
20003	11	2.5	0.8	--	3.4	2.5	--	--
20016	12	1.5	0.8	--	5.3	--	0.8	--
20101	13	0.9	1.9	--	2.8	1.9	--	--
20115	14	1.4	--	--	2.9	--	--	--
20129	15	1.0	--	--	4.0	--	--	--
20212	16	2.2	--	--	2.2	<u>2.2</u>	2.2	--
20226	17	5.3	--	--	4.7	5.3	0.7	--
20309	18	--	--	--	3.5	7.1	--	--
20328	19	0.5	0.5	--	2.1	3.1	0.5	--
20407	20	0.8	3.1	--	3.1	1.6	--	1.6
20421	21	5.0	0.8	--	1.7	1.7	--	--
20504	22	0.7	0.7	0.7	2.2	5.2	0.7	--
20519	23	3.8	0.6	--	3.2	1.9	--	1.3
20604	24	4.7	0.8	--	3.1	0.8	--	--
20618	25	0.9	--	--	1.8	4.4	--	--
20806	26	0.8	0.8	--	2.3	1.5	--	--

Figure 9 Chunyat's Two-element NPs

Age	Sess	A+N %	N+N %	Pr+N %	D+Cl %	m+Cl %	Cl+N %	Wh+Cl %
11008	1	--	<u>8.3</u>	--	<u>8.3</u>	--	--	--
11100	2	--	2.4	--	11.9	--	7.1	--
11121	3	--	2.4	--	4.9	--	4.9	--
20110	4	0.8	3.8	--	4.6	<u>1.5</u>	--	--
20117	5	1.4	1.4	--	6.8	1.4	--	--
20206	6	--	1.0	--	6.3	3.1	--	--
20213	7	--	--	0.8	5.0	0.8	--	--
20307	8	<u>4.7</u>	--	--	9.3	1.9	--	--
20323	9	2.3	--	--	2.3	4.7	--	--
20410	10	1.9	2.5	--	2.5	1.2	0.6	--
20507	11	3.2	--	--	2.7	3.8	0.5	--
20523	12	0.5	--	0.5	2.8	2.8	--	--
20608	13	1.8	1.8	--	4.7	1.8	--	--
20624	14	0.5	7.1	--	2.9	3.3	--	--
20706	15	0.5	0.5	--	3.0	2.5	--	--
20713	16	1.1	<u>3.4</u>	0.4	1.5	3.0	0.4	--
20800	17	--	1.4	--	1.9	--	--	--
20817	18	0.7	1.4	0.7	2.1	2.1	--	--
20907	19	--	0.9	0.4	2.2	2.7	--	<u>1.3</u>
20923	20	--	2.0	--	2.0	4.8	--	1.2
21013	21	1.0	1.4	1.4	1.7	--	--	1.4
21027	22	2.1	2.7	2.4	2.4	2.1	0.3	6.6

Figure 10 Gakei's Two-element NPs

Age	Sess	A+N %	N+N %	Pr+N %	D+Cl %	m+Cl %	Cl+N %	Cl+Cl %	W+Cl %
11101	1	1.3	<u>2.6</u>	--	2.6	--	--	--	--
11108	2	--	4.0	--	--	2.0	--	--	--
11122	3	--	4.3	--	<u>4.3</u>	--	--	--	--
11129	4	--	2.4	--	1.2	--	--	--	--
20008	5	--	--	--	1.3	--	--	--	--
20207	6	--	1.5	--	1.5	--	1.5	1.5	1.5
20221	7	<u>1.3</u>	0.6	--	3.2	--	--	--	--
20228	8	1.5	--	--	--	3.0	--	--	--
20304	9	--	--	--	2.5	0.5	0.5	--	--
20311	10	1.6	2.4	1.6	0.8	--	--	--	--
20318	11	0.8	0.8	0.8	0.8	0.8	--	--	--
20325	12	--	1.3	--	--	--	--	--	--
20408	13	0.6	1.2	0.6	1.2	<u>1.2</u>	1.8	--	--
20430	14	0.8	0.4	--	1.5	0.4	--	--	0.4
20503	15	--	2.0	<u>0.7</u>	--	0.7	--	--	--
20711	16	--	--	2.1	0.7	--	--	--	--
20808	17	1.0	--	0.5	1.5	--	1.0	--	0.5
20818	18	--	--	--	2.1	--	1.3	0.4	--
20909	19	1.2	0.4	--	1.7	0.4	--	--	0.4

Figure 11 Tsuntsun's Three-element NPs

Age	Sess	D+Cl+N %	m+D+Cl %	V+D+Cl+N %	D+Cl+N+V %	V+m+Cl+N %
10522	1	--	--	--	--	--
10703	2	--	--	--	--	--
10710	3	--	--	--	--	--
10800	4	--	--	--	--	--
10807	5	--	--	--	--	--
10821	6	--	--	--	--	--
10907	7	--	--	--	--	--
10914	8	--	--	--	--	--
10929	9	--	--	--	--	--
11030	10	--	--	--	--	--
11113	11	--	--	--	--	--
11127	12	--	--	--	--	--
20009	13	--	--	--	--	--
20016	14	--	--	--	--	--
20108	15	--	--	--	--	--
20205	16	--	--	2.5	--	--
20215	17	1.9	--	1.3	--	--
20303	18	--	--	--	--	--
20317	19	--	--	--	--	--
20400	20	0.7	--	--	--	--
20414	21	--	0.6	--	--	--
20500	22	--	0.8	--	--	--
20514	23	0.7	--	--	--	0.7
20618	24	--	--	--	--	--
20702	25	0.4	--	0.4	--	--

Figure 12 Bernard's Three-element NPs

Age	Sess	D+Cl+N %	m+Cl+N %	V+D+Cl+N %	D+Cl+N+V %	V+m+Cl+N %
10700	1	--	--	--	--	--
10800	2	--	--	--	--	--
10814	3	--	--	--	--	--
10828	4	--	--	--	--	--
10904	5	--	--	--	--	--
10918	6	--	--	--	--	--
10925	7	--	--	--	--	--
11010	8	--	--	--	--	--
11023	9	2.0	--	--	--	--
11106	10	--	--	1.3	--	--
20003	11	0.8	--	0.8	--	--
20016	12	--	--	--	--	--
20101	13	4.7	--	0.9	--	--
20115	14	--	--	--	--	--
20129	15	5.1	--	--	--	--
20212	16	2.2	1.1	--	--	--
20226	17	2.0	2.7	0.7	--	--
20309	18	1.2	--	--	--	--
20328	19	--	0.5	--	--	--
20407	20	4.7	--	--	0.8	--
20421	21	0.8	2.5	--	--	0.8
20504	22	0.7	4.5	0.7	0.7	0.7
20519	23	0.6	--	--	1.9	--
20604	24	1.6	2.4	0.8	2.4	--
20618	25	--	0.9	3.5	--	--
20806	26	3.8	2.3	0.8	--	--

Figure 13 Chunyat's Three-element NPs

Age	Sess	D+Cl+N %	m+Cl+N %	V+D+Cl+N %	D+Cl+N+V %	V+m+Cl+N %
11008	1	2.8	--	--	--	--
11100	2	--	--	--	--	--
11121	3	<u>2.4</u>	--	--	--	--
20110	4	7.7	0.8	--	0.8	--
20117	5	4.1	--	--	--	--
20206	6	--	--	--	--	--
20213	7	5.8	--	--	--	--
20307	8	0.9	--	--	--	--
20323	9	--	--	--	--	--
20410	10	--	--	--	--	--
20507	11	2.7	--	--	--	--
20523	12	2.8	--	--	--	--
20608	13	2.4	--	0.6	--	--
20624	14	1.9	<u>1.9</u>	1.9	1.0	--
20706	15	4.0	0.5	0.5	--	2.0
20713	16	1.1	2.6	--	0.7	0.7
20800	17	6.0	--	--	--	--
20817	18	0.7	0.7	<u>0.7</u>	<u>1.4</u>	--
20907	19	7.1	0.9	2.2	0.9	--
20923	20	5.2	0.8	0.8	0.4	--
21013	21	3.7	0.3	1.0	4.8	0.3
21027	22	5.7	1.5	2.4	--	--

Figure 14 Gakei's Three-element NPs

Age	Sess	D+Cl+N %	m+Cl+N %	V+D+Cl+N %	D+Cl+N+V %	m+Cl+N+V %	V+m+CL+N %
11101	1	--	--	--	--	--	--
11108	2	4.0	2.0	--	--	--	--
11122	3	--	--	--	--	--	--
11129	4	1.2	1.2	--	--	--	--
20008	5	--	--	--	--	--	--
20207	6	1.5	--	--	<u>1.5</u>	--	--
20221	7	3.2	0.6	--	0.6	--	--
20228	8	20.9	--	--	1.5	--	--
20304	9	1.5	0.5	--	--	--	--
20311	10	2.4	--	<u>0.8</u>	1.6	--	0.8
20318	11	2.5	--	0.8	3.3	--	--
20325	12	1.3	--	1.3	--	--	--
20408	13	4.3	--	--	--	--	--
20430	14	1.5	--	3.4	0.4	--	--
20503	15	1.4	--	--	2.7	--	0.7
20711	16	2.1	--	5.0	1.4	--	--
20808	17	1.5	<u>1.0</u>	2.6	--	--	1.5
20818	18	1.3	0.8	2.1	1.3	0.4	0.8
20909	19	3.3	0.8	--	0.8	--	--

**Figure 15 Bare Nouns with Verbs
Tsuntsun & Bernard**

Tsuntsun				Bernard			
Age	V+N %	N+V %	N+V+N %	Age	V+N %	N+V %	N+V+N %
10522	—	—	—	10700	—	—	—
10703	—	—	—	10800	4.5	—	—
10710	—	—	—	10814	—	—	—
10800	4.8	—	—	10828	<u>4.3</u>	—	—
10807	—	4.8	—	10904	17.1	2.4	—
10821	5.6	—	—	10918	14.0	—	—
10907	8.0	—	—	10925	32.8	3.3	—
10914	—	—	—	11010	22.7	4.5	—
10929	<u>18.8</u>	—	—	11023	25.5	2.0	<u>5.9</u>
11030	16.7	—	—	11106	25.0	9.2	13.2
11113	17.4	—	4.3	20003	21.8	5.9	6.7
11127	13.3	<u>26.7</u>	2.2	20016	38.3	6.0	7.5
20009	11.6	11.6	—	20101	14.0	9.3	9.3
20016	3—	15.0	<u>3.3</u>	20115	15.7	8.6	5.7
20108	30.4	10.8	2.9	20129	21.2	8.1	3.0
20205	29.2	7.5	5.0	20212	26.1	5.4	1.1
20215	28.2	10.3	1.3	20226	18.7	2.7	0.7
20303	35.2	10.3	1.4	20309	27.1	5.9	5.9
20317	35.0	10.8	5.7	20328	22.3	15.5	3.6
20400	28.4	12.8	6.1	20407	18.8	4.7	7.0
20414	28.6	15.5	8.3	20421	22.5	5.0	0.8
20500	31.7	7.9	6.3	20504	18.7	3.7	6.0
20514	34.0	10.6	5.0	20519	11.5	2.6	1.3
20618	24.1	8.2	5.9	20604	21.3	2.4	0.8
20702	18.9	6.2	4.9	20618	22.8	3.5	—
				20806	20	3.8	3.1
AVG	17.99	6.76	2.51	AVG	18.72	4.41	3.14

**Figure 16 Bare Nouns with Verbs
Chunyat & Gakei**

Chunyat				Gakei			
Age	V+N	N+V	N+V+N	Age	V+N	N+V	N+V+N
11008	<u>11.1</u>	—	—	11101	<u>25.6</u>	<u>16.7</u>	<u>9.0</u>
11100	7.1	<u>9.5</u>	4.8	11108	40.0	10.0	4.0
11121	14.6	14.6	—	11122	26.1	21.7	8.7
20110	12.3	9.2	0.8	11129	28.6	25.0	8.3
20117	8.2	15.1	—	20008	32.5	11.7	6.5
20206	14.6	4.2	<u>1.0</u>	20207	34.8	9.1	3.0
20213	11.6	14.9	0.8	20221	14.3	7.8	7.8
20307	5.6	23.4	0.9	20228	10.4	4.5	—
20323	15.5	16.3	—	20304	26.1	6.0	2.5
20410	11.7	35.8	3.1	20311	26.6	3.2	6.5
20507	8.6	24.3	2.2	20318	18.0	9.0	3.3
20523	17.7	20.9	9.3	20325	14.5	3.9	1.3
20608	17.8	20.1	7.7	20408	11.6	6.7	2.4
20624	15.2	20.5	9.5	20430	14.0	2.3	1.9
20706	19.1	14.6	3.5	20503	12.8	0.7	1.4
20713	18.0	13.1	4.9	20711	5.7	—	0.7
20800	17.7	12.6	3.7	20808	8.8	3.6	2.1
20817	20.6	4.3	0.7	20818	9.3	1.3	—
20907	17.7	3.5	0.9	20909	9.9	2.9	0.4
20923	15.9	4.4	0.4				
21013	12.2	3.7	1.4				
21027	7.8	0.9	0.6				
AVG	13.67	12.99	2.55	AVG	19.45	7.69	3.67

Figure 17 Classifiers with Verbs
Tsuntsun

Age	D+CL		NUM+CL		CL+N		D+CL+N		NUM+CL+N	
	obj %	subj %	obj %	subj %	obj %	subj %	obj %	subj %	obj %	subj %
10522	--	--	--	--	--	--	--	--	--	--
10703	--	--	--	--	--	--	--	--	--	--
10710	--	--	--	--	--	--	--	--	--	--
10800	--	--	--	--	--	--	--	--	--	--
10807	--	--	--	--	--	--	--	--	--	--
10821	--	--	--	--	--	--	--	--	--	--
10907	--	--	--	--	--	--	--	--	--	--
10914	--	--	--	--	--	--	--	--	--	--
10929	--	--	--	--	--	--	--	--	--	--
11030	--	--	--	--	--	--	--	--	--	--
11113	--	--	--	--	--	--	--	--	--	--
11127	<u>2.2</u>	--	--	--	--	--	--	--	--	--
20009	2.9	--	--	--	--	--	--	--	--	--
20016	3.3	--	--	--	--	--	--	--	--	--
20108	1.0	<u>1.0</u>	--	--	--	--	--	--	--	--
20205	11.7	5.8	--	--	--	--	2.5	--	--	--
20215	4.5	2.6	--	--	--	--	1.3	--	--	--
20303	4.1	4.1	1.4	--	--	--	--	--	--	--
20317	3.8	1.3	--	--	<u>0.6</u>	--	--	--	--	--
20400	2.0	2.0	--	--	0.7	--	--	--	--	--
20414	4.8	1.2	<u>0.6</u>	--	1.2	--	--	--	--	--
20500	2.4	0.8	2.4	--	1.6	--	--	--	--	--
20514	2.1	3.5	0.7	--	--	--	--	--	0.7	--
20618	4.1	2.9	2.4	--	2.9	--	--	--	--	--
20702	8.6	5.3	1.6	0.8	2.5	--	0.4	--	--	--
AVG	2.30	1.23	0.36	0.03	0.38	--	0.17	--	0.03	--

Figure 18 Classifiers with Verbs
Bernard

Age	D+CL		NUM+CL		CL+N		D+CL+N		NUM+CL+N	
	obj %	subj %	obj %	subj %	obj %	subj %	obj %	subj %	obj %	subj %
10700	--	--	--	--	--	--	--	--	--	--
10800	--	--	--	--	--	--	--	--	--	--
10814	--	--	--	--	--	--	--	--	--	--
10828	--	--	--	--	--	--	--	--	--	--
10904	--	--	--	--	--	--	--	--	--	--
10918	<u>6.0</u>	--	--	--	--	--	--	--	--	--
10925	4.9	1.6	--	--	--	--	--	--	--	--
11010	1.5	--	--	--	--	--	--	--	--	--
11023	3.9	<u>2.0</u>	1.0	--	--	--	--	--	--	--
11106	5.3	5.3	--	--	--	--	1.3	--	--	--
20003	10.1	2.5	--	--	0.8	--	0.8	--	--	--
20016	2.3	--	--	--	--	--	--	--	--	--
20101	4.7	0.9	--	--	--	--	0.9	--	--	--
20115	4.3	2.9	--	--	1.4	--	--	--	--	--
20129	2.0	5.1	1.0	--	1.0	--	--	--	--	--
20212	1.1	--	--	--	--	--	--	--	--	--
20226	8.0	2.7	0.7	--	--	0.7	0.7	--	--	--
20309	2.4	2.4	2.4	--	<u>1.2</u>	--	--	--	--	--
20328	4.7	1.0	--	--	1.6	1.0	--	--	--	--
20407	2.3	--	--	--	0.8	--	--	0.8	--	--
20421	1.7	1.7	--	--	--	0.8	--	--	0.8	--
20504	6.7	5.2	<u>0.7</u>	--	3.7	0.7	0.7	<u>0.7</u>	0.7	--
20519	4.5	1.9	1.9	--	3.2	--	--	1.9	--	--
20604	--	4.7	--	0.8	0.8	--	<u>0.8</u>	2.4	--	--
20618	3.5	5.3	--	--	0.9	--	3.5	--	--	--
20806	3.1	9.2	1.5	1.5	0.8	--	0.8	--	--	--
AVG	3.19	2.09	0.35	0.09	0.62	0.13	0.37	0.22	0.06	--

Figure 19 Classifiers with Verbs
Chunyat

Age	DET+CL		NUM+CL		CL+N		DET+CL+N		NUM+CL+N		subj
	obj %	subj %	obj %	subj %	obj %	subj %	obj %	subj %	obj %	subj %	
11008	--	2.8	--	--	--	--	--	--	--	--	--
11100	<u>7.1</u>	--	--	--	--	--	--	--	--	--	--
11121	4.9	<u>2.4</u>	--	--	--	--	--	--	--	--	--
20110	13.1	3.8	0.8	0.8	--	--	--	0.8	--	--	--
20117	6.8	2.7	--	--	--	--	--	--	--	--	--
20206	4.2	6.3	2.1	1.0	--	--	--	--	--	--	--
20213	5.0	8.3	--	--	--	--	--	--	--	--	--
20307	9.3	4.7	1.9	--	--	--	--	--	--	--	--
20323	3.1	29.5	--	--	--	--	--	--	--	--	--
20410	3.1	11.7	--	--	--	--	--	--	--	--	--
20507	16.8	22.7	--	--	--	--	--	--	--	--	--
20523	7.0	10.7	<u>0.9</u>	--	0.5	0.5	--	--	--	--	--
20608	10.1	5.9	0.6	<u>0.6</u>	--	--	0.6	--	--	--	--
20624	11.4	3.3	1.9	1.0	--	0.5	1.9	1.0	--	--	--
20706	6.5	13.1	1.5	2.0	--	--	0.5	--	2.0	--	--
20713	7.1	11.2	2.6	5.2	0.4	--	--	0.7	0.7	--	--
20800	7.4	14.0	2.3	--	--	--	--	--	--	--	--
20817	14.2	7.8	--	--	--	--	0.7	<u>1.4</u>	--	--	--
20907	7.5	8.8	0.9	0.4	0.4	--	2.2	0.9	--	--	--
20923	7.5	2.0	0.8	0.4	0.4	0.4	0.8	0.4	--	--	--
21013	7.5	10.2	--	--	--	--	1.0	4.8	0.3	--	--
21027	10.2	13.8	--	--	0.6	0.3	2.4	--	--	--	--
AVG	7.72	8.90	0.74	0.52	0.10	0.07	0.46	0.45	0.14	--	--

Figure 20 Classifiers with Verbs
Gakei

Age	D+CL		NUM+CL		CL+N		D+CL+N		NUM+CL+N	
	obj %	subj %	obj %	subj %	obj %	subj %	obj %	subj %	obj %	subj %
11101	3.8	--	--	1.3	--	--	--	--	--	--
11108	2.0	--	--	--	--	--	--	--	--	--
11122	--	--	4.3	--	--	--	--	--	--	--
11129	--	--	--	--	--	--	--	--	--	--
20008	2.6	--	--	--	2.6	--	--	--	--	--
20207	--	<u>1.5</u>	1.5	--	1.5	1.5	--	<u>1.5</u>	--	--
20221	3.9	7.1	--	--	--	--	--	0.6	--	--
20228	--	7.5	--	--	3.0	--	--	1.5	--	--
20304	<u>7.5</u>	1.5	--	--	5.0	--	--	--	--	--
20311	7.3	11.3	--	--	--	--	<u>0.8</u>	1.6	0.8	--
20318	7.4	2.5	2.5	1.6	<u>9.0</u>	0.8	0.8	3.3	--	--
20325	7.9	3.9	--	2.6	5.3	--	1.3	--	--	--
20408	11.6	7.3	--	--	6.7	--	--	--	--	--
20430	10.6	3.8	--	--	8.7	0.8	3.4	0.4	--	--
20503	7.4	7.4	--	--	6.8	--	--	2.7	0.7	--
20711	16.3	10.6	--	--	5.7	<u>2.1</u>	5.0	1.4	--	--
20808	10.3	6.2	<u>0.5</u>	--	4.1	3.6	2.6	--	1.5	--
20818	9.3	7.2	0.8	--	8.0	2.5	2.1	1.3	0.8	0.4
20909	5.8	5.4	0.4	1.2	2.9	1.2	--	0.8	--	--
AVG	5.98	4.38	0.53	0.36	3.64	0.66	0.84	0.80	0.20	0.02

Figure 21 Pronouns with Verbs
Tsuntsun & Bernard

Tsuntsun				Bernard			
Age	V+Pr %	Pr+V %	Pr+V+Pr %	Age	V+Pr %	Pr+V %	Pr+V+Pr %
10522	—	—	—	10700	—	—	—
10703	—	—	—	10800	—	—	4.5
10710	—	—	—	10814	—	—	—
10800	—	—	—	10828	—	—	—
10807	4.8	—	—	10904	—	—	—
10821	5.6	11.1	—	10918	—	6.0	—
10907	—	—	—	10925	—	4.9	—
10914	—	—	—	11010	—	—	—
10929	—	—	—	11023	—	—	—
11030	—	2.8	—	11106	1.3	<u>2.6</u>	—
11113	—	—	—	20003	0.8	1.7	—
11127	—	<u>2.2</u>	—	20016	—	3.0	—
20009	1.4	1.4	—	20101	0.9	4.7	—
20016	—	1.7	—	20115	—	1.4	—
20108	<u>1.0</u>	2.0	—	20129	<u>3.0</u>	2.0	—
20205	1.7	1.7	—	20212	1.1	—	—
20215	3.8	3.2	0.6	20226	0.7	2.7	—
20303	—	2.1	—	20309	—	2.4	—
20317	2.5	1.9	—	20328	0.5	6.2	—
20400	1.4	2.7	—	20407	—	4.7	0.8
20414	4.2	3.0	1.2	20421	3.3	2.5	—
20500	0.8	6.3	—	20504	0.7	7.5	—
20514	0.7	6.4	0.7	20519	—	25.0	—
20618	7.6	12.9	—	20604	3.1	18.9	—
20702	10.7	10.3	2.1	20618	—	20.2	—
				20806	3.8	11.5	—
AVG	1.85	2.87	0.18	AVG	0.75	4.92	0.20

Figure 22 Pronouns with Verbs
Chunyat & Gakei

Chunyat				Gakei			
Age	V+Pr %	Pr+V %	Pr+V+Pr %	Age	V+Pr %	Pr+V %	Pr+V+Pr %
11008	—	—	—	11101	1.3	<u>11.5</u>	—
11100	—	—	—	11108	—	2.0	—
11121	—	—	—	11122	—	8.7	—
20110	—	3.1	—	11129	<u>1.2</u>	6.0	—
20117	—	—	—	20008	5.2	7.8	—
20206	2.1	<u>6.3</u>	—	20207	3.0	6.1	—
20213	0.8	1.7	—	20221	9.1	26.6	—
20307	0.9	4.7	—	20228	11.9	7.5	—
20323	—	4.7	—	20304	8.0	26.1	0.5
20410	—	0.6	—	20311	6.5	15.3	—
20507	<u>0.5</u>	—	—	20318	2.5	21.3	<u>1.6</u>
20523	0.5	0.9	—	20325	7.9	32.9	2.6
20608	1.2	1.2	—	20408	4.3	24.4	2.4
20624	0.5	1.0	—	20430	3.0	25.3	6.4
20706	—	0.5	—	20503	4.7	31.8	4.1
20713	1.5	—	—	20711	5.0	31.2	0.7
20800	4.2	1.9	—	20808	6.2	21.1	1.5
20817	5.0	12.8	—	20818	7.2	25.7	2.5
20907	2.2	8.0	<u>0.4</u>	20909	5.8	32.2	3.3
20923	9.1	12.3	0.8				
21013	2.0	13.6	0.3				
21027	2.7	15.9	0.9				
AVG	1.51	4.04	0.11	AVG	4.88	19.13	1.36

**Figure 23 Development of NPs Expressing Possession
Chunyat**

Age	N+N %	N+D+Cl %	PossM %	Pr+N %	N+Cl+N %	Pr+Cl+N %
11008	<u>8.3</u>	0.0	0.0	0.0	0.0	0.0
11100	2.4	4.8	0.0	0.0	0.0	0.0
11121	2.4	0.0	0.0	0.0	0.0	0.0
20110	3.8	0.0	0.0	0.0	0.0	0.0
20117	1.4	<u>9.6</u>	0.0	0.0	0.0	0.0
20206	1.0	1.0	0.0	0.0	0.0	0.0
20213	0.0	5.0	0.0	0.8	0.0	0.0
20307	0.0	1.9	0.9	0.0	0.0	0.0
20323	0.0	2.3	0.0	0.0	0.0	0.0
20410	2.5	0.0	0.0	0.0	0.0	0.0
20507	0.0	0.0	0.0	0.0	0.0	0.0
20523	0.0	0.0	<u>2.8</u>	0.5	0.5	0.0
20608	1.8	0.6	3.0	0.0	0.0	0.0
20624	7.1	1.4	1.9	0.0	0.0	0.0
20706	0.5	1.0	0.5	0.0	0.0	0.0
20713	3.4	0.0	0.7	0.4	0.4	0.0
20800	1.4	0.0	0.0	0.0	1.4	0.0
20817	1.4	0.0	0.0	0.7	0.0	0.0
20907	0.9	0.9	0.0	0.4	0.4	0.0
20923	2.0	1.6	0.4	0.0	0.4	0.8
21013	1.4	3.1	1.4	1.4	0.0	0.0
21027	2.7	1.2	0.3	2.4	0.0	0.3
AVG	2.02	1.56	0.54	0.30	0.14	0.05

**Figure 24 Development of NPs Expressing Possession
Gakei**

Age	N+N %	N+D+Cl %	PossM %	Pr+N %	N+Cl+N %	Pr+Cl+N %
11101	<u>2.6</u>	0.0	0.0	0.0	0.0	0.0
11108	4.0	0.0	0.0	0.0	0.0	0.0
11122	4.3	0.0	0.0	0.0	0.0	0.0
11129	2.4	0.0	0.0	0.0	0.0	0.0
20008	0.0	0.0	0.0	1.3	0.0	0.0
20207	1.5	0.0	0.0	0.0	1.5	0.0
20221	0.6	0.0	0.0	0.0	0.0	0.0
20228	0.0	0.0	0.0	0.0	0.0	0.0
20304	0.0	0.0	0.0	0.0	0.0	0.5
20311	2.4	<u>1.6</u>	1.6	0.0	0.8	0.0
20318	0.8	2.5	0.8	0.0	0.0	0.0
20325	1.3	1.3	0.0	0.0	0.0	0.0
20408	1.2	1.2	0.6	0.6	0.0	0.0
20430	0.4	3.4	0.0	0.4	0.0	0.0
20503	2.0	4.1	<u>0.7</u>	0.0	0.0	0.0
20711	0.0	0.7	2.1	0.0	0.7	0.7
20808	0.0	0.5	0.5	1.0	0.0	0.0
20818	0.0	2.5	0.0	0.0	0.0	0.8
20909	0.4	0.4	0.0	0.0	0.0	0.4
AVG	1.27	0.96	0.33	0.17	0.16	0.13

The Acquisition of *Bei2* as Verb and Preposition in a Cantonese-speaking Child

Colleen H. Wong

1. Category membership: verb, coverb or preposition?

Description of syntactic regularities is based on the characterization of syntactic categories; and the characterization of syntactic categories depends on the respective theoretical approaches adopted by linguists. Criteria have been proposed to justify category assignment to words. And all linguists realize too well the indeterminacies of category assignments.

On the indeterminacy in characterizing words into classes, a case in point in the Chinese language is with a group of words that some linguists call 'coverbs'. It is observed that coverbs behave like verbs and prepositions. Ding Shengshu et al. (1961:95) shows how coverbs can be conceived in relation to verbs and prepositions: 'A verb, together with its complement if present, can be the predicate of the clause. Coverb is a kind of verb, but it differs from the verb in two respects: it generally doesn't function as the main entity in the predicate; and it obligatorily takes an object.'

Chao (1968:749-752) acknowledges the 'transitional nature of Chinese prepositions, both in the classificatory and in the historical sense' (ibid. p.749) and defines prepositions by enumeration. He notes that there is a listable number of verbs that occur as first verbs in 'verbal expressions in series' ('V-V series') with the same order of frequency as they occur in other positions. He calls this class of verbs 'coverbs' or prepositions. He describes the formal properties of prepositions as follows: (a) they do not, as a rule, have aspects; (b) they do not usually function as centres of predicates; (c) they do not usually omit their objects.

Indeed the traditional view on the origin of prepositions is that verbs are grammaticalized to become prepositions through serial verb constructions, i.e. "V1 + NP1 + V2 (+NP2)" --> "Prep + NP1 + V + (NP2)". 'Grammaticalization' refers to 'the process whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions, and, once grammaticalized, continue to develop new grammatical functions' (Hopper and Traugott 1993:xv). In terms of grammaticalization, V1 is usually the preposition and V2 is the verb. (For an explication of cases of grammaticalization processes involving verbs and prepositions, see, for example, Peyraube 1994).

Sentences (1), (2) and (3) below illustrate the distinction between a verb and a preposition. (These and the following example sentences written in Chinese characters are presented in Appendix 1.)

In (1), *zai* is a verb -- the sentence is not a V-V series and *zai* is the sole predicate:

- (1) *zai* jianqiao
at Cambridge

(2) and (3) each contains a V-V series. Chao (1968:754) notes that 'in a V-V series the second verb is the centre', with the V1 expression modifying the V2 expression.

- (2) *zai* jianqiao shang xue (ibid. p.754)
at/in Cambridge study
(studying at Cambridge)

zai occupies the first verb position in the V-V series *zai* - *shang xue* to express 'place where' (ibid. p.337). *zai* is a preposition in this sentence.

In (3) below, the prepositional phrase *zai jianqiao* ('at Cambridge') occupies the position of a complement:

- (3) zhu *zai* jianqiao
live at Cambridge

Prepositional phrases as bound phrase complements in V-V series are limited to *dao*, *zai*, *gei* (in Mandarin Chinese), all translatable as 'to' or 'at' (ibid. pp. 352-3, 753-4). The preposition in a complement position is usually in the neutral tone and is very often compounded enclitically with the verb. Furthermore, Li Jianxi (1955:199) points out that preposition *zai* denoting 'location' can occur in the pre-verbal as well as the post-verbal position if the verbs are 'static' intransitive verbs -- for example: *wo zai tai shang zhan zhe* ('I - on the platform - standing') and *wo jian zhe zai tai shang* ('I - standing - on the platform'). But they can only occur in the pre-verbal position if the verbs are 'motion/continuous' intransitive verbs -- for example: *tamen zai chaochang shang saipao* ('they - on the playground - racing') and commonly not *tamen saipao zai chaochang shang* (they - racing -on the playground). When *zai* denotes 'arrival point', then the prepositional phrase is put post-verbally (Li 1955 pp. 203-4). Chao (1968:754) points out that in wenyao prepositions for 'place or time toward or at' as well as 'arrive at' are in the complement position -- *xu yu jianqiao* ('prefaced at Cambridge')

It has also been pointed out that the prepositional complement denoting 'place' can be preposed (e.g. Li Jianxi 1955:199; Chao 1968:754):

- (4) *zai jianqiao zhu*
 at Cambridge live

And Chao (1968:754) notes that prepositions for 'place from' or 'time from' occupy either position -- *lai zhi he chu* -- *zhi yuan fang lai* 'come from where -- from far away come'.

So what are some diagnostic syntactic frames to test whether a form is a verb or a preposition (cf. Zhu 1982:174)? There are some frames for prepositions that can be summarized from the above discussion. Prepositions cannot serve alone as predicates. Prepositions can serve as the first constituents in predicate-series, or V-V, constructions. Also prepositions cannot reduplicate. Prepositions cannot take aspectual suffixes *le*, *guo*, *zhe*.

But there are some syntactic environments which are common to verbs and prepositions. Both verbs and prepositions can take auxiliaries. Take sentence (2) for example. We can add an auxiliary *hui* ('will') to the preposition *xian* or to the verb *shang xue*. Both can take adverbs. In (2) again, we can add the adverb *sien* ('first') to the preposition or the verb. Both verbs (transitive) and prepositions can take objects. And both can be negated.

Indeed, syntactic indeterminacy is evidenced in negatively specified criteria on the one hand and overlapping criteria on the other. In view of this, let us examine the semantic criteria for prepositions as well.

2. A semantic description of prepositions

Prepositions are often described as well in terms of semantic criteria. This section is a semantic description of prepositions. Preposition is a listable word class. For Mandarin Chinese, Li (1955:197-215) divides prepositions into four major groups each with a number of sub-groups based on the meanings of the things or events that the prepositions introduce. They are prepositions of: (a) time-place, (b) reason or motive, (c) means-association-comparison, and (d) special prepositions. The first three groups see the prepositions preceding their complements, while the special prepositions occur after their complements. And (d) includes the prepositions which have the function of object preposing and those that introduce adverb phrases.

Chao's list consists of 57 prepositions (1968: 754-766). Some meanings of prepositions that Chao describes are: (a) time when, (b) place where, (c) interest and benefit, (d) purpose, cause and reason, (e) manner, instrument, (d) comparison, (e) prepositions associated with a subsection of pretransitives (object-preposing verbs) (pp.336-343).

2.1 A meaning-based grouping of Cantonese prepositions

In this section I will first describe Cantonese prepositions by grouping them in terms of meaning. Example sentences of Cantonese will be given. Mandarin preposition equivalents are provided for reference purposes. I will then discuss the syntactic environments of Cantonese prepositions and point out how they differ from Mandarin prepositions.

Based on the meanings of the things or events that the prepositions introduce (Cf. Li 1955), I shall classify Cantonese prepositions into the following: (Note 1)

(a) prepositions of time-place

i. being (introducing a static relationship within a specific place or time)

e.g. *hai2*, *heung2*, *lam4*, *can3*.

'at, in, while'

(5) *keoi hai/heung nguk kei zou gung fo*

's/he at home do homework'

(6) *keoi lam zau sung bei ngo ge*

's/he when leaving give me (something) PART'

(7) *can mei hak hou zau la*

'while not yet dark better go PART'

Mandarin examples: *zai*, *yu*, *dang*, *lin*

ii. origination (introducing the start point or distance)

e.g. *hai2*, *jau4*, *lei4*, *gaak3*

(8) *ngo hai(/jau) ngoumun lei ge*

'I from Macau come PART'

(9) *waan-zai gaak (/lei) saiwaan jau gei jyun?*

'Wanchai from Western District how far'

(10) *jiga gaak zo da zeung go zan si ji ging ngsap gei nin la*

'now from ASP the war already 50 some years PART'

Mandarin: *cong*, *qi*, *zi-cong*, *li*, *ju*

iii. passage (introducing the route)

e.g. *aai1-zyu6*, *seon6-zyu6*, *gan1-zyu6*, *kaau3-zyu6*, *ziu3-zyu6*

'along, according to'

(11) *aai-zyu tiu daai keoi haang zau wui dou ga la*

'along the nullah walk then will get there PART PART'

(introducing the time lapse: postpositions)

e.g. *(ji5-)noi4*, *zi1-hau6*, *ji5-hau6*

'from that time onward, since'

(12) sap nin *noi* mei gin gwo keoi
'for 10 years not meet him/her'

(13) sik zo *zi-hau* zau tou tung lo
'eat ASP after then stomach ache PART'

Mandarin: *guo, jing, kao, ai-zhe, shun, yan*

iv. target (introducing the target or orientation of an action)

e.g. *heung3(-zyu6), deoi3(-zyu6), tung4, deoi3(-jyu1), zi3-jyu1, kwaan1-jyu1*

'toward, to, about, as for'

(14) *heung(/deoi)-zyu* ngo co
'facing me sit'

(15) nei jau *tung* keoi gong mat a!?
'(now) you to him talk what PART'

(16) *deoi* nei gin si ngo mou *jigin*
'about this matter I have no opinion'

Mandarin: *xiang, wang, chao, dui, dui-yu*

v. reach (introducing the end point, or the goal of an action),
with three sub-types:

v-1 reaching a place

e.g. *dou3, heung2* (to precede the predicate)

(17) keoi haang *dou* nei dou ji ging mou saai hei
's/he walk up to here already out of breath'

(18) ngo da syun *heung* jinggwok hoi caan gun
'I plan to go to Britain open a restaurant'

Mandarin: *dou*

v-2 reaching a time, or from a certain time to a certain time

e.g. *dou3, jau4 ... dou3*

(19) keoi zou *dou* sap dim
's/he work till 10 o'clock'

(20) ngo *jau* ziu zou *dou* maan a
'I from morning work till night PART'

Mandarin: *dou, cong...qi*

v-3 reaching a person

bei2

(21) ngo sung bun syu *bei* keoi
'I give CL book to him/her'

(22) ngo sung bun syu *bei* keoi tai
'I give CL book to him/her read'

- (23) bun syu ngo ji ging gaau zo *bei* keoi
 'the book I already give ASP (to) him/her'

Mandarin: *gei*

(b) prepositions of benefit (or harm)

- i. beneficiary (the prepositions introduce the persons or things that receive the benefit or harm that the actions may bring)

e.g. *wai6, tung4* 'for'

- (24) keoi gam san fu dou hai *wai* zo di sai lou
 's/he (work) so hard all for the children'

- (25) sin saang *tung* di hok saang bou fo
 'the teacher for the student (give) extra lesson'

Mandarin: *wei-le, ti, gei*

- ii. benefactor (the prepositions introduce the initiators of the benefit or harm)

e.g. *bei2, jau2(-dak1)*

- (26) ngo *bei* keoi laau zo jat caan

'I *bei* him/her reprimand' 'I was reprimanded by him/her'

- (27) gin saam *bei* keoi zing wu zou zo

'the clothes *bei* him/her make dirty' 'The clothes was made dirty by him/her'

- (28) *bei* keoi gaau dim zo

'*bei* him/her do (something) right' 'S/He made it!'

e.g. *bei2, jau2(-dak1)* (with some concession)

- (29) *bei*(/*jau*/*jau-dak*) keoi gong la

'*bei* him/her talk PART' 'Let him/her talk.'

e.g. *jan-wai* 'because of'

- (30) ngo m wui *jan-wai* keoi ngok zau pa zo keoi

'I not because him/her fierce then fear him/her'

'Just because s/he is fierce I won't be afraid of him/her'

Mandarin: *bei, you, ren, gei, rang*

(c) prepositions of means

- i. instrument (the prepositions introduce the tools or material needed for an action)

e.g. *jung6, lo2, ling1, dong3*

'use, take hold of'

- (31) seon fung hai *jung/lo* da zi gei da ge

'the envelopment is written with a type-writer'

- (32) ngo *jung/lo* saau sik

'I with hand eat'

- (33) ngo *lo/ling* heoi lai zou baan
 'I take hold of him/her use as an exemplar'
 'I use him/her as an exemplar'
- (34) keoi *dong* ngo hai so gwa
 's/he takes me as a fool'

Mandarin: *ba, na, yong*

- ii. reference (the prepositions introduce the reference or criteria)
 e.g. *gan1-(zyu6)*, *ziu3-(zyu6)*
 (35) *gan-zyu* keoi gong gam zou
 'follow him/her say likewise do'
 'do as s/he says'

Mandarin: *yi, zhau, an, gen*

- iii. exception (the prepositions introduce the exceptions)
 e.g. *ceoi4*
 (36) *ceoi zo* keoi go go dou m zaan sing
 'except for him/her nobody approves (of it)'

Mandarin: *chu*

- iv. association (the prepositions introduce the entities that accompany the action)
 e.g. *tung4, gan1*
 (37) nei *tung/gan* ngo jat cai heoi la
 'you with me together go PART'
 (38) nei *tung* ngo hou, ngo jau wui tung nei hou ze
 'you to me good, I also will to you good PART'
 'if you are good to me I'll be good to you too'

Mandarin: *he, tong, yu, gen*

- v. comparison (the prepositions introduce the entities that are in comparison with the subject)
 e.g. *tung4* 'on a par with'
 (39) keoi *tung* nei jat jeung gam gou
 's/he and you same height'
 e.g. *guo3* 'comparative'
 (40) keoi gou *gwo* nei
 's/he taller than you'
 e.g. *bei2-gaa4* 'comparative'
 (41) nei jeung *bei-gaa* hou
 'this better'
 e.g. *hou2-ci3* 'like'

- (42) *hou-ci keoi gam mei dak lo*
 `like him/her then OK PART'
 `be like him/her then it's OK'

Mandarin: *he, yu, gen, bi, guo*

(d) special prepositions

i. object-preposing

It can be seen that these prepositions e.g. *zeung1, lo2* are also those preposition of `means: instrument' (Type (c) i). They are special in a sense that they can prepose the objects to precede the verbs, or to topicalize a constituent.

e.g. *zeung1, lo2* `take hold of'

- (43) *ngo zeung bun sy sung/lo zo bei go hok saang*
 `I *zeung* CL book give ASP bei CL student'
 `I gave/took the book to a student'

- (44) *keoi zeung di tau faat yim zo hung sik*
 `s/he *zeung* hair dye ASP red'
 `s/he has the hair dyed red'

- (45) *ngo yi ging zeung fan bou gou gong dak hou cing co*
 `I already *zeung* CL report explain very clearly'
 `I already explained the report very clearly'

e.g. *lin4*

- (46) *nei lin gam cin ge zi dou m sik*
 `such an easy word you don't even know'

e.g. *deoi3-jyu1, zi3-jyu1, gwaan1-jyu1* `about, as for'

- (47) *deoi-jyu tin man hok ngo jat di dou m sik*
 `as for astrology I know nothing'

Mandarin: *ba, dui-yu, lian*

ii. introducing adverbials of degree or result

These prepositions e.g. *dou3* originate from the prepositions of `reach: place' (Type (a) v. v-1). (Cf. Li 1955:237). They introduce adverbial phrases to show the degree or result of the actions.

e.g. *dak1, dou3*

- (48) *gong dak hou*
 `speak well'

- (49) *geng dou ngo m sik juk*
 `so scared I can't move'

Mandarin: *de*

2.2 The syntactic environments of Cantonese prepositions

So what are some diagnostic syntactic frames to test whether a form is a verb or a preposition in Cantonese? Like Mandarin, Cantonese prepositions cannot serve alone as predicates. Sentence (1) in Cantonese will be: *hai gim kiu* ('at/in Cambridge') where *hai* is a verb. But unlike Mandarin, Cantonese prepositions can be one-word sentences functioning as answers to questions:

- (50) *nei gan-m-gan ngo heoi? gan*
'you with-NEG-with me go?' 'with'
(for *gan*, refer to type (c) iv, sentence (37))

Like Mandarin, Cantonese prepositions can serve as the first constituents in predicate-series, or V-V, constructions. Sentence (2) in Cantonese: *hai gim kiu dok syu* ('at/in Cambridge study') the *hai* is a preposition (cf. Type (a) i, sentence (5)). Cantonese and Mandarin prepositions cannot reduplicate. Also stranding of prepositions is not allowed; for example,

- (51) preposition stranding:
(English) That's the student I gave a book *to*.
(Mandarin) **jiu shi zhe ge xue shang wo song le yi ben shu gei*
(Cantonese) **zau hai ni go hok saang ngo sung zo bun syu bei*

Unlike Mandarin, Cantonese prepositions can take aspectual suffixes. The following sentence is good in Cantonese:

- (52) *hai zo gim kiu zyu ji ging leung nin la*
'in ASP Cambridge live already two years PART'
hai is a preposition, *zo* is a perfective marker, and *zyu* is a verb.

In Cantonese, like in Mandarin, the verbs and prepositions can take auxiliaries, adverbs, and objects, and can be negated.

A summary of the syntactic behaviour of Cantonese prepositions:

- 1). They usually cannot serve alone as predicates. But they can occur alone when they act as one-word answers to questions like (50); this is unlike Mandarin.
- 2). They can serve as the first constituents in the predicate-series, or V-V, constructions.
- 3). Preposition stranding is not allowed (cf. (51)).
Like verbs, they
- 4). can take aspectual suffixes (cf. (52)),
- 5). can take auxiliaries,
- 6). can take adverbs,
- 7). can be negated,

8). can be in the A-not-A frame (cf. (50).

Unlike verbs, they

9). cannot reduplicate.

We shall summarize the verb environments and the preposition environments of *bei2* with sentences (53), (54), (55) and (56) below.

bei2 is the centre of predication in (53) and it is a verb with the meaning of 'give'.

(53) *bei zeze*

'give Sister' 'give something to Sister'

bei2 is the first verbal predicate in (54) and it is a preposition introducing the benefactor of the action (cf. Type (b) ii 'benefactor', sentences (26), (27), (28) and (29)).

(54) *bei Gakei sik zo*

'bei Gakei eat ASP' 'something has been eaten by Gakei'

bei2 is a preposition in the complement position when it is in the post-verbal position (cf. sentences (3) and (4) and their description). In terms of meaning, complement-*bei2* belongs to Type (a) v-3. 'reach a person': introducing the end point or the goal of an action. (cf. (21), (22) and (23). An additional example is (55) below:

(55) *lo go zi doi bei ngo a*

'get hold of CL paper bag *bei* me PART'

'give me a paper bag'

Complement-*bei2* can also introduce the entity that receives benefit/harm of an action, depending on the meaning of the verb; in this sense complement-*bei2* belongs to Type (b) i 'beneficiary'. For example,

(56) *hoi bei nei*

'open *bei* you' 'open for you'

The V-*bei2* combination as in (55) and (56) are not V-V combination but V-P combination, contrary to McCawley (1992). He proposes (for adult Mandarin) that combinations of v-*gei* such as *songgei* and *fugei* are compound verbs that takes two NP objects, and *gei* is not a preposition. He uses the placement of the perfect suffix -*le* to support his point:

wo fugei-le ta \$200 'I have paid him \$200'

**wo fu-le gei ta \$200*

In Cantonese, however, we can always place the perfect suffix *zo* after the first item, the V, and not after the second item. An example:

- (57) caak zo *bei* ngo tai ha sin
 dissemble-ASP *bei* me see-COMPL Adv

Also, with (58) below, we can extract the NP following *mai* 'to buy' but not the NP following *bei2*:

- (58) mami maa toto *bei* gakei
 'mother buy rabbit *bei* gakei'
 'Mother bought Gakei a toy rabbit'
 (59') toto, mami maa 0 *bei* gakei
 (59'') *gakei, mami maa toto *bei* 0

bei2 can occur between two verbs, for example,

- (59) caak zo *bei* Gakei tai sin
 'dissemble ASP *bei* Gakei see first'
 'dissemble it to let Gakei see first'

Here a clause follows *bei2*. This post-*bei2* clause is the purpose of doing the main clause, so the post-*bei2* clause can be considered a 'purposive clause'. In this way, an inter-verbal *bei2* functions as a complementizer introducing a purposive clause. It has the meaning of 'so that'. In fact, *gei* in Mandarin as a clausal complementizer is suggested in Huang:1982 (pp.106-7, Footnote #36) and Tsao:1989.

So this study posits four structures of *bei2*:

- (a) verb (no other verbs in utterance)
- (b) preposition (*bei2* in pre-verbal position)
- (c) complement (*bei2* in post-verbal position)
- (d) complementizer (*bei2* in between verbs)

3. The development of *bei2* in a Cantonese-speaking child

Against the above description of the syntactic contexts and semantics of Cantonese prepositions, this paper reports on the development of *bei2* as 'verbs' and 'prepositions' in a Cantonese-speaking child, aged 1;11 to 2;9 (1-year-11-months to 2-years-9-months). The development of *bei2* will be described in terms of the four structures posited above.

3.1 Acquisition of prepositions

The acquisition of prepositions in Mandarin and English have been studied by some researchers. For example, Jepson (1989) studies the acquisition of three Modern

Standard Chinese coverbs *gen*, *yong* and *ba* by twelve children aged 2;2 to 4;2 and observes how the categories develop from prototypical members of a class to more mature adult-like categories in a drop-add process of features. Tomasello (1987) documents a 2-year old's differential acquisition pattern of English prepositions. Valian (1986), working on a range of syntactic categories with six English-speaking children aged from 2 years to 2;5 finds that the children distinguished between prepositions and verbs (among many other findings).

In this study some occurrences of *bei2*, i.e. 'in between verbs' are called 'complementizers' (cf. structure (d) at the end of Section 2 above). Such analysis could have implications on a debate on the grammatical representations in early child grammar. There are two opposing hypotheses about phrase structure in early grammar: the discontinuity hypothesis (e.g. Radford 1990) and the strong continuity hypothesis (e.g. Whitman, Lust and Lee 1990; Valian 1991; Deprez 1992). The main arguments of the discontinuity hypothesis are that (i) the stage in which the child produces predominantly single-word utterances does not show categorial distinctions; (ii) once children move into the categorial stage, there is usually a stage in which their grammar consists of only lexical categories (N, V, A, P and their projections; around 20 months); functional categories (inflections, complementizers, and determiners are developed later (around 24 months), e.g. English-speaking children's early grammar lack CP, IP, DP -- '*Hayley draw boat / want mummy come / bottle milk*'. The strong continuity hypothesis argues for positing grammatical representations in early grammar to be fundamentally identical to the representations proposed by syntacticians for the corresponding adult structures. In each case the appropriate representations have included functional categories and/or their projections. Although the present study does not have the theoretical underpinning of grammatical representations, it hopes to find out at what age young Cantonese-speaking children can produce complex complementation structure involving *bei2*.

3.2 Data and analysis

The data used in this study are based on the language produced by a Cantonese-speaking child, coded here as 'Gakei', when she was aged 1;11 - 2;9. The data consist of 19 audio-taped recorded sessions (average 45 minutes each session) of the child interacting with at least one adult at her home. The data were collected over a period of ten months.

The data are spontaneous utterances produced by the child Gakei. No utterances in the *bei2* data here are imitation of sentences produced by the adults interacting with her. She seldom repeated herself either. What looks like consecutive similar sentences invariably vary in speech tone and sentence final particles. There are no frozen expressions involving *bei2*.

The following situation can illustrate:

age: 2;3;11 (2 years; 3 months; 11 days)

situation: The child's grandma is sitting on the sofa.

CHI = child, INV = investigator

CHI: mama zou san

'grandma good morning'

INV: hai laak

'yes ASP'

CHI: <bei mama sin> [/] bei mama sin

'bei grandma first' 'bei grandma first'

CHI: ngo bei mama jap heoi wo

'I bei grandma enter-go ASP'

CHI: hoi mun bei mama jap heoi wo

'open door bei grandma enter-go ASP'

CHI: gei sing muk

'how smart'

The occurrences of *bei2* are charted in the following 4 patterns:

- (a) verb (no other verbs in utterance)
- (b) preposition (*bei2* in pre-verbal position)
- (c) complement (*bei2* in post-verbal position)
- (d) complementizer (*bei2* in between verbs)

Occurrences of *bei2* in Gakei

YR;MO DAY	1;11 01	1;11 08	1;11 22	1;11 29	2; 08	2;2 07	2;2 21	2;2 28	2;3 04	2;3 11	2;3 18	2;3 25	2;4 08	2;4 30	2;5 03	2;7 11	2;8 08	2;8 18	2;9 09	TOTAL
Pattern																				
(a)	5	6	2		2	1	3		6	19	6	1	1	4	6		1	8	5	60
(b)				1	2	1	2	1	2	4	1		5	1	1	1	1	4	5	28
(c)					1		3	1	1	3	2	1	1	1	4	2		2	4	25
(d)						1	3		5	3	1				3	1	3			19
	5	6	2	1	5	3	11	2	14	29	10	2	7	6	14	4	5	14	14	

The data show that Pattern (a) 'verb' appears earliest in the child Gakei's language. Pattern (b) 'preposition' appears a month later, at around age 2. Patterns (c) and (d), 'complement' and 'complementizer' respectively, appear around 2;2 (aged 2 years 2 months).

All four patterns of *bei2* are found at 2;2. In the last 13 audio-recording sessions (beginning from session 2;2;21), the number of occurrences of Patterns (a), (b), (c) and (d) is 60, 28, 25 and 19, respectively.

For ease of description of the development of the four Patterns of *bei*2 over ten months in 19 recording sessions, I will describe the Patterns separately. Features associated with each Patterns will be pointed out; example sentences produced by Gakei can be found in the Appendix 2. These sentences are referred to by means of numbers in square brackets ([]).

Pattern (a) 'verb'

The first appearances of verb-*bei* can occur alone as commands [1], can take object noun phrases [2], '*bei*+NP' can be followed by the adverb *sin* 'first' [3], and can be negated [4]. Later, (aged 2;0;8) verb-*bei* can be a two-place predicate with subject and object NPs [5]. From age 2;2;21 onwards, it is found that *bei* takes an auxiliary *jiu* 'want' [6]. Longer sentences are produced since verb-*bei* can be a three-place predicate [7], can take a complement [8] and can be modified by temporal adverbial phrases such as *jat-zan* 'in a while', *tau-sin* 'just now', *dang-zan-gaan* 'in a while', *dai-jat* 'another day' [9], [10], [11], [12] respectively. [10] looks like a relative clause. [11] involves double objects. In [13] *bei zo cin ho ji gwo* 'give-ASP can pass' (age 2;3;04) the *bei*-phrase is the first of the two verb phrases. The second verb phrase denotes the result of the action of the first phrase. In adult Cantonese the adverb *zau* 'then' can be added to link up the two verb phrases to show the 'if... then' relationship.

Pattern (b) 'preposition'

Preposition-*bei* in the data belongs to the type of 'prepositions of benefit' (cf. Type (b) ii). It develops later than verb-*bei* and appears in the child's language at around age two. Examples of the early preposition-*bei* are [14] and [15] -- *m bei mami tai* 'not *bei* mother see' and *ni di bei zeze tai zo la* 'this *bei* Sister see-ASP PART'. Additional examples are [16] - [19]. In [20] *bei ngo tung heoi jam* '*bei* me for him drink; let me help him drink (something)' we find two prepositional phrases before the verb; one is *bei ngo*, and the other is *tung heoi* where *tung* is also a 'preposition of benefit' (cf. Type (b) i) (age 2;2;21). In [21] the *bei*-phrase is followed by two verb phrases -- *nei bei ngo ceot lei gong je* 'you *bei* me come out say something' (age 2;8;18).

Pattern (c) 'complement'

Complement-*bei* denoting 'reaching a person' and 'beneficiary' largely develops from age 2;2 onwards (25 out of 26 instances in the data). The first instance (age 2;0;08) is a short utterance *lo bei* 'get *bei*' [22] which the child produced immediately after her own utterance *mei bun syu* 'buy CL book'. But she did not follow up on the beneficiary when the adult pursued.

The more developed complement-*beis* can take objects [23], and are found in longer sentences [24] (with subject and object of the verb and object of the preposition) and [25] (with object of the verb and object of the preposition).

Pattern (d) 'complementizer'

Like the complement-*bei*, complementizer-*bei* introducing a purposive clause develops from age 2;2 onwards. There are a total of twenty occurrences in the data. The utterances are longer ([26]-[31] contain two verbs in each sentence). The structure of [31] is quite advanced. It consists of three verbs 'yau', 'gong' and 'teng' ('existential-be', 'tell' and 'listen' respectively), and the complementizer occurs after the second verb.

4. Conclusions

The form *bei*₂ starts being used in a restricted set of environment (the 'verb' environment) and then extends to include a larger set of environments (the 'preposition', 'complement', and 'complementizer' environments).

*bei*₂ is acquired early. From shortly before age two to around two-and-a-half, the child has acquired a repertoire of the functions of *bei*₂: as a verb, a preposition, a complement, and a complementizer.

Cantonese-speaking children at a very early age, by two-and-a-half, produce complex complementation structures involving *bei*₂ such as those in Pattern (d) where *bei*₂ occurs between verbs.

More needs to be done in order to obtain a more complete and generalized picture of the development of prepositions in Cantonese-speaking children, for example, to describe the development of all prepositions in a larger number of Cantonese-speaking children.

(I wish to thank my friends and colleagues again for their comments and help, especially Thomas Lee and Patricia Man.)

Notes

¹The tone marks are based on the 'Linguistic Society of Hong Kong Cantonese Romanization Scheme' (1993). 1(55/53) 2(35) 3(33/3) 4(21) 5(13) 6(22/2)

References

- Chao, Y.R. (1968) A Grammar of Spoken Chinese. U. of California Press
- Deprez, V. (1992) 'Variations in parameter setting'. A paper presented at the Conference on Syntactic Theory and First Language Acquisition, Cornell U. April 24-26, 1992

- Ding, Shengshu (1961) *Xiandai Hanyu Yufa Jianghua* (Lectures on Modern Chinese Grammar). Beijing, Commercial Press
- Hopper, P and E. Traugott. (1993) *Grammaticalization*. Cambridge CUP
- Huang, James C-T. (1982) 'Logical relations in Chinese and the theory of grammar', doctoral dissertation, MIT
- Jepson, J. (1989) 'The acquisition of categories in Chinese'. *Lingua*, 78, 193-216
- Lee, Thomas; Sam Leung and Colleen Wong. (1992) 'Syntactic categories in Cantonese-speaking two-year-olds'. A paper presented at the LSHK Annual Research Forum, 1992
- Li, Jianxi. (1955) *Guoyu Wenfa* (Chinese Grammar). Shanghai: Commercial Press
- McCawley, J. (1992) 'Justifying part-of-speech assignments in Mandarin Chinese'. *J. of Chinese Linguistics*, 20, 2, 211-245
- Peyraube, A. (1994) 'Mechanisms of syntactic change in Chinese and in other languages'. Seminar paper presented at the English Department, CUHK, April 1994
- Radford, A. (1990) *Syntactic Theory and the Acquisition of English Syntax*. London: Blackwell
- Tomasello, M. (1987) 'Learning to use preposition: a case study'. *J. Child Language*, 14, 79-98
- Tsao, Feng-fu. (1989) 'Serial verb constructions in Chinese', *Proceedings on the Second International Conference on Sinology*. Taipei: Academia Sinica
- Valian, V. (1986) 'Syntactic categories in the speech of young children'. *Developmental Psychology*, 22, 4, 562-579
- Valian, V. (1991) 'Syntactic subjects in the early speech of American and Italian children'. *Cognition*, 40, 21-81
- Whitman, J; K. Lee and B. Lust. (1990) 'Continuity of universal grammar in first language acquisition: the issues of functional categories'. *Proceedings of NELS 1990*
- Zhu Dexi (1982) *Yufa Jiangyi* (Lecture on Syntax). Beijing: Commercial Press

Appendix 1: Example Sentences

- (1) 在劍橋。
- (2) 在劍橋上學。
- (3) 住在劍橋。
- (4) 在劍橋住。
- (5) 佢喺/响屋企做功課。
- (6) 佢臨走送俾我嘅。
- (7) 趁未黑好走啦。
- (8) 我喺/由澳門嚟嘅。
- (9) 灣仔隔/離西環有幾遠?
- (10) 而家隔咗打仗嗰陣時已經五十幾年啦。
- (11) 挨住條大渠行就會到㗎啦。
- (12) 十年內未見過佢。
- (13) 食咗之後就肚痛囉。
- (14) 向/對住我坐。
- (15) 你又同佢講乜嘢啊?
- (16) 對呢件事我冇意見。
- (17) 佢行到呢度已經有晒氣。
- (18) 我打算响英國開餐館。
- (19) 佢做到十點。
- (20) 我由朝做到晚啊。
- (21) 我送本書俾佢。
- (22) 我送本書俾佢睇。
- (23) 本書我已經交咗俾佢。
- (24) 佢咁辛苦都係為咗啲細路。
- (25) 先生同啲學生補課。
- (26) 我俾佢鬧咗一餐。
- (27) 件衫俾佢整污糟咗。
- (28) 俾佢攪掂咗。
- (29) 俾/由/由得佢講啦。
- (30) 我唔會因為佢惡就怕咗佢。
- (31) 信封係用/擺打字機打嘅。
- (32) 我用/擺手食。
- (33) 我擺/拎佢嚟做板。
- (34) 佢當我係傻瓜。
- (35) 跟住佢講咁做。
- (36) 除咗佢個個都唔贊成。
- (37) 你同/跟我一齊去啦。
- (38) 你同我好，我又會同你好啫。
- (39) 佢同你一樣咁高。
- (40) 佢高過你。
- (41) 呢樣比較好。
- (42) 好似佢咁咪得囉。
- (43) 我將本書送/擺咗俾個學生。
- (44) 佢將啲頭髮染咗紅色。
- (45) 我已經將份報告講得好清楚。
- (46) 你連咁淺嘅字都唔識。
- (47) 對於天文學我一啲都唔識。
- (48) 講得好。

- (49) 驚到我唔識嘅。
- (50) 你跟唔跟我去? 跟。
- (51) M: 就是這個學生我送了一本書給。
C: 就係呢個學生我送咗一本書俾。
- (52) 喺咗劍橋住已經兩年啦。
- (53) 俾姐姐。
- (54) 俾家琦食咗。
- (55) 攞個紙袋俾我啊。
- (56) 開俾你。
- (57) 拆咗俾我睇吓先。
- (58) 媽咪買兔兔俾家琦。
- (59') 兔兔, 媽咪買 0 俾家琦。
- (59'') 家琦, 媽咪買兔兔俾 0。
- (59) 拆咗俾家琦睇先。

Appendix 2: Child Data

- [1] 俾。
- [2] 俾姐姐。
- [3] 俾家琦先啊。
- [4] 唔俾啊。
- [5] 媽咪俾紙。
- [6] 我要俾。
- [7] 玉卿姐姐呢個家琦。
- [8] 俾鹽先好。
- [9] 一陣俾。
- [10] 我頭先俾佢啲錢錢呢?
- [11] 等陣間, 我俾公仔你。
- [12] 第日俾。
- [13] 俾咗錢可以過。
- [14] 唔俾媽咪睇。
- [15] 呢啲俾姐姐睇咗啦。
- [16] 俾我接住咗。
- [17] 唔俾睇。
- [18] 一陣俾爹咁食。
- [19] 我俾呢隻你食。
- [20] 俾我同佢飲。
- [21] 你俾我出嚟講嘢啦。
- [22] 攞俾。
- [23] 開俾家琦啊。
- [24] 媽咪買兔兔俾家琦。
- [25] 攞紙袋俾我啊。
- [26] 攞俾家琦飲。
- [27] &bat1 啲俾我飲啊。
- [28] 切瓜瓜俾家琦食瓜瓜哦。
- [29] 開門俾嫲嫲入去喎。
- [30] 拆咗俾家琦睇先。
- [31] 有個故仔講俾媽咪聽囉。

Appendix 3

English gloss of the Cantonese sentences

- [1] give
- [2] give sister
- [3] give Gakei first
- [4] not give
- [5] mummy give paper
- [6] I want to give
- [7] Sister Yuk-hing give this to Gakei
- [8] you'd better put in salt first
- [9] will give in a while
- [10] where is the money I have just given him
- [11] in a while I will give you the doll
- [12] another day give
- [13] given money can go
- [14] don't let mummy see
- [15] this is already seen by sister
- [16] caught by me
- [17] not let you see
- [18] in a while let daddy eat
- [19] I give this thing to you to eat
- [20] let me help it drink
- [21] you let me come out to talk
- [22] get give
- [23] open for Gakei
- [24] mummy bought a toy rabbit for Gakei
- [25] get a bag for me
- [26] get for Gakei to drink
- [27] scoope some for me to drink
- [28] cut melon for Gakei to eat melon
- [29] open door for grandma to go in
- [30] dissemble for Gakei to see first
- [31] have a story to tell for mother to hear

Focus Adverbs in Child Cantonese*

Thomas Hun-tak Lee

Introduction

Focus particles such as *even* and *only* have been extensively studied for languages like English and German, and their semantic properties are relatively well understood (cf. inter alia, Horn 1969, Fraser 1971, Jackendoff 1972, Karttunen and Peters 1979, Rooth 1985, Kay 1990, König 1981, 1991). Focus devices contribute to the meaning of the sentence by highlighting a particular constituent, and at the same time relating it to alternatives that may replace the constituent in the sentence. The highlighted constituent represents the focus whereas the set of alternatives brought into consideration by the focussing device constitutes the domain of values the focus is related to.

To illustrate this notion, consider (1). In this sentence, the particle *only* relates the individual denoted by the focus (*John*) to other individuals who may have come, by asserting that if the latter exist at all, they must be identical to the former, hence the uniqueness claim of (1'') (cf. Horn 1967, Rooth 1985). In (2), adopting the analysis of Karttunen and Peters (1979), one can think of the particle *even* as relating the individual denoted by the focus (*John*) to its alternatives by ranking them on a scale of likelihood of their coming. The formula (2'') suggests that John is the least likely of the individuals who could have come.

(1) Only John came.

(1') John came.

(1'') For all x such that x came, $x = \text{John}$.

(2) Even John came.

(2') John came.

(2'') For all $x \neq \text{John}$, such that x came, the likelihood of x 's coming exceeds the likelihood of John's coming.

The mechanism of highlighting achieved by focus particles can be viewed as a process of variable introduction. On this view, focus particles are operators that introduce variables in the focus position of the sentence, thus creating an open sentence. The open sentence is called the scope of the operator (Karttunen and Peters 1979), and serves to restrict the latter's domain of quantification (Rooth 1985). For example, the effect of *only* in (1) is to create an open sentence ' x came' by replacing the focussed constituent *John* with a variable. This open sentence, which forms the scope of *only*, restricts the universal quantifier in (1'') by limiting the domain to the set of individuals who satisfy the open sentence ' x came'. Similarly, the focus particle *even* creates an open sentence ' x came'. It restricts the operator 'for all $x \neq \text{John}$ ' to the set of individuals distinct from John who came. It has been observed that the contextual restriction of focus operators would need to be further delimited by discourse. Thus, in

'only John came', one normally may not consider all individuals who came, but only those relevant to the discourse at hand (cf. Rooth 1985).

As observed in König (1991), a focus particle never reverses the truth value of the sentence it gets attached to. A sentence containing a focus particle always entails the same sentence without the particle. Thus (1) entails (1'), and (2) entails (2'). The main function of focus particles is to convey other meanings by association with a constituent in the sentence, as illustrated above. The meaning contributed by the focus particle may underlie the assertion of the sentence or may form part of the presupposition. With focus adverbs like *only*, the uniqueness condition it imposes, given in (1''), clearly counts as the truth condition of the sentence. If (1'') is violated, (1) will be falsified. On the other hand, for adverbs like *even*, the implication of relative likelihood indicated in (2'') does not play as important a role as (2') and is thus analysed as a presupposition or conventional implicature (cf. Karttunen and Peters 1979).

König (1991) has proposed a useful taxonomy for subdividing focus particles. Some particles can be termed inclusive or additive in that the alternatives brought into consideration may also satisfy the open sentence when substituting for the variable. Other particles are termed exclusive or restrictive, since they exclude alternatives from satisfying the open sentence. In this taxonomy, *only* will be considered an exclusive operator, and *even* an inclusive operator. One may also divide focus particles into scalar particles, ie. those that induce a pragmatic scale for ranking alternatives, and non-scalar particles, which do not rank alternatives. In such a subclassification, *even* will be considered a scalar operator, since it introduces a scale of likelihood into the picture and ranks the focussed constituent in relation to its alternatives on that scale. On the other hand, *only* introduces no such scale of ranking and is thus non-scalar.

The acquisition of focus particles is of interest to the child language researcher, because as shown in (1'') and (2''), focus particles carry with them bound variable structures. Given the fact that the age at which bound variable structures emerge has been a matter of controversy (cf. Roeper 1986), children's knowledge of focus particles will provide another window on the issue of bound variables. Focus particles involve logical properties: they impose specific conditions on the range and domain of quantification of the operator, and enter into scope relations with other operators in complex ways (cf. Karttunen and Peters 1979, Rooth 1985). Whether children handle scope phenomena involving focus particles in ways similar to or different from how they handle quantifier scope will enrich our understanding of the development of logical scope. Third, as we have seen, the bound variable structures encoded by some focus particles are expressed in the form of presuppositions or conventional implicatures (e.g. *even*), and scalar focus particles imply pragmatic scales that are dependent on speaker/hearer's knowledge of the world. Thus the acquisition of focus particles could also be seen as a component of the acquisition of pragmatics. It would be interesting to see, for instance, if the logical properties of focus particles develop differently from properties related to pragmatic scales.

In this paper, I report preliminary findings on the acquisition of four inclusive focus adverbs: *zoi3* (again), *jau6* (also/again), *zung6* (also/still), and *dou1* (also) by two Cantonese-speaking children between two and three years of age. As we will see, while these adverbs impose intricate conditions on direction and range of quantification, three of them are used productively before or by three years of age. I will first review the work that has been done on Chinese focus adverbs. Then I will outline the semantic properties of the Cantonese inclusive adverbs relevant to the acquisition research, and then present the longitudinal findings.

The Study of Chinese Focus Adverbs

In the study of Chinese focus adverbs, a great deal of attention has been paid to *jiu* (then; only) and *cai* (only-if; only-then), as the meanings of these adverbs are complex and involve pragmatic scales that cut across the dimensions of time, quantity and conditional implication. Thus, it is well known that *jiu* denotes recency, smaller-than-expected quantity, and sufficient condition, whereas *cai* is used to mark lateness, larger-than-expected quantity, and necessary condition (cf. Ding 1961).¹ Apart from lexicographic studies of these two adverbs (notably Ding 1961, Tsao 1976, Lu 1980), efforts have also been made to explicitly characterize their semantic representations. This line of research is best exemplified by Paris (1987) and Biq (1987), who provide a useful account of the meanings of *jiu* and *cai* in terms of the interaction of the representations and pragmatic principles governing implicature. These researchers have also attained some success in explaining the distributional characteristics of the two adverbs, in particular why the contrast between them becomes neutralized in some contexts and why the ordering in the scale may be reversed depending on the relative order of the focus adverb and the focussed constituent.

A number of studies have examined the semantics of inclusive adverbs such as *ye3* (also), *you4* 'also/again', and *zai4* (again). In her study of the *lian..ye/dou* construction, Paris (1979) proposes that a sentence of the form 'NP *ye3* VP', where NP is the focus which denotes an individual, asserts 'NP VP', and presupposes that there is some x other than the referent of the subject NP such that 'x VP'. Ma (1982) examines a range of examples in which *ye3* serves as a constituent connector, and attributes its various uses to the property of 'similarity of kind', a notion which is not further clarified. Biq (1989) argues that the juxtaposition, inclusion and downtoning uses of *ye3* can be analyzed as manifestations of the core property of *ye3* at three different levels: the levels of propositions, epistemic modality and speech acts. (Biq 1988) argues that the inclusive adverb *you4* (also/again), when applied to an entity, should be viewed as an operator that accumulates different yet relevant properties of the entity. Thus a sentence like *Lisi you4 shengqi le* (Lisi-again-angry-asp) is said to assert that Lisi is angry, and conveys the implicature that Lisi was also angry at an earlier point in time. The generalization given by Biq is shown in (3):

$$(3) \text{ you4 } v(s) \iff v_t(s) \ \& \ \exists v_{t-1} v_{t-1}(s)$$

(s=entity; v=property; t=speech time; t-1= time prior to speech time)

Ma (1985) examines the reiterative and non-iterative uses of the adverb *zai4* (again), and challenges the claim that the reiterative use of the adverb is bound to irrealis situations. Ma makes interesting observations about the location of stress and the scope of the adverb: a sentence of the form 'NP *zai4* VP' will have stress falling on the adverb if VP is part of the presupposition; otherwise stress will shift to the VP. In their detailed analysis of *zai4*, Renaud and Luo (1987) observe that the reiterative use of *zai4* requires the subject of the clause to be agentive; however, this agentivity requirement can be suspended if the adverb falls within the scope of a modal or negation.

Below, I present an analysis of the four Cantonese adverbs drawing from the insights of earlier analyses of Mandarin Chinese as well as studies of Cantonese (cf. Cheung 1972, Matthews and Yip 1994). The Cantonese adverbs *zoi3* (again), *jau6* (also/again), *zung6* (also/still), *dou1* (also) are concerned with the reiteration or addition of events.² All of them are nonmovable adverbs occurring in the position between the subject and the verb phrase.³ Semantically, they show an interesting distribution with regard to the range of positions they focus on. *Zoi3* and *jau6* are used to focus on time to signal reiteration of events; *dou1* can only focus on the subject or topic of the sentence; *zung6* focusses on the VP and anything within it; *jau6* is subject to the least restriction in being able to focus on time, the subject, the object, the verb phrase, as well as a full clause.

Zoi3 (again)

Zoi3 is an operator that focuses on time intervals to mark reiteration. In (4a), the use of the adverb presupposes or conventionally implicates that Johnny threw in a bet at an earlier time than speech time. The sentence asserts that he is doing it again tomorrow. The presupposition of the sentence is preserved under negation and conditional modality, as (4b) and (4c) also presuppose that Johnny threw in a bet at an earlier time.

- (4) a. Ting1jat6, Johnny zoi3 lok6-zyu3.
 tomorrow Johnny again throw-bet
 'Johnny will bet again tomorrow'.
 b. Ting1jat6, Johnny m4 zoi3 lok6-zyu3.
 tomorrow Johnny not again throw-bet
 'Johnny will not bet again tomorrow'.
 c. Jyu4gwo2 Johnny zoi3 lok6-zyu3, ngo5 zau6 m4 coi2 keoi3.
 if Johnny again throw-bet I then not pay-attention-to him
 'If Johnny bets again, I will ignore him'.

The sentences in (4) show that the adverb can be used to denote reiteration of events at a future reference time. (5) shows that the adverb can also be used to signal reiteration of events at an earlier time than speech time. In this situation, the sentence presupposes that an identical event occurred at a time earlier than the reference time.

Thus (5) presupposes that Johnny had opened a book before the recent past when he opened it again. The adverb *zoi3* is thus not restricted to irrealis contexts.⁴

- (5) Tau4sin1, Johnny zoi3 da2-hoi1 bun2 syu1.
 Just-now Johnny again open cl book
 "Just now, Johnny opened the/a book again".

A peculiarity should be noted about the reiteration of events by *zoi3*: it demands that the type of asserted event and the agent of the asserted event be identical to those of the presupposed event. However, it requires only identity of type for the object participant. In other words, (5) is ambiguous in having a second reading in which the object 'CL-book' is non-specific. The sentence then would mean 'Just now Johnny opened a book again', which is equivalent to 'Just now, Johnny opened another book'.

As in Mandarin (cf. Renaud and Luo 1987), an agentivity condition is imposed on the subject of the sentence by *zoi3*. In (4-5), the subject of the sentence refers to an agentive or volitional entity. The sentences in (6), however, contain non-agentive subjects and are thus ungrammatical. The agentivity restriction can be relaxed if the adverb comes under the scope of a modal or irrealis operator, as can be seen from the well-formedness of the corresponding examples in (7).

- (6) a. *kam4jat6, keoi5 zoi3 beng6 zo2.
 yesterday, s/he again ill asp
 'Yesterday, s/he fell ill again'.
 b. *soeng6 sing1kei4, ji1 tou3 hei3 zoi3 soeng2jin2 zo2 gei2 jat6.
 last week this cl film again show asp several day
 'Last week, this film was shown again for several days'.
 (7) a. keoi5 wui3 zoi3 beng6.
 s/he will again ill
 'S/he will fall ill again'.
 b. ji1 tou3 hei3 ha6 sing1kei4 (wui3) zoi3 soeng2jin2.
 this cl film next week (will) again show
 'Next week, this film will be shown again'.

Given the agentivity requirement of *zoi3*, it is not surprising that the adverb can occur in imperative sentences whose subjects must represent volitional agents, as in (8). The examples in (9) indicate that the adverb is generally not compatible with states, unless the latter are understood in an inchoative sense.

- (8) nei zoi3 da2-hoi1 bun2 syu1 al.
 you again open cl book sfp
 '(Would you) open the book again?'.
 (9) a. *keoi5 zoi3 sau3.
 s/he again skinny
 'She is skinny again'.

- b. *keoi5 zoi3 zung1ji3 jau4wa2.
 s/he again like oil painting
 'She likes oil paintings again'.

It should be noted that *zoi3* can be used without a reiterative sense, in which case it signals temporal succession. Thus, the sentences in (10) do not denote reiteration of the event identified by the second clause. Rather, the second clause signals an event subsequent to that denoted by the first clause. As the child language data do not evidence any such use of the adverb, our discussion will be limited to the reiterative uses of *zoi3*.⁵

- (10) a. Johnny laau6-jyn4 Mary, Paul zoi3 laau6 keoi5.
 Johnny scold-finish Mary, Paul also scold her
 'After Johnny finished Mary, Paul-also scolded her'.
 b. Sin1saang1 gaau3jyn4 jing1mān2, zoi3 gaau3 sou3hok6.
 teacher teach-finish English, also teach maths
 'After the teacher finished teaching English, he taught maths'.

Based on the preceding discussion, it is clear that the adverb *zoi3* quantifies a time variable and relates it to alternative times or time intervals. One may therefore give the representation (11) for the focus adverb in the manner of Parsons (1990).⁶

- (11) A sentence of the form 'NP_j *zoi3* VP' carries with it an assertion and a presupposition:

assertion: $\exists e$, agent (j, e), VP (e), culminate (t, e)

presupposition: $\exists t' < t \quad \exists e' [\text{agent (j, e'), VP (e'), culminate (t', e')}]$

The above representation says that a sentence of the form 'NP *zoi3* VP' asserts that there is an event *e*, whose agent is the referent of the subject of the sentence (*j*), the event is of the type denoted by the VP, and the event culminates at time *t*. The presupposition or conventional implicature conveyed by the sentence is that there is another time *t'*, prior to *t*, at which an event *e'* of identical type to that of the asserted event occurred, whose agent was also identical to the agent of the asserted event. The adverb *zoi3* introduces time and event variables into the sentence. It restricts the domain of quantification to the time intervals before the time *t* at which the asserted event occurs. In this way, it relates the time of the asserted event to other time intervals at which events of identical type with identical agents could have been performed.

Jau6 (also/ again)

As remarked earlier, *jau6* is the least constrained of the four adverbs in that it imposes no restriction on the direction of quantification and the type of constituent to be focussed on. It can focus on time to mark reiteration of events, as illustrated in (12). The examples also show that *jau6* could be used to signal events reiterated either

at a time prior to or following speech time. The adverb is free from any agentivity constraint, since (12b) and (12c) involve non-volitional subjects.

- (12) a. Ting1jat6, Johnny jau6 lok6-zyu6.
tomorrow, Johnny again throw-bet
'Johnny will bet tomorrow again/'JOHNNY will also bet tomorrow'.
b. Kam4jat6, keoi5 jau6 beng6 zo2.
yesterday, s/he again ill asp
'Yesterday, s/he fell ill again/'Yesterday, S/HE also fell ill'.
c. Soeng6 sing1kei4, jau6 lok6 zo2 coeng4 syut6.
last week again fall asp cl snow
'last week, a snow fell again'.

Jau6 can focus not only on time, but also on the subject, the object, the verb phrase or the entire subject predicate construction. Consider the following sentences. In (13a), the subject is focussed, as alternatives are introduced in the subject position. *Jau6* focusses on the object in (13b), and the verb phrase in (13c). In (13d), the alternative consists of both the subject and the verb phrase, with the adverb functioning very much like a sentential connector. The data suggest that of the four adverbs studied, *jau6* has maximal range: it can focus on any constituent of the sentence.

- (13) a. ngo5 da2 zo2 go3 din6wa2, keoi5 jau6 da2 zo2 go3 din6wa2.
I make asp cl phone-call, s/he also make asp cl phone-call
'I made a phone call, s/he also made a phone call'.
b. Johnny maai2 zo2 zi1 bat1, jau6 maai2 zo2 go3 bat1hap2.
Johnny buy asp cl pen, also buy asp cl pen-box
'Johnny bought a pen, and also bought a pen box'.
c. keoi5dei6 cel1tin1cel1dei6, jau6 daai6seng1 coeng3go1.
they bull-shit also loudly sing
'They had a bull session, and also sang loudly'.
d. Johnny ci4 zo2 gung1, Paul jau6 beng6 zo2.
Johnny resign asp job, Paul also ill asp
'Johnny resigned from his job, Paul has also fallen ill'.

Jau6 differs from *zo13* in being compatible with any type of verb. It can cooccur with stative as well as non-stative verbs (cf. 14a, b). If *jau6* occurs in a sentence with a stative verb, it will not signal reiteration unless an inchoative meaning is intended. This can be seen from (14c).

- (14) a. Keoi5 hai6 ji1sang1, ngo5 jau6 hai6 ji1sang1.
s/he be doctor, I also be doctor
'S/he is a doctor, I am also a doctor'.
b. Johnny hou2 cute, jau6 hou2ci2 Yves Montand
Johnny very cute, also very like Yves Montand
'Johnny is very cute, and also resembles Yves Montand'.

- c. Johnny jau6 sau3 zo2.
 Johnny also skinny asp
 'Johnny has become skinny again'.

The semantic representation of the focus operator *jau6* will depend on which constituent is being focussed. If we consider a case in which the verb phrase is focussed on (e.g. 13c), the following would be a representation of the meaning of the focus adverb.

(15) A sentence of the form 'NP_i *jau6* VP', where VP is the focus, carries an assertion and a presupposition:

assertion: $\exists e$ agent (j, e) and VP (e) and culminate (t, e)
 presupposition: $\exists VP' \exists e'$, such that agent (j, e') and VP' (e') and culminate (t', e')
 where $VP' \neq VP$

It should be pointed out that the adverb has other non-focus uses. One such use which is relevant to our child language analysis is its use to mark unexpected new situations. (16) can be uttered in a situation in which the truth of the sentence without the adverb (ie 'her/his ignoring me') comes as unexpected.

- (16) Keoi5 jau6 m4 coi2 ngo5
 s/he JAU not pay-attention-to me
 'S/he does not pay attention to me'

Zung6 (also/still)

The focus adverb *zung6* is used in two main ways. It functions as an inclusive operator similar to *jau6*, but with its range limited to the verb phrase. (17 a, b) show that the adverb may focus on the verb phrase or the object, and (17c) is evidence that it cannot focus on the subject.

- (17) a. Johnny caau2 zo2 dip6 gaai3laan2, zung6 zyu2 zo2 bou1 tong1.
 Johnny fry asp dish broccoli, also cook asp pot soup
 'Johnny fried a dish of broccoli (and) also cooked a pot of soup'.
 b. Johnny caau2 zo2 dip6 coi3, zung6 caau2 zo2 dip6 ngau4juk6.
 Johnny fry asp dish vegetable, also fry asp dish beef
 'Johnny fried a dish of vegetables, (and) also a dish of beef'.
 c. *Johnny sik6 zo2 joek6, Paul zung6 sik6 zo2 joek6.
 Johnny eat asp medicine, Paul also eat asp medicine
 'Johnny took (his) medication, Paul also took (his) medication'.

Zung6 shows the properties of a non-scalar operator, and should not be interpreted as something akin to 'even', since it does not signal either lesser or greater degree of likelihood or even smaller or greater informational value (cf. Karttunen and Peters 1979, Kay 1990). Consider the (a) sentences in (18-19) in which there is good reason

to consider the event in the second conjunct to rank higher than that in the first conjunct with respect to some scale of seriousness or value. For instance, one may consider 'breaking a leg' more serious than 'hurting one's finger'; one may also see 'buying a flat' as involving greater monetary value than 'buying a car'. The corresponding (b) sentences show, however, that it does not matter how one orders the conjuncts, suggesting that the adverb does not carry with it a scale that ranks alternatives in relation to the focussed element.⁷

(18) a. Johnny zing2 can1 zek3 sau2zi2, zung6 tyun2 zo2 zek3 goek3.

Johnny hurt asp cl finger, also break asp cl leg

'Johnny hurt his finger, and also broke his leg'.

b. Johnny tyun2 zo2 zek3 goek3, zung6 zing2 can1 zek3 sau2zi2

Johnny break asp cl leg, also hurt asp cl finger

'Johnny broke his leg, and also hurt his finger'.

(19) a. Keoi5 maai2 zo2 bou6 ce1, zung6 maai2 zo2 can4 lau2.

s/he buy asp cl car, also buy asp cl flat

's/he bought a car, and also bought a flat'.

b. Keoi5 maai2 zo2 can4 lau2, zung6 maai2 zo2 bou6 ce1.

s/he buy asp cl flat, also buy asp cl car

'S/he bought a flat, and also bought a car'.

When *zung6* occurs in imperfective sentences, it functions as an aspectual operator signalling continuation of an activity or state. As an aspectual operator, *zung6* presupposes a transition from a temporal phase in which a proposition *p* is true to a phase in which *p* is false. The sentence containing *zung6* asserts that the reference time falls within the phase in which *p* is true; that is to say, the transition has not occurred (cf. Lobner 1989). *Zung6* is compatible with activities and states, as shown in (20a,b), but generally not with accomplishment or achievement sentences, unless the latter can be modified by imperfective aspect markers. Thus, the accomplishment sentence given in (21a) cannot have a continuative reading, whereas the activity sentence in (21b), which does not have the boundedness signified by the classifier, is compatible with the aspectual reading. The continuative reading is permitted in accomplishment sentences if an imperfective marker such as progressive aspect *gan2* is included, as evidenced in (21c).

(20) a. Johnny zung6 jam2 zau2.

Johnny still/also drink wine

'Johnny still/also drinks wine'.

b. Johnny zung6 hai6 hai6zyu6jam6.

Johnny still/also be chairperson

'Johnny is still/also a chairperson'.

(21) a. *Ngo5 zung6 wan2 go3 gung1jan4 ga3

I still find cl worker sfp

'*I am still looking for a worker'.

- b. Ngo5 zung6 wan2 gung1jan4 ga3
 I still find worker sfp
 'I am still looking for a worker'.
- c. Ngo5 zung6 wan2 gan2 go3 gung1jan4 ga3
 I still find prog ct worker sfp
 'I am still looking for a worker'.
- (22) a. Keoi5 zung6 jau2 saam1 man1.
 s/he still have three dollar
 'S/he still has three dollars/ she has three dollars/left'.
- b. Ngo5 zung6 jiu3 saam1 man1.
 I still want three dollar
 'I want three more dollars'.

When aspectual *zung6* attaches to verbs like *jau2* 'have' or *jiu3* 'want' taking an amount object, the object signifies the amount that delimits the continuation of the state. For example in (22a), the adverb signals that the situation of the his/her owning the object (i.e. money) persists, but the continuation of the state is delimited by the amount indicated in the object NP (i.e. three dollars). If that amount is measured out, ie 'three dollars', then the continuation of the state will cease. Similarly, in (22b), the amount phrase in the object circumscribes the continuation of the state of wanting. Once the three dollars are given, then by implication, the state of wanting will cease.⁸

Dou1 (also)

As an inclusive operator, the focus adverb *dou1* quantifies only leftward, and focusses on the subject of the sentence. This is shown by (23): the (a) sentence is grammatical because the focus falls on the subject. In contrast, the (b) sentence is ruled out because the intended focus falls on the verb phrase.

- (23) a. Ngo5 jam2 zo2 ca4, keoi5 dou1 jam2 zo2 ca4.
 I drink asp tea, s/he also drink asp tea
 'I've been to the teahouse, s/he has also been to the teahouse'.
- b. *Ngo5 jam2 zo2 ca4, dou1 maai2 zo2 sung3.
 I drink asp tea, also buy asp food
 'I have been to the teahouse, and has also bought food'.

As *dou* can only quantify an argument to its left, what is to the right of it falls outside its range. It is therefore not subject to any of the event type restrictions the other three adverbs are subject to.

The semantic representation of *dou1* is given in (24).

(24) A sentence of the form 'NP_j dou1 VP' carries an assertion and a presupposition:

assertion: NP_j VP
 presupposition: $\exists x \neq j$, such that NP_x VP is true.

The adverb *dou1* has other major uses which should be mentioned. Like its cognate in Mandarin, it is a universal quantifier quantifying the subject. It can also be used in emphatic negation contexts such as (25). The sentences in (25a) can be uttered if the speaker denies an earlier expectation that the film may be good. Likewise, (25b) denies an earlier expectation that the speaker may like the bouquet of roses s/he is presented with.⁹ Table 1 summarizes the focus features of the four adverbs.

- (25) a. *ji1 bou6 hei3 dou1 m4 hou2 tai2 ge2.*
 this cl film emp not good watch sfp
 'This film is not good to watch' (denial of expectation).
 b. *Ngo5 dou1 m4 zung1yi3 mui4gwai3fa1 ge2.*
 I emp not like rose sfp
 'I don't like roses' (denial of expectation).

Table 1: Semantic Properties of the Cantonese Focus Adverbs
Zoi3, Jau6, Zung6, Dou1

Adverb	Range of focus	Temporal /atemporal meanings	Agentivity of subject	Realis	Activity (sik6 jin1 'smoke')	Accomplishment (sik6 zi1 jin1 'smoke a cigarette')	Achievement (zing2 laan6 'do-break'= damage)	States (leng3 'pretty'; sik1 sik6 jin1 'know how to smoke')
<i>Zoi1</i>	time	again	+	±	+	+	+	-
<i>Jau6</i>	S	again	±	±	+	+	+	-
		also	±	±	+	+	+	+
<i>Zung6</i>	VP	also	±	±	+	+	+	+
		still	±	±	+	+(with imperfective aspect)	-	+
<i>Dou1</i>	Subject or topic	also	±	±	+	+	+	+

Focus adverbs in Cantonese child language

The data come from the longitudinal recordings of two Cantonese-speaking children: Gakei, observed during 1;11-2;09 for 19 sessions, and Tinfaan, observed from 2;02 through 3;01 for 12 sessions. The analysis is based on the spontaneous utterances of the two children, the frequencies of which are given in Table 2. In this tabulation, an occurrence was considered spontaneous if in the preceding two conversational turns the adverb was not produced by the adult with the same predicate or argument, or with the same predicate and argument if both were present, unless the combination was initiated by the child in that stretch of conversation. An occurrence

was also considered spontaneous if it was a grammatically appropriate response to an A-not-A or wh-question. All adult tokens were assumed to be spontaneous.

It is clear that Cantonese-speaking children before three were productive in the use of focus adverbs denoting reiteration or inclusion (cf. Tse et al 1991 for similar remarks on early occurrence of presuppositional adverbs in Mandarin). *Zoi3* ('again') is the least productive of the four adverbs. It did not occur at all in the speech of Gakei. With the other child, Tinfaan, the frequency of occurrence of *zoi3* vis-a-vis other adverbs was the lowest. Table 3 gives the age at which the adverb has been used with four or more distinct verbs in spontaneous utterances. On this count also, *zoi3* proves to be the least productive, with both children not reaching criterion. As for the other adverbs, Gakei used *jau6* ('also/again') and *zung6* ('also/still') before *dou1* ('also'), whereas Tinfaan used *jau6* and *dou1* before *zung6*. One may conclude from this that *jau6* was the easiest for the children.

The difficulty posed by *zoi3* may have to do with the fact that semantically it is the most restricted of the four adverbs, as it requires identity of agency, event type and theme type, and also imposes restrictions on the cooccurrence of aspect and verb semantics. By the same reasoning, one could see why *jau6*, semantically the least restricted adverb, was the earliest to emerge for both children. While it is true that a degree of correlation can be established between the frequency of adult tokens and that of child tokens in our data (cf. Table 2), input frequency cannot in general be a decisive factor from what we know about the acquisition of abstract linguistic structures.

Table 2: Number of spontaneous occurrences of the focus adverbs *zoi3*, *jau6*, *zung6*, *dou1* in two Cantonese-speaking children

	<i>Zoi3</i>	<i>Jau6</i>	<i>Zung6</i>	<i>Dou1</i>
Gakei (1;11-2;09)				
child	0	54	32	38
adult	20	340	87	173
Tinfaan (2;02-3;01)				
child	3	48	20	48
adult	33	140	180	126

Table 3: Age at which the spontaneous use of focus adverbs *zoi3*, *jau6*, *zung6*, *dou1* is productive (cooccurrence with 4 or more distinct verbs)

Child	<i>Zoi3</i>	<i>Jau6</i>	<i>Zung6</i>	<i>Dou1</i>
Gakei	-	2;03;18	2;02;28	2;08;08
Tinfaan	-	2;09;07	3;00;20	2;07;05

To get a clearer picture of how each child used these focus adverbs, their utterances were classified into a number of categories depending on the type of constituent they focus on, based on contextual information: focus on time (realis, volitional), focus on time (realis, non-volitional), focus on time (irrealis), focus on subject, focus on object, focus on VP, aspectual use (continuation of activity), aspectual use (continuation of state), and other uses. Below, I define and illustrate each of the focus and aspectual categories.

Focus on time (realis, volitional)

An event occurred prior to speech time involving a volitional agent and a theme object. The child utters a sentence containing a focus adverb signalling that an event of identical type has occurred with the subject being the same volitional agent and the theme object being of the same type.

(26) (Gakei and the investigator are playing with a toy phone. The child puts some coins into it and then tries to take some out. Investigator helps the child take out the last coin from the toy phone. Gakei at 2;03;18).

INV: &e1 lo2dou2 lo1. hai6 mai6?
 excl. take-succeed sfp. be not-be
 'Wow, we've taken this out, haven't we?'
 CHI: jau6 lo1dou2
 again take-succeed
 '(we)'ve taken one out again'.

Focus on time (realis, nonvolitional)

An event occurred at an earlier time, with the subject being a non-volitional entity. The child utters a sentence containing a focus adverb, asserting that an event identical in type has occurred in which the subject NP refers to the same non-volitional entity.

(27) (Gakei wants to open the door. A few seconds later, some books fall on the floor. Investigator helps child put books back onto the table. Some books fall down again. Gakei at 1;11;29).

INV: &i1, dit3 zo2
 excl. fall asp
 '(the books) fell.
 exp: The books fall again.
 INV: &ou3: [=! puts them back on the table].
 CHI: jau6 dit3 zo2
 again fall asp
 '(they) dropped again'.

Focus on time (irrealis, volitional)

An event occurred at an earlier time involving a volitional agent and a theme object. The child utters a sentence containing a focus adverb, asserting that an event of identical type will occur with the same agent and the same object or an object of identical type.

(28) (Investigator was asking Tinfaan questions about picture cards earlier. They are now talking about the Christmas tree in the child's home. Child says she wants to pinch leaves off the tree. Investigator tells her not to. Child takes out some picture cards asking the investigator to ask her questions again. Tinfaan at 2;11;16).

CHI: nei2 zoi3 man6 ngo5 a1

you again ask me sfp

'(will) you ask me again?'

com: child is getting out some picture cards.

INV: jidi1 tai2 zo2 ga3 la3, m4 man6 ga3 la3

these see asp sfp sfp, not ask sfp sfp

'(we)'ve seen these already, (I'm) not asking (about these)'

Focus on subject

A property encoded in a verb phrase has been predicated of an entity in the preceding discourse. The child utters a sentence containing a focus adverb asserting that the property encoded by the same verb phrase is predicated of a different entity represented by the subject of the sentence.

(29) (Tinfaan and investigator are looking at pictures. Tinfaan at 2;09;07).

INV: jyut6loeng6 hai2 dou6 zou6 mat1je2?

moon be there do what

'What is the moon doing there?'

CHI: fan3gaau3

sleep

INV: hai6 wo3

be sfp

'yes'

CHI: keoi5 jau6 fan3gaau3 [!= pointing at the little boy in the picture].

he also sleep

'he is also sleeping'.

(30) (Gakei and investigator are making objects of different shapes out of play-doh. Gakei at 2;09;09)

INV: ngo5 go3 jyun4 gwo3 nei5 wo3 daan3hai6

I cl round than you sfp but

'but mine is rounder than yours'

CHI: ngo5 jilgo3 dou1 jyun4 a3

I this cl also round sfp

'Mine is also round'.

(31) (Tinfaan tells investigator what she is afraid of. Tinfaan at 2;08;24).

CHI: ngo5 geng1 mou4po4 ga3 ngo5

I fear witch sfp I

'I am afraid of witches'

INV: ngo5 m4 geng1 ga3

I not afraid sfp

'I am not afraid (of witches)'

CHI: malmi4 dou1 m4 geng1

mom also not afraid

'Mommy is also not afraid (of witches)'

Focus on object

An event has occurred in which an agent has carried out an action on a theme object. The child utters a sentence containing a focus adverb, asserting that the same agent represented by the subject of the sentence has carried out or carries out the same action (indicated by the verb) on another theme object.

(32) (Tinfaan and investigator are playing with toy babies. Child is taking the toys to eat various things. Tinfaan at 2;02;10).

INV: nei5 daai3 bi4bil heoi3 sik6 mat1je5 a3?

you take baby to eat what sfp

'what are you taking the baby to eat?'

CHI: &ei6, syut3gou1, zung6 jau2 gwo2zap1, syu4tiu2, hai6 m4 hai6?

excl. icecream, also have juice, icecream-stick be not be

'(eat) icecream, also juice, icecream bar, right?'

Focus on verb phrase

An event has occurred in which an agent has performed an action signified by a verb phrase. The child utters a sentence containing a focus adverb, asserting that the same agent has performed or performs another action signified by a different verb phrase.

(33) (Investigator describing a story based on a picture of a boy kicking a ball at another boy. Tinfaan at 2;11;16).

CHI: da2 laam4kau4 zi1cin4 fei4zai2 lo2 keoi5 go3 bo1
play basketball before fat-boy take he cl ball
'Before playing basketball, the fat boy took his ball'

INV: &m6, gam2 keoi5 dim2joeng2 tek3 keoi3 ga3?
so he how kick him sfp

'So how does he kick him?'

CHI: gam2 fei3zai2 da2 keoi5 go3 tau4
so fat-boy hit he cl head
'so the fat boy hit his head'

INV: &haak6
excl.

CHI: zung6 da2 keoi5 go3 +... mo2 zyu6 keoi5 go3 tau4
also hit he cl touch asp he cl head
'(the fat boy) also hit...touched his-head'

Aspectual use (continuation of activity)

An activity has been going on. The child utters a sentence using a focus adverb indicating the activity still goes on or she would like the activity to go on.

(34) (Gakei watching a television program part one of which has finished. Mother asks child to turn off the TV. Child refuses. Gakei at 2;02;07).

CHI: m4 sik1 a3
not turn-off asp
'(I'm) not turning off (the TV)'

CHI: zung6 tai2 a3
still watch sfp
'(I) still (want to) watch (the TV)'

Aspectual use (continuation of state)

A state has been in existence. The child utters a sentence using a focus adverb indicating that the situation is still true at speech time.

(35) (Tinfaan tells the investigator that she has some sore spots on her mouth and it hurts. Tinfaan at 3;00;20)

INV: tung3 m4 tung3 ga3?
painful not painful sfp
'is it painful? /does it hurt?'

CHI: 0 [=! nods her head].

INV: ap6 tau2 a4?
nod head sfp
'(so you are) nodding (your) head?'

CHI: ji4ga1 zung6 tung3 ga3
 now still painful sfp
 'it still hurts'

Tables 4 and 5 show the spontaneous uses of various focus adverbs in each of the categories for the two children. The profile of Tinfaan given in Table 4 shows that when focussing on time to signal reiteration of events, she uses the two adverbs *jau6* ('again/also') and *zoi3* ('again'), the former for realis situations and the latter for irrealis situations. As we have seen earlier (cf. Table 1), the adult grammar does not limit *jau6* to realis contexts and *zoi3* to irrealis contexts. The child's conservatism in drawing this distinction may reflect some subtle semantic difference between the two adverbs in relation to the realis/irrealis distinction, one which finds expression in other Chinese dialects like Mandarin (cf. Renaud and Luo 1987). To focus on the subject, Tinfaan relies on two adverbs, *jau6* ('again/also') and *dou1* ('also'). The adverb *zung6* is used exclusively to focus on the object and the verb phrase, and to a limited extent for aspectual marking of continuation of activity or state. Non-focus uses are also found: Tinfaan uses *jau6* to signal unexpected events, and *dou1* for universal quantification or emphatic negation. The overall pattern suggests that the child sees a division of labor between the various adverbs and adheres to this division fairly rigidly.

Turning to the use of the adverbs by Gakei in Table 5, we do not see as neat a complementary distribution of the adverbs as we observed in Tinfaan. As *zoi3* ('again') has not yet emerged, the child uses *jau6* to signal reiteration of realis and irrealis events. To focus on the subject, Gakei uses primarily the adverb *dou1* ('also') and to a lesser extent the adverb *jau6* ('again/also'). She uses *zung6* ('also/still') and *jau6* ('again/also') for focussing on objects and verb phrases. Aspectual marking of continuation is achieved mainly by the use of *zung6* ('also/still') and minimally by *jau6* ('again/also'). Some non-focus uses of the adverbs are found in this period of Gakei's development: she uses *jau6* for unexpected events, and *dou1* to signal universal quantification and emphatic negation.

A number of differences can be observed between the development of the two subjects. First, the adverb *zoi3* ('again') has appeared in Tinfaan but not in Gakei. The former child associates this adverb with irrealis contexts and the adverb *jau6* ('again/also') with realis contexts. Second, Table 5 indicates clearly that Gakei's use of the adverb *jau6* ('again/also') is considerably more diverse than Tinfaan's. While this adverb is used to focus only on time and on the subject NP by Tinfaan, it is used additionally to focus on the object, the verb phrase and even to mark aspect. A third difference lies in their use of the adverb *zung6* ('also/still'): while Gakei shows quite a number of aspectual uses of the adverb, Tinfaan employs it mainly as an inclusive rather than aspectual operator.

Despite these differences between the two children, similar patterns can be found in their acquisition of the focus adverbs. Our findings have established beyond doubt that Cantonese speaking children are capable of using focus adverbs before three years of age. The adverb *zoi3*, the most restricted of the four adverbs, was the latest to appear,

Table 4: The Focus adverbs *zoi3*, *jau6*, *zung6*, *dou1* in Tinfaan (2;2-3;1)*

Type of Focus	Focus on time (realis; volitional)	Focus on time (realis: non-volitional)	Focus on time (irrealis; volitional)	Focus on subject	Focus on object	Focus on VP	Aspectual use (continuation of activity)	Aspectual use (continuation of state)	other uses
Focus adverb									
<i>zoi3</i> 'again'			2 (2;11;16)						
<i>jau6</i> 'also/again'	2 (2;10;18)	1 (2;06;01) 1 (2;10;18)		1 (2;03;02) 2 (2;08;24) 10 (2;09;07) 1 (2;10;18) 5 (2;11;16) 3 (3;00;20)					unexpected event 1 (2;10;18) 2 (3;01;21)
<i>zung6</i> 'also/still'					1 (2;02;10) 1 (2;06;10) 3 (2;07;05) 1 (2;10;18) 1 (3;00;20) 9 (3;01;21)	1 (2;11;16)		2 (3;00;20)	
<i>dou1</i> 'also'				2 (2;06;01) 5 (2;07;05) 2 (2;08;24) 3 (2;09;07) 2 (3;00;20) 8 (3;01;21)					emphasis/ univ. quant. 1 (2;04;27) 1 (2;08;24) 1 (2;09;07) 1 (2;10;18) 6 (3;01;21)

*The number in each cell represents the number of tokens for the particular adverb and category concerned, with the age of occurrence indicated in parenthesis (year; month; day). The number of categorizable tokens is smaller than the number of spontaneous occurrences of the adverbs.

Table 5: The Focus adverbs *zoi3*, *jau6*, *zung6*, *dou1* in Gakei (1;11-2;9)*

Type of Focus	Focus on time (realis; volitional)	Focus on time (realis; non-volitional)	Focus on time (irrealis; volitional)	Focus on subject	Focus on object	Focus on VP	Aspectual use (continuation of activity)	Aspectual use (continuation of state)	other uses
Focus adverb									
<i>zoi3</i> 'again'									
<i>jau6</i> 'also/again'	1 (2;03;18)	1 (1;11;29) 1 (2;02;07) 1 (2;02;21) 3 (2;04;30) 3 (2;08;18)	3 (2;04;30)	2 (2;03;25) 1 (2;04;08) 2 (2;05;03 topic)	1 (2;05;03) 1 (2;08;08)	1 (2;04;30)	1 (2;08;18)	1 (2;03;11) 1 (2;04;30)	unexpected event 1 (2;03;18) 1 (2;04;30) 3 (2;05;03) 1 (2;08;08)
<i>zung6</i> 'also/still'			1 (2;03;25)		1 (2;03;25) 1 (2;08;18)	1 (2;02;21) 1 (2;02;28)	3 (2;02;07) 1 (2;02;21) 1 (2;03;04) 1 (2;03;18) 2 (2;07;11)	1 (2;00;08) 1 (2;02;21) 1 (2;03;04) 1 (2;03;18) 5 (2;08;18)	
<i>dou1</i> 'also'				2 (2;07;11) 1 (2;08;08) 13 (2;09;09)				2 (2;08;18)	emphasis/ univ. quant. 1 (2;04;08) 5 (2;08;08) 1 (2;08;18) 5 (2;09;09)

*The number in each cell represents the number of tokens for the particular adverb and category concerned, with the age of occurrence indicated in parenthesis (year; month; day). The number of categorizable tokens is smaller than the number of spontaneous occurrences of the adverbs.

whereas *jau6*, the least constrained of the group, was among the earliest to be used productively by the two children. With regard to the semantic distribution of the adverbs, *dou1* ('also') is exclusively used to focus on the subject in the two children. Reiteration of realis events is only signalled by means of the adverb *jau6* ('also'). Children are aware early on of the aspecual properties of *zung6*, which is one of the adverbs used by both children to signify continuation of activities and states.

Conclusion

The theoretical bearing of the focus adverbs in child Cantonese can be summarized as follows. As lexical items that encode variables bound by restricted existential or universal quantifiers, focus adverbs were not produced before two years of age, but began to be used spontaneously between two and three years of age. The frequencies indicated in Table 2 and illustrations of their use given in (26-35) demonstrate that the occurrence of focus adverbs in child speech is by no means infrequent, and is grounded in naturalistic discourse. While we have only a few examples of these focus adverbs interacting with other operators, the examples in the data (e.g. 31) in which focus adverbs are used in negative predicates suggest that children can handle scope. It is a striking fact that the children did not err at all with regard to the direction of quantification specified by particular focus adverbs. For example, they never used *zung6* to focus on the subject; nor did they use *dou1* to focus on the verb phrase or the object. This shows that children become sensitive to logical properties such as the direction of quantification in very early stages of their language development. Future research will tell us more whether these bound variable structures are related at all to bound variable structures created by quantifier phrases or wh-phrases.

References

- Biq, Y.O. (1987) *The Semantics and Pragmatics of Cai and Jiu in Mandarin Chinese*, Bloomington, Indiana, Indiana University Linguistics Club.
- Biq, Y.O. (1988) "From objectivity to subjectivity: the text-building function of *you* in Chinese", *Studies in Language* 12.1, pp. 99-122.
- Biq, Y.O. (1989) "*Ye* as manifested on three discourse planes: polysemy or abstraction?", in J. Tai and F. Hsueh (eds) *Functionalism and Chinese Grammar*, South Orange, New Jersey, Chinese Language Teachers' Association.
- Cheng, R. (1983) "Focus devices in Mandarin Chinese", in T.C. Tang, R. Cheng, and Y.C. Li (eds) *Studies in Chinese Syntax and Semantics: universe and scope*, Taipei, Taiwan, Student Book Co.
- Cheung, S. H.-N. (1972) *Hoeng1gong2 Jyut6Jyu2 Jyu2faat3 dik1 jin4gau3* (Cantonese as spoken in Hong Kong), Hong Kong, Chinese University of Hong Kong.
- Ding, Shengshu (1961) *Xiandai Hanyu Yufa Jianghua* (Lectures on Modern Chinese Syntax), Beijing, Chinese Social Science Academy Press.
- Fraser, B. (1971) "An analysis of 'even' in English", in C. Fillmore and D. T. Langendoen (eds) *Studies in Linguistics Semantics*, New York, Holt Rinehart and Winston.

- Horn, L. (1969) "A presuppositional approach to only and even", *Papers from the 5th Regional Meeting*, Chicago Linguistics Society, pp. 98-107.
- Jackendoff, R. (1972) *Semantic Interpretation in Generative Grammar*, Cambridge, Massachusetts, MIT Press.
- Karttunen, L. and S. Peters (1979) "Conventional implicature", in C. Oh and D. Dineen (eds) *Syntax and Semantics 11: Presupposition*, New York, Academic Press.
- Kay, P. (1990) "Even", *Linguistics and Philosophy* 13, pp. 59-111.
- König, E. (1981) "The meaning of scalar particles in German", in H. Eikmeyer and H. Rieser (eds) *Words, Worlds, and Contexts*, Berlin, de Gruyter.
- König, E. (1991) *The Meaning of Focus Particles*, London, Routledge.
- Lobner, S. (1989) "German *schon-erst-noch*: an integrated analysis", *Linguistics and Philosophy* 12, pp. 167-212.
- Lu, Shuxiang (1980) *Xiandai Hanyu Babai Ci* (Eight hundred words in Modern Chinese), Beijing, Commercial Press.
- Ma, Xiwen (1985) "Gen fuci 'zai' youguan de jige jushi" (several constructions related to the adverb *zai*), *Zhongguo Yuwen*, 1985, no. 2, pp. 105-114.
- Ma, Zhen (1982) "Shuo ye" (On *ye*), *Zhongguo Yuwen*, 1982, no. 4.
- Matthews, S. and V. Yip (1994) *Cantonese: a comprehensive grammar*, London, Routledge.
- Paris, M.C. (1979) "Some aspects of the syntax and semantics of the 'lian...ye/dou' construction in Mandarin", *Cahiers de Linguistique Asie Orientale* no. 5, pp. 47-70.
- Paris, M.C. (1987) "Xiandai Hanyu zhong *jiu* he *cai* de yuyi fenxi" (A semantic analysis of *jiu* and *cai* in Modern Chinese), *Zhongguo Yuwen* vol. 200.
- Parsons, T. (1990) *Events in the Semantics of English*, Cambridge, Massachusetts, MIT Press.
- Renaud, F. and Luo Shenyi (1987) "Etude Lexicographique de 'zai' (a nouveau)", *Cahiers de Linguistique Asie Orientale* 16.1, pp. 81-108.
- Roeper, T. (1986) "How children acquire bound variables", in B. Lust (ed) *Studies in the Acquisition of Anaphora*. Dordrecht, D. Reidel.
- Rooth, M. (1985) *Association with focus*, Ph.D. dissertation, University of Massachusetts, Amherst.
- Tsao, Feng-fu (1976) "'Expectation' in Chinese: a functional analysis of two adverbs", *Proceedings of the Berkeley Linguistics Society* 2, pp. 360-374.
- Tse, Kwok-ping, Tang Ting-chi, Shi Yuwei, Li Eng (1991) *Chinese Children's Language Acquisition and Development*, National Science Council Research Report. Dept of English, National Taiwan Normal University.

Notes

* The research reported in this paper is supported by a grant from the University and Polytechnic Research Grants Committee RGC #CUHK/02/91. The data for this study were collected by Alice Cheung and Patricia Man, whose work is hereby acknowledged. I am grateful to Sam Leung, Audrey Li, Feng-hsi Liu, Marie-Claude Paris, Neil Smith, and participants of NACCL 6 for comments on an earlier version of the paper, and to participants of the LSHK Chinese syntax and semantics workshop held on April 16, 1994 at CPHK for criticisms of my analysis of Cantonese focus adverbs. This paper was written while I was a visitor at the Dept. of Brain and Cognitive Science, MIT. I wish to thank Ken Wexler for many of the stimulating thoughts he shared with me on language acquisition issues. In this paper, the following abbreviation conventions are used: asp=aspect; cl=classifier; sfp=sentence final particle; excl.=exclamation; nom=nominalizer; emp=emphatic marker.

¹ Focussing devices other than focus adverbs (e.g. wh phrases, the existential verb *shi* 'be', and parallel structures) have been examined at some length in Cheng (1983).

² In terms of the meanings that will be discussed in this paper, the four Cantonese adverbs *zoi3*, *jau6*, *zung6*, *dou1* correspond approximately to the Mandarin adverbs *zai4*, *you4*, *hai2*, and *ye3* respectively. Cantonese romanizations here follow the Linguistic Society of Hong Kong Cantonese romanization scheme published in 1993 (see LSHK Newsletter no. 17, Dec. 1994).

³ They are adverbs in that they are bound forms that modify verb phrases and not noun phrases; they do not take aspectual marking; they cannot undergo A-not-A reduplication, and cannot serve as centers of predication.

⁴ This distinguishes Cantonese from dialects like Mandarin, in which a realis sentence like (5) would be ungrammatical, and the same meaning would be expressed by another adverb: *you4*.

⁵ The Mandarin counterpart of *zoi3* (ie *zai4*) has a conditional use illustrated below (cf. Lui 1980):

Mingtian zai qu kan yisheng.
tomorrow then go see doctor
'(I) will go to see the doctor tomorrow'.

Here the constituent before the adverb specifies the condition under which the event denoted by the verb phrase will happen (ie 'seeing the doctor'). Cantonese *zoi3* cannot be used to denote sufficient condition.

⁶ This representation ignores complexities introduced by the object NP. As we have seen from (5), identity of the referent of the object is not obligatory in the use of *zoi3*: only type identity is required.

⁷In examples like (19b), the sentence would sound odd if the sentence final particle *tim1* 'even' is added, as in (19c). One could of course make (19c) acceptable by pragmatic accomodation, taking the second conjunct to denote something greater in value than that denoted by the first conjunct.

(19) c. ?? Keoi5 maai2 zo2 can4 lau2, zung6 maai2 zo2 bou6 ce1 tim1
s/he buy asp cl flat, also buy asp cl car even
'S/he bought a flat, and even bought a new car'.

⁸*Zung6* has a rhetorical use which is not documented in our data. For example, in a sentence of the form 'S1 *zung6* S2', the second clause could signal something contrary to what is entailed or predicted by the first clause, as in the following sentence.

keoi5 git3 zo2 fan1, zung6 seng4 jat6 paak3 to1
s/he marry asp marry, still always date
'(why is it that) although s/he got married, s/he still dates a lot?'

⁹The focus adverb *dou1* has two additional uses which have not shown up in the child language data. It can be used as an exclusive focus operator quantifying rightward when used with the focus copula *hai6* (be), as illustrated below.

jam tong dou hai loupo zyu ge hou
drink soup DOU be wife cook nom good
"As for soup drinking, it is those soups prepared by one's wife that are good"

Dou1 can also convey the 'even' reading, especially in conjunction with certain sentence final particles:

Johnny dou1 lik1-dak1-hei2 go3 gong3kam4 a4
Johnny also lift cl piano sfp
"Even Johnny can lift the piano?"

The Realization and Omission of Subject and Object in Early Cantonese

Patricia Man

In this paper, we will discuss some phenomena related to the syntactic categories 'subject' and 'object' in the data of two Cantonese-speaking children, Tsuntsun and Bernard. In the first section, we will look at three aspects in their data before the age of 2. These aspects converge on the assumption that the syntactic concept 'object' exists in the early grammar, although there is no strong evidence for the emergence of 'subject'. In the second section, we will discuss a prominent phenomenon in Cantonese, that is, the drop of subject and object. The data suggest that children take argument-drop as a grammatical option early in their speech.

1. 'Subject' and 'object' before the age of 2

'Subject' and 'object' in this study are defined as the sister node of VP¹ and V respectively. Saying that a child has the concepts 'subject' and 'object' is to say that he has the following rules in his grammar:

S → NP VP
VP → V NP

We will look at the second rule first, that is, the concept 'object'. We will focus on the verbs which are defined as transitive in the adult standard.

Table 1a verb-noun combination: Bernard
(the examples are stripped of sentence final particle, negator and preverbal nominal)

10700	-
10800	daap3 cel (travel car) daai3 ze1 (bring umbrella)
10814/10828	-
10904	sik6 caang2 caang2 (eat orange) hoi1 dang1 (open light) sai2 tau4 (wash head)
10918	ke4 maa5 maa5 (ride horse) co5 cel (sit car) daap3 celcel baai2 geilgeil (put recorder) daap3 nei1 gaa3 (travel this CL ² -car) daap3 go2 gaa3 (travel that CL-car) lo2 nei1 di1 (take this-plural)

Table 1b verb-noun combination: Tsuntsun
(the examples are stripped of sentence final particle, negator and preverbal nominal)

10522/10703/10710 -	
10800	co5 celcel (sit car)
10807	co5 cel
10821	hoil cel (open car = drive)
10907	co5 cel taan4 kam4 (play piano)
10914 -	
10929	tai2 syul (see book = read) co5 cel dai3 cel (bring car) cai3 cel (stack car) tai2 cel (see car)
11030	taan4 kam4 jiu3 gwail (want turtle) daap3 baal1si2 (travel bus) daap3 dik1si2 (travel taxi) ngaak1 sau2 (shake hand) jam2 seoi2 (drink water)
11113	daap3 baal1si2 daai3 nei1-dou6 (wear here) daai3 sau2 (wear hand) ngoi3 syul (want book) co5 gwail gwail (sit turtle) taan4 kam4 zoek3 saam1 (wear clothes) sik6 gwo2gwo2 (eat fruit) sik6 caang2 (eat orange) ngoi3 bol (want ball) daa2 jan3 (hit people)

1.1 Object

In the data of Tsuntsun up to 10710, all 24 utterances containing a transitive verb occur in bare verb sequence. In the first session of Bernard's data (10700), the two utterances with a transitive verb also occur in bare sequence.

Table 1 is a list of verb-noun sequences in the data of Tsuntsun and Bernard. In Tsuntsun 11030 and 11113, and Bernard 10904 and 10918, an explosion of verb-noun combinations is observed. What is happening in the children's speech during this time?³ 10904 is the first session in which Bernard answered a V-not-V questions with a verb-noun sequence.

(Bernard 10904)

INV: hoil-m4-hoil dang1 a3?
open-not-open light SFP⁴
'Turn on the light?'

CHI: hoil # hoil dang1.
open open light

(Bernard 10904)

MOT: jam2-m4-jam2 'milk milk' a3?
drink-not-drink milk SFP
'Drink some milk?'

CHI: jam2 # 'milk milk'.
drink milk

Before then, he used a bare verb sequence as response to this kind of questions. Before 10904, there are 5 utterances containing V as responses to V-not-V questions in Bernard's data. All 5 utterances are bare Vs, in one instance the bare occurs in reduplication.

In 11030, we found the first time Tsuntsun answered a V-not-V question with elements other a bare verb.

(Tsuntsun 11030)

INV: nei5 zung1-m4-zung1ji3 sik6 syut3gau1 a3?
you like-not-like eat ice-cream SFP
'Do you like ice-cream?'

CHI: zung1-ji3 sik6.
like eat

(Tsuntsun 11030)

INV: saan1-m4-saan1-maai4 a3?
close-not-close-complete SFP
'close the door?'

CHI: saan1-maai4.
close-complete

Before 11030, there are 25 utterances containing V as responses to V-not-V questions. All of these 25 utterances are bare Vs, sometimes reduplicated, with or without a sentence final particle.

In 11113, we found the first verb-noun (henceforth V-N) answer to V-not-V questions in Tsuntsun's data:

(Tsuntsun 11113)

INV: nei5 jiu3-m4-jiu3 go3 bolbol a3?
you want-not-want CL ball SFP
'Do you want the ball?'

CHI: m4 ngoi3 bo1 a3.
not want ball SFP

In adult Cantonese, answers to V-not-V questions should not contain the object NP, unless it is to serve contrastive function.

question: tai2-m4-tai2 nei1 ceot1 a3?
see-not-see this CL-film SFP
'Want to see this film?'

answer A: m4-tai2 nei1 ceot1, tai2 go2 ceot1.
not-see this CL-film, see that CL-film
'I don't want to see this one, I want to see that one.'

answer B: *tai2 nei1 ceot1.
see this CL-film

Both Tsuntsun and Bernard initially answered all V-not-V questions with a bare V, which apparently conform to the adult standard. However 11113 and 10904 onwards, Tsuntsun and Bernard respectively alternate between using V or V-N answers to V-not-V questions. There must be some significant change in children's analysis of V-not-V questions; otherwise, they would not turn from the correct bare V response to the 'incorrect' V-N response. One may suggest that before these two sessions, Tsuntsun and Bernard had difficulties in attending to more than one 'word' in the V-not-V questions and thus were unable to repeat more than one 'word' in the response. It is reasonable to assume higher processing capacity as children grow older. But why do they alternate between two response patterns instead of sticking to one? One may go on arguing that the expansion of processing capacity only is manifested in some occasions but not others and thus results in an alternation between one-word and two word responses.

An isolated phenomenon may receive various interpretations. We now turn to another phenomenon which may help to show that the children have developed a rather rigid verb-noun structure around the period discussed above. The phenomenon we now look at is children's answers to object questions (that is, wh-questions with the wh-phrase in the object position). Unlike V-not-V questions, there is no way for a child to rely on imitation to answer an object question correctly. ('correct' here means 'structurally acceptable', therefore an answer containing V-N; we ignore bare nominal answers because it does not show the structure) He must provide an NP in place of the wh-phrase in the object position, whether that NP makes sense or not. The first correct answers to object questions are found in 11030 and 10918 in Tsuntsun and Bernard respectively.

(Tsuntsun 11030)

INV: nei5 jiu3 mat1je5 a3?
you want what SFP
'What do you want?'

CHI: jiu3 gwai1 a3.
want turtle SFP

(Bernard 10918)

INV: lo2 mat1je5 a3 Bernard?
take what SFP
'What are you taking?'

CHI: nei1 # lo2 nei1-di1.
this take this-plural

On or before 11030 and 10918, Tsuntsun and Bernard either repeat the verb in the object questions or offer a bare nominal answer. For instance:

(Tsuntsun 10929)

INV: ngoi3 mat1je5 a3?
want what SFP
'What do you want?'

CHI: ngoi3 a3.
want SFP

(Bernard 10918)

INV: wa3, keoi5-dei6 waan2 mat1je5 a3?
they play what SFP
'Wow, what are they playing?'

CHI: &m3 # jiu4 # jiu4jiu4-baan2.
seesaw

We might imagine that a child does not need to possess any syntactic structure in order to answer object questions. What he has to do is simply to follow the word order in the question and replace that wh-phrase with a nominal. In fact, we found the following interesting answers to object questions:

(Tsuntsun 11127)

INV: nei5 zung1ji3 waan2 mat1je5 a3?
you like play what SFP
'What do you want play?'

CHI: <waan2> [>] + ...
play

INV: <nei5> [<] zung1-m4 +...
you like-not
'Do you want +...'

CHI: baa4baa4 waan2 je5 a3.
Daddy play thing SFP

Two other instances like this are found in Tsuntsun's data after 2;0:

(Tsuntsun 20108)

INV: nei5 wan2 mat1je5 a3?
you seek what SFP
'What are you looking for?'

CHI: wan2 le1 je5 a3.
seek SFP thing SFP

(Tsuntsun 20205)

INV: me1je5 dit3-zo2 a3?
what drop-ASP SFP
'What has dropped?'

CHI: dit3-zo2 je5 a3.
drop-ASP thing SFP

Tsuntsun really replaces the wh-phrase by a nominal, but in these occasions, the nominal is the indefinite 'je5'. In fact, he could choose not to answer or to reply with a bare V. We could hardly know why he answered like that. But the structure of the answer is appealing.

In Bernard's data, we do not find this kind of uninformative answers to object questions. But there is one thing we would like to point out. In his answers in the early stage, we found the following two instances:

(Bernard 10918)

INV: lo2 mat1je5 a3 Bernard?
take what SFP
'What are you taking?'

CHI: neil # lo2 neil-dil.
this take this-plural

(Bernard 10925)

INV: Bernard tai2-gan2 mat1je5 a3?
see-ASP what SFP
'What are you reading at?'

CHI: syu1 # tai2 # tai2 syu1syu1.
book see see book

The highlighted part in the answer, an incomplete nominal expression, suffices to serve communication purpose, that is, to provide relevant information. However, the child, instead of completing the nominal expression, abandoned it and uttered a V-N sequence.

Our discussion of children's answers to questions aims at showing that children use V-N sequences in occasions where they need not be so. And these occasions happen to cluster in two consecutive sessions for each child. Based on the

above discussion, we wanted to show that children before 2 years have developed structural unit composed to a verb and a postverbal nominal.

Lastly, we want to show that the postverbal nominal in that structural unit, V-N, has some internal structure. This, we have to rely on the emergence of determiner. The first instance of determiner in Bernard's data is 10904 and its first use in a postverbal nominal is 10918. For Tsuntsun, if we treat *nei-dou* as analyzable, the first instance of determiner is 11113, which is also in a postverbal nominal. Our definition of 'object' is the sister node of V, which is an NP. Based on the emergence of determiner we suggest that postverbal nominal by 10918 and 11113 for Bernard and Tsuntsun are NP. Together with our previous with response patterns, we suggest that around 11030-11113 and 10904-10918 the syntactic category 'object' has emerged in the grammar of Tsuntsun and Bernard respectively.

Below is a summary of the emergence point of the three phenomena being discussed:

Table 2 Emergence point of three phenomena related to the emergence of 'object'

1st instance of	Tsuntsun	Bernard
determiner in postverbal nominal expression	11113	10918
V-N answer to object question	11030	10918
V-N answer to V-not-V question	11113	10904

1.2 Subject

'Subject' is more independent than 'object' in relation to their head verb. Semantic relations (e.g. agent) and pragmatic roles (eg. speaker and hearer) highly constrain the referential content of subject. Due to these properties specific to subject, the phenomena discussed so far tell little about the emergence of 'subject'. First, NPs in subject position are usually proper names or pronouns, which do not take determiner. Even a preverbal nominal expression takes determiner, it can be something other than a subject (eg. 'topic'). Second, pragmatic factors play such a strong role in the omission of subject, namely, first and second person referents are usually omissible, so that it is unclear whether subject-drop in the answer to V-not-V questions is a structural constraint or a pragmatic one.

We are left with answers to subject question, that is, questions with wh-phrase in the subject position.

Tsuntsun and Bernard's first correct answers to subject questions are found in 20009 and 11023 respectively. For Tsuntsun, the first instance of N-V sequence is 4

months ahead of his first correct response to subject questions. And his first instance of V-N sequence is 3 months ahead of his first correct response to object questions. For Bernard, the two time gaps are more less the same, around one and a half month:

Table 3 Correct responses to questions and the emergence of V-N and N-V sequences

1st instance of	Tsuntsun	Bernard
V-N sequence	10800	10800
correct answer to object question	11030	10918
N-V sequence	10807	10994*
correct answer to subject question	20009	11023

* we exclude 'nei5 bong1 ngo5 aak1' (you-help-I-SFP) in 10800 since it looks like a frozen string when we consider the child's grammar by that session.

Unable to provide more evidence, we may rely on the correct answer to subject question to set the emergence point. In that case, we have to allow a transitional stage in which grammatical categories are defined on different basis. Suppose the pre-syntactic categories are X, Y, Z etc. A stage in which 'object' but not 'subject' has emerged implies the coexistence of the following rules:

$$\begin{aligned} S &\rightarrow XY \\ Y &\rightarrow \begin{matrix} VP \\ Z \end{matrix} \\ VP &\rightarrow VNP \end{aligned}$$

X is not defined by Y. But VP happens to be a candidate of Y. As long as Y is restricted to VP, X may be said of having the status of 'subject'. Alternatively, one may assume that subject and object emerge simultaneously. Thus a transitional stage as described above may be avoided.

1.3 Conclusion

In this section, we have discussed children's V-N answers to V-not-V questions and object questions, and the emergence of determiner in postverbal nominal expressions. We found that the first instance of each of these three cases appear in 11030 or 11113 in Tsuntsun's data and 10904 or 10918 in Bernard's data. We suggest that these converge on the assumption that 'object' has emerged during this period. We do not have much to say about the emergence of subject since the only phenomenon related to this, namely, children's answers to subject questions, tells little on its own. We have pointed out the consideration one needs to take in assuming different points of emergence point for subject and object.

The above conclusion has implications on the interpretation of other structures in the data. First, utterances which have left-dislocated patient-nominal may not be analyzed as topicalization if it is found before the emergence of object in the child's grammar, therefore 11030 for Tsuntsun and 10904 for Bernard. For instance, in Tsuntsun 10914, we found:

ping4gwo2 # ngau5 ngau5.
apple bite

It should not be treated as topicalization since 'object' may not have emerged by that time. Nevertheless, we do not rule out the possibility of treating it as a topic-predicate structure, in that case, topic is base-generated in that position.

The second implication concerns the omission of subject and object. If 'subject' and 'object' have not yet emerged in the child's grammar, there is no way to talk of 'omitted' subject and object. Omission of subject and object is the topic which we will turn to in the next section.⁵

2. Argument-drop

The 'telegraphic' nature of the speech of young children has been recognized early in the 60s, and is confirmed in a wide range of studies and languages (see Brown 1973). Twenty years later, researchers working with the principles and parameters theory reinterpret the same corpus of child data and raise different questions. One of these insightful reinterpretations is found in Hyams (eg. 1986)' work on argument drop. She argues that children's early utterances without subject NPs is a manifestation of the subject-drop option available in the universal grammar, although that option may not match the one with the child's target language. She exemplifies this point using child English data (Brown's corpus) which has a considerable percentage of utterances lacking subject, although adult English does not have this option in finite clauses, unlike languages such as Italian or Spanish. Some other languages allow both subject and object to drop in finite clauses. Cantonese, like Mandarin, German and Dutch, falls under this category. Later works (eg. Hyams 1992, Lillo-Martin 1992) cater for both subject-drop and object-drop. But researchers like Valian (eg. 1991) and Bloom (1990) provide evidence arguing that 'argument-drop' in child English is an epiphenomenon of factors such as age, MLU and processing limitation.

When did the Cantonese-speaking children find out that their target language allows both subject and object to drop?

2.1 Subcategorization and object-drop

Possessed with the relevant phrase structure rules, a child can proceed to learn the subcategorization frame for each verb and to find out the argument-drop options in his language so that he can realize subject and object in legitimate ways. Suppose the

child start with learning subcategorization of each verb. He has to determine the transitivity of a verb, whether it is transitive, intransitive or optionally transitive. A Cantonese child is faced with apparently two types of verbs in the input:

- i) verbs which do not occur with object
- ii) verbs which occasionally occur with object

He may derive hypothesis (A) or (B) from the above observation:

(A) there are two subclasses of verbs in Cantonese

- i) optionally transitive verbs
- ii) intransitive verbs

(B) there are two subclasses of verbs in Cantonese

- i) transitive verbs (with object-drop option in the language)
- ii) intransitive verbs

Under hypothesis (A), there is no subclass of obligatorily transitive verbs. There is only a subclass of optionally transitive verbs which make no reference to specified object if they do not occur with any overt one. Under hypothesis (B), transitive verbs without overt objects are taken as referring to some specified object, the content of which can be recovered by linkage to discourse topic.⁶

Suppose a child in his early interaction with adult already knows whether the language allows a verb with unrealized argument to make reference to specified object. If this is the case, he would not derive hypothesis (A). We suggest that this is very likely the case, since the data show that children very often use bare V utterance to make reference to specified object even after the emergence of object.

(Tsuntsun 11127)

INV: sik6 hoeng1ziu1 a1.
eat banana SFP
'Eat some banana.'

CHI: sik6 ziu1.
eat banana

INV: &haak6, sik6 ziu1 ##.
eat banana
'right, "eat banana"'

(a moment later)

INV: sik6 hoeng1ziu1.
eat banana

CHI: sik6.
eat

(Bernard 11023)
 grandma:ceong2-zai2-baaul ngoi3-m4-ngoi3?
 sausage-bun want-not-want
 'Do you want a sausage bun?'
 CHI: sik6.
 eat

In effect, both children are using *unrealized* object to refer to specified entity.

We also assume that a child, when he utters the verb for the first time, he already knows the basic meaning of a verb, including the information that at least how many arguments the verb has to take. Based on this and the above assumption, that is, the child know that unrealized object can make reference to specified entity, we suppose that the Cantonese-speaking children take *transitivity* as the default subcategorization of each new verb which is transitive according to adult standard. Thus we have extracted simple sentences with transitive verbs defined on adult standard, and counted the rate of subjects and objects, and from it infers the rate of omitted subjects and objects. This is done from 11127 and 10925 onwards for Tsuntsun and Bernard respectively, since we have suggested that 'object' has emerged by the previous session, and we assume that 'subject' emerged simultaneously with 'object'.

2.2 Subject-drop and object-drop in the data

2.2.1 Rate of argument-drop in simple transitive sentences⁷

In Table 4 and 5 we can see the great difference between the rate of overt subject and object. While overt subject rate has undergone a steady growth in both children, overt object rate has a sudden growth between the first and second period in Tsuntsun's data, but has a slight gradual decrease in Bernard's data.

**Table 4 percentage of overt subjects and objects
in simple transitive sentences: Tsuntsun**

	no.	% S	% O
A	382	2	5
B	595	3	7
I	729	9 (91)	30 (70)
II	773	16 (84)	51 (49)
III	576	23 (77)	56 (44)
I-III	2078	15 (85)	45 (55)

A: 1; 5; 22-1; 9; 29 B: 1; 5; 22-1; 11; 13

(%S and %O in A and B are % of preverbal and postverbal nominals)

I: 1;11;27 - 2; 2; 05

II: 2;2;15 - 2; 4;14

III: 2; 5; 00 - 2; 7; 02

(numbers in brackets are inferred rate of argument-drop)

**Table 5 Percentage of overt subjects and objects
in simple transitive sentences: Bernard**

	no.	%S	%O
A	18	0	11
B	105	4	34
I	354	19 (81)	57 (43)
II	329	21 (79)	54 (46)
III	351	22 (78)	47 (53)
IV	300	29 (71)	48 (52)
I-IV	1334	22 (78)	51 (49)

A: 1; 7; 00-1; 8; 28; B: 1; 7; 00-1; 9; 18

(%S and %O in A and B are % of preverbal and postverbal nominals)

I: 1; 9; 25 - 2; 0; 03

II: 2; 0; 16 - 2; 2; 12

III: 2; 2; 26 - 2; 4; 21

IV: 2; 5; 04 - 2; 8; 06

(numbers in brackets are inferred rate of argument-drop)

When we compare the overt argument rate in our data with the child English data provided by Valian (1991), we find that children's knowledge of the argument-drop options in their target develop very early. Valian reports that American children drop only 30% of subjects between MLU 1.50 to 2.00 (1;10-2;2).⁸ Subject-drop rate in Tsuntsun period (1;11;27-2;2;05) is three times to that of his American counterparts around the same age. Bernard has a slightly lower subject-drop rate than Tsuntsun, but still it is more than double of the English-speaking children. The subject-drop rates for both children fall just slightly in period II and is still more than two times to that of the American children.

Although subject-drop rate decrease gradually over times, even in the last session of Tsuntsun (2;7;02) and Bernard (2;8;06), the subject-drop rate is 65% and 76% respectively, which are still more than double of the American children in a younger age. 30% is the highest subject-drop in Valian's data. As the American children grow older, they use more and more subjects. Valian's data show that, after MLU 2.00 (over 2;3) subject use is never lower than 84%.

Table 6 Subject-drop in early speech

American children (1;10-2;2)		Tsuntsun	Bernard
30%	I	91%	81%
	II	84%	79%

Tsuntsun and Bernard also show great difference from their American counterpart in object-drop. The data shows that object-drop is a grammatical option in the speech of Tsuntsun and Bernard. In Valian's study, the highest individual omission rate in the youngest group (1;10-2;2) is 14%. Again, period I of Tsuntsun (70%) makes the

greatest difference from his American counterparts. Even in his period II and both periods of Bernard, the object-drop rate is more than three times of that of the highest individual rate from the American children.

Table 7 **Object-drop in early speech**

American children (1;10-2;2)		Tsuntsun	Bernard
7%	I	70%	43%
	II	49%	46%

Valian argues that the argument-drop phenomenon in child English can be better explained by age and MLU than grammatical factors. Her data support these arguments. Although Cantonese-speaking children clearly contrast with their English counterpart by taking subject-drop and object-drop as grammatical option, the factor of age seems to be at work for subject-drop, especially in Tsuntsun's data. However, object-drop seems to be more resistant to the age factor, particularly in Bernard's data.

2.2.2 Processing limitations

Another non-grammatical account of the subject-drop phenomenon in early child English comes from Bloom (1990) who argues that it is a result of processing limitation rather than a grammatical option in the children's grammar. He shows that the longer the VP length, the higher the probability of dropping the subject of the utterance. We would like to know if processing factors also affect the rate of using this option by Cantonese-speaking children. We have compared the rate of overt subject and object of the ditransitive verb 'bei2' and some other transitive verbs in the data of Tsuntsun and Bernard. If processing factors do affect the rate of argument-drop, ditransitive verbs should have lower overt subject rate than other transitive verbs, since ditransitive verbs need to take two objects while other transitive verbs only need one.

We found that 'bei2', used as a verb,⁹ really has the lowest overt (matrix subject rate as compared to other five transitive verbs. 'daap3' in Tsuntsun's data also has a very low subject rate, however, this does not happen to Bernard's data. It is interesting to note that 'daap3' has a very high overt object rate, while that of 'bei2' is close to the average of simple transitive sentences.¹⁰¹¹ If we assume that unrealized arguments, like overt arguments, also contribute to processing load, then the fact that 'bei2' has the lowest subject rate match the prediction that the longer the VP length, the more probable the subject is omitted.

Table 8 'Bei2' and selected transitive verbs

	Tsuntsun*			Bernard*		
	total	S	O	total	S	O
bei2 (give)	79	5%	47%	139	4%	59%
daap3 (travel)	85	7%	86%	61	17%	85%
hoi1 (open)	306	14%	35%	70	17%	57%
zaa1 (hold)	86	17%	42%	24	33%	92%
lo2 (take)	45	29%	73%	27	30%	44%
cai3 (stack)	30	23%	23%	46	50%	43%

*Tsuntsun 1;11;27-2;7;02

*Bernard 1;9;25-2;8;06

Another testing ground for the effect of processing factors is subject-drop in simple sentences versus that in the matrix clause of complex sentences. This time, we would like to know about the strength of processing factor in interaction with semantic and pragmatic factors. We have confined complex sentence to those containing a verb which only takes a VP complement (eg. *sik1* 'know', *jiu3* 'want', *zung1ji3* 'like'). Most of these verbs are 'mental' verbs, in contrast with verbs which denote actions which are visualizable.

Table 9 Overt subject in matrix clause

	Tsuntsun		Bernard	
	total	matrix S	total	matrix S
simple	2078	15%	2078	15%
complex sentence ('mental' verb in matrix clause)	25	20%	25	20%

We found that the subject-drop rate in matrix clause is not necessarily higher in complex sentences than in simple sentence. The property of the matrix verb plays an important role. In Bernard's data, the overt subject rate in the matrix clause of complex sentences with a matrix 'mental' verbs is exceedingly higher than that in simple sentences. Bernard seems to be sensitive to the semantic and pragmatic properties of these verbs. Initiating (in contrast with responding to) an utterance with a mental verb without overt subject would make the addressee difficult to figure out the reference of that unrealized subject, since unlike other verbs, the event denoted by these kind of verbs is not visualized in the context. Although there is no clear difference in the data of Tsuntsun between this kind of complex sentences and simple sentences, his data show a positive inclination to the same direction.¹²

2.2.3 Restrictions on argument-drop

In Cantonese, some syntactic positions do not allow argument-drop (V_t = transitive verb):

- i) V_{t1} NP' V_{t2}
(both V_{t1} and V_{t2} take NP' as object: eg. *maai5 tong2 sik6*)
- ii) *jau5* NP' V_t
(V_t takes NP' as object: eg. *jau5 tong2 sik6*)
- iii) V_{t1} NP' V_2
(NP' is the object of V_{t1} but not V_2 . V_2 may be transitive or intransitive: eg. *ling1 je5 zau2*)

The above three patterns can be collapsed into a general one:

V_1 NP' V_2

In this structure, NP' cannot be dropped. Only a small number of sentences in the data are in this structure. There are only 15 in the data of Tsuntsun and 37 in that of Bernard. All of these 52 sentences, except one in Bernard's data, have overt NP'. Although the frequency of relevant data is low, the rate of conformity with adult standard is strikingly high.

2.2.4 Conclusion

We have discussed argument-drop in the data of two Cantonese-speaking children, Tsuntsun and Bernard. Their data contrast strongly with that of the English-speaking American children as reported in Valian 1991. Tsuntsun and Bernard drop high proportion of subject and object as compared to their English-speaking counterparts. Non-grammatical factors such as age and VP length seem to have effect on the use of subject-drop option by Tsuntsun and Bernard. This is shown in the gradual increase of overt subject rate over time, and the contrast between ditransitive 'bei2' and some selected transitive verbs on subject-drop rate. Semantic and pragmatic factors also play a role in argument-drop. For instance, subject-drop rate in the matrix clause of complex clause with matrix 'mental' verbs is lower than that of simple transitive sentences, a result unexpected from a pure processing account. We have also pointed out that Tsuntsun and Bernard show high conformity to adult's standards concerning the constraints on argument-drop.

All of the frequency count is based on the assumption that subject and object emerge at the same time, although we do not have strong evidence for the emergence of subject. It turns out that this arbitrary decision fits the data of Bernard. From period I through IV, the growth of overt subject (or preverbal nominal) rate is so smooth as compared to the great leap between period B and period I, which we suggest to be the onset of subject. For Tsuntsun, overt subject rate grow gradually nomatter when we set the emergence point of subject.

3. Concluding remarks

This paper has focused on the simplest cases of argument realization or argument-drop. Realization (or drop) of arguments of complex predicate has not been touched on. As a closing remark, we would like to point out that Tsuntsun and Bernard have shown knowledge of inflected predicates versus simple predicates with respect to argument realization. In simple sentences, transitive verbs which are inflected with the modal suffixes 'dak1' or 'dou2' take more or less the same proportion of subject and object. In Tsuntsun's 55 simple sentences with 'V_t-dak1/dou2', 4% have overt subject and 7% overt object. For the 52 sentences of the same type in Bernard's data, overt subject and object are both 15%. This clearly contrasts with the other simple transitive sentences in which overt object rate is always higher than overt subject rate. It is left for future research to study into argument realization for complex predicates in child language.

References

- Bloom, P. (1990) 'Subjectless sentences in child language.' *Linguistic Inquiry* 21: 491-504.
- Brown, R. (1973) *A First Language: the early stages*. Cambridge, MA: Harvard University Press.
- Hyams, N. (1986) *Language Acquisition and the Theory of Parameters*. D. Reidel Publishing Co.
- Hyams, N. (1992) 'Null subjects in child language and the implications of crosslinguistic variation.' *Paper presented at the Cornell symposium on syntactic theory and first language acquisition: crosslinguistic perspective*, Ithaca, New York.
- Lillo-Martin, D. (1992) 'Setting the Null Argument Parameters: evidence from American Sign Language and other languages.' *Paper presented at the Cornell symposium on syntactic theory and first language acquisition: crosslinguistic perspective*, Ithaca, New York.
- Valian, V. (1991) 'Syntactic subjects in the early speech of American and Italian children.' *Cognition* 40:21-81.

Appendix I

Structures involving V before 2;0

	Tsuntsun	Bernard
V	before 10800	before 10800
neg-V	10800	10800
V-N	10800	10800
N-V	10807	10904
V-comp/asp	10800/10907	10904
V ₁ -V ₂ (-N)	10800	10918
V-[det-N]	11113	10918
N-V-N	11113	10925
N-neg-N	11127	11106
V-comp/asp-NP	11127	11106

Notes

¹'subject' should be defined as the sister node of 'predicate' which may or may not be a VP. Since the discussion in this paper is confined to transitive sentences, we decided to cut short the definition.

²CL = classifier

³see Appendix 1 for structures involving V before 2;0

⁴SFP = sentence final particle

⁵We will take up the assumption that subject and object emerge at the same time in the discussion in the next section.

⁶In languages with object agreement, content of omitted object can be recovered by verbal inflection.

⁷simple transitive sentences as discussed in this section exclude those with compound predicates (eg. V-Adj.) or inflected predicates (eg. V-dak1).

⁸Valian (1991)'s data on subject-drop and object-drop as discussed in this section concern simple sentences only.

⁹Instances where 'bei2' is used as a preposition, as in the following sentence, is excluded.

sung3 bun2-syu1 bei2 nei5.
 give CL-book you
 'give a book to you'

¹⁰In simple transitive sentences, the average overt subject rate from period I onwards are 15% and 22% for Tsuntsun and Bernard respectively, and the average overt object rate are 45% and 51% respectively.

¹¹'bei2' is counted as having an overt object if any one of the two objects is realized.

¹²The argument still holds if we compare the percentage of this type of complex sentences with the selected transitive sentences listed in Table 8.

Acquisition of Wh-words by Cantonese-speaking Children

Alice Shuk-yee Cheung

1.0 Introduction

Children are exposed to a vast number of wh-words at a very early age in the communication with their caretakers, and these wh-words can have both interrogative and non-interrogative uses. When do children recognise whether a particular utterance consisting of a wh-word is an interrogative question or not? How do they distinguish between them?

In the past, there have been numerous studies on children's acquisition of wh-questions across different languages, including English (Ervin-Tripp 1970, Tyack & Ingram 1977, Cairns & Hsu 1978, Bloom et al 1982, Stromswold 1988), Serbo-Croatian (Savic 1975), Korean (Clancy 1988), Tamil (Vaidyanathan 1988), Chinese (Erbaugh 1982, Lee 1989), and French (Sinclair & van Gessel 1990), etc. There were several findings based on these studies. Firstly, the order of acquisition of the different wh-questions across languages has been observed to be basically *what-*, *where-* questions before *who-*, *why-* questions, which is followed by *how-* and *when-* questions, i.e. *wh-* interrogatives in argument positions are acquired before *wh-* interrogatives in adjunct positions. Secondly, it was observed that wh-words occupying a different syntactic position were acquired at different times, namely, *who*-subject questions were acquired before *who*-object questions, and *what*-object questions before *what*-subject questions. Thirdly, verb has been observed to play a role in the acquisition of the different wh-questions, depending on its type, transitivity and pattern of use.

Different explanations have been put forward to account for the above findings. Contributing factors proposed include cognitive development (Tyack & Ingram 1977; Cairns & Hsu 1978; Miao 1986; Clancy 1989), syntactic complexity (Bloom et al 1982), semantic complexity (Bloom et al 1982), government theory (Stromswold 1988), functional distribution (Ervin-Tripp 1970), contextual distribution (Tyack & Ingram 1977; Bloom et al 1982), functional need (Sinclair & Gessel 1990), input frequency (Erbaugh 1982; Clancy 1989; Heath 1982), language socialization pattern (Heath 1982) and semantic probability (Ervin-Tripp 1970). As for the role of verbs, its importance had not been discussed in detail in earlier studies. Each of the different factors succeeded in accounting for some of the observations made of the acquisition order of wh-questions. However, there is not as yet a single account that can fully explain all the observations.

Studies have concentrated in the order of acquisition of interrogative wh-words. There has been no study in the acquisition of wh-words as a whole, which includes both their interrogative and non-interrogative uses, and in the case of interrogatives, their occurrence in direct and indirect questions. Thus, the present study aims to explore the following four things:- (1) the order of acquisition of interrogative wh-words in Cantonese-speaking children; (2) the acquisition pattern of non-interrogative wh-words in Cantonese-speaking children; (3) the

relation between the acquisition of interrogative and non-interrogative wh-words and (4) the role of verb in the wh-word acquisition.

The study has its significance in providing basic information as to when, compared to other crosslinguistic studies, Cantonese-speaking children acquire different types of wh-words, and it adds to the existing literature on wh-word acquisition areas that have seldom been explored (viz.. interrogative use of wh-words in indirect questions; non-interrogative use of wh-words). Besides, it establishes the importance of verb in accounting for the whole acquisition pattern. In addition, it provides a detailed description of the different uses and distribution of the various wh-words in Cantonese, which has not been fully explored in the past.

1.1 Wh-words in Cantonese

The different wh-words in Cantonese have not been studied in detail before in any of the previous studies. Literature in this area was scarce. Li (1968) gave a list of the common wh-questions found in child Cantonese, and Yue-hashimoto (1993) listed with examples a number of interrogative wh-words, indefinite wh-words and their position of occurrence in adult Cantonese. In this study, we have provided a detailed description of the different uses and distribution of the various wh-words found in Cantonese. Table 1 shows the distribution of the interrogative wh-words.

Table 1: Distribution of Cantonese interrogative wh-words

<u>Wh-word</u>		<u>Place of occurrence</u>
邊個 binlgo3	'who'	subject, object, in NP containing a nominaliser
邊個 binlgo3	'which'	subject, object
邊+CL(+N) binl+CL(+N)		subject, object, as specifier of NP
邊個 binlgo3	'whose'	in NP containing a nominaliser
[邊個㗎 binlgo3 gaa3		object
邊個嘅 binlgo3 ge3		(as above)
(gaa3 & ge3 are sentence final particles)]		
乜嘢 matlje5	'what'	subject, object, inside compound noun, in NP containing a nominaliser
乜 matl		(as above)
咩嘢 melje5		(as above)
咩 mel		(as above)
邊度 binldou6	'where'	subject, object, adjunct, in NP containing a nominaliser
邊處 binlsyu3		(as above)
邊 binl		object, adjunct

點樣	dim2joeng2	'how/what-manner'	adjunct
點	dim2		(as above)
點樣	dim2joeng2	'what-like'	subject, object, in NP containing a nominaliser
點	dim2		(as above)
點樣	dim2joeng2	'what-do'	subject, object, adjunct
點	dim2		(as above)
點解	dim2gaai2	'why'	adjunct
幾時	gei2si4	'when'	adjunct, in NP containing a nominaliser
			(subject - when main verb is copula or existential;
			object - when main verb is <i>hai2</i> "be in/at")
幾多	gei2dol	'how many/much'	specifier of NP
幾	gei2	'what-extent'	specifier of AdjP

The verb 做 *zou6* 'do' combines with 乜(嘢)/咩(嘢) *mat1(je5)/me1(je5)* 'what' to form different wh- questions equivalent to the meaning of 'what-do', 'what-for', 'what-matter' and 'why' etc.

'what-do'	<i>zou6</i> 'do' + 'what' as subject,
	'what' as object of <i>zou6</i> 'do'
'what-for'	'what' as object of <i>zou6</i> 'do'
'what-matter'	'what' as object of <i>zou6</i> 'do'
'why'	<i>zou6</i> 'do' + 'what' as adjunct

We can see that some interrogative wh-words can have more than one meaning e.g. *bin1go3*, taking the meaning of either 'who', 'which', or 'whose'. Some have more than one form e.g. *bin1dou6* and *bin1* both means 'where'. Some wh-words have the same meaning e.g. *dim2gaai2* and *zou6mat1* both mean 'why'. These wh-words can either occur in direct or indirect questions.

The wh-words listed in Table 1 also occurred as non-interrogatives. They occur as indefinite pronouns, intensifiers or in rhetorical questions. Their distributions are as illustrated in Table 2.

Table 2: Distribution of Cantonese non-interrogative wh-words

(1) as indefinite pronouns

(a) in negative environments (only allowed in object position)

e.g.	冇	邊	個	想	去	。
	<i>mou5</i>	<i>bin1go3</i>	<i>soeng2</i>	<i>heoi3</i>		
	have-not	anybody	want	go		
		(Nobody wants to go.)				

(b) in yes-no questions (only allowed in object position)

e.g. 佢 識 唔 識 邊 個 醫生 呀 ?
keoi5 sik1 m4 sik1 binlgo3 jilsang1 aa3
he know not know anyone doctor SFP
(Does he know of any doctor?)

(c) in conditional sentences (must be in the antecedent of the conditional)

e.g. 邊 個 知 就 話 俾 我 聽 。
binlgo3 zil zau6 waa6 bei2 ngo5 teng1
whoever know then tell I listen
(Whoever knows, tell me.)

(d) as non-factive complements

e.g. 我 以爲 你 執 咗 乜 嘢 。
ngo5 ji5wai4 nei5 zap1 dil matlje5
I think you pick CL something
(I thought you were picking up something.)

(e) co-occurrence with expressions denoting uncertainty or tentativeness

e.g. 佢 好 似 見 到 乜 嘢 。
keoi5 hou2ci5 gin3 dou2 matlje5
he seem see ASP something
(He seems to have seen something.)

(f) occurrence with sentence final particles such as 㗎 *laak3*, 㗎㗎 *gaa3laak3*, 㗎㗎 *gaa3laa3*, which denote the speaker's realization through his making of inference that an event has happened.

e.g. Situation: B is searching for something. A looks at him from a distance. B stops searching and looks relieved.

A: 佢 搵 到 乜 嘢 㗎 。
keui5 wan2 dou2 matlje5 laak3
he find ASP something SFP
(He finds something.)

Although A does not see the actual thing that is being searched for, but from the reaction of B, A can infer that B has found the thing.

(g) occurrence before the universal quantifier 都 *dou1* 'all'

e.g. 邊 個 都 識 佢 。
binlgo3 dou1 sik1 keoi5
who all know him
(Everyone knows him.)

(h) labelling of entity whose name is not known or can't be recalled

e.g. 有 個 邊 個 頭 先 搵 你 。
jau5 go3 binlgo3 tau4sin1 wan2 nei5
have CL who just-now find you
(Someone looked for you just now.)

(i) used in listing of entities of the same category

e.g. 媽咪去咗買衫定裙定乜嘢呀。
maalmi4 heoi3 zo2 maai5 saaml ding6 kwan4 ding6 matlje5 aa3
mummy go ASP buy blouse or dress or what SFP
(Mummy has gone to buy a blouse or dress or the like.)

(2) as intensifiers

(a) occur in both positive and negative environments (cf. interrogatives: only in positive environment)

e.g. 條裙幾長！
tiu4 kwan4 gei2 coeng4
CL dress INT long
(The dress is very long!)

e.g. 條裙冇幾長。
tiu4 kwan4 mou5 gei2 coeng4
CL dress no INT long
(The dress is not very long.)

(b) presence of sentence final particles which denote certainty such as *wo3*, *gaa3*, *gaak3*:

e.g. 呢條繩幾長喎。(㗎 / 咯)
nil tiu4 sing2 gei2 coeng4 wo3 (gaa3 / gaak3)
this CL string INT long SFP (SFP / SFP)
(This piece of string is quite long.)

(3) in rhetorical questions

(a) asserting the impossibility of the action expressed by the verb

e.g. 咁多碟點洗呀？
gam3 dol dip2 dim2 sai2 aa3
so many dish how wash SFP
(How am I going to wash all these dishes?)

(b) emphasizing the disagreement with the preceding statement:

e.g. A: 又話食咗飯嘅？
jau6 waa6 sik6 zo2 faan6 ge2
also say eat ASP rice SFP
(It's said that you have eaten.)

B: 邊有食呀？
bin1 jau5 sik6 aa3
Where have eat SFP
(I have not eaten!)

(c) emphasizing the disagreement with the action/state expressed by the verb

e.g. 知呀。咩唔知呀？
zil aa3. mel m4 zil aa3
know SFP what not know SFP
(I know! Of course I know!)

(d) repeating and emphasizing the wh-word used in the preceding question

e.g. A: 點解冇汽水嘅?
dim2gaai2 mou5 hei3seoi2 ge2
why no soft-drink SFP
(Why is there no soft drink?)

B: 冇買囉, 點解呀.
mou5 maai5 lol, dim2gaai2 aa3
no buy SFP why SFP
(Why? I haven't bought it.)

(e) recounting of events to arouse the interest of the hearer

e.g. Situation: A is recounting an accident to B.

A: 我頭先落街, 點知呢有架
ngo5 tau4sin1 lok6 gaai1, dim2 zil nel jau5 gaa3
I just-now go-down street how know SFP have CL

車衝埋我度 ...
cel cung1 maai4 ngo5 dou6 ...
car rush towards me place ...
(I went down the street just now, God knows why a car
rushes towards me...)

Interrogative and non-interrogative wh-words appear in the same form. As seen from Tables 1 & 2, some of them can be distinguished by the particular environment that they occur in or the particular function that they serve in the utterance. Other linguistic cues, such as the stress placed on the wh-word, prolongation of the vowel of the wh-word and the absence of pause after the utterance containing the wh-word, may also help to distinguish between interrogative wh-words and wh-words used as intensifiers or in rhetorical questions.

2.0 Methodology

Longitudinal and cross-sectional studies have been the most common form of study in child language research. Previous studies on the order of production of wh-questions by children were mostly longitudinal studies (Ervin-tripp 1970, Savic 1975, Tyack & Ingram 1977, Bloom et al 1982, Vaidyanathan 1988, Clancy 1989, Sinclair & Van Gessel 1990). As for the order of comprehension of wh-questions, the studies were mainly experimental studies (Tyack & Ingram 1977, Cairns & Hsu 1978, Miao 1986). The present study includes both a longitudinal and a cross-sectional study.

2.1 Longitudinal study

The subjects in the longitudinal study were eight Cantonese-speaking children, four males and four females, aged between 1;05;22 and 2;08;10 at start

of study, from eight monolingual families living in Hong Kong. The age and sex distribution of the eight subjects under study are as given in Table 3.

Table 3 : Age and Sex Distribution of the eight children in the longitudinal study

Child	Age	Sex
Tsuntsun	1;05;22-2;07;02	Male
Bernard	1;07;00-2;08;06	Male
Chunyat	1;10;08-2;10;27	Male
Gakei	1;11;01-2;09;09	Female
Tinfaan	2;02;10-3;02;18	Female
Bohuen	2;03;22-3;04;08	Female
Johnny	2;04;08-3;04;14	Male
Jenny	2;08;10-3;08;09	Female

The children were visited for one year on a regular basis, usually at bi-weekly intervals, during which the investigator¹ played with the individual child for one hour and the session was audiotaped. The sessions were normally conducted in home settings, but they also included places outside homes, such as the parks, fast food shops, restaurants, toy shops and supermarkets. The investigator talked and played with the child for most of the time in the session, while the child's parent or siblings joined in occasionally. Activities included playing with toys, picture identification, reading books, etc. The data collected were transcribed, following the format of CHILDES (MacWhinney 1991), and the transcriptions were computerised.

2.2 Cross-sectional study

In the cross-sectional study, thirty Cantonese-speaking children, aged between 2;6 and 5;0, were tested on their comprehension of wh-questions. They were evenly distributed into five age groups, each consisting of three boys and three girls. During the test sessions, the subjects were tested individually. They listened to an audiotaped story while looking at two pictures which depicted the story. At the end of the story, the pictures were put away and the subject was asked a set of questions related to the story. The answers were recorded by hand and were also audiotaped. The subject was then tested with the next story.

The test included twelve short stories, each centred on one verb, which belongs to one of the four verb types specially chosen for the examination of their

¹ Data used in this study were drawn from the language corpus collected by the research group on *The Development of grammatical Competence in Cantonese-speaking Children* (a joint UPGC funded project done by six researchers of the Chinese University of Hong Kong, Hong Kong Polytechnic University and the University of Hong Kong). The data were collected and transcribed by three of the investigators, one of which was the author, who was responsible for the data collection and transcription of three (CKT, MHZ and LTF) of the eight children.

effects on the comprehension of wh-questions. The verb types included transitive actional verbs (e.g. 踢 *tek3* kick, 打 *daa2* hit, 切 *cit3* cut), intransitive actional verbs (e.g. 跳 *tiu3* jump, 爬 *paa4* crawl, 飛 *fei1* fly), transitive stative verbs (e.g. 諗住 *nam2zyu6* have-in-mind, 驚 *geng1* fear, 載住 *zoi3 zyu6* hold) and functional verbs (e.g. 有 *jau5* exist/have, 喺 *hai2* be-at/in, 係 *hai6* be). All possible wh-questions that could be asked of the verbs were tested on. The same wh-word was placed in all syntactic positions possible of the wh-word, that is, in the subject, object or adjunct position. The relation between the syntactic position of the wh-word and the comprehension of wh-questions was explored.

3.0 Findings

The findings of both the longitudinal and cross-sectional studies show an order in the acquisition of direct wh-questions, the pattern of which resembles those of previous studies across different languages. Subject/object asymmetry was observed. Verb type was found to play an important role in the acquisition. Besides, as found from the longitudinal study, there is a similar order in the acquisition of indirect wh-questions. As for the acquisition of non-interrogative wh-words, no conclusion can yet be drawn from the limited data as to whether there is an order in the acquisition. However, the distribution of the different non-interrogatives was found and possible linguistic cues for the distinction of interrogative and non-interrogative wh-words, as indicated from the data, were suggested.

3.1 Longitudinal study

3.1.1 Occurrence and distribution of wh-words

Data from the longitudinal study showed that children started to acquire some of the wh-words at this early stage of language development. They occurred as interrogatives in direct questions, indirect questions and non-interrogatives. Of the wh-words that have been acquired, the positions where they occurred in a question resembled those of the adults'. However, the wh-words did not appear in the full range of positions that they could occur in. Their occurrence and distribution are summarised in Tables 4, 5 and 6 below.

Table 4 : Interrogative wh-words used in direct questions in early child Cantonese

邊個 <i>binlgo3</i>	'who'	subject, object
	'which'	subject, object, SPEC of NP
	'whose'	object
邊+CL(+N) <i>binl+CL(+N)</i>	'which'	subject, object, SPEC of NP
乜嘢 <i>matlje5</i>	'what'	subject, object,
乜 <i>matl</i>		compound noun
咩嘢 <i>melje5</i>		
咩 <i>mel</i>		
邊度 <i>binldou6</i>	'where'	subject, object, adjunct
邊處 <i>binlsyu3</i>		

邊	binl		
點樣	dim2joeng2	'what-manner/how'	adjunct
點	dim2	'what-do'	
點解	dim2gaai2	'why'	adjunct

The verb 做 *zou6* 'do' combines with 乜(嘢)/咩(嘢) *mat1(je5)/mel(je5)* 'what' to form different wh- questions equivalent to the meaning of (1) 'what-do', (2) 'what-for', (3) 'what-matter' and (4) 'why'. They occur in adjunct position.

(wh-words not yet acquired:

點樣 *dim2joeng2* 'what-like', 幾時 *gei2si4* 'when',
幾多 *gei2dol* 'how many', 幾 *gei2* 'what extent')

(distribution:

no occurrence of wh-words in NP containing a nominaliser was found
no occurrence of wh-words in AdjP was found.)

Table 5 : Interrogative wh-words used in indirect questions in early child Cantonese

邊個	binlgo3	'who'	subject, object
		'which'	subject, object
邊+CL(+N)	binl+CL(+N)	'which'	subject, object
乜嘢	matlje5	'what'	object
乜	matl		
咩嘢	melje5		
咩	mel		
邊度	binldou6	'where'	subject, object
邊處	binlsyu3		
邊	binl		

(wh-words not yet acquired:

點解 *dim2gaai2* 'why', 點樣 *dim1joeng2* 'what-like',
幾時 *gei2si4* 'when', 幾多 *gei2dol* 'how many',
幾 *gei2* 'what extent')

(distribution:

no occurrence of 咩 *mel* 'what' in the subject position of the embedded clause)

Table 6 : The emergence of non-interrogative wh-words by Cantonese-speaking children

Subject	Age	Wh-word	Use
Chunyat	2;06;08	幾多 <i>gei2dol</i>	intensifier
	2;08;00	幾 <i>gei2</i>	intensifier
	2;10;23	乜 <i>matl</i>	indefinite (in Yes-No question)
Gakei	2;03;18	幾 <i>gei2</i>	intensifier
		乜 <i>matl</i>	indefinite (in negative environment)
	2;03;25	點 <i>dim2</i>	rhetorical (emphasize impossibility)
	2;08;08	點 <i>dim2</i>	rhetorical (arouse interest)

Tinfan	2;05;18	幾 <i>gei2</i>	intensifier
	2;06;01	幾多 <i>gei2dol</i>	intensifier
	2;08;02	邊 <i>bin1</i>	rhetorical (emphasize disagreement with preceding statement)
	2;11;16	咩 <i>me1</i>	indefinite (labelling entity whose name is unknown)
	3;01;21	邊個 <i>bin1go3</i>	indefinite (in conditional sentence)
Bohuen	3;03;12	點解 <i>dim2gaai2</i>	rhetorical (repeat and emphasize the wh-word said previously)
	3;04;08	乜嘢 <i>mat1je5</i>	indefinite (labelling entity whose name is unknown)
Johnny	2;05;03	幾 <i>gei2</i>	intensifier
	2;09;30	幾多 <i>gei2dol</i>	intensifier
Jenny	2;11;08	邊 <i>bin1</i>	rhetorical (emphasize disagreement with preceding statement)
		點 <i>dim2</i>	rhetorical (emphasize impossibility)
	3;01;30	邊 <i>bin1</i>	indefinite (labelling entity whose name is unknown)
	3;02;13	咩 <i>me1</i>	rhetorical (emphasize disagreement with action/state expressed)
	3;04;22	咩 <i>me1</i>	rhetorical (emphasize disagreement with preceding statement)
		幾多 <i>gei2dol</i>	intensifier

(wh-word not yet acquired:

幾時 *gei2si4* 'when')

(distribution²:

indefinite: 乜(嘢) *mat1(je5)* / 咩 *me1* 'something ; anything' ;

邊個 *bin1go3* 'whoever'

- in negative environments
- in yes-no questions
- in conditional sentences
- labelling of entity whose name is not known or can't be recalled (vs. 9 adult use)

intensifier: 幾多 *gei2dol*, 幾 *gei2* 'very'

- in positive environment only (cf. adult - both pos. and neg.)
- only gives the meaning of 'very', but not 'fairly'

rhetorical: 點 *dim2* 'how' ; 邊 *bin1* 'where' ; 點解 *dim2gaai2* 'why' 咩 *me1* 'what' (only the truncated form is used)

- asserting the impossibility of the action expressed by the verb
- emphasizing the disagreement with the preceding statement
- emphasizing the disagreement with the action / state as expressed by the verb
- repeating & emphasizing the wh-word used in the preceding question

² Although detailed suprasegmental analysis was not done, audiorecordings of the children show that there were also the presence of stress in the wh-word, the prolongation of the vowel of the wh-word and the absence of pause after the utterance containing the wh-word.

- recounting events to arouse the interest of the hearer
- 點 *dim2* 'how'; 邊 *bin1* 'where': mostly positive in form and negative in meaning
- 咩 *me1* 'what': mostly negative in form and positive in meaning)

3.1.2 Order of acquisition

As to the order of acquisition, argument questions are found to be acquired earlier than adjunct questions for both direct and indirect questions. In direct questions, the following order was observed:

what-identification / where-identification < what- / where- < who/ why/ how (when-,how-many-,what-extent-,how(what-like) questions not yet acquired; where- argument questions acquired before where-adjunct questions)

In indirect questions, what-, who- and where- questions are the earliest acquired. Of the wh-words occurring in adjunct position, only the wh-word *how* was found. In both direct and indirect questions, subject/object asymmetry was observed for what- and which- questions, with their object questions acquired before subject questions. As for where- and who- questions, subject/object asymmetry was observed only in direct questions, the former having object questions acquired before subject questions while vice versa for the latter. Tables 7 and 8 below summarise the order of acquisition of interrogative wh-words.

No conclusion can yet be drawn as to whether there is a regular pattern in the order of acquisition of non-interrogative wh-words as their occurrence in the one-year study was limited.

Table 7: The order of emergence of interrogative wh-words in direct questions

Subject	Wh-word	1st interrogative production
Tsuntsun	what(-identification)	1;08;21 咩 <i>me1</i>
	who(-identification)	1;10;30 邊個 <i>bin1go3</i>
	where(-identification)	2;00;09 邊 <i>bin1</i>
	where-object	2;05;14 邊 <i>bin1</i>
	what-do	2;05;14 點 <i>dim2</i>
	what-do	2;06;18 做乜 <i>zou6mat1</i>
	what-object	2;06;18 咩 <i>me1</i>
Bernard	where(-identification)	2;03;09 邊度 <i>bin1dou3</i>
	what-do	2;03;28 做乜嘢 <i>zou6mat1je5</i>
	what(-identification)	2;05;04 乜嘢 <i>mat1je5</i>
	which-object	2;05;19 邊個 <i>bin1go3</i>
	why	2;06;04 點解 <i>dim2gai2</i>
	where-object	2;06;18 邊度 <i>bin1dou6</i>
Chunyat	what(-identification)	2;01;10 乜嘢 <i>mat1je5</i>
	what(-clarification)	2;01;10 咩 <i>me1</i>
	where(-identification)	2;07;06 邊度 <i>bin1dou6</i>
	which(-identification)	2;07;06 邊 <i>bin1 CL</i>
	why	2;07;06 做咩 <i>zou6me1</i>
	how	2;07;13 點樣 <i>dim2joeng2</i>
	which-object	2;08;17 邊個 <i>bin1go3</i>

	where-adjunct	2;08;00 邊度 binldou3
	what-object	2;08;00 乜嘢 matlje5
	why	2;08;00 點解 dim2gai2
	where-object	2;08;17 邊度 binldou6
	who-subject	2;09;07 邊個 binlgo3
	whose	2;10;13 邊個gaa3 binlgo3 SFP
	who(-identification)	2;10;27 邊個 binlgo3
	what-matter	2;10;27 做乜嘢 zou6matlje5
	what-do	2;10;27 做乜嘢 zou6matlje5
	where-subject	2;10;27 邊度 binldou6
	which-subject	2;10;27 邊個 binlgo3
Gakei	what(-identification)	2;02;21 咩 mel
	where(-identification)	2;02;07 邊度 binldou6
	what(-clarification)	2;02;21 咩 mel
	why	2;02;21 做乜 zou6matl
	which(-identification)	2;02;28 邊 binl CL N
	where-object	2;03;11 邊 binl
	how	2;03;18 點 dim2
	what-object	2;04;30 咩 mel
	what-matter	2;04;30 做乜 zou6matl
	what-do	2;05;03 做乜 zou6matl
	why	2;05;03 點解 dim2gai2
	where-adjunct	2;08;08 邊度 binldou6
	who-subject	2;09;09 邊個 binlgo3
Tinfaan	what(-identification)	2;02;10 乜 matl
	where(-identification)	2;02;10 邊度 binldou6
	which(-identification)	2;03;30 邊個 binlgo3 N
	who-subject	2;04;27 邊個 binlgo3
	where-object	2;04;27 邊 binl
	what(-clarification)	2;04;27 咩 mel
	who(-identification)	2;05;18 邊個 binlgo3
	what-do	2;05;18 做咩 zou6mel
	what-object	2;06;01 咩 mel
	where-adjunct	2;06;01 邊 binl
	which-object	2;06;01 邊 binlgo3
	why	2;06;01 點解 dim2gai2
	how	2;07;20 點 dim2
	whose	2;07;20 邊個gaa3 binlgo3 POSS
	what-matter	2;08;02 做咩 zou6mel
	who-object	2;09;07 邊個 binlgo3
	where-subject	3;02;18 邊度 binldou3
Bohuen	what(-identification)	2;03;23 咩 mel N
	where(-identification)	2;05;06 邊 binl
	what-object	2;06;09 乜嘢 matlje5
	what(-clarification)	2;06;09 乜嘢 matlje5
	how	2;07;03 點 dim2
	where-object	2;10;02 邊 binl
	who(-identification)	2;10;02 邊個 binlgo3
	who-object	2;10;23 邊個 binlgo3
	what-for	3;01;01 做乜嘢 zou6matlje5
	what-matter	3;02;20 做乜嘢 zou6matlje5
	what-d o	3;02;20 做乜 zou6matl
	why	3;03;12 點解 dim2gai2
Johnny	what(-identification)	2;06;24 咩 mel
	why	2;07;21 點解 dim2gai2
		3;03;11 做咩 zou6mel
	where(-identification)	2;09;30 邊度 binldou6
	how	3;00;08 點 dim2
	who(-identification)	3;01;16 邊個 binlgo3

	what-object	3;01;16 咩(嘢) <i>me1(je5)</i>
	who-object	3;01;16 邊個 <i>binlgo3</i>
	who-subject	3;02;16 邊個 <i>binlgo3</i>
	where-adjunct	3;02;16 邊度 <i>binldou6</i>
	which-object	3;02;16 邊 <i>binl CL</i>
	which(-identification)	3;02;16 邊 <i>binl CL</i>
	what(-clarification)	3;03;11 咩 <i>me1 N</i>
	whose	3;04;14 邊個 <i>binlgo3 POSS</i>
	where-object	3;04;14 邊(度) <i>binl(dou6)</i>
	what-for	3;04;14 做乜嘢 <i>zou6matlje5</i>
	what-do	3;04;14 做乜嘢 <i>zou6matlje5</i>
	what-matter	3;04;14 做乜嘢 <i>zou6matlje5</i>
Jenny	what(-identification)	2;08;10 咩 <i>me1</i>
	what-for	2;08;22 做乜嘢 <i>zou6matlje5</i>
	what-do	2;09;14 做乜嘢 <i>zou6matlje5</i>
		2;11;08 點 <i>dim2</i>
	where-identification	2;09;28 邊度 <i>binldou6</i>
	what-matter	2;09;28 做乜嘢 <i>zou6matlje5</i>
	why	2;11;01 做乜 <i>zou6matl</i>
	where-adjunct	2;11;08 邊度 <i>binldou6</i>
	where-object	2;11;29 邊度 <i>binldou6</i>
	what-object	3;00;11 咩 <i>me1</i>
	which(-identification)	3;01;13 邊 <i>binl CL</i>
	who-subject	3;01;13 邊個 <i>binlgo3</i>
	what(-clarification)	3;02;13 咩 <i>me1</i>
	which-object	3;03;15 邊 <i>binl CL</i>
	how	3;03;15 點樣 <i>dim2joeng2</i>
	who(-identification)	3;08;09 邊個 <i>binlgo3</i>
	what-subject	3;08;09 咩 <i>me1</i>

Table 8 : The order of emergence of interrogative wh-words in indirect questions

Subject	Wh-Word	1st interrogative production
Tsuntsun	what-object	2;06;18 咩 <i>me1</i>
Chunyat	how	2;07;13 點樣 <i>dim2joeng2</i>
	what-extent	2;09;23 幾 <i>gei2</i>
Gakei	what-object	2;03;25 乜鬼 <i>matl gwai2</i>
		2;04;08 咩 <i>me1</i>
		2;09;09 乜(嘢) <i>matl(je5)</i>
	what-extent	2;08;08 幾 <i>gei2</i>
Tinfan	what-extent	2;02;10 幾 <i>gei2</i>
	who-subject	2;08;24 邊個 <i>binlgo3</i>
	what-object	2;08;24 咩 <i>me1</i>
	where-object	3;01;21 邊度 <i>binldou6</i>
Johnny	who-object	3;04;14 邊個 <i>binlgo3</i>
	what-object	3;04;14 咩 <i>me1</i>
Jenny	who-subject	2;11;29 邊個 <i>binlgo3</i>
	what-object	3;04;22 咩(嘢) <i>me1(je5)</i>
	what-extent	3;06;16 幾 <i>gei2</i>
	which-object	3;07;25 邊 <i>binl CL (N)</i>
	where-object	3;07;25 邊度 <i>binldou6</i>
	where-subject	3;08;09 邊度 <i>binldou6</i>

3.1.3 Relation between the verb type and the wh-word

In this study, it is observed that children use certain verb types more frequently with certain question types. Table 9a shows the common verb types and question types the children used in direct questions. Particular verb types are found to co-occur with particular question types. Table 9b indicates all the possible types of question which the verbs can be asked of according to adult grammar.

Table 9a: Verb used by children: co-occurrence with question type

a) Transitive actional (47 different verbs)		b) Transitive stative (20 different verbs)	
Q. type	Verbs	Q.type	Verbs
What-object	23	Why	17
Why	18	What-object	5
Who-subject	17	Who-subject	3
How	7	Which-object	1
Which-object	5	Who-object	1
Where-adjunct	5	What-subject	1
Who-object	4	Which-subject	0
Where-object	1	How	0
Which-subject	0	Where-object	0
		Where-subject	0
c) Intransitive actional (9 different verbs)		d) Intransitive stative (18 different verbs)	
Q.type	Verbs	Q.type	Verbs
Why	6	Why	11
Who-subject	2	How	3
How	1	Who-subject	2
Where-adjunct	1	Which-subject	1
Which-subject	0	What-subject	0
What-subject	0	Where-subject	0
Where-subject	0		
e) Locative actional (6 different verbs)		f) Locative stative (4 different verbs)	
Q.type	Verbs	Q.type	Verbs
Where-object	5	Where-object	2
Why	3	Why	2
How	1	What-object	1
What-object	1	Which-object	1
Who-subject	0	Where-adjunct	1
Which-subject	0	Who-subject	0
		What-subject	0
		Which-subject	0
f) Existential verb (1)		g) Copula verb (1)	
Q.type	Verb	Q.type	Verbs
What-object	1	What-object	1
Why	1	Who-object	1
Who-object	0	Which-subject	1
Who-subject	0	Where-subject	1
Which-subject	0	Why	1
Where-subject	0	Who-subject	0
What-subject	0	What-subject	0
		Where-object	0

Table 9b: Possible question types asked of the verb in adult grammar

a) <u>Transitive actional</u> (47 different verbs)		b) <u>Transitive stative</u> (20 different verbs)	
Q. type	Verbs	Q.type	Verbs
What-object	45	Why	20
Why	47	What-object	19
Who-subject	47	Who-subject	19
How	47	Which-object	16
Which-object	45	Who-object	2
Where-adjunct	47	What-subject	4
Who-object	13	Which-subject	19
Where-object	26	How	18
Which-subject	47	Where-object	9
		Where-subject	4
c) <u>Intransitive actional</u> (9 different verbs)		d) <u>Intransitive stative</u> (18 different verbs)	
Q.type	Verbs	Q.type	Verbs
Why	9	Why	18
Who-subject	8	How	5
How	9	Who-subject	16
Where-adjunct	9	Which-subject	14
Which-subject	7	What-subject	13
What-subject	4	Where-subject	10
Where-subject	2		
e) <u>Locative actional</u> (6 different verbs)		f) <u>Locative stative</u> (4 different verbs)	
Q.type	Verbs	Q.type	Verbs
Where-object	6	Where-object	4
Why	6	Why	4
How	6	What-object	1
What-object	3	Which-object	4
Who-subject	6	Where-adjunct	3
Which-subject	6	Who-subject	4
		What-subject	4
		Which-subject	4
f) <u>Existential verb</u> (1)		g) <u>Copula verb</u> (1)	
Q.type	Verb	Q.type	Verbs
What-object	1	What-object	1
Why	1	Who-object	1
Who-object	1	Which-subject	1
Who-subject	1	Where-subject	1
Which-subject	1	Why	1
Where-subject	1	Who-subject	1
What-subject	1	What-subject	1
		Where-object	1

Of the direct wh-questions produced by the children, it was observed that for verbs which had two arguments, different verb types had preference for particular type of argument questions. As can be seen from the tables above, for every type of verb, not the full range of questions were asked by the children. The children had different 'preferences' for different verb type. For example, out of all the possible argument questions that could be asked using these verbs, transitive actional verbs were mainly used with what-object questions and who-subject questions; for locative actional verbs, where-object questions were mostly

asked; for the existential verb 有 *jau5* 'have', it was mostly used with what-object questions.

3.2 Cross-sectional study

3.2.1 Order of acquisition

In the cross-sectional study, the effect of the verb type on the order of acquisition of wh-questions was examined, with the following orders being observed:

- (a) Order (irrespective of the verb type contained in the question and the syntactic position of the wh-word):

who- / what- < where- / why- < how- < when- questions

- (b) Order (depending on the verb type contained in the questions)

trans. actional verb: what- < who- / where- < why- < how- < when-

intr. actional verb: who- / what- < where- / why- < how- < when-

trans. stative verb: who- < where- / why- < what- < how- / when-

functional verb: who- / what- < where- / why- < when-

The argument questions, what-, who- and where- questions, were acquired before the adjunct questions, why-, how- and when- questions. Where-argument questions were acquired at approximately the same time as where-adjunct questions³.

Subject/object asymmetry was observed. What-object questions were acquired before what-subject questions. Who-object questions were before who-subject questions⁴.

³ This finding is different from the previous studies, wherein where-argument questions were acquired earlier than where-adjunct questions. This is related to the choice of verb used in the experiment. Clancy (1989) found that for where- questions, those containing different verbs would be acquired at different times. The same observation was also made in our longitudinal study. The acquisition pattern was as follows:

where_{is} < where_{go} < where_{loc act.} < where_{non-loc act.}

In the cross-sectional study, the choice of verbs was those of non-locative actional verbs, thus 邊度 *bin1dou6* where-argument questions containing these verbs were acquired at a later stage, that explained why these questions were not earlier comprehended than the adjunct questions in the study.

⁴ The finding that who-subject questions were more difficult than 邊個 *bin1go3* who-object question was different from results of the other studies, including that reported in the longitudinal study. However, it was found that the poorer performance on 邊個 *bin1go3* who-subject was contributed to mainly by the question containing the verb 打 *daa2* 'hit'. A plausible explanation for this poor performance might be that in this particular story, the boy in the story, 小明 *siu2ming4* Siu Ming was both the agent and the patient of the action expressed by the verb 打 *daa2* 'hit'. He hit his sister and was in turn hit by his mother. This might have affected the comprehension of the children. The other explanations are yet to be explored.

3.2.2 Relation between the verb type and the wh-word

Children's performance on the same question type differs according to the nature of verb type used. Questions containing transitive actional verbs were amongst the earliest to be comprehended by the children. Children's performance on questions containing transitive actional verbs either surpassed (e.g. 乜嘢 *mat1je5* what- and 邊度 *bin1dou6* where- questions first reached criterion in transitive actional verb) or equalled to that on other verb types (e.g. 點解 *dim2gaai2* why- questions first reached criterion in two categories, one of which was transitive actional verb).

A one-way ANOVA was carried out with each subject's percentage-correct score as the dependent variable and verb type as the repeated-measures independent variable. It was found that verb type was a significant factor ($F(3, 596)=3.2, p<0.05$). The tukey test revealed that across the sample as a whole, performance on transitive actional verbs was significantly better than on transitive stative verbs (mean 70.0 vs 58.0; $p<0.05$).

To find out how the relative difficulty of questions types depends on the type of verb contained in the question, a one-way ANOVA was carried out for each question type. Verb type was a significant factor for 乜嘢 *mat1je5* what- and 邊度 *bin1dou6* where- and 點樣 *dim2joeng2* how- questions ($p<0.005$) but not 邊個 *bin1go3* who-, 點解 *dim2gaai2* why- or 幾時 *gei2si4* when-questions.

For 乜嘢 *mat1je5* what-questions, performance on transitive stative verbs was significantly worse than that on the other verb types. 點樣 *dim2joeng2* how-questions for the various verb types reveals that transitive stative verbs were significantly more difficult than transitive actional and intransitive actional verbs ($p=0.000$). As for 邊度 *bin1dou6* where- questions, comprehension was significantly poorer on functional verbs than on transitive actional and intransitive actional verbs. The effect of verb type on the comprehension of these wh-questions is clearly illustrated in Figure 1 to 3 below.

Figure 1. Cantonese-speaking Children's Comprehension of Mat1ye5 'what'-questions

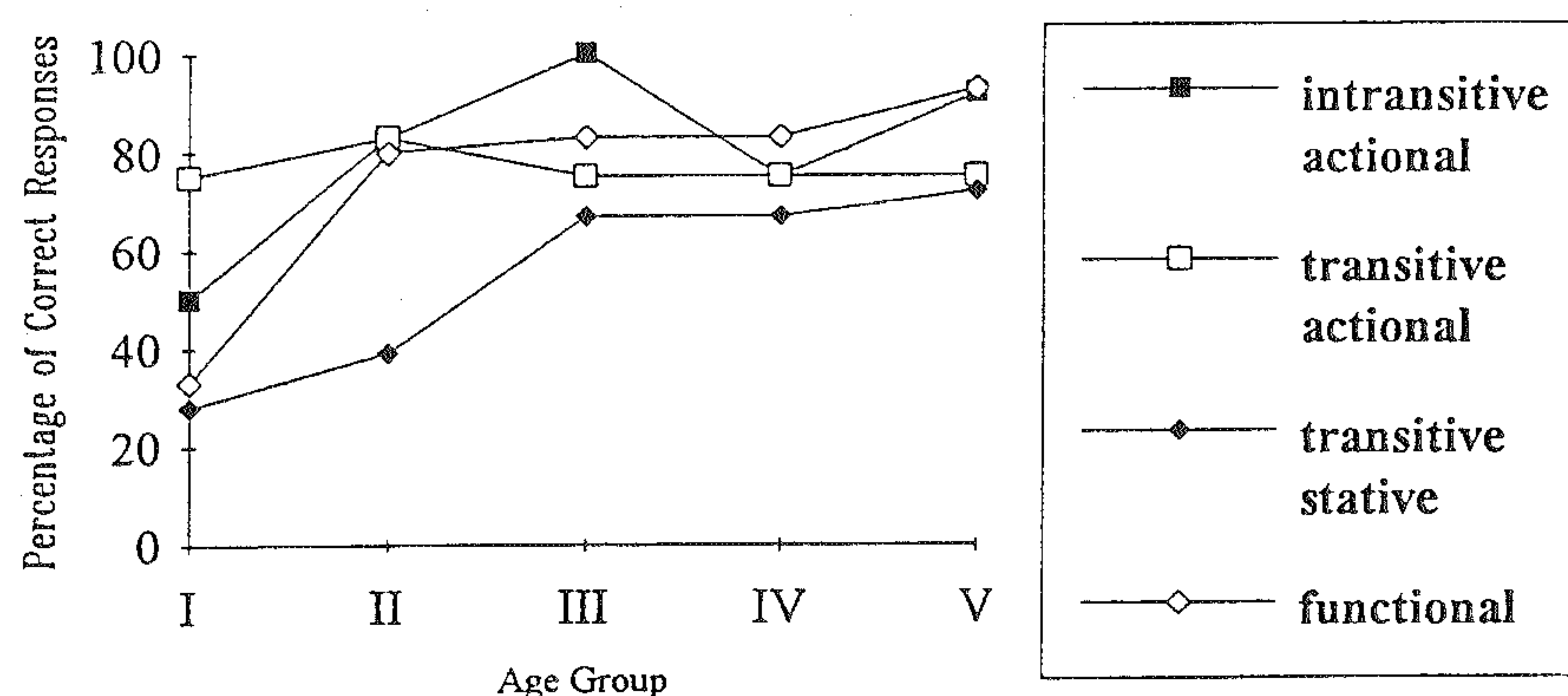


Figure 2. Cantonese-speaking Children's Comprehension of Bin1dou6 'where'-questions

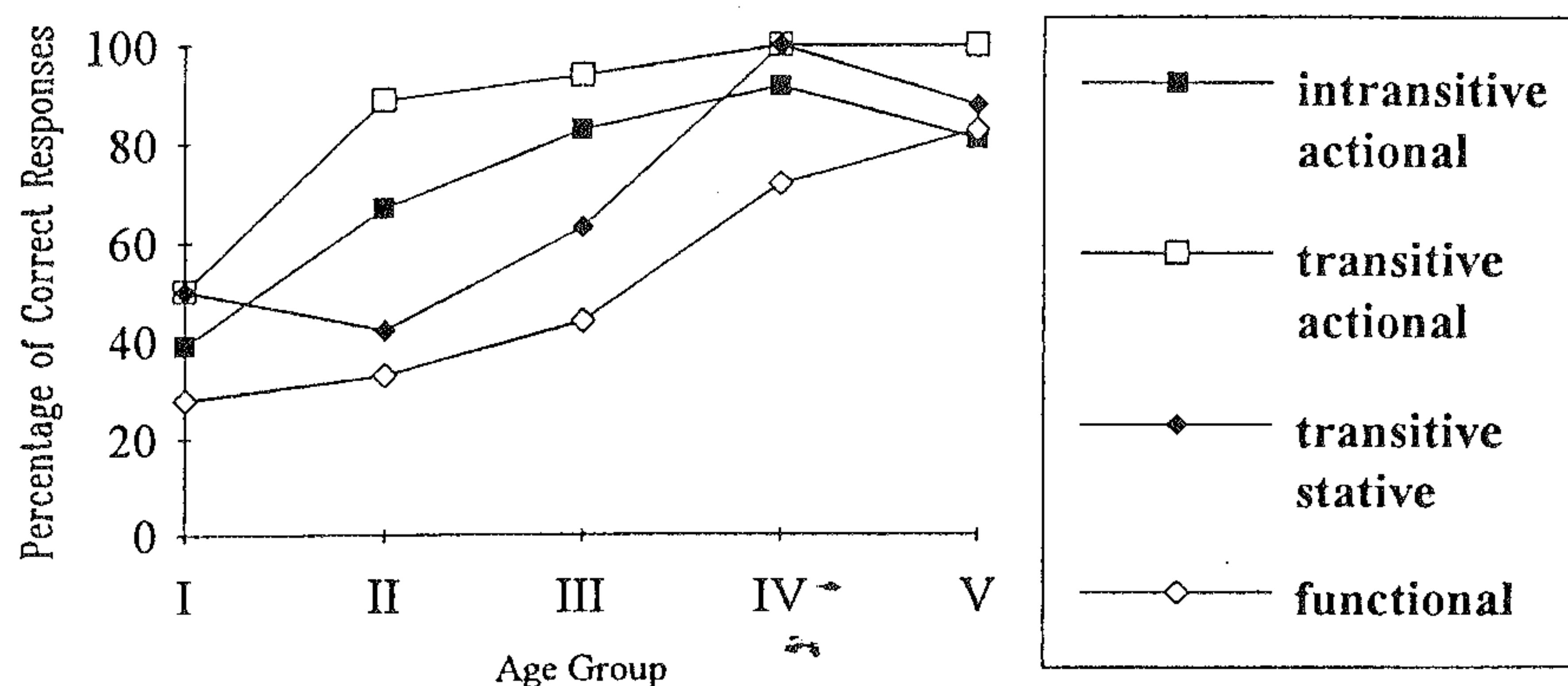
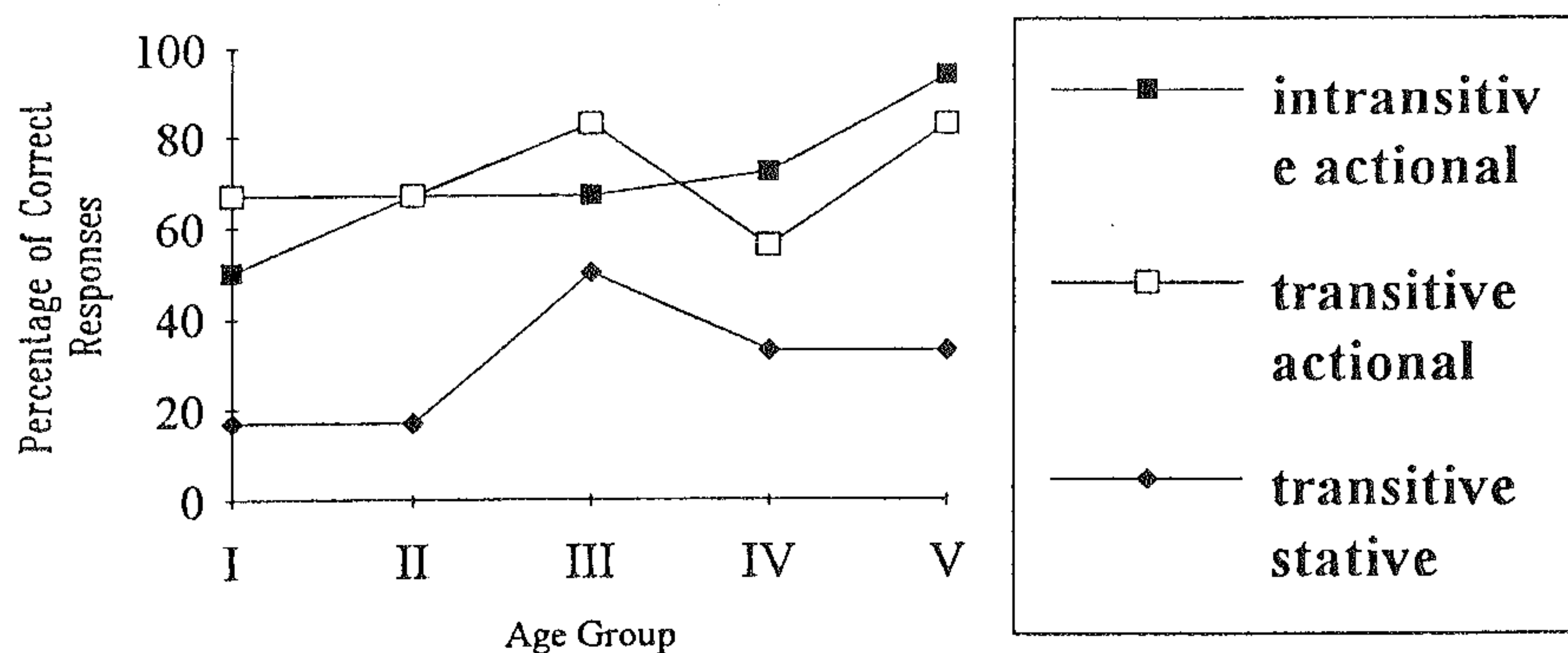


Figure 3. Cantonese-speaking Children's Comprehension of Dim2yeung2 'how'-questions



As can be seen from Figures 1 and 3, performance on 乜嘢 *mat1je5* what- and 點樣 *dim2joeng2* how-questions was consistently lower for transitive stative verbs than that for the other verb types. Figure 2 shows that performance on 邊度 *bin1dou6* where-questions was much poorer for functional verbs than for other verb categories.

4.0 Discussion

Based on the analysis of data collected from the longitudinal and cross-sectional studies, as presented in Sections 3.1 and 3.2, this study proposes that the semantic properties of the verb and the order of acquisition of verbs in children contribute to the acquisition of wh-words.

The obligatory constituents of a sentence are determined by the semantic properties of the verb. Each verb has its argument structure and this structure determines which constituents of a sentence are obligatory and which are optional. The arguments of a verb are not arbitrarily selected. A semantic relationship, in terms of thematic role, holds between the argument and the verb. Different verbs have different argument structures and the theta roles that the arguments are assigned also differ. According to theta criterion (Chomsky 1981), each argument is assigned one and only one theta role and each theta role is assigned to one and only one argument.

Bloom, Lightbrown & Hood (1975) and Bloom, Miller & Hood (1975) studied children's acquisition of argument structure and emphasized the importance of verbs for acquisition. Besides, there is also increasing prominence in linguistic theory of the contribution of the lexical verb to the syntax of sentences (Bresnan 1978, Bresnan & Kaplan 1982, Chomsky 1982, Grimshaw 1990).

The early child language data provide some evidence for the prominence of the verb in a sentence. From the children's earliest errors made in the comprehension of wh-questions, it is observed that children normally repeat the verb of the wh-question in responding to questions that they have not yet acquired. In the errors made in the earliest stage of the child's language development, it is the typical that only the verb or verb phrase, not the wh-word, is given as the answer to the question.

Relationship between the verb and the acquisition of wh-questions has been observed in a number of studies (Tyack & Ingram 1977, Cairns & Hsu 1978, Bloom et al 1982). In this study, it is observed that children use certain verb types more frequently with certain question types (see Table 9a in Sect. 3.1.3).

Such kind of preference indeed throws some light on the distinction between interrogative questions and some of the rhetorical questions. The use of a verb type with a "non-preferred" wh-word to a certain extent suggests the possibility of the wh-utterance being a rhetorical question rather than an interrogative question. Examples of such types are like 邊 *bin1* 'where' + 有 *jau5* 'have' (where 'where' is mostly used with locative verbs and 'have' is mostly used with object questions, 咩 *me1* 'what' + 冇 *mou5* 'not have' (where existential verb is used mainly in object questions), 點 *dim2* 'how' + 知 *zi1* 'know' (where 'how' is mostly used with intransitive stative verbs)), etc.

4.1 Verb Semantics and the Acquisition of Wh-words

As discussed in Section 4.0, different verbs have different argument and thematic structure. We suggest here that verbs that have multiple theta role mappings are more difficult than verbs with unique theta role mappings. Thus questions that contain the former type of verb would be acquired later. This is illustrated in the acquisition of 邊度 *bin1dou6* where- questions.

In the children's production of argument questions containing locative verbs, 86.36% of these were 邊度 *bin1dou6* where-object questions. That is to say, whenever locative verbs appeared in questions, they mainly occurred in 邊度 *bin1dou6* where-object questions. On the other hand, 邊度 *bin1dou6* where-object questions containing locative verbs were earlier to be comprehended by the children than other questions, such as 邊個 *bin1go3* who-subject questions, containing the same verb type. Besides, 邊度 *bin1dou6* where-object questions containing non-locative verbs were acquired later than 邊度 *bin1dou6* where-object questions containing locative verbs.

The above observations can be accounted for by looking at the theta role mappings. Locative verbs assign the role of location to the internal argument. In the case of locative actional verbs, for Chinese, the role assigned to the internal argument is locative goal. For locative stative verbs, the role of location is assigned to the internal argument. As for the external argument of locative verbs, it can be assigned the role of theme or patient⁵, and of those NPs assigned with the role of theme, they either carry the semantic feature of [+animate] or [-animate]. This is illustrated in (a)-(c) below.

- (a) 媽咪 喺 街 市。
maalmi4 hai2 gaailsi5
mummy at market
(Mummy is at the market.)
- (b) 媽咪 去 街 市。
maalmi4 heoi3 gaailsi5
mummy go market
(Mummy goes to the market.)
- (c) 的 蘋 果 喺 雪 櫃。
dil ping4gwo2 hai2 syt3gwai6
CL apple at refrigerator
(The apples are in the refrigerator.)

The internal argument of the locative verb is always assigned the role of location as shown in (a) to (c). The external argument in (a) and (b) is assigned the role of theme and patient respectively and the NP assigned with the above thematic roles carries the semantic features of [+animate, +human]. As for the external argument in (c), it is assigned the role of theme but unlike that in (a), the NP carries the semantic features of [-animate]. The semantic property of the internal argument of the verb, as compared to the external argument, is unique and without the need to make any choices, it should be easier for the child and thus 邊度 *bin1dou6* where-object questions with locative verbs are earlier to be acquired than other wh-questions, like those of 邊個 *bin1go3* who-subject and 乜嘢 *mat1je5* what-subject questions, containing a locative verb.

⁵ Theme refers to the entity affected by the action or state expressed by the predicate. Patient refers to the person or thing undergoing the action expressed by the predicate.

It is also observed that where-object questions containing non-locative verbs were acquired later than those containing locative verbs. The order of acquisition of 邊度 *bin1dou6* where-object questions containing different verb types is as follows: where_{is} < where_{go} < where_{loc act} < where_{non-loc act}. Similar order of acquisition is found in Clancy (1989). It can be seen that even within the same type of wh-question, questions containing different verb types may be acquired at different times. The internal argument of the locative verb is assigned the theta role of location, while non-locative verb does not necessarily or commonly assign a locative role to the internal argument. Verbs that assign only one type of theta role to the argument are easier for the child. For verbs that may optionally assign different theta role to the argument, e.g. 揭 *kit3* "turn over", which can assign the theta role of patient (in 乜嘢 *mat1je5* what-object questions) or location (in 邊度 *bin1dou6* where-object questions), where-object questions containing these non-locative verbs are relatively more difficult for children as a choice need to be made. The particular semantic role that a locative verb expects is location while it is not the case with non-locative verbs. Some earlier studies have shown that there is in general a particular semantic role that a verb expects (Carlson and Tanenhaus (1988) Stowe (1988), Tanenhaus and Carlson (1989), Tanenhaus, Boland, Garnsey, and Carlson (1989) as in Stowe, Tanenhaus and Carlson (1991)). Thus if what is asked for matches with what the verb expects, the particular question should be easier for the child to comprehend.

By looking at the theta role mappings, we can also account for the earlier acquisition of 邊度 *bin1dou6* where-object questions than 邊度 *bin1dou6* where-subject questions. All locative verbs can assign a locative role to its internal argument but not a locative role to the external argument. These verbs cannot be asked of in 邊度 *bin1dou6* where-subject questions; only intransitive and functional verbs can occur in these questions. The semantic possibility for these verbs assigning a locative role to the external argument is very low. With intransitive verbs, the external argument is either assigned the role of agent or theme or on rare occasions, the role of location. Functional verbs do not have a particular preference to assign a locative role to the external argument. With the multiple theta mapping on the external argument and with location being the 'non-preferred' role assigned by these verb types, they explained the difficulty in the acquisition of 邊度 *bin1dou6* where-subject questions, hence their later acquisition than 邊度 *bin1dou6* where-object questions.

As can be seen, the argument and thematic structure of the verb do not only give suggestions to which particular verb type will likely occur with which wh-word type, they also give some suggestions as to the acquisition order of the different wh-words.

Identification questions such as 'what is this' and 'where is it' are among the earliest to be acquired by children. In the 'what is this' questions, the verb is usually the copula 係 *hai6* 'is', and it either appears in the question, or it may not appear on the surface.

Apart from 係 *hai6* 'is', other naming verbs such as 叫做 *giu3zou6* 'named as' or 等於 *dang2jyul* 'equal' may also appear in these questions. This class of naming verb equates or defines a person or a thing. Verbs belonging to this class have two arguments and the roles assigned to these arguments are fixed. The internal argument is assigned the role of identificational reference object while the external argument is assigned the role of theme (Jackendoff 1990). With no other options possible, children need not choose from among the options and thus it is easier for the children.

The verb in the 'where is it' questions is the locative stative verb 喺 *hai2* 'be in'. This verb assigns a locative role to its internal argument and a theme role to its external argument. Like 'what is this' questions, the roles assigned are fixed with no other options. That is why it is also easier for the children. This explains why questions containing verbs such as actional and stative verbs are acquired later. For example, for the transitive actional verb 踢 *tek3* 'kick', the internal argument can be assigned the role of patient or location. Questions containing such classes of verbs are more difficult than the identification questions.

We have suggested earlier on that the asymmetry in the acquisition of 邊度 *bin1dou6* where-object and 邊度 *bin1dou6* where-subject questions can be accounted for by looking at the theta role mappings. The subject-object asymmetry observed in the order of acquisition of 邊個 *bin1go3* who-questions and 乜嘢 *mat1je5* what-questions is also accounted for when we look into the theta role mappings involved and the semantic features required of the argument as assigned by the verb.

邊個 *bin1go3* who-subject questions are acquired before 邊個 *bin1go3* who-object questions while 乜嘢 *mat1je5* what-object questions are acquired before 乜嘢 *mat1je5* what-subject questions. As can be seen from Table 9a, who- and what- questions mostly contain actional verbs. The external argument of transitive actional verbs is assigned the role of agent and most of these verbs require a [+animate, +human] agent. Transitive actional verbs allowing a [+animate, -human] agent are limited. As for the internal argument, it can be assigned the role of either theme or location. In looking at the theta role mappings, we can anticipate that subject questions should be easier than object questions, thus 邊個 *bin1go3* who-subject questions are acquired before 邊個 *bin1go3* who-object questions. This alone cannot account for the earlier acquisition of 乜嘢 *mat1je5* what-object questions than 乜嘢 *mat1je5* what-subject questions. Here, the semantic features required of the argument as assigned by transitive actional verbs is important in accounting for this asymmetry. Of the 47 transitive actional verbs referred to in Table 9a, none of them allows for an external argument having the semantic feature of [-human], while 45 of them allow for the assignment of the semantic feature of [-human] to the internal argument. With actional verbs being acquired earlier than other types of verbs, which will be discussed in 4.2, and their requirements of the semantic features of the arguments, we can account for why 乜嘢 *mat1je5* what-object questions are acquired before 乜嘢 *mat1je5* what-subject questions. The case of 邊個 *bin1go3* who-questions can also be further explained as transitive actional verbs that allow

for an internal argument with the semantic feature of [+human] is far fewer than those which allow for the semantic feature of [-human]. 45 of them allow for a non-human object but only 13 of them allow for a human object. The semantic possibility in this case favours that of a [+animate, +human] agent and a [-human] patient for transitive actional verbs. This explains why 邊個 *bin1go3* who-subject questions occur before 邊個 *bin1go3* who-object questions and 乜嘢 *mat1je5* what-object questions before 乜嘢 *mat1je5* what-subject questions.

Apart from the theta role mappings and semantic features associated with the verbs, the type of complement that the verb subcategorises for explains why children showed different performance on the comprehension of the same wh-questions which only differ in the verb type contained. From the cross-sectional results put forward in Section 3.2.2, it was shown that 乜嘢 *mat1je5* what-object questions containing transitive stative verbs were much more difficult than those containing transitive actional verbs. Only children of GpV (4;6-5;0) have started to acquire 乜嘢 *mat1je5* what-object questions containing transitive stative verbs. For 乜嘢 *mat1je5* what-object questions containing transitive actional verbs, children of GpI (2;6-3;0) have already acquired them. Transitive stative verbs can either subcategorise for an NP or, in the usual case, a clausal complement while transitive actional verbs subcategorise for an NP complement. An NP complement is simpler and easier for the child, thus questions containing transitive actional verbs are also easier. Besides, of the three stative verbs, 諗住 *nam2zyu6* 'have in mind', 驚 *geng1* 'fear' and 載住 *zoi3zyu6* 'hold', chosen in the experiment, the first two questioned a clausal complement while the third one questioned an NP complement. It is found that what-object questions with stative verbs that subcategorise for an NP complement were easier than those that subcategorise for a clausal complement. Children of Gp II (3;0-3;6) have already acquired the former type of questions, while for the latter type, children of GpV have only started to acquire them.

The difference in the performance of adjunct wh-questions which only differ in the verb contained in the question is also found to be a result of the relation between the verb and the wh-word questioned. 點樣 *dim2joeng2* how-questions with actional verbs were acquired earlier than how-questions with stative verbs. In how-questions with actional verbs, *how* held a manner relationship with the verb and described the action of the verb. In those with stative verbs, a manner relationship is not held between *how* and the verb. For the stative verb 載住 *zoi3zyu6* 'hold', *how* shows a state as a result of the action described by the verb. This made the comprehension of how-questions with actional verbs easier than those with stative verbs.

4.2 Verb acquisition and wh-word acquisition

The order of acquisition of the different verb types is observed to be playing a role in the order of acquisition of wh-questions as well. Previous studies have shown that actional verbs are amongst the earliest to be acquired (Bloom 1991, Tomasello 1992), while stative and mental verbs are acquired later. With actional

verbs being acquired at an earlier stage, the early wh-questions usually contain actional verbs. As discussed in the last section, the verb semantics of actional verbs favours their occurrence with what-object questions and who-subject questions, thus the order of acquisition of verbs and the semantic properties of different verb types together contribute to the orders of acquisition of what-object questions before what-subject questions and who-subject questions before who-object questions.

Different verbs bear different semantic relations with the other constituents of the sentence. The order of acquisition of wh-words is also related to the order of acquisition of semantic relations. The earliest questions acquired are the identification questions, e.g. 呢個係咩 What is this? 喺邊 Where is it? These questions contain the functional verbs 係 *hai6* "is" and 喺 *hai2* "be in" respectively. The copula verb 係 *hai6* "is" has the function of identifying and equating entities in the subject and object position of the sentence. The locative verb 喺 *hai2* "be in" has the function of identifying the location of certain entities. The early occurrence of these questions are in line with the observation that functional relations are acquired before verbal relations (Brown 1973, Bloom 1991, Tomasello 1992), which is related to the earlier acquisition of functional verbs than descriptive verbs. The semantic relation of existence is acquired before the relation of agent-action-object.

點解 Why- questions in the early stage are why-questions asking about the existence or non-existence of certain state or phenomenon, followed by why-questions of motivation e.g. 你做乜飲呀? *nei5 zou6mat1 jam2 aa3* 'Why did you drink?' The motivation for the drinking act is being queried. The early acquisition of the functional relation of existence enhance the earlier production of why-questions compared to the other adjunct questions, and for some children even before the 邊個 *bin1go3* who-questions. The later acquisition of actional verbs than functional verbs explains why questions of motivation are acquired later than questions of existence and states, since the former questions contain actional verbs.

Mental verbs and certain stative verbs subcategorise for clausal complements. These verbs are acquired later than actional verbs, and indirect wh-questions with these verbs as matrix verbs are observed to be acquired later in comparison to direct questions. So the order of verb acquisition does play a role in the order of acquisition of both direct and indirect questions.

The order of acquisition of verbs not only helps to account for the acquisition of wh-words in direct questions, but also accounts for the earlier acquisition of direct questions than indirect questions for the same interrogative wh-word, as well as the similar acquisition order of wh-words in direct and indirect questions. Verbs that can subcategorize for a clausal complement include mental verbs e.g. 知 'know' and some stative verbs e.g. 驚 *geng1* 'fear'. Besides, actional verbs followed by the particle 下 *haa5* can also subcategorize for a clausal complement e.g. 揭下 *kit3 haa5* 'flip through' (The particle 下 *haa5* means something like 'for a short duration'). The verb types that subcategorise for a

clausal complement are acquired later than actional verbs and stative verbs that subcategorize for a noun phrase complement, thus indirect questions with these verbs as the matrix verb are acquired later than direct questions. The acquisition of indirect questions whose matrix verbs are actional verbs followed by the particle 下 *haa5* is at about the same time as, or even earlier than, that of direct questions containing this verb type. Thus the relative order of acquisition of direct and indirect questions lies not so much in the difference between their syntactic complexity, but rather on the verb type involved. Besides, the orders of acquisition of wh-words inside direct questions or indirect questions are the same. This shows that the same effect of the semantic properties of the verb on the acquisition of wh-words holds in both matrix and embedded clauses.

4.3 Conclusion

This study on the acquisition of wh-words by Cantonese-speaking children is an attempt to give a detailed description of the pattern of acquisition of the different wh-words, both their interrogative and non-interrogative uses, and to explore on the factors which have contributed to such a pattern.

The seldom explored non-interrogative wh-words and wh-words used in indirect questions have also been looked into in the study, throwing light on their relation with their interrogative counterparts used in direct questions. Such an exploration adds much to the existing literature wherein studies have concentrated on the acquisition of interrogative wh-words as used in direct questions.

A number of observations have been made. In relation to the order of acquisition, we have found that there is an order in the acquisition of wh-questions. 乜嘢 *mat1je5* 'what-' and 邊度 *bin1dou6* 'where-' identification questions (i.e. questions containing the copula 係 *hai6* 'is') are the earliest to be produced, followed by 乜嘢 *mat1je5* 'what-' and 邊度 *bin1dou6* 'where-' questions using verbs other than that of the copula 係 *hai6* 'is', and later on who-, why- and how- questions. 幾時 *gei2si4* 'when-', 幾多 *gei2dol* 'how-many-', 幾 *gei2* 'what-extent' and 點樣 *dim2joeng2* 'what-like' questions are amongst the latest to be acquired; they have not yet been acquired by any one of the children under study. It is generally observed that argument questions are acquired before adjunct questions. The order is observed to occur in both the acquisition of direct and indirect questions.

In relation to the subject-object asymmetry found in the argument questions, we have found that for 乜嘢 *mat1je5* 'what-' and 邊度 *bin1dou6* 'where-' questions, object questions are acquired before subject questions, while for 邊個 *bin1go3* 'who-' questions, subject questions are acquired before object questions.

As for the acquisition of non-interrogative wh-words, we have observed that for children of this young age, they have started to use non-interrogative wh-words in environments as found in adult speech. They are used as indefinite pronouns, intensifiers and they appear in rhetorical questions. However, they are

used far less than the interrogative wh-words. The limited data did not reveal a regular pattern in their order of acquisition and further exploration is needed to draw a conclusion as to whether there is any specific order. We observe that interrogative wh-words which occur in the argument positions are acquired before their non-interrogative counterparts. For those that occur in adjunct position, the interrogative and non-interrogative uses appeared at about the same time.

Verb is a prominent constituent in a sentence and it determines which other constituents should occur in a sentence. Wh-words, as one of the constituents of the utterance, have thus a close relation with the verb. We propose in this study that verbs indeed play an important role in the acquisition of wh-words and that it is the main factor that contributes to such an acquisition. The argument and thematic structure of the different verbs and the order of acquisition of verbs together contribute to the acquisition pattern of the wh-words. They accounted for the order of acquisition of the interrogative wh-words both in direct and indirect questions, the subject-object asymmetry as observed in argument questions, as well as some of the differences that distinguish between interrogative and non-interrogative wh-words and that help children make the distinction. As the semantic and syntactic properties of the verb and their acquisition are same across different languages, the similarities in the acquisition of wh-words across different languages are thus accounted for.

The findings of the study show a similar trend as found in other crosslinguistic studies of both movement and non-movement languages, thus adding evidence to the universality across languages and the process of language acquisition.

Our proposal of the importance of verb in accounting for the acquisition pattern of wh-words goes beyond the other proposals in previous studies as it can account for most of the different patterns observed. The problems in accounting for some patterns left unexplained in earlier proposals have been tackled. Of course, earlier proposals such as the cognitive development, input frequency, contextual situation, etc. have their importance though they may only account for some of the patterns observed. Besides, the role of pragmatics in the child's language acquisition cannot be undermined. The different proposals may not be mutually exclusive; they can indeed be treated as complementary to one another. Yet, how the different factors actually interact to bring about the acquisition and the level of importance of each of these factors will be an interesting topic to pursue in future studies.

References

- Alleton, Viviane (1988). "The so-called "rhetorical interrogation" in Mandarin Chinese" *Journal of Chinese Linguistics* 16(2):278-296.
- Bloom, Lois, Patsy Lightbrown & Lois Hood (1975). "Structure and variation in child language". *Monographs of the Society for Research and Child Development*, 40 (serial no.160).

- Bloom, Lois, Peggy Miller & Lois Hood (1975). "Variation and reduction as aspects of competence in language development". In A. Pick (ed.), *Minnesota symposia on child psychology* (vol.9, pp.3-55). Minneapolis: University of Minnesota Press.
- Bloom, Lois, Susan Merkin & Janet Wootten (1982). "Wh-questions: linguistic factors that contribute to the sequence of acquisition". *Child Development* 53:1084-92.
- Bloom, Lois (1991). *Language Development from Two to Three*. Cambridge University Press.
- Bresnan, J. (1978). "A realistic transformational grammar". In M. Halle, J. Bresnan & G. Miller (eds.), *Linguistic theory and psychological reality* (pp.1-59). Cambridge, MA: MIT Press.
- Bresnan, J. & Kaplan, R. (1982). "Grammars as mental representations of language". In J. Bresnan (ed.), *The mental representation of grammatical relations* (pp.xvii-lit). Cambridge, MA: MIT Press.
- Brown, Roger (1968). "The Development of Wh Questions in Child Speech". *Journal of Verbal Learning and Verbal Behavior* 7:279-290.
- Brown, Roger (1973). *A First Language: the early stages*. Cambridge, MA: Harvard University Press.
- Cairns, Helen & Jennifer Hsu (1978). "Who, Why, When and How: a developmental study". *Journal of Child Language* 5:477-488.
- Carlson, G.N. & Tanenhaus, M.K. (1988). "Thematic roles and language comprehension" in W. Wilkins (ed.) *Thematic Relations* (pp.263-288). New York: Academic Press.
- Cheng, Robert L. (1984). "Chinese question forms and their meanings" *Journal of Chinese Linguistics* 12(1):86-145.
- Cheung, Shuk-yee A. (1992). The Acquisition of Interrogatives by Cantonese-speaking Children. Paper presented at the 1992 Linguistic Society of Hong Kong Annual Research Forum.
- Cheung, Shuk-yee A. & Thomas Hun-tak Lee (1993). "Cantonese-speaking Children's Comprehension of Wh-questions" in Clark, Eve (ed.) *The Proceedings of the Twenty-fifth Annual Child Language Research Forum*. Stanford: CSLI.
- Chomsky (1982). *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: MIT Press.

- Clancy, Patricia (1989). "Form and function in the acquisition of Korean wh-questions" *Journal of Child Language* 16:323-347.
- de Villiers, Jill (1991). "Why Questions" in T. Maxfield and B. Plunkett (eds) *The Acquisition of Wh*. University of Massachusetts Occasional Papers in Linguistics, Amherst: GLSA.
- Erbaugh, Mary (1982). *Coming to Order: natural selection and the origin of syntax in the Mandarin-speaking child*. Doctoral dissertation, University of California at Berkeley.
- Ervin-Tripp, Susan (1970). "Discourse agreement: how children answer questions" in Hayes J.R. (ed.), *Cognition and the Development of Language*. New York: Wiley.
- Ferguson, Charles A. (1989). "Individual differences in language learning" in M. Rice & R. Schiefelbusch (eds.), *The Teachability of Language*. Baltimore, MD: Paul Brooks.
- Grimshaw, J. (1990). *Argument Structure*. Cambridge: The MIT Press.
- Grimshaw J. & S. Vikner (1990) "Obligatory adjuncts and the structure of events" ms., Brandeis University and University of Geneva.
- Harris, Zellig (1978) "The interrogative in a syntactic framework" in Hiz, H. (ed), *Questions* (pp.1-35). Dordrecht: Reidel.
- Heath, S.B. (1982). "Questioning at home and at school: a comparative study" in G. Spindler (ed.), *Doing the ethnography of schooling: educational anthropology in action*. New York: Holt, Rinehart & Winston.
- Holzman, Mathilda (1972) "The use of interrogative forms in the verbal interaction of three mothers and their children" *Journal of Psycholinguistic Research* 1(4):311-336.
- Huang, J. (1982a) *Logical Relations in Chinese and the Theory of Grammar*. Doctoral dissertation, MIT.
- Huang, J (1982b) "Move wh in a language without wh-movement" *The Linguistic Review* 1:368-416.
- Hyams, Nina (1984) *Language Acquisition and the Theory of Parameters*. Dordrecht: Reidel.
- Jackendoff, Ray (1990). *Semantic structures*. Cambridge, MA: MIT Press.
- Klima, E.S. & Bellugi, U. (1966). "Syntactic regularities in the speech of children" in J. Lyons & R.J. Wales (eds), *Psycholinguistics papers*. Edinburgh: Edinburgh University Press.

- Lee, Hun-tak Thomas (1989). "Development of a Mandarin-speaking Child's Comprehension of Wh-questions" *Cahiers de Linguistique Asie Orientale* 18 (1):29-62.
- Li, Charles N. & Sandra A. Thompson (1981). *Mandarin Chinese: a Functional Reference Grammar*. Berkeley & Los Angeles: University of California Press.
- Li, Charles N. & Sandra A. Thompson (1984). "Mandarin" in Chisholm Jr., W.S. (ed) *Interrogativity: a colloquim on the grammar, typology and pragmatics of questions in the seven diverse languages* (pp. 47-61) Amsterdam: John Benjamins.
- Li, Lai-ha (1968) *The Development of the Syntax of a Cantonese-speaking Child from 1 1/2 to 2 1/2 Years of Age*. M.A. dissertation. University of Malaya.
- Li, Yen-hui Audrey (1992). "Indefinite *Wh* in Mandarin Chinese" *Journal of East Asian Linguistics* 1:125-155.
- 李宇明、唐志東「四歲前兒童「誰」字句的發展」《語言研究》1990年第2期
(Li & Tang (1990) "Children's development of 'who'-sentences before four" *Yuyan Yanjiu* 2: 46-51)
- MacWhinney, B. & C. Snow (1985). "The Child Language Exchange System" *Journal of Child Language* 12:271-296.
- MacWhinney, B. (1991). *The CHILDES Project: Tools for Analyzing Talk*. Hillsdale, New Jersey: Lawrence Erlbaum.
- Maxfield, T. & B. Plunkett (eds.) (1991) *The Acquisition of Wh*. University of Massachusetts Occasional Papers in Linguistics. Amherst: GLSA.
- 繆小春「幼兒對疑問詞的理解－幼兒回答特殊疑問句的發展特點」朱曼殊主編《兒童語言發展研究》華東師範大學出版社1986年12月
(Miao (1986) "Young children's comprehension of wh-words: development of young children's response to wh-questions" in Zhu Man-shu (ed.), *Ertung Yuyan Fajan Yinjiu*. Huatung Shifan Daxue Chubanshe)
- Perez-Leroux, A.T. (1991). "The acquisition of wh-movement in Caribbean Spanish" in Maxfield, T. & B. Plunkett (eds.) (1991) *The Acquisition of Wh*. University of Massachusetts Occasional Papers in Linguistics. Amherst: GLSA.
- Prideaux, Gary D. (1976). "A functional analysis of English question acquisition" *Journal of Child Language* 3(3): 417-422.
- Rizzi, Luigi (1990). *Relativized minimality*. Cambridge: MIT Press.

- Shatz, Marilyn (1978). "Children's comprehension of their mothers' question-directives" *Journal of Child Language* 5:39-46.
- Shatz, Marilyn (1979). "How to do things by asking: form-function pairings in mother's questions and their relation to children's responses". *Child Development* 50:1093-1099.
- Sinclair, Anne and Ruth Van Gessel (1990). "The form and function of questions in children's conversations" *Journal of Pragmatics* 14: 923-944.
- Stowe, Laurie A. (1988). "Thematic structures and sentence comprehension" in G. Carlson and M.K. Tanenhaus (eds.) *Linguistic Structure in Language Processing* (pp.319-357). Dordrecht: Reidel.
- Stowe, Laurie A. , Michael K. Tanenhaus & Greg N. Carlson (1991). "Filling gaps on-line: use of lexical and semantic information in sentence processing" *Language and Speech* 34(4):319-340.
- Stromswold, Karin (1988). "Linguistic Representations of Children's Wh-questions. *Papers and Reports in Child Language Development* 27:107-114.
- Tanenhaus, M.K., Burgess, C., Hudson-D'Zmura & Carlson, G.N. (1987). "Thematic roles in language comprehension." *Proceedings of the Ninth Annual Cognitive Science Society Meetings* (pp.587-596). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Tanenhaus, M.K. & Carlson, G.N. (1989). "Lexical structure and language comprehension" in W.D. Marslen-Wilson (ed.) *Lexical Representation and Process* (pp.529-561). Cambridge, M.A.: MIT Press.
- Tomasello, Michael (1992). *First Verbs: a case study of early grammatical development*. Cambridge University Press.
- Tyack, Dorothy & David Ingram (1977). "Children's Production and Comprehension of Wh-questions" *Journal of Child Language* 4:211-224.
- Vaidyanathan, R. (1988) "Development of Forms and Functions of Interrogatives in Children: a longitudinal study in Tamil. *Journal of Child Language* 15:533-549.
- Yue-hashimoto, Anne (1993). *Comparative Chinese Dialectal Grammar: Handbook for Investigators*. Paris: Ecoles Des Hautes Etudes En Sciences Sociales.
- Zhu, De-xi 朱德熙 (1982). *Yufa jiangyi 語法講義*. Beijing: The Commercial Press.

Functional Categories in Child Cantonese

Thomas Hun-tak Lee, Colleen H Wong, Cathy S P Wong

Two divergent views have been proposed on the issue of functional categories in child language. The strong continuity hypothesis argues for the availability of these categories right from the very early stage of syntactic acquisition (e.g. Whitman, Lee and Lust 1991; Deprez and Pierce 1994). According to this view, the categorial inventory of child grammar is the same as that of adult grammar. Functional categories like inflection, complementizer and determiner are present in early grammar. An alternative view, represented by Radford (1990) and other researchers, regards the initial grammars of children (at around 2 years old) as differing from those of adults in lacking functional categories. Only lexical categories such as nouns, verbs, and adjectives are said to be present at the earliest stage. Lexical categories are distinguished from functional categories in their being able to assign thematic roles to their sister constituents; functional categories lack this ability of thematic role assignment.

Evidence such as the following data from two-year-olds are typically used to support the absence of determiners, modals, tense and agreement inflection, and complementizers like 'to' in early syntax.

- (1) read book ("I will read the book")
see cow ("I can see a cow")
man drive truck ("The man drives the truck")
want lady get chocolate ("I want the lady to get the chocolate")
Mummy doing dinner ("Mummy is doing the dinner")

The later onset of lexical categories and thus of a lexical-thematic stage is attributed either to maturation or to the idea that functional categories are the locus of parametric variation, and it takes time for the child to set the parameters.

On the other hand, recent research has also produced data that point to the necessity of positing functional categories. A good example of this type of evidence comes from Deprez and Pierce (1993, 1994) related to the order of negation and finite and nonfinite verbs in French. It is well known that French *pas* occurs after a verb if the verb is finite, but is positioned before it if the verb is nonfinite. The researchers found that children were sensitive to this distinction early on in their development. Therefore the inflectional category must be present right from the start. As the data below show, children before two know where to place *pas*, i.e. before an infinitive verb (as in 2), and after a finite verb (as in 3).

- | | | |
|-----|-------------------------------|--|
| (2) | pas casser (1;8)
not break | pas la poupee dormir (1;9)
not the doll sleep |
| (3) | marche pas (1;8)
walks not | elle a pas la bouche (1;10)
she has not the mouth |

Deprez and Pierce further show that while functional categories are available from the earliest stage of syntactic development, they may not be equally active. Thus, the data

from German suggest that the functional category of C(omplementizer) is not consistently utilized as a landing site for verb movement. Movement of the verb to I(nflection) as reflected in preverbal negation (as in 4a) is much more common in German children than postverbal negation, which indicates verb movement to C (as in 4b).

- (4) a. nein auch hause gehn (1;10)
 no also house go "I'm not also going home"
- b. mag nicht (1;9)
 like not

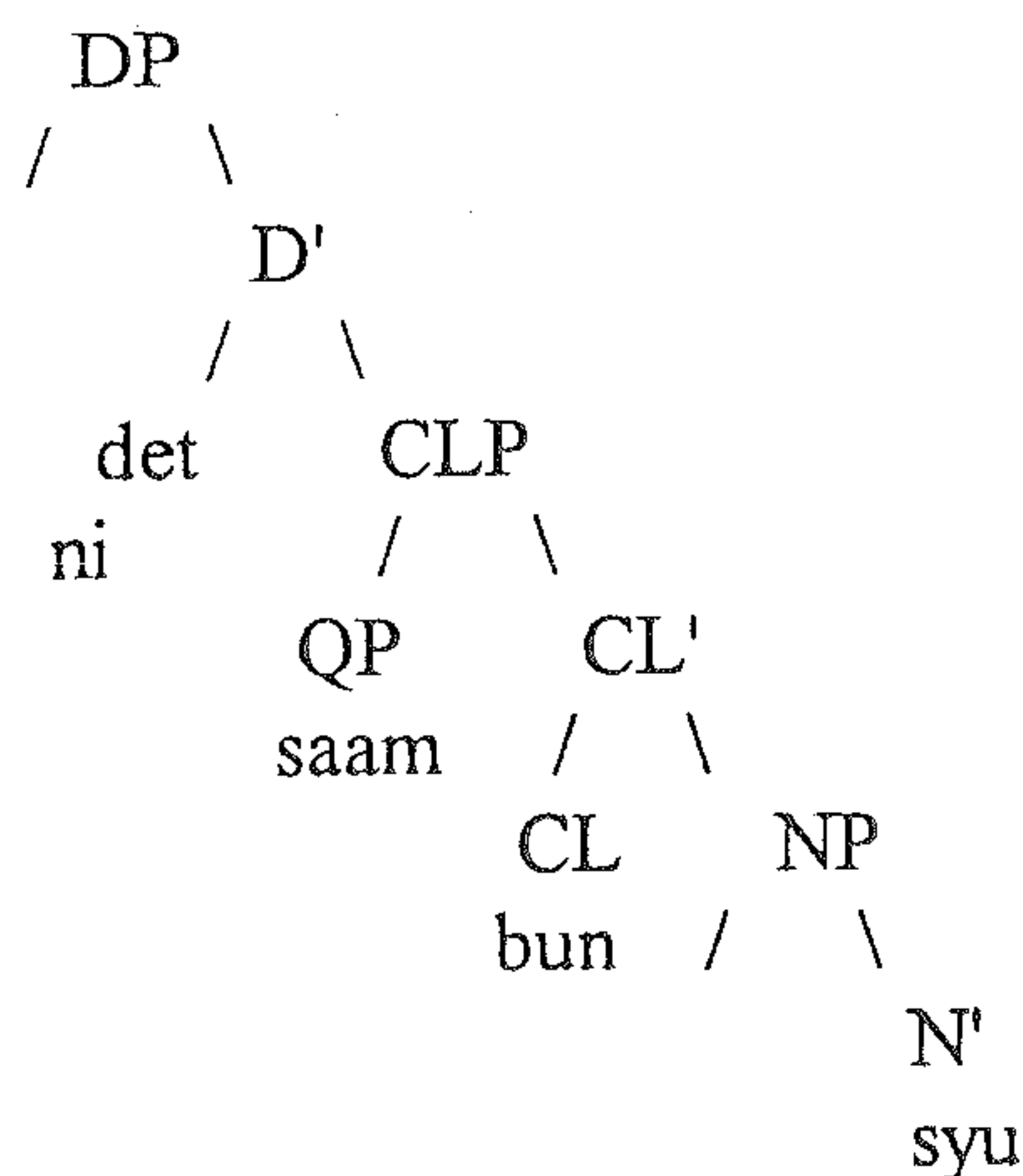
This paper reports findings on the early development of the functional categories of classifier, aspect, modal auxiliary and sentence final particles in Cantonese-speaking children. We argue that the Cantonese data provide support for a strong continuity view, since functional categories are evidenced before two years old, at around 1 year 8 to 9 months. If our argument is valid, it would be a demonstration of strong continuity in a language that is morphologically impoverished and whose functional categories are not reflected in a variety of overt movement processes.

The data for our analysis come from two children, Bernard and Tsuntsun, observed longitudinally for one year. Bernard was recorded from 1;7 to 2;8 in 26 sessions, whereas Tsuntsun was observed from 1;5 to 2;7 in 25 sessions. The MLU of Bernard is computed to be 0.95 morphemes at the beginning of the period of observation; this increased to 2.10 near the end of the period studied. The corresponding MLU range for Tsuntsun was 1.88 to 2.52 morphemes. The relevant data for our discussion of functional categories are based on the sessions before two years of age, since this is the crucial period bearing on the issue whether functional categories exist.

1. Classifiers

The first functional category we look at is the classifiers. The NP structure of Cantonese is assumed to be (5). Possible combinations of constituents within the NP are given in (6).

(5) NP structure in Cantonese:



It is well known that in a classifier language like Cantonese, the CL is necessary in delimiting and enumerating nouns. If a determiner or a numeral is present, then the classifier is obligatory. Cantonese differs from Mandarin in allowing the combination 'CL-noun'. The classifier in Cantonese is a bound form that must attach to another constituent. It cannot stand alone by itself. That is, one cannot have an utterance consisting of a single classifier.

- (6)
- | | |
|-----------------|--|
| bare noun: | syu1 'book' |
| CL noun | bun2 syu1 'CL-book' |
| det CL noun: | ni1 *(bun2) syu1 'det-CL-book' |
| num-CL-noun | saam1 *(bun2) syu1 'three-CL-book' |
| det-num-CL-noun | (ni1) (saam1) bun2 syu1 'this-three-CL-book' |

Earlier reports on Mandarin (e.g. Erbaugh 1986) have suggested that while Chinese children do not develop a sizable inventory of classifiers until well after 4 years old, they are nonetheless sensitive to the distribution of classifiers? When do Cantonese-speaking children become sensitive to classifier as a syntactic category?

Table 1 gives a tabulation of the spontaneous utterances of Bernard's containing classifiers, before the age of two. For Bernard, it appears that 1;09;18 is a turning point in his knowledge of classifiers. A number of observations can be made about the data. First, the general pattern of his classifiers is that they never occur alone. The classifier is a bound form for the child (with only one or two exceptions). Secondly, by 1;09;18, there are already 4 classifiers- the fuzzy classifier *di1*, the general individual classifier *go3*, the classifier for vehicles *ga3*, and the locative classifier *dou6*. They combine with the demonstrative *ni1/ go2* 'this/that'. In other words, one can think of the classifier as the category of elements that combine with 'det' to form a constituent. At this stage of development, it seems that classifiers do not generally combine with a head noun. Either a 'det CL' noun phrase or a 'num CL' noun phrase is found. If one takes into account the fact that bare nouns begin to emerge in Bernard only at 1;08, it is clear that classifiers emerge right from the beginning as a central part of the noun phrase for the Cantonese child.

The pattern of development for Tsuntsun is more complicated. For this child, bare nouns also emerge at around 1;08. However, classifiers for this child do not become productive until later, at around 1;11, as can be seen from Table 2. Tsuntsun has a curious pattern that distinguishes him from Bernard in that he used single classifier. Apparently, Tsuntsun did not fully realize at the early stage that the classifier should be a bound form.

Despite the individual variation between Tsuntsun and Bernard, we would still propose that the classifier is productive for him at 1;11. This is supported by two pieces of evidence. First, by that stage, we have already observed two different classifiers combining with distinct determiners. Secondly, Tsuntsun shows an odd behavior at this time, prefixing the negative *m4* 'not' to noun classifiers, as seen from the table. So it is reasonable to postulate that he is treating these as a category. Admittedly, the evidence for establishing the category of classifiers is not as strong for Tsuntsun as for Bernard. Nonetheless, we see some sensitivity to the 'det-cl' structure. It should be observed that for the children, there were no occurrences of 'det noun' or 'num noun' without the mediation of the classifier. For the child, the classifier is a set of elements that can occur alone, or be prefixed by *m4* 'not' or can suffix to the determiner.

Table 1: Classifiers in the speech of Bernard (1;7-1;11)*

Age	det-CL	CL-N	num-CL	det-CL-N	det-num-CL	CL	others
1;07;00	nilgo3(1) 'this-CL'						
1;08;00							
1;08;14						go3(1)	
1;08;28							
1;09;04	nilgo3(1)						
1;09;18	nildil(3) nilgo3(2) nilga3(1) go2ga3(1) 'that-CL' nildou6(1)					(hai2) dou6	
1;09;25	nildou6(2) nildil(2) nilgo3(2) nilbun2(1)		loeng2go3(2) 'two-CL' jat1go3(1) 'one-CL'				
1;10;10	nildou6(16) nilgo3(2) nilzek3(3)		saam1seoi3(1) 'three-CL' loeng2seoi3(1) loeng2zek3(1)				
1;10;23	nildou6(13) nilgo3(2)						NPnildou6
1;11;06	nildou6(9) nilgo3(2) nildil(2) nilzek3(1)			nilgo3 cinlcaul 'this-CL- swing'			

*The number in parenthesis indicates the number of occurrences of the classifier indicated; only non-repetitions are included; identical tokens occurring within the same conversational episode are counted as one.

Table 2: Classifiers in the speech of Tsuntsun (1;5-1;10)*

Age	det-CL	CL-N	num-CL	det-CL-N	det-num-CL-N	CL	Others
1;05;22	nilgo3(2) 'this-CL'						
1;07;03							
1;07;10							
1;08;00							
1;08;07	binlgo3(1) 'which-CL'					manl(1) zek3(1) zek3zek3(1)	
1;08;21							
1;09;07						(hai2)dou6(1)	
1;09;14	nilgo3(1)						
1;09;29						go3(1)	
1;10;30	binlgo3(1)					dou6(1) tiu4(1)	
1;11;13	nildou3(2) nilgo3(2)						m4lap1(1) m4go3(1) m4dou6(1) m4gin6(1) m4gau6(1) m4di1(1)
1;11;27	go2go3(2) 'that-CL' nildou6(1)					bun2(1) go3(1)	m4go3(1) m4jat1go3(1)

*The number in the parenthesis indicates the number of occurrences of the classifier indicated; only non-repetitions are included; identical tokens occurring in the same conversational episode are counted as one.

Some instances of doubling of classifiers were observed, e.g. a combination like *nil go3 zek3 goek3* 'det-cl-cl' N' (this foot) was produced by Tsuntsun at 2;04. An utterance like *nil go3 zek3 ngaan2* 'det-cl-cl-N' (this eye) was observed in the data of Bernard at 2;02. In each case, the general sortal classifier *go3* was coupled with another classifier, but these occurrences were not common (see Szeto, this volume, for further details). With regard to the NP structure, children seemed to have developed first bare nouns, then determiner - classifier sequences, followed by numeral - classifier combinations. As we will see below, the other functional categories of the two children also begin to emerge at around the ages

at which we observe classifiers, lending support to our view of their classifier development.

2. Aspect

The second functional category we examine is aspectual markings. Two distributional properties of aspect markers in Cantonese are (a) they are suffixed to the verb stem; (b) they do not cooccur with one another (with few exceptions). The three aspect markers observed in the data are the perfective aspect *zo2*, the durative aspect *zyu6*, and the progressive aspect *gan2*.

An outline of the properties of these aspect markers relevant to our discussion is in order. *Zo2* signals completion of event. It cooccurs with activity, achievement and accomplishment verbs, illustrated by (7). Generally *zo2* does not cooccur with states unless an inchoative meaning is possible. Thus *sap1 zo2* would mean 'become wet', and *laat6taat3 zo2* would mean 'become dirty'.

- (7) activity (or process) e.g. *fan3gaau3* 'sleep'; *da2bo1* 'play ball'
achievement (inc. punctual verbs) e.g. *da2laan6* 'break'; *m4gin3* 'disappear';
dit3 'fall down'; *zou2* 'leave'.
accomplishment e.g. *sik6faan6* 'eat meal'; *sai2tau4* 'wash hair';
lik1 hang4lei2 'carry luggage'; *tai2 syu1* 'read book'

- (8) states e.g. *sap1* 'wet'; *laat6taat3* 'dirty'

Gan2 is a progressive marker signalling that an action is being carried out (cf. Cheung 1972). It is compatible with activity verbs and accomplishment verbs. However, it generally does not cooccur with achievement verbs, since the latter do not admit of duration, as illustrated in (9). *Gan2* is incompatible with states (cf. 10). Nor is it used in imperatives as seen from (11).

- (9) **da2 gan2 laan6* 'breaking'
hit PROG bad
**m4 gin3 gan2* 'disappearing'
not PROG see
(10) **sap1 gan2* 'wetting'
**laat6 taat3 gan2* 'dirtying'
(11) *hei2 (*gan2) san1* 'stand up'
rise PROG body

In contrast to *gan2*, the durative marker *zyu6* which has at least the following meanings. According to Cheung (1972), it does not signify the progress of an action, but rather signals the representation of an action as a motionless state. This is reflected in the contrast between the sentences in (12).

- (12) a. lik1 gan2 di1 hang4lei2 'carrying the luggage'
 carry PROG some luggage
 b. lik1 zyu6 di1 hang4lei2 'holding the luggage'
 carry DUR some luggage

The former (12a) indicates the motion of carrying the luggage, whereas the latter signals a state of carrying the luggage without motion. Since it marks the durative aspect of an action, *zyu6* is incompatible with achievement verbs or states. It is not compatible with all activity and accomplishment verbs. Which verbs of these two categories *zyu6* can cooccur with will depend on whether the action signified can be represented as a state. Consider the following contrast:

- (13) a. ?Siu2ming4 da2 zyu6 bo1
 Siuming hit DUR ball
 b. Siu2ming4 da2 gan2 bo1
 Siuming hit PROG ball

If the above are independent utterances, the former sounds odd, whereas the latter is quite natural. This may be due to the fact that 'playing ball' cannot normally be represented as a state. If one changes the VP to *zip3 bo1* 'catch ball', then the sentence is acceptable. Similarly, a verb like *mong6* 'look' is fine with the aspect marker.

- (14) a. Siu2ming4 zip3 zyu6 bo1 'Siuming caught the ball'
 Siuming catch DUR ball
 b. Mong6 zyu6 ngo5. 'Look at me'
 look DUR me

Tables 3 and 4 give the statistics for the use of the three aspect markers by the two children before two years of age. If one uses the criterion that an aspect marker is considered acquired if it cooccurs with at least three different verbs in successive sessions, then clearly the perfective aspect marker *zo2* is acquired by 1;09 for Bernard, and by 1;11 for Tsuntsun. By this criterion, the ages at which Bernard and Tsuntsun show productive uses of the three aspect markers are shown as below. Our data indicate that perfective aspect *zo2* appeared first, followed by the durative aspect *zyu6*. The progressive aspect marker *gan2* was the last to appear in this early stage.

(15)	<i>zo2</i>	<i>zyu6</i>	<i>gan2</i>
	<i>PERF</i>	<i>DUR</i>	<i>PROG</i>
Bernard	1;09	2;03	--
Tsuntsun	1;11	2;01	2;06

A strong correlation can be observed between the distribution of verb aspect and verb semantics. Note that none of the aspect markers were attached to the copula or the verb *jau5* 'have'. That is, none were attached to functional verbs. In addition, for Bernard, none of the aspect markers were attached to adjectives. For Tsuntsun, there were two tokens of *zo2* used with inchoative meaning with the adjectives for 'wet' and 'dirty'.

The Cantonese data before 2;2 show that there is an overwhelming tendency for the perfective marker *zo2* to cooccur with achievement verbs or punctual verbs. This tendency

persists throughout the period, though in later stages, non-bounded events are also marked by *zo2*. The following are representative of the verbs marked by the perfective.

- (16) *laan6* 'break' *m4 gin3* 'disappear'
 dit3 'fall down' *lat1* 'fall off'
 syn2 'wound' *waai6* 'turn bad'
 tyun2 'split/break' *ting4* 'stop'
 mit1 'pinch/tear off' *mang1* 'pull off'

Some instances of other verbs which are not inherently bounded, e.g. *fan3 gaau3* 'sleep', *maat3* 'wipe', and *sai2* 'wash', but these constitute a small minority of the tokens.

In short, we have strong evidence that aspect as a syntactic category is established between 1;9 and 1;11. We have evidence that the earliest productive aspect markers are *zo2*, *zyu6*, and *gan2*, in that order. The children observed the distribution of aspect markers without fault. Children never used an aspect marker following a noun, a classifier or a sentence final particle.

**Table 3: Distribution of aspect markers following
adjectives, transitive and ergative verbs in Bernard (1;7-2;0)***

Age	Adj	Ergative verbs	Transitive verbs
1;07;00			
1;08;00			
1;08;14			
1;08;28			
1;09;04			
1;09;18		<i>zo2: laan6 zo2 (a3) [3]</i> 'break'	
1;09;25		<i>zo2: m4 gin3 zo2</i> 'disappear'	<i>zo2: gam6 zo2</i> 'press'
1;10;10		<i>zo2: laan6 zo2</i> 'break' <i>m4 gin3 zo2</i> 'disappear'	
1;10;23			
1;11;06			
2;00;03		<i>zo2: m4 gin3 zo2 go3 hau2</i> 'disappear'	<i>zyu6: ze2zel nei5baai2 zyu6</i> 'put'
2;00;16		<i>zo2: m4 gin3 zo2 [2]</i> 'disappear'	<i>zo2: sik6 zo2 faan6 na3</i> 'eat'

*identical utterances of the child within the same conversational episode are counted as one. The number in square brackets indicates the number of tokens of the utterance in different episodes with or without sentence final particles. Adj=adjective.

Table 4: Distribution of aspect markers following
adjectives, transitive and ergative verbs in Tsuntsun (1;5-2;0)*

Age	Adj	Ergative verbs	transitive verbs
1;05;22			
1;07;03			
1;07;10			
1;08;00			zyu6: za1 zyu6 [3] 'hold'
1;08;07			zyu6: za1 zyu6 'hold'
1;08;21			zyu6: za1 zyu6 'hold'
1;09;07			zyu6: za1 zyu6 (sfp) [3] 'hold'
1;09;14			zyu6: za1 zyu6 (sfp) [2] 'hold'
1;09;29			zyu6: za1 zyu6 (sfp) [3] 'hold'
1;10;30			zyu6: za1 zyu6 (a3) [3] 'hold' zo2: hoil zo2 'open'
1;11;13		zo2: laan6 zo2 a3 'break' dit3 zo2 a3 'fall' m4 gin3 zo2 'disappear'	zyu6: za1 zyu6 a1 [7] 'hold'
1;11;27		zo2: dit3 zo2 (sfp) [3] 'fall' dit3 zo2 bol 'fall/drop'	zyu6: za1 zyu6 (a3) [5] 'hold' zo2 zyu6 'block-the-path'
2;00;09		zo2: dit3 zo2 (sfp) [3] 'fall' baba m4 gin3 zo2 'disappear'	zyu6: za1 zyu6 a3 [3] 'hold'
2;00;16	zo2: lyun1 zo2 a3 'become-curly'	zo2: dit3 zo2 'fall' laan6 zo2 'break' waai6 zo2 'turn-bad/break'	zyu6: za1 zyu6 (a3) [10] 'hold' ze2zel za1 zyu6 a3 'hold'

*identical utterances of the child within the same conversational episode are counted as one. The number in square brackets indicates the number of tokens of the utterance in different episodes with or without sentence final particles. Adj=adjective.

3. Modal auxiliaries

Cantonese has a number of preverbal modal auxiliaries: *ho2yi3* 'can'; *wui3* 'will'; *hang2* 'willing'; *nang4gau3* 'able-to', and *m4 sai2* 'no need=don't'. Preverbal modal auxiliaries are like verbs in that they can occur alone as an independent utterance; they can be negated,

and can participate in A not A structures. They differ from most verbs in taking VP complements (i.e. clausal complements with a null subject). In addition, there is a postverbal modal auxiliary *dak1* 'can'. This modal has only deontic meanings, indicating either ability or permission. It can be suffixed to the verb in any of the following environments:

(17) V NP; V-resultative; V-dir complement; V NP NP

e.g. tiu4 si4 lik1 dak1 ceot1lei4

CL key take AUX out

"One is allowed to take the key out/ one is able to take the key out"

For our purpose, the postverbal aux is important because it patterns in different ways than the main verb. So its occurrence is a clear signal that the child is becoming sensitive to a different category than aux.

Table 5: Age of first occurrence of preverbal and postverbal auxiliaries in Cantonese children*

	preverbal aux		postverbal aux <i>dak1</i>	
child	affirmative clause	negative clause	affirmative clause	negative clause
Tsuntsun	-----	2;03	1;11	1;11
Bernard	2;04	2;00	1;11	2;00

*only occurrences in sentences containing both the auxiliary and the main verb are considered.

Table 6: Auxiliary verbs in the early speech of Cantonese children*

child	preverbal aux	postverbal aux <i>dak1</i>
Tsuntsun	<i>wui3, ho2ji3, (m3) sai2</i>	<i>dak1</i>
Bernard	<i>wui3, ho2ji3, (m3) sai2</i>	<i>dak1</i>

*only occurrences in sentences containing both auxiliary and main verb are considered.

Tables 5 and 6 show beyond doubt that at around 1;11, modal auxiliaries begin to develop in children. Their onset is slightly later than that of the functional categories of classifier and aspect. Secondly, the data from both children suggest that the postverbal aux occurs at the same time as or earlier than the preverbal modal auxiliaries. In the case of Tsuntsun, the postverbal aux *dak1* appeared in 1;11, whereas the first preverbal modal did not occur until 2;03. As for the affirmative clauses of Bernard, it is also the case that the postverbal *dak1* occurred at 1;11, and the preverbal aux did not emerge until several months later. It remains for us to explain why modal aux is acquired somewhat later than the functional categories of classifier and aspect.

4. Sentence final particles

The category of sentence final particles deserves special attention partly because Cantonese is particularly rich in this category having dozens of sfps differing in the type of modality they exhibit, and the syntactic constraints they are subject to (see Kwok 1984 for a basic description). The early emergence of this category also makes the speech of the Cantonese-speaking child visibly different from that of their English counterparts.

The main fact relevant to our discussion of the categorial status of sentence final particles is the distinction between non-interrogative particles and interrogative particles. An example of a non-interrogative particle is the neutral particle *a3*, whose function is to make the utterance sound not so blunt, as illustrated in (18). This particle can be attached to the end of an affirmative utterance without changing the latter's declarative nature. It can also be attached to the end of a question without changing the latter's interrogative status, as in (19-20).

- (18) ngo5 wui3 lei4 a3
I will come sfp
"I will come"

- (19) nei5 lei4 m4 lei4 a3?
you come not come sfp
"Are you coming?"

- (20) bin1go3 lei4 a3?
who come sfp
"Who is coming?"

Interrogative particles differ from non-interrogative particles in their being able to contribute turn a statement into a question. Thus the sentence *keoi5 lei4* 'he come' can be understood as 'he is coming'. However, when the interrogative particle *a4* is added to it, the utterance becomes an yes-no question, as in (21). When the wh-interrogative particle *le1* is added to a NP, the utterance signals a where-question (cf. 22).

- (21) keoi5 lei4 a4?
s/he come q-sfp
"Is it the case s/he is coming?"

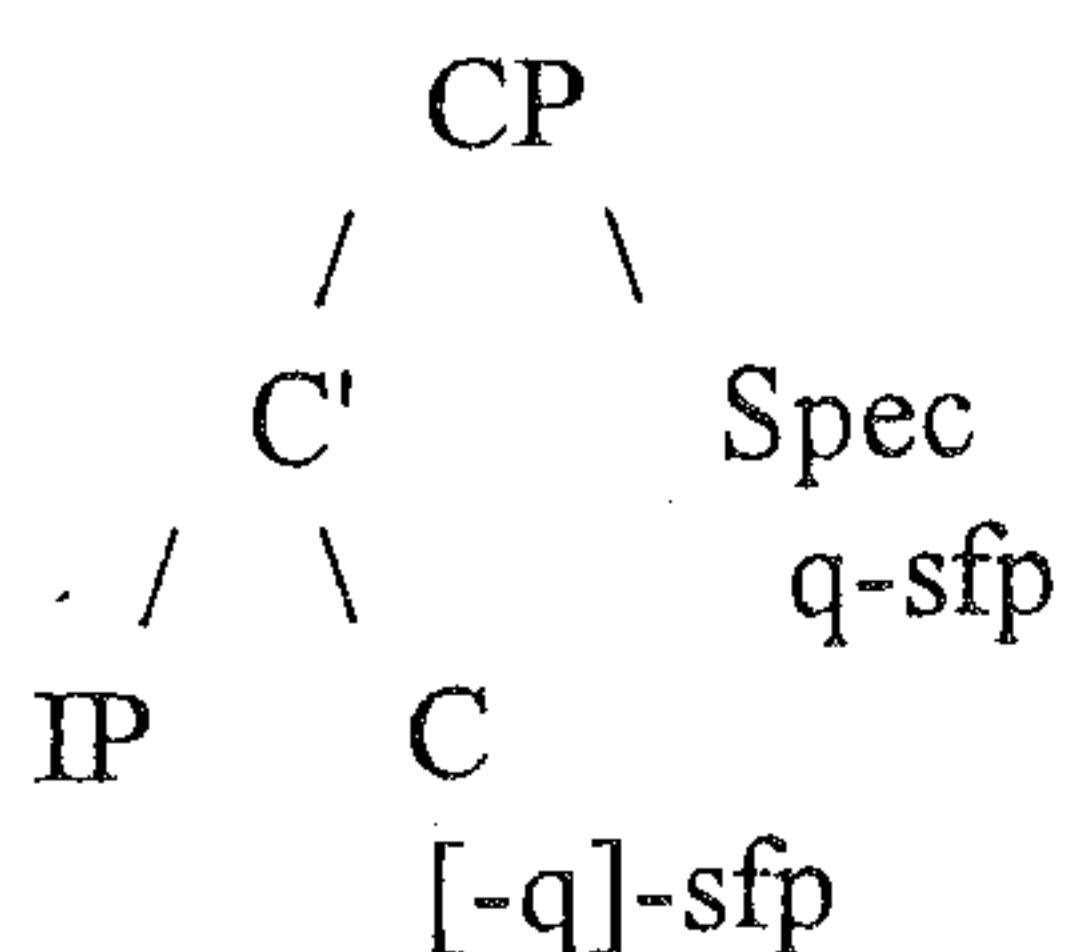
- (22) ba4ba1 le1?
dad q-sfp
"Where is dad?"

Some sentence final particles signify focus meanings, such as 'only', and they can occur in concatenation with other sfps in a restricted order. Thus, (23) shows the focus particle *za3* occurring before the interrogative particle *a4*. As is clear from the English gloss, the question particle has the focus within its scope.

- (23) keoi5 tai2 zo2 saam3 bun2 syu1 za3 a4?
 s/he read asp three CL book sfp (=only) sfp
 "Did he only read three books?"

We assume the structure of (24) in our representation of sentence final particles, following Law and Neidle (1991) and Cheng (1991). Non-interrogative sfps would occupy the head position of the complementizer phrase, and interrogative sfps the specifier position of the CP, which occurs after the head. This tree can predict the ordering of question sfps after the noninterrogative sfps. It also has the advantage of ruling out sentences with two interrogative sfps, as in (25c), which contains the interrogative sfp *a4*, as well as the interrogative particle *me1*.

(24)



- (25) a. keoi5 lei4 a4
 s/he come q-prt
 'Is s/he coming?'
 b. keoi5 lei4 me1
 s/he come q-prt
 'Isn't she coming?'
 c. *keoi5 lei4 a4me1?
 s/he come q-prt q-prt

Tables 7 and 8 reflect the use of sfps in the speech of Bernard and Tsuntsun before two. A number of observations can be made about the sfps of Bernard. First of all, the particles always appear in utterance final position. Secondly, generally only single sfps are used, except perhaps for the use of the particle combination *lei4ga3*. Thirdly, sfps are clearly evidenced at 1;09;18, if we assume the cooccurrence of 3 distinct sfps in the same session as a criterion of mastery. For Bernard at least, sfps initially follow NPs mostly, but this nominal bias disappears by 1;9, with sfps by then attaching to NP, VP, S productively.

Table 7: Distribution of sentence final particles in Bernard (1;7-1;11)*

Age	Number of sfp	sfp following VPs	sfp following APs	sfp following NPs	sfp following S
1;07;00					
1;08;00	wo3(2); a3(2) =4	a3(1); wo3(1)		a3(1); wo3(1)	
1;08;14	a3(12) =12	a3(1)	a3(1)	a3(10)	
1;08;28	a3(12) =12	a3(1)	a3(1)	a3(10)	
1;09;04	a3(11); a1(2) =13	a3(1)		a3(10); a1(2)	
1;09;18	a3(47); ga3(3); a1(2); a4(1); bo3(1); lei4(1); laak3(1); no4(1); wo4(1) =57	a3(30); ga3(3); a1(2); no4(1); a4(1); wo4(1)	a3(1); laak1(1)	a3(15); bo3(1)	a3(2)
1;09;25	a3(13); o3(1) =15	a3(3); o3(1)	a3(1)	a3(10)	
1;10;10	a3(19); la3(1); a1(1) =21	a3(6); la3(1)		a3(11); a1(1)	a3(2)
1;10;23	a3(11); ga3(2); o3(1); lei4ga3(1); bo3(1); nok1(1); naak3(1) =18	a3(6); ga3(2); lei4ga3(1); nok1(1)	a3(1)	a3(4); o3(1)	naak3(1); bo3(1)
1;11;06	a3(17); ga3(7); la3(4); lu3(3); la1(3); lo1(3); lok1(2); lei4ga3(1); a1(1); wo3(1); lo3(1); na3(1) =44	a3(9); ga3(3); la3(1); lu3(1); na3(1)	a3(1); la3(1)	a3(3); lo1(3); ga3(2); lok1(2); la3(1); lu3(1); a1(1); la1(1)	a3(4); ga3(2); la1(2); la3(1); lu3(1); lei4ga3(1); lo3(1); wo3(1)

*identical utterances of the child within the same conversational episode are counted as one.

Table 8: Distribution of sentence final particles in Tsuntsun (1;5-1;10)*

Age	Number of sfp	sfps following VPs	sfp following APs	sfp following NPs	sfp following S
1;05;22	a3(6); a4(1) =7	a3(2)	a3(1)	a3(3); a4(1)	
1;07;03	a3(12); a4(1) =13	a3(5); a4(1)	a3(3)	a3(3)	a3(1)
1;07;10	a3(12) =12	a3(8)	a3(1)	a3(3)	
1;08;00	a3(18); a4(6); la3(1) =25	a3(9); a4(5); la3(1)	a3(5)	a3(4); a4(1)	
1;08;07	a3(29); a1(3); a4(2); le1(1) =35	a3(15); a1(1); a4(2)	a3(6); a1(2)	a3(8); le1(1)[Q]	
1;08;21	a3(25); a4(1); no3(2); wo4(1); lei4(1) =30	a3(12); a4(1); no3(1); wo4(1)	a3(2)	a3(10); no3(1); lei4(1)[Q]	a3(1)
1;09;07	a3(47); no3(1); lei4a3(2) =50	a3(29)	a3(3); no3(1)	a3(15); lei4a3(2)[Q]	
1;09;14	a3(27); lei4(1) =28	a3(17)	a3(3)	a3(7); lei4(1)[Q]	
1;09;29	a3(121); a4(5); lei4(1); lo1(1); a3ma3(1) =129	a3(70)	a3(15); a4(2); lo1(1)	a3(36); lei4(1)[Q] a4(1)[Q]; a4(2); a3ma3(1)	
1;10;30	a3(77); lei4a3(1) =78	a3(48); lei4a3(1)	a3(2)	a3(25); a3(1)[Q]	a3(1)

*identical utterances of the child within the same conversational episode are counted as one.

Turning to the data from Tsuntsun in Table 8, we observe that his sfps occur predominantly in final position. As in Bernard's speech, generally only single sfps are used, with the exception of *lei4+sfp*, and the concatenation *a3ma3*. Again, if we assume the same criterion for mastery as that adopted for Bernard, it is clear that sfps appear at 1;08;00. The sfp's of Tsuntsun attach to all categories, but mostly to VPs, then NPs, then APs.

In order to probe further the productivity of sfps as a category, we examine the utterances of Bernard containing sfps before the age of two, in comparison to the preceding utterances of the adults. In Table 9, the preceding utterances of the adults are classified according to whether they are statements or questions. Further, the particular sfp used by the adult in the preceding utterance, if any, and the utterance used by the child are indicated with a slash notation. The number of tokens of each type of pairing is indicated in parenthesis. The underlinings show pairings that involve both the adult and the child using the same sfp.

Table 9: Distribution of sentence final particles in Bernard (1;7-1;11) according to type of preceding utterance produced by adults*

Age	preceding utterance is declarative	preceding utterance is A-not-A question	preceding utterance is YNQ	preceding utterance is wh-question	no preceding adult utterance
1;07;00					
1;08;00		<u>a3/a3</u> (1)		wa2/a3(1)	wo3(2)
1;08;14	<u>a3/a3</u> (1); wo3/a3(1)			<u>a3/a3</u> (1) ha2/a3(1) lei4ga3/a3(1)	a3(7)
1;08;28		<u>a3/a3</u> (1)		<u>a3/a3</u> (1); lei4ga3/a3(3); ne1/a3(1)	a3(6)
1;09;04	<u>/a3</u> (3); wo3/a3(2); a1/a3(1); laak3/a3(1)			<u>a3/a3</u> (1) lei4ga3/a3(1)	a1(2) a3(2)

1;09;18	<u>/a3(6);</u> wo3/a3(4); a1/a3(4); la3/a3(2); laak3/a3(2); laak3/ga3(1); <u>/wo4(1);</u> <u>/a1(1);</u> la1/a3(1); <u>laak3/laak3(1);</u> aak3/a3(1); ga4/ga3(1); lei4ga3/lei4(1); wo4/bo3(1); lo1/a3(1)	<u>a3/a3(6);</u> ga3/a3(2)		<u>a3/a3(6);</u> lei4ga3/a3(2); <u>/ga3(2);</u> le1/a3(1); a3/a4(1)[Q]; a3/a1(1); ge2/a3(1)	a3(8); no4(1)
1;09;25	<u>/a3(3);</u> wo3/a3(2); lei4ga3/a3(1)	a3/o3(1); ga3/a3(1); e6/a3(1)		<u>a3/a3(2);</u> wa2/a3(2); lei4ga3/a3(1)	a3(1)
1;10;10	a1/a3(2); <u>/a3(2);</u> <u>a3/a3(1);</u> wo3/a3(1); lu3/a3(1)	<u>a3/a3(3);</u> a3/la3(1); <u>/a3(1);</u>	a4/a3(1)	<u>a3/a3(3);</u> le1/a3(1); wa2/a3(1); a3/a1(1); <u>/a3(1)</u>	a3(1)
1;10;23	<u>/a3(1);</u> wo3/bo3(1)	<u>a3/a3(5);</u> <u>/a3(1)</u> ga3/ga3(1); ga3/lei4ga3(1); a3/ga3(1); a3/nok3(1);	a4/a3(1)	a3/o3(1); ha2/naak3(1); ga3/a3(1)	a3(3)
1;11;06	<u>/ga3(2);</u> ga3/ga3(1); <u>a1/a1(1);</u> a4/la3(1); <u>/la3(1);</u> la1/a3(1); e3/a3(1); a3/la3(1); <u>/la1(1);</u> <u>wo3/wo3(1);</u> <u>/lo3(1);</u> <u>/a3(1)</u>	<u>a3/a3(2);</u> ga3/a3(1); a3/la3(1); a3/na3(1)	ga4/ga3(1); la4/a3(1)	a3/lo1(3); a3/lok1(2); <u>a3/a3(1);</u> ga3/ga3(1); a3/ga3(1); a3/lu3(1); a3/lei4ga3(1)[Q]; ha2/a3(1)	la3(8); lu3(2); ga3(1); a1(1); la1(1)

*identical utterances of the child within the same conversational episode are counted as one. Following each number is the sfp used by adults and the sfp used by the child in the same adjacency pair, separated by a slash. '_' means there is no sfp used in the preceding adult utterance. If only one sfp is listed, it is the one produced by the child.

The data in Table 9 show clearly that the use of sfps by Cantonese-speaking children is clearly not a process of imitation, because of the extremely low percentage of identity between adult and child sfps (ie the ones indicated by underlining). This is especially so if the preceding adult utterance is declarative. There is a greater degree of sfp match in A not A and wh questions. The data also show that there is some degree of adult-child match in terms of tone register, with the sfps occurring mostly in the upper tonal register (ie tones marked by 1, 2, and 3).

For our purpose, the data show that by 1;9, children reanalyze sfp as a category that is not equal to a question particle; since some cooccur with wh-questions and some with A-not-A questions, and still others occur in declaratives. The first sfps acquired are non-interrogative sfps. In the speech of Bernard before two years old, only two occurrences of a sfp are found in a question (cf. 26-27)

(26) Bernard (1;09;18)

INV: si1zi2 hai2dou6 zou6 mat1je2 a3?
 lion at-here do what sfp
 'What's the lion doing here?'

CHI: fan3gaau3 a4?
 sleep sfp
 'Is he sleeping?'

(27) Bernard (1;11;06) (line 1015):

INV: jau5 mat1je2 sik6 a3, hou6zeon3?
 have what eat sfp, Houzeon
 'What's there to eat, Houzeon?'

CHI: ni1di1 mel lei4ga3?
 these what sfp
 'what are these?'

The fact that children rule out interrogative meanings for sfps cooccurring with a question particle can be taken as evidence that they are sensitive to a structure of the complementizer phrase as that indicated in (24).

Conclusions

Our review of the early utterances of the two youngest observed children in our corpus indicate that the functional categories of classifier, aspect, and modal auxiliaries occur somewhere between 1;09 and 1;11 for Cantonese-speaking children. There is also indirect evidence from the preponderance of noninterrogative sentence final particles in the children that children have a grasp of the incompatibility of two question particles. This lends indirect support to the early sensitivity to the structure of the complementizer phrase. The review taken as a whole provides strong support for the strong continuity view, contrary to the claims of Radford (1990).

We should mention, however, that not all categories that are considered to be functional categories develop at the same pace. A fifth category we have examined, namely the category of genitive marking *ge3*, shows a very low frequency of use. Only three occurrences were found in the entire corpus of Bernard:

- (28) *ze3ze1 ge3* (sister's) [1;10;23]
 ngo5 ge3 doi2doi2 (my bag) [2;05;19]
 nei5 ge3 syut3 tiu2 (you lollipop) [2;05;19]

Only several instances of the genitive marker *ge3* were found in the entire corpus of Tsuntsun, all of the form:

- (29) *ngo5 ge3* (mine)

This kind of pattern corroborates the observation of Deprez and Pierce (1993, 1994) about the uneven development of the inflection and complementizer categories in the early speech of German children.

References

- Cheng, Lisa (1991) On the Typology of Wh Questions. Doctoral Dissertation. MIT.
- Cheung, Hung-nin Samuel (1972) [in Chinese] Cantonese as Spoken in Hong Kong. Hong Kong: The Chinese University of Hong Kong Press.
- Deprez, V., and A. Pierce (1993) "Negation and functional projections in early grammar", *Linguistic Inquiry* 24.1: 25-67.
- Deprez, V., and A. Pierce (1994) "Crosslinguistic evidence for functional projections in early child grammar", in T. Hoekstra and B. Schwartz (eds) *Language Acquisition Studies in Generative Grammar*. Amsterdam: John Benjamins.
- Erbaugh, M. (1986) "Taking stock: the development of Chinese noun classifiers historically and in young children", in C. Craig (ed) *Noun Classes and Categorization*. Amsterdam: John Benjamins.
- Hyams, N. (1992) "The genesis of clausal structure", in J. Meisel (ed) *The Acquisition of Verb Placement*. Dordrecht: Kluwer.
- Kwok, H. (1984) *Sentence Particles in Cantonese*. Center for Asian Studies. University of Hong Kong.
- Law, Sam Po and C. Neidle (1991) "A syntactic analysis of Cantonese sentence-final particles", ms. Johns Hopkins University and Boston University.
- Radford, A. (1990) *Syntactic Theory and the Acquisition of English Syntax*. Oxford: Blackwell.
- Whitman, J., K. Lee, and B. Lust (1991) "Continuity of the Principles of Universal Grammar in First Language Acquisition: the issue of functional categories", *NELS* 21

Part B: Cross-sectional Experimental Studies

The Acquisition of (In)definiteness in Cantonese

Thomas Hun-tak Lee and Kitty Szeto

The acquisition of the semantic distinction between definiteness and indefiniteness presents an interesting puzzle, because the literature contains sharply conflicting findings on when the distinction is acquired. The early studies of Brown (1973) and Maratsos (1976) argue that the distinction is mastered by age 4. Other studies (notably Warden 1976, 1981), Karmiloff-Smith 1979, 1981; Hickman 1980) report that even 7-year-olds experience difficulty with this distinction, and argue that it is not fully acquired until after 7 years old. Some people see this late onset as a genuine linguistic deficiency for the young child, while others believe that the child has actually acquired the rule, but it is a pragmatic factor that intervenes in the child's performance.

In this paper, we try to suggest a way of looking at the acquisition data in terms of the formulations of file change semantics of Heim (1982, 1983) and its development by Avrutin (1994). We'll be claiming that there are a number of generalizations that can be made about the data, and that they can be made sense of in terms of the rules for representing definite and indefinite NPs in file change semantics. In the second part of this paper, we'll report on an experiment on the acquisition of (in)definiteness by Cantonese-speaking children, Cantonese being a language that uses word order rather than the determiner system to signal definiteness.

1. File change semantics and (in)definiteness

The semantic facts that concern us are illustrated in the sentences below:

- (1) A boy came and sat down on a chair.
- (2) The chair suddenly collapsed.
- (2') A chair suddenly collapsed.
- (1') The mailman came and sat down on the chair.

Generally when one introduces a referent into discourse, one does not presuppose that the hearer can locate the referent in the universe of discourse, and the indefinite form is used, hence the use of 'a boy' and 'a chair' in (1). After the referent has been established, one could assume that the hearer can locate it in the restricted domain of discourse, and one typically uses a definite description to refer to it, as 'the chair' in (2). In fact, if an indefinite NP, it will not be interpreted as referring to the established referent, but rather to a different referent. This novelty may arise because of scalar implicature. Whether one can assume that the hearer has adequate referential knowledge to justify the use of a definite description depends on the set of objects shared by the speaker and hearer. It is perfectly acceptable to use the definite NP to introduce a referent into discourse if one can assume the hearer can uniquely identify the referents. (1') would be acceptable if both speaker and hearer know that it is the same mailman who comes by every day, and there is just one chair in the place where the mailman sat. (1') would be infelicitous if there is more than one mailman who comes by and there is more than one chair. This corresponds to what Hawkins (1978) calls the 'larger situation use' and the 'immediate situation use'.

These semantic facts are captured in file change semantics (Heim 1982, 1983, Avrutin 1994) in the following way. Let us outline very briefly the assumptions of the framework and how they apply to the sentences in (1-2). In communication, we introduce referents into discourse and we attribute properties to these referents. One way of conceptualizing this process is to think of participants keeping in their mental store a file of cards which register the discourse referents and their properties. As the discourse unfolds, and referents are introduced and talked about, the file changes. New files may be created and old files may be updated. The way they are constructed and modified are given by the following rules. In Heim's theory, only non-quantificational NPs are represented by file cards.

(3) Rules for interpreting indefinite and definite NPs (hearer-based)

- (a) for every indefinite NP, start a new card.
- (b) for every definite NP, update an old card.
- (c) if a definite NP can be accommodated by bridging it to an old card, then start a new card.

Applying these principles to the basic data, we can open a file with two new cards after hearing (1), in accordance with (3a). Each card will have a number on it. eg..

F1:

1	2
boy (1)	chair (2)
came (1)	sat-on-by-1 (2)
sat-on-2 (1)	

After hearing (2), we update the existing card 2 in accordance with (3b), since there is only one card that fits that description. The file will be like the following after updating.

F2:

1	2
boy (1)	chair (2)
came (1)	sat-on-by-1 (2)
sat-on-2 (1)	collapsed (2)

Now suppose we hear (2'). According to (3a) we will need to start a new card. (3a) is a way of capturing the novelty condition.

F2':

1	2	3
boy (1)	chair (2)	chair (3)
came (1)	sat-on-by-1 (2)	collapsed (3)
sat-on-2 (1)		

Since communication is always carried out against a background of assumptions shared by speaker and hearer, these presuppositions must be represented in the file system. It is permissible, therefore, for cards to crop up because of perception of salient objects in the context of utterance. One can represent this knowledge of referents which are not introduced by utterances but prompted by the nonlinguistic context by means of cards. In addition, cards are needed to represent the knowledge of the world shared by speaker and hearer. As suggested by Avrutin (1994) there may be different cards that represent real world knowledge. Because of these file cards, it is possible to start a new file card when hearing a definite NP, as given in Rule (3c).

For example, if instead of hearing (1) we hear (1'), these noun phrases do not match any of the existing file cards established by discourse, since 'the mailman' and 'the chair' are mentioned for the first time here. So the utterance is infelicitous. But hearers may accomodate for the missing presuppositions by supplying them, a process called accomodation by Lewis (1979). Accomodation is possible because of the presence of 'communtiy cards' and 'visual situation cards'. The new file cards can be crosslinked to the cards that represent deictic information or real world knowledge. The file for (1') may be the following:

F1':

0.1	0.2
community	visual situation
1	2
mailman of 0.1 (1)	chair of 0.2 (2)
came (1)	sat-on-by-1 (2)
sat-on-2 (1)	

Heim's and Avrutin's formulations look at indefinite and definite descriptions from the hearer's perspective. One could extend their formulations to include the speaker's perspective. I suggest that something like (4) would be necessary.

(4) Rules for the use of (in)definite/definite NPs (speaker-based)

- (a) to start a new file card, use an indefinite NP
- (b) to update an existing file card, use a definite NP
- (c) to start a new file card, use a definite NP only if it can be crosslinked to an old file card (which is assumed to be also an old card of the hearer's)

(4a) captures the intuition that to introduce a referent into discourse, the referent must be a new discourse representation and correspond to a new file card. Clearly for referent introduction, an indefinite NP should be used. If one wants to say something about a referent that has been established in discourse, that entity must have an existing file card corresponding to it. Therefore this is equivalent to updating an existing file card, and in such a case, a definite NP should be used. In cases where a new referent is

introduced, and a new card needs to be started, if one is sure that the new card can be linked to real world knowledge cards, and one can safely assume that the hearer can do so. Applying the principles in (4) to the sentences (1,2, 2' and 1') will give the right results.

2. Generalizations based on the acquisition literature

Having stated the rules for the interpretation and use of indefinite and definite NPs, we turn to the acquisition facts, which are summarized in Table 1 and Table 2. To evaluate the validity of the findings, we will need to look at the methodology and the results in some detail. Let us observe some factors that have been considered by researchers in evaluating the findings.

Consider some of the data suggesting early acquisition. In Brown (1973) (cf. results column) it is reported that children reach criterion in their use of definite and indefinite articles by 3;5. But if you examine the data listed, it is not entirely clear that the evidence is strong, since a lot of the indefinite NPs are predicate nominals and not referring expressions. Thus the fact that the child uses "That a jeep. I put some in the jeep." (Adam) doesn't mean that there is an anaphoric link between 'a jeep' and 'the jeep'. This utterance could arise without any grasp of anaphora if 'a jeep' is interpreted predicatively and 'the jeep' is used deictically. We know that the predicative nominal use of indefinites is acquired extremely early (cf. Warden 1976, MacNamara 1982). One could also be sceptical of some of data from Maratsos (1976). For example, one well-known method used by Maratsos is elicited production based on narratives, shown in (3) in Table 1. In one paradigm, referents are introduced into discourse, and children are asked to refer to the established referents. The idea is that if the established referent is unique, the child should continue with the definite article; if the established referent is one of a plural set of identical objects, then the child should continue with the indefinite article. A sample story is like the following: "Someone with a wooden box goes to a pond to get an animal. He sees {two animals, a frog and a turtle/ two bunches of animals, frogs and turtles}. So he puts one of them into his box. What did he put in?". Maratsos reports high accuracy rates for the 3-and 4-year-olds (68% and 98% respectively). But the finding may not be a reliable indicator because children are not required to respond only with noun phrases. If the child uses 'a N' there is no guarantee that this is a referring expression; it could well be a predicative nominal. Similarly, 'the N' could be used deictically. Further this task is a simple task in the sense that the referent is established for the child, who only has to handle subsequent reference. Some of the other experiments that report late acquisition typically require the child to both introduce the referent and refer to it subsequently.

Consider also some of the findings that suggest late acquisition. A good example is Warden (1976) (cf. 1 in Table 2). In this study, a picture description task was used, and children were tested in pairs, with one member of the pair telling the story to the other, who could not see the picture. As can be seen from the results, the performance was poor, with the 7 year-olds and under reaching only around 60% accuracy in referent introduction. Now, one could challenge this finding by saying that although the other child cannot see the picture, the pictures are none the less present in the context of the utterance for both speaker and listener, and this may induce the high percentage of definite noun phrases for referent introduction. Some researchers have

Table 1. Studies suggesting the early acquisition of (in)definiteness

Study	Subjects & method	Results
1. Brown (1973) Lg: English	1;6-4;0 ; naturalistic observation	indefinite and definite articles acquired by 3;5 in terms of accuracy in obligatory contexts. 'That a jeep. I put some in the jeep.' (Adam) 'This was a big rabbit. And scared the rabbit' (Sarsh) 'It's a gun.' (Eve) Errors: 'Let me see you ride a bike' (Adam, of unique object in setting) 'We saw them in a zoo.' (Sarah) 'The cat's dead.' (Mother: what cat?)
2. Maratsos (1976) Lg: English	3, 4 year-olds; comprehension act out. Experimenter and child act out lead portion of story. Child acts out critical sentence. e.g. "Now this boy came and sat down in one of the chairs. And just as he sat down, suddenly {a/ the} chair fell over." 'the N' --> same object; 'a N' --> different object.	3- and 4-yr-olds interpreted 'a N' to refer to a different object 78% and 75% of the time. They interpreted 'the N' to refer to the same object 94% of the time.
3. Maratsos (1976, 1974)	3, 4 year-olds; elicited production based on stories; only subsequent reference tested. Paradigm A: aX--> the N. two unique referents are introduced in the story, followed by a question. Paradigm B: Xs--> a N. two plural sets are introduced in the story, followed by a question. "Someone with a wooden box goes to a pond to get an animal. He sees {two animals, a frog and a turtle/ two bunches of animals, frogs and turtles}. So he puts one of them into his box. What did he put in?	The 4 year-old (high) group performed accurately 98% of the time; the 4 year-old (low) group and the 3 year-old group responded accurately 68% of the time.

4. Emslie & Stevenson (1981) Lg: English	2, 3, 4 year-olds, and adults; picture description using cartoon stories consisting of 3 pictures each. Children tested in pairs with one member of the pair telling the story to the other, who cannot see the pictures. e.g. #1: a girl sits in a chair holding a teddy bear. A dog watches. #2: the dog runs off with the teddy bear in its mouth; the girl looks startled. #3: the girl chases after the dog, which has dropped the teddy bear.	in introducing referents, 3-4 year-olds used indefinite NPs 68-85% of the time; for subsequent reference, 3-4 year-olds used definite NPs 96-100% of the time. Findings also show that an entity which is introduced into the picture relevant to the story, but not yet introduced into the discourse would induce the use of definite NPs.
5. Power & Dal Martello (1986) Lg: Italian	3, 4, 5 year-olds; picture description task as in Emslie & Stevenson; children tested in pairs.	in introducing referents, 3-4 year-olds used indefinite NPs 60% of the time; 5 year-olds used indefinite NPs 82% of the time. For subsequent reference, children used definite NPs 85-90% of the time. Findings also show that in introducing referents, children would use indefinite NPs less when they tell the story again to a second listener, ie when they are already familiar with the referent being introduced.

Table 2. Studies suggesting the late acquisition of (in)definiteness

Study	Subjects & method	Results
1. Warden (1976) Lg: English	3, 5, 7, 9 year-olds, and adults; picture description based on stories each consisting of three pictures. Children tested in pairs, with one member of the pair telling the story to the other, who cannot see the picture. e.g. #1: a dog is chasing a hen. #2: a cow stops the dog, and the hen is hiding behind the door; #3: the hen has laid an egg.	For referent introduction, 3, 5, 7, and 9-year-olds used indefinite NPs 46%, 62%, 61% and 82% of the time; for subsequent reference, the same children used definite NPs 90-100% of the time. 70% of the 3-year-olds' and 45% of the 5-year-olds' indefinite NPs are predicate nominals.
2. Warden (1981) Lg: English	5, 6, 7, 8 year-olds and adults; children asked to describe a short video film in 4 conditions, depending on whether the narration begins during or after the viewing, and on whether the listener is co-present or on the phone line.	for referent introduction, 5, 6, 7, 8 year-olds used indefinite NPs 57%, 61%, 62% , 60% of the time. All children used the definite NP for subsequent reference. Children did not use more indefinite NPs when the referents or the interlocutor were absent.
3. Karmiloff-Smith (1979) Lg: French	(a) Experiment #6: 8 age groups studied (3-10 year-olds); replication of Maratsos (1976)'s elicited production task based on stories.	For the 'aX--> the N' condition, 3 and 4 year-olds used the definite NP 40% and 63% of the time; children 5 years or older used definite NPs 80% or more of the time. For the 'Xs--> a N' condition, 3 to 7 year-olds used indefinite NPs only 44% or less of the time; 8, 9, and 10 year-olds used them 64%, 86% and 89% of the time.
	(b) Experiment #15: 4 to 10 year-olds tested; replication of Maratsos (1976)'s comprehension act-out task.	except for the 7 year-olds, children interpreted an indefinite NP as referring to a different object 71-95% of the time. However 4-8 year-olds picked the same object when hearing a definite NP for 55% or less of the time.

	(c) Experiment #1: elicited production based on a verbal instruction related to a collection of objects. e.g. a girl, a singleton set containing X, and a plural set containing three identical Ys. Experimenter says to child: ask the girl to lend you that (touching X or a Y).	For singleton sets, children 4 years or older used definite NPs on 84% or more of the trials. For plural sets, 3-4 year-olds used indefinite NPs 14% or less of the time; 5 to 7 year-olds used indefinite NPs 57% or less of the time; 8 & 9 year-olds used them 82% and 100% of the time.
	(d) Experiment#5: elicited production; children asked to describe sequences of actions enacted by experimenter. In the singleton condition, there is a boy, a girl and three different objects. e.g. boy jumps into X; girl jumps into X. In the plural condition, there is a boy a girl and a set of three identical objects. boy acts on X; girl acts on X.	In describing the first action, children used definite NPs for the singleton condition on 80% or more of the trials; they used indefinite NPs for the plural condition on 55% or less of the trials. In describing the second action on the same object in the singleton condition, 4 to 9 year-olds used definite NPs 79% to 92% of the time. In describing the second action on the same object in the plural condition, 4 to 6 year-olds used definite NPs 37% or less of the time; 7, 8, 9, and 10 year-olds used them on 52-72% of the trials.
4.Karmiloff -Smith (1981) Lg: English & French	4 to 9 year-olds; picture description task based on sets of pictures each consisting of 6 drawings. e.g. #1: a little boy is walking along a road; #2: the boy sees a balloon vendor; #3: the balloon vendor gives the boy a green balloon; #4: the boy walks off alone with the balloon; #5: the balloon flies off into the sky; #6: the boy starts to cry.	children under 6 years had great difficulty introducing referents and distinguishing referents in narratives. a typical narrative of an under 6 year-old: "He (B)'s walking along, and he (B) sees a balloon man, and he (M) gives him a green one, and he (B) walks off home.. and it (GB) flies away into the sky. So he (B) cries." a typical narrative of a more advanced child: "A litle boy is walking along on a sunny day. He (B) sees a balloon man. The balloon man gives him a balloon. He (B) walks off home. The balloon suddenly flies off into the sky, and he (B) starts to cry.

5.Hickmann (1980) Lg: English	7, 10 year-olds and adults; children asked to retell a story based on film strips to someone who has not seen them. e.g. Film: "a dog and a frog appear on the screen. The frog invites the dog to his birthday party. The dog refuses to go, because he has bought a candy bar and a flower for him but they are ruined and he is afraid that the frog will not be his friend any more. The dog reassures him and explains they are still friends, and they go off to the birthday party together.	7 and 10 year-olds used appropriate forms for referent introductions 55% and 85% of the time. Appropriate forms include indefinite NPs, demonstrative existential clause (e.g. 'there was this dog..') and definite NP with sufficient description (eg. 'the candy bar and the flower he was gonna give to the frog for his birthday'). Inappropriate forms include definite articles and third person pronouns.
6.Hickmann & Liang (1990) Lg: Chinese	4, 5, 6, 7, 10 year-olds and adults. Picture description task.	For referent introductions, 4 year-olds used mostly the preverbal position; 5 and 6 year-olds used preverbal and postverbal positions equally frequently; 7 and 10 year-olds preferred the postverbal position. For subsequent reference, children of all ages showed a preference for preverbal position.
7. Kail & Hickmann (1992) Lg: French	6, 9, and 11 year-olds; elicited narratives based on a book of 24 pictures depicting 3 main characters. Children asked to narrate the story to an adult who looked at the pictures with the child and also to a blindfolded adult.	In introducing referents, 6, 9, and 11 year-olds used indefinite NPs 48%, 31% and 68% of the time in the seeing condition, and 64%, 93% and 89% of the time in the blindfolded condition.
8. Bamberg (1986) Lg: German	3, 5, 9 year-olds and adults; elicited narratives based on a book of 24 pictures.	Adults used indefinite NPs and definite NPs to introduce referents 47% and 45% of the time; children used indefinite NPs and definite NPs to introduce referents 9% and 75% of the time respectively.

tried to avoid this problem by showing short films and have the child tell the story after viewing, or by having the children in separate rooms connected by telephone, or by introducing entities that cannot be seen in the film but are only talked about by the main protagonists in the story (e.g. Hickman 1980, Warden 1981).

2.1 Comprehension task should be easier than production task

Despite all these caveats, one can arrive at a number of generalizations based on the literature, and these can be made sense of in terms of the file change rules that we introduced earlier. If one goes back to the rules in (3), the first observation I'd like to make is that comprehension tasks should be easier than production tasks, the reason being that when interpreting indefinite and definite NPs as a listener, one could assume that felicity conditions are guaranteed by the speaker. If there is something that sounds infelicitous, the hearer can accommodate for missing presuppositions. In speaking, on the other hand, the burden of ensuring felicity rests squarely on the speaker. In the child language experiments, accommodation is not tested, so the rules for comprehending NPs are simple in that the child does not need to refer to the state of referential knowledge of the interlocutor at all. If a definite NP is heard, just update a file card. In the case of an indefinite NP, start a new file card. This can be illustrated with respect to 1 in Table 1, the comprehension experiment by Maratsos. The text the child hears is something like "Now this boy came and sat down in one of the chairs. And just as he sat down, suddenly {a/the} chair fell over." If the child hears 'the chair', he simply updates a card, and the only one that will fit the description is the card started by 'one of the chairs'. So the child acts on the chair sat upon by the boy. If the child hears 'a chair', the child starts a new file card, and will act on a different chair than the one mentioned previously. The results showed that the kids did well on this task. The other comprehension experiment in the literature is a replication of Maratsos by Karmiloff-Smith (1979), as shown in Table 2, 3(b). In similar experimental conditions, French children picked a new referent for the indefinite 95% of the time. They did not do so well with the definite NP, partly because as suggested by Karmiloff-Smith, French children at some stage of their development confused uniqueness with singular cardinality, since the indefinite article in French is also the form for the numeral 'one'.

2.2 Introducing referents with indefinite NPs is more difficult than referring to established referents with definite NPs

The second generalization is that the literature contains overwhelming evidence that introducing referents with indefinite NPs is much more difficult than referring to established referents with definite NPs. If one looks at Warden (1981) which uses a methodology that aims at eliminating the deictic bias, children still felt reluctant to use indefinite NPs to introduce referents. Even 8-year-olds only used indefinite NPs around 60% of the time. This pattern of findings is also confirmed in a number of studies of narrative production, notably Karmiloff-Smith (1981), Hickman (1980) and Bamberg (1986), covering English, French and German. Again the file change rules may give us some clue as to why this may be so. Recall that rule (4) has a very simple rule for updating files (ie referring to established referents), and that is to use a definite NP. This is something that does not have to make reference to the hearer's consciousness. However, in starting a new file, one has two options, one can use an indefinite NP or one can use a definite NP depending on one's assessment of the hearer's referential

knowledge. Perhaps it is this complexity related to file card creation that underlies the pervasive problem in establishing referents. The evidence we have from production experiments all point in this direction. The only exception is the results of Emslie and Stevenson (1981), which are unaccounted for in the present explanation. Further experimental replications may resolve the discrepancy.

2.3 Using an indefinite NP to refer to a member of a plural set of identical objects is more difficult than using a definite NP to refer to a unique referent

The third generalization bears on the link between uniqueness and anaphora in the semantics of indefinite and definite NPs. It has been observed (Hawkins 1978) that the definite/indefinite contrast in English has an additional dimension, what he calls inclusiveness vs exclusiveness. It seems that 'the N' refers to the totality of a set, whereas 'a N' refers to a member of a set. Hawkins draws attention to contrasts such as those in (5).

(5) John lost a hand in the war/ ?John lost a head in the war.

If one is dealing with a singleton set, one refers to the element of the set by 'the N'. On the contrary if it is a member of a plural set, one uses the definite NP. This semantic contrast may be independent of the anaphoric contrast based on whether file cards exist in one's mind, and is not always encoded along with the (in)definiteness distinctions in other languages. For example, word order differences signal a definiteness contrast in Chinese, but it does not seem to have any bearing at all on uniqueness. So it is not the case that preverbal NPs will mark unique objects and postverbal NPs will mark non-uniqueness. In the acquisition literature the inclusiveness dimension and the anaphoric dimensions are sometimes conflated. A case in point is the elicited production tasks of Maratsos (1976), where the prompt for the definite article is both an established referent, and also a unique referent. This is also true of the experiments of Emslie and Stevenson (1981) and Power and Dal Matello (1986). In other words, one can have a paradigm like the following (6):

(6)		referent introduction	subsequent reference
X	I	II	
(singleton)			
X X X	III	IV	
(plural)			

Maratsos only tested the possibilities II and IV, while Emslie and Stevenson and Power and Dal Matello tested the possibilities I and II.

The researcher who first tried to sort out this issue is Karmiloff-Smith. In one of her experiments, given in (3c) of Table 2, she asked the child to refer to either a unique object or a member of a plural set of identical objects, in an array as in (7):

(7) boy girl X Y Y Y

The experimenter touches one of the objects X or Y, and then instructs the child to ask either the boy or the girl to lend him/her the object. If it is X, then the expected answer will contain the definite NP. If it is a Y, then the expected answer should contain an indefinite article. As can be seen from the results, children had no problem referring to the unique object with the definite article, but had difficulty in referring to a member of the plural set using the indefinite article. In another of her experiments, Karmiloff-Smith tries to tease apart the anaphora factor and the uniqueness factor, but asking children to describe sequences of actions done on singleton sets and plural sets of identical objects, shown in (3d) of Table 2 (Experiment #5 of Karmiloff-Smith 1979). The findings are interesting, because they show (i) that in introducing referents, children again had no problem using the definite article to refer to unique objects, but had difficulty in using the indefinite article to refer to a member of a plural set; (ii) at the same time, once the referent has been introduced into discourse, and children refer back to them, they used definite NPs correctly for the singleton condition 79% to 92% of the time, but had a great deal of difficulty doing the same for the plural condition.

How can we make sense of this finding? we suggest that children may not fully distinguish file cards created by discourse and file cards created by real world knowledge or deixis in the same way. For adults, even if one is dealing with a set of several objects, which can trigger creation of file cards, the object that has been mentioned in discourse assumes prominence. So when subsequently referring to the established referent, the adult is clear about which file card to update, and encodes it with the definite form. However, children may not distinguish these two types of file cards sharply, and as a result they are indecisive in the choice of the determiner.

With these generalizations stated, let us return to the issue whether (in)definiteness is acquired early or late. The data suggest to us that it is more meaningful to ask which of the pragmatic rules are firmly in place. Thus it would seem that children by five years or so have acquired Rules (3a-b) and (4a-b). However, rules (3c) and (4c) are still difficult for them. It is difficult to claim that children have all the rules, including the rules in (3c, 4c), but that they cannot apply it properly because of the difficulty of assessing hearer knowledge. This is because these are inherently pragmatic rules, and it is hard to claim that one has the rule in a lot of contexts of use, one cannot apply it properly. It would be like saying one has grasped the norms of politeness, but in most situations of social interaction, one turns out to be impolite to others. The conclusion that I wish to draw from this brief survey is that children by 5 have acquired some of the pragmatic rules but some are not yet firmly in place.

Below we present the Cantonese acquisition data in the light of the general pattern of findings on the acquisition of (in)definiteness.

3. Acquisition of (in)definiteness by Cantonese-speaking children

3.1 Word order and (in)definiteness in Cantonese

It is well known that languages vary in how they encode this distinction. As reported in Kramsky (1972), while many languages use an article system (e.g. English and French), other languages use case inflections (e.g. in Altaic languages such as Turkish, the definite object is marked by accusative case if it is definite; it is marked by

nominative case if it is indefinite; Korean also makes use of case endings to signal definiteness cf. Chungmin Lee; Scandinavian languages show similar facts, cf. Diana). There are other languages that use syntagmatic rather than paradigmatic contrasts, notably languages like Finnish and Chinese (cf. Chesterman 1991) which use word order to signal the distinction. Cantonese is a Sinitic language which lends itself to the study of the role of word order in the acquisition of definiteness, because of particular facts about the distribution of NPs in the language.

In Cantonese, a minimal contrast can be shown by placing singular noun phrases of the form [CL N] before and after ergative verbs. In these constructions, when preverbal the NPs are definite; when postverbal they only take on an indefinite interpretation. The contrast is illustrated in (8a) and (8b). (8a) would be infelicitous unless the speaker presupposes that the hearer has a file card for the badge that dropped, whereas no such presupposition is present for (8b).

(8a) (Tau4sin1) go3 zoeng1 lat1 zo2
(just-now) CL badge drop asp "(Just now) the badge dropped"

(8b) (Tau4sin1) lat1 zo2 go3 zoeng1
(just-now) drop asp CL badge "(Just now) a badge dropped"

A similar contrast is shown in the existential construction (cf. Huang 1987) between the singular noun phrase preceded by the existential verb 'yau2' ('have') and the same noun phrase when it is not so preceded. As can be seen from (9a) and (9b), the presence of an existential verb forces an indefinite reading, whereas its absence will put the NP in preverbal position and force a definite reading.

(9a) (Tau4sin1) go3 bo1 luk1 zau2 zo2
(just-now) CL ball roll away asp "(Just now) the ball rolled away"

(9b) (Tau4sin1) jau5 go3 bo1 luk1 zau2 zo2
(just-now) have CL ball roll away asp "(Just now) a ball rolled away"

The main reason why one has to resort to this type of [CL N] noun phrase to show a minimal contrast is because ordinary numeral phrases are defective in their distribution and are generally prohibited from the subject position. This is a general fact about Sinitic languages and is true of both Mandarin and Cantonese. Thus (9c) is not acceptable without the existential verb.

(9c) (Tau4sin1) ?*(jau5) jat1 go3 bo1 luk1 zau2 zo2
just-now have one CL ball roll away asp "(just now) a/one ball rolled away"

This fact has been noted for Chinese grammarians for a long time (cf. Chao 1968), and recently the phenomenon has been analyzed in terms of the failure of the subject numeral phrase to get existential closure since the Chinese subject phrase lies outside the nuclear scope of the logical form (cf. Tsai 1994). The function of the existential verb is precisely to license existential closure. Now the types of noun phrases that can take on both definite or indefinite meanings depending on position are bare noun

phrases, numeral phrases and [CL-N] phrases. To examine the role of word order in signalling definiteness for singular NPs, one must use phrases of the type [CL N].

(10) N/ [num CL N]/ [CL N] (Cantonese)

3.2 Comprehension experiment

Aims and objectives

The broad aim of the study is to find out when Cantonese children become sensitive to the link between word order and definiteness. One might expect the correlation between definiteness and preverbal position to be acquired early for two reasons. (a) First, the property is a dominant feature of the language, and abundant positive evidence should be available to the child for the acquisition task; (b) secondly, we know from naturalistic findings on word order in Chinese child language (cf. the work of Erbaugh 1982) that young children keep a very rigid word order template, which means they are sensitive to the mapping between word order and thematic roles very early on. This finding is supported by the experimental study of Chang and Tseng (1988). They showed that when semantic information contradicted the word order, three- to five-year-old children relied more heavily on word order than adults in interpreting the thematic roles of arguments [e.g. interpreting the ball as agent in 'ball kick doll']. If children tune in to word order for the purpose of thematic role mapping, would it not be possible that they do the same for distinguishing between definite and indefinite reference?

On the other hand, child language data that more directly bear on the issue of definiteness suggest that the acquisition may be a drawn-out process. In a study of Beijing Mandarin children, Hickman and Liang (1990) asked 4-, 5-, 6-, 7- and 10-year-olds to do a picture-based narration task. Their results indicate that most of the referent introductions of 4-year-olds were preverbal; 5- and 6-year-olds use preverbal and postverbal positions equally frequently. The 7- and 10-year-olds, however, showed a preference for post-verbal position in introducing the referent.

We designed a comprehension experiment modeled after Maratsos' design to find out:

(a) how Cantonese children interpret the definiteness of NPs marked by the existential verb *jau* as opposed to identical sentences in which the relevant NP is not preceded by the existential verb [ie the difference between (9a) and (9 b)];

(b) how Cantonese children interpret the definiteness of NPs that follow ergative verbs such as verbs of appearance and disappearance as opposed to NPs that precede such verbs [ie the contrast between (8a) and (8b)];

(c) whether the existential verb *jau* is perceived differently from verbs of appearance and disappearance in their marking of indefiniteness;

Subjects and method

The experiment involved an act-out comprehension task. The subjects were 40 kindergarteners; 20 boys and 20 girls. They were evenly distributed into 5 age groups, with 8 subjects in each group. Group I: 3;7 - 4;0, II: 4;1 - 4;6, III: 4;7 - 5;0, IV: 5;1 - 5;6, V: 5;7 - 6;0. In addition, 8 adults served as control.

A simple toy-moving task was used in the comprehension test in which children were asked to listen to recorded story episodes which averaged around 50-60 syllables in length. As they listen to the episodes, the child subjects were to enact the actions described by moving some toys. The toys were coloured drawings of objects that were made on cardboard paper and were mounted on magnets. They were placed on a white board so they can be moved around easily.

The test sentences included 2 types of contrasting sentences. The first type was made up of an intransitive clause with a subject NP of the form [CL-N] introduced by the existential verb *jau5*; and an identical sentence with the same subject NP but without the existential verb (cf. 9a,b). The second type of contrasting sentences was made up of sentences with an ergative verb taking a theme [CL-N] argument in post-verbal position; and an identical sentence with the same [CL-N] argument placed in pre-verbal position (cf. 8a,b).

Six verbs were selected for each of the two types of contrasting sentences, giving a total of 24 test sentences. In order to see if there would be any animacy effects on the interpretation of definiteness, 3 of the verbs for each contrasting sentence type were used with inanimate NP arguments and the other 3 with animate NP arguments. All the verbs chosen were single argument verbs. The verbs used in the ergative constructions were *pa4-cut1-lei4* 'climb-out'; *cam4* 'sink'; *lat1* 'drop'; *lam3* 'collapse' and *zau2maai4lei4* 'come-near' and *zau2cut1lei4* 'come-out'. The verbs used in existential sentences were *haang4* 'walk', *luk1* 'roll'; *zau2* 'go'; *hoilzau2* 'drive off'; *lau2gai2* 'be-in-tantrum'; *dou2zyn3* 'turn over'. 3 of the verbs for each contrasting type were used with inanimate NP arguments and the other 3 for animate arguments. A total of 24 test sentences were used.

There were 12 pairs of episodes corresponding to 12 pairs of contrasting sentences. Test sentences were embedded in the episodes in such a way that each episode ended with a test sentence. Members of each pair differed only in the test sentence used. To minimize the cognitive load of the experimental task, six prop settings, each shared by two episodes, were used.

Each episode had a similar event structure: there was one protagonist and a set of four identical objects (either animate or inanimate). The first part of the episode was intended to linguistically establish a discourse referent from among the four identical objects. It was done by an action carried out by the protagonist related to one of the four identical objects. One of the four identical objects would either be physically isolated from the rest or be marked with something which would make it visually salient. The latter part of the episode would be a test sentence with a [CL-N] in a particular syntactic position signifying definite or indefinite reference. Upon hearing the test sentence, the child subjects were required to move one of the four identical objects. Which of the four identical objects (an object that had prior-mention or not) that the children moved was marked as their response. An example of a pair of

episodes occurring in the same setting can be illustrated with reference to sample episodes 1 and 2 below.

Sample episode #1 (for *jau*-sentences)

Siuming gamjat maai kei kui hai gaai gindou
 today sell flag he at street see-ASP

sei go popo haaidou dang ce
four CL granny PROG wait car

"Siuming is selling flags today and goes out. He sees four elderly women waiting for a car."

Siuming maai kei bei keizung-jatgo popo
 sell flag to one-of granny

zeung zi kei tip hai kui gin saam-dou
take CL flag attach on her CL clothing-LOC

"Siuming sells a flag to one of the elderly women, and pins a flag on her dress."

gozo hounoi, zung meijau ce
after long-while, still not-have car

"After a while, there is still not a car"

(version 1)

go popo zau zo

CL granny leave ASP "The elderly woman left."

(version 2)

jau go popo zau zo

have CL granny leave ASP "An elderly woman left."

Sample episode #2 (for sentences with ergative verbs)

Siuming gamjat zeukdzo gin san saam cutgai
 today wear CL new clothing go-street

gin saam soengbin jau sei go zoeng
CL clothing on-top have four CL buttons

"Siuming is wearing a new outfit today and goes out. On his outfit are four buttons."

kui zeung keizung-jatgo zeung kau haai deng mou-dou
he take one-of button pin on CL hat-LOC

"He pins one of the buttons on top of his hat."

Siuming zau haang faan ukkei la, haang haang ha,
 then walk back home SFP walk walk ASP,

"Then he walks back home. As he is walking, "

(Version 1)

go zoeng lat zo

CL button drop ASP "a button drops."

(Version 2)

lat zo *go zeung*

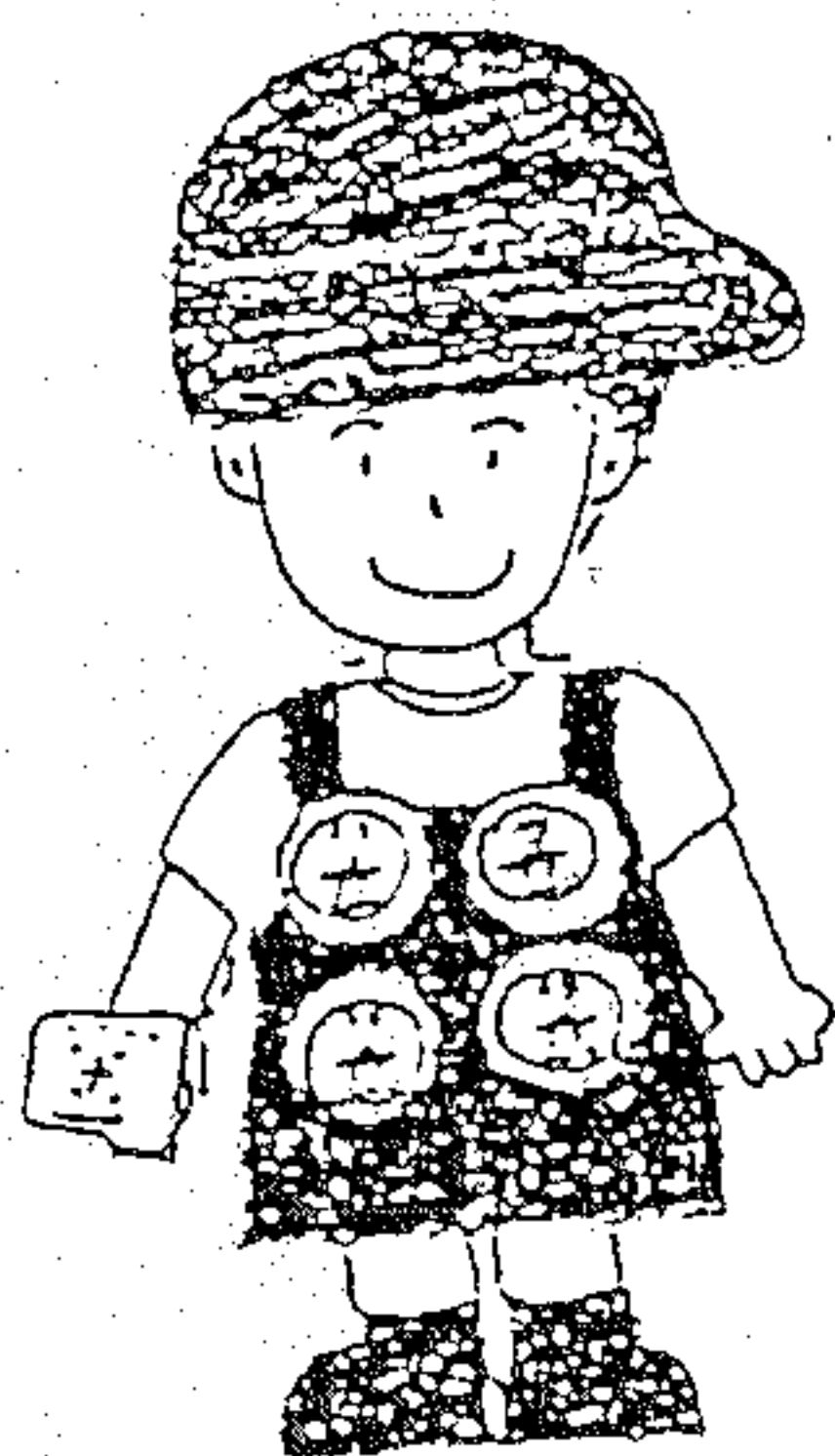
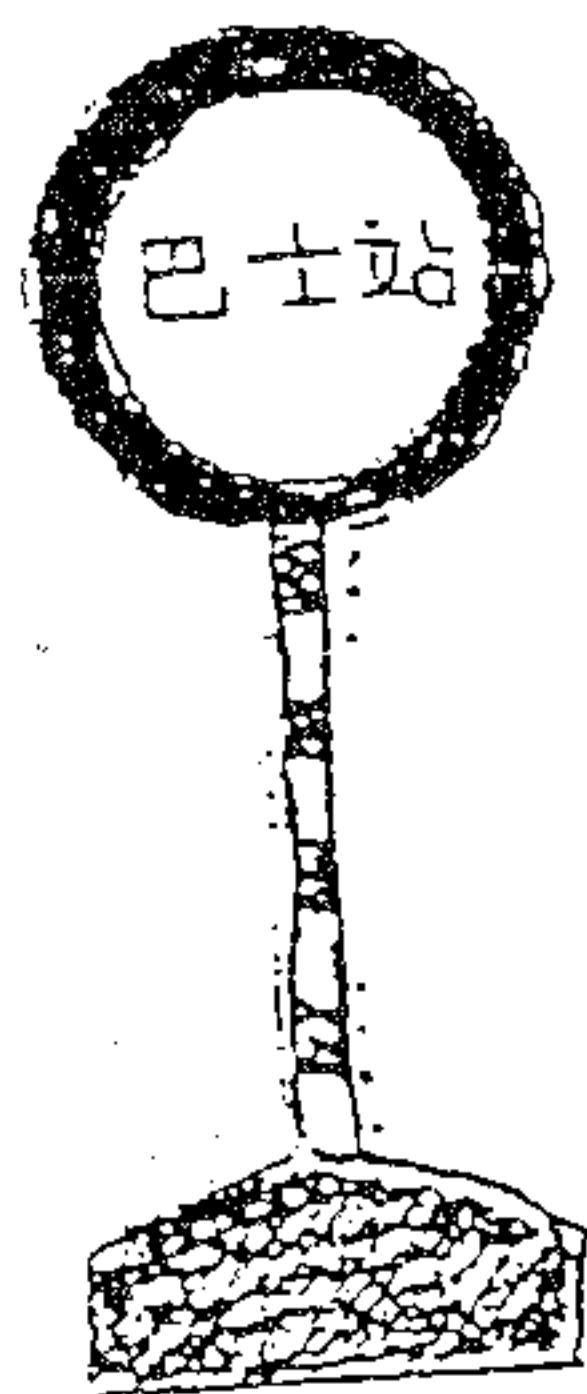
drop ASP CL button " the button drops"

sample episode 1

Siuming gamyat maai kei
today sell flag

kui hai gaai gindou sei go popo haaidou dang che
he at street see-ASP four CL granny PROG wait car

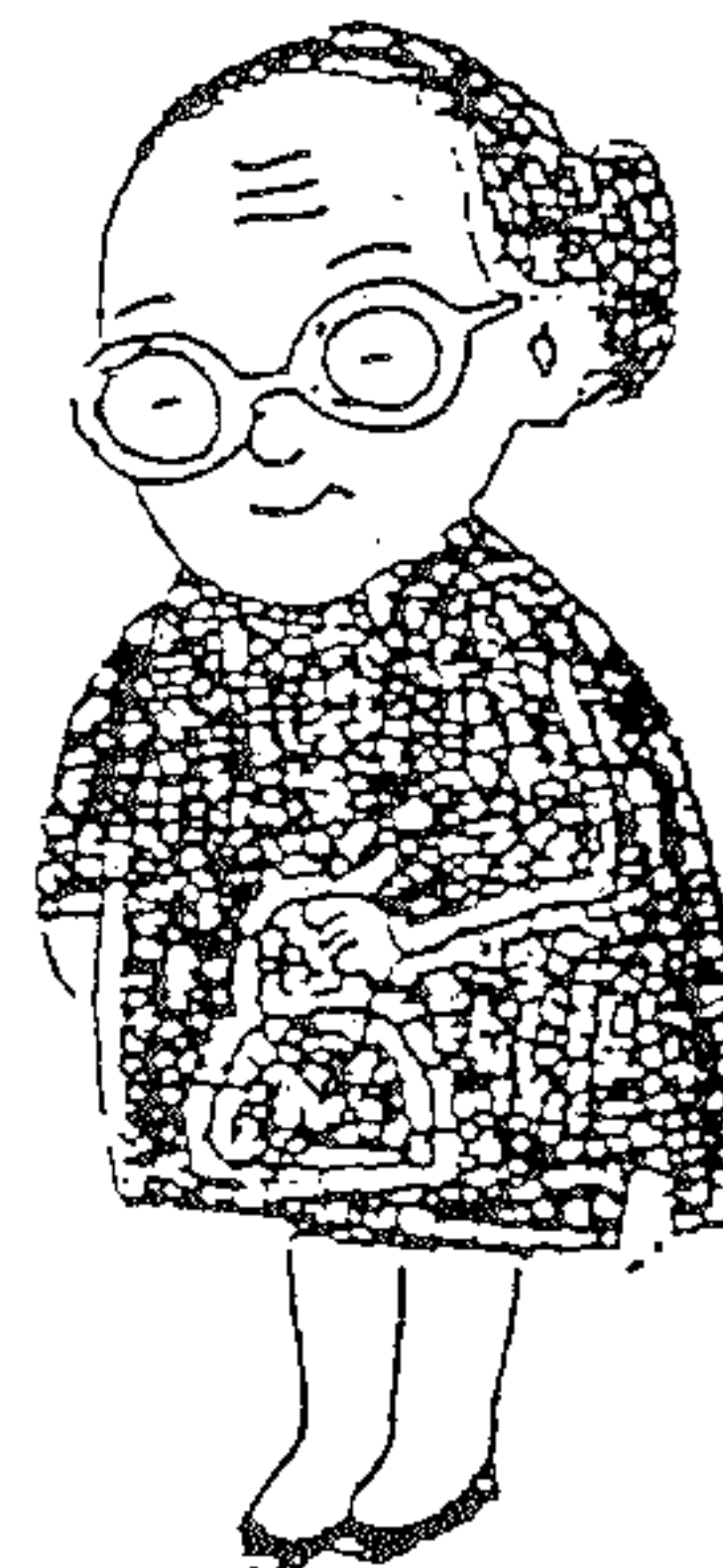
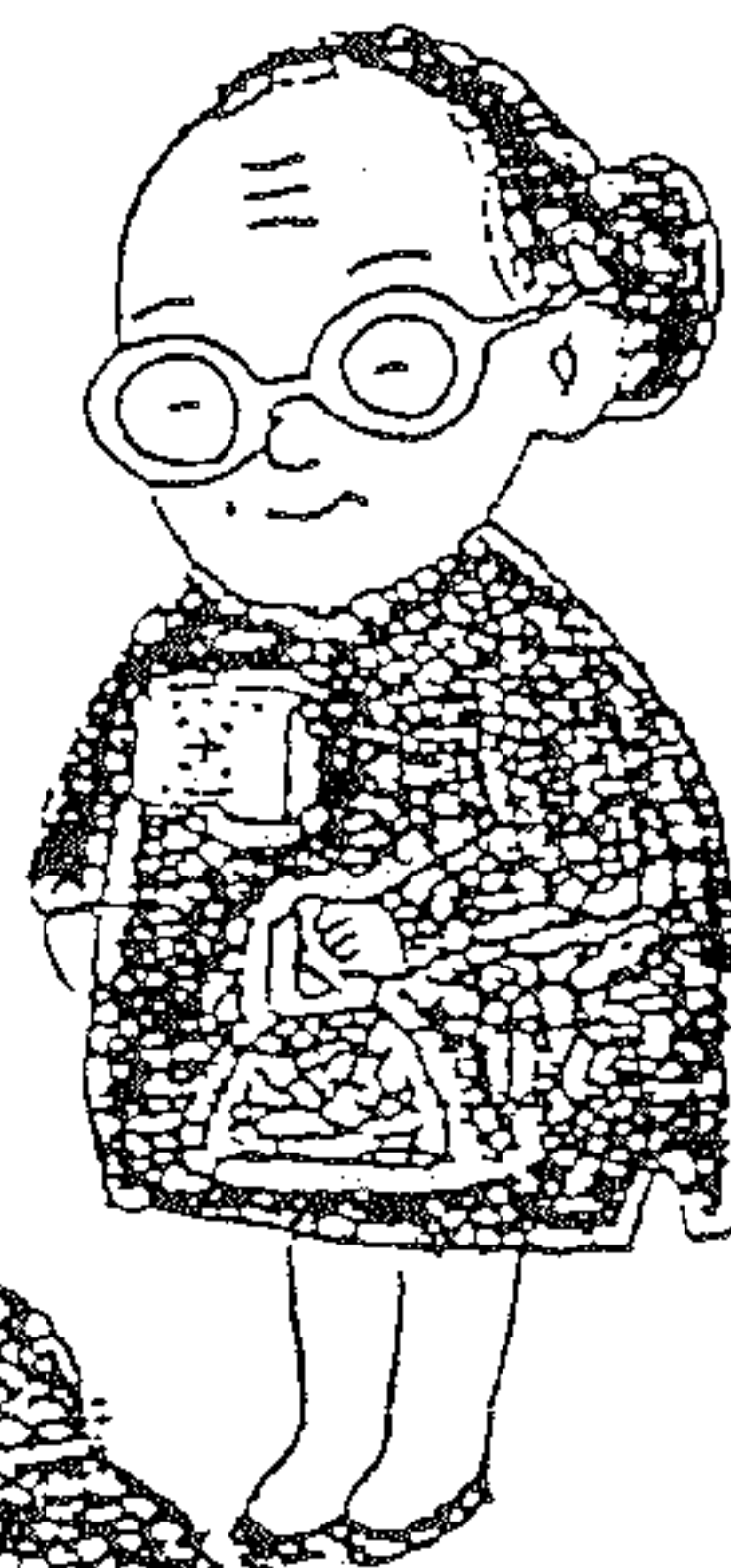
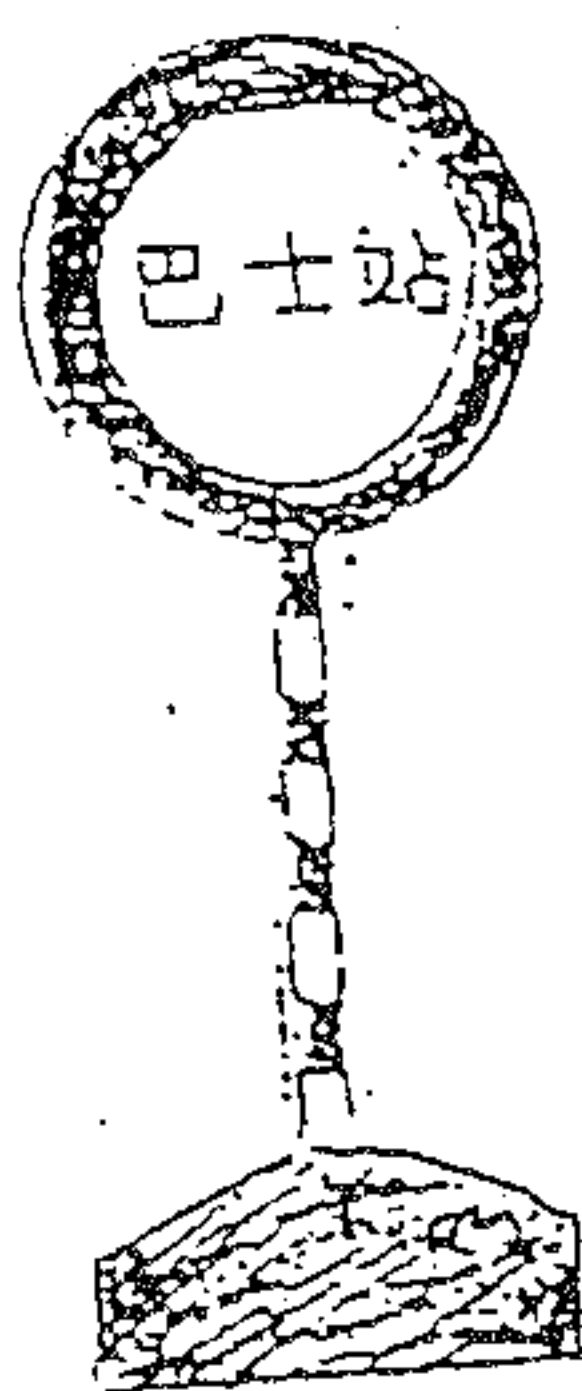
"Siuming is selling flags today and goes out. He sees four elderly women waiting for a car."



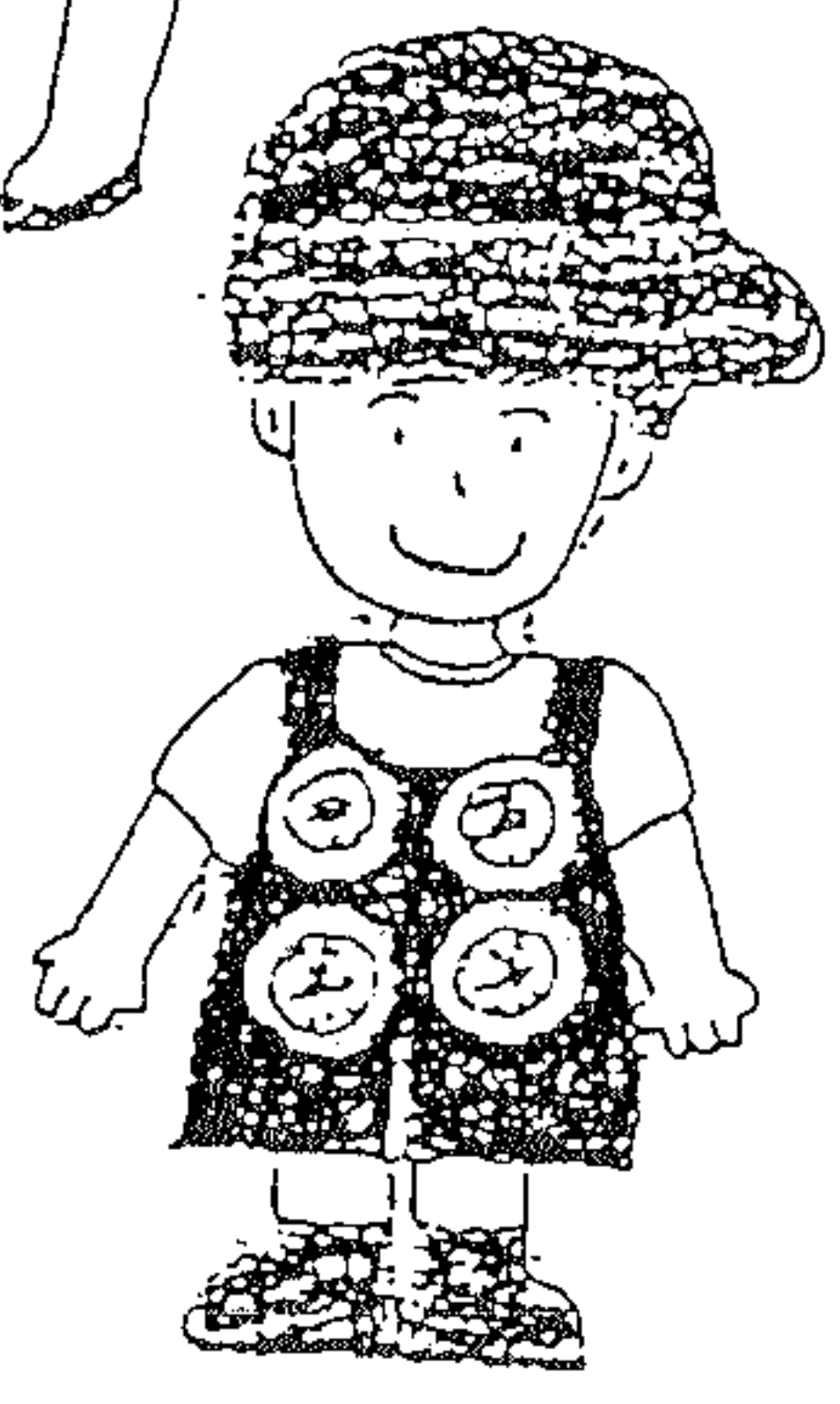
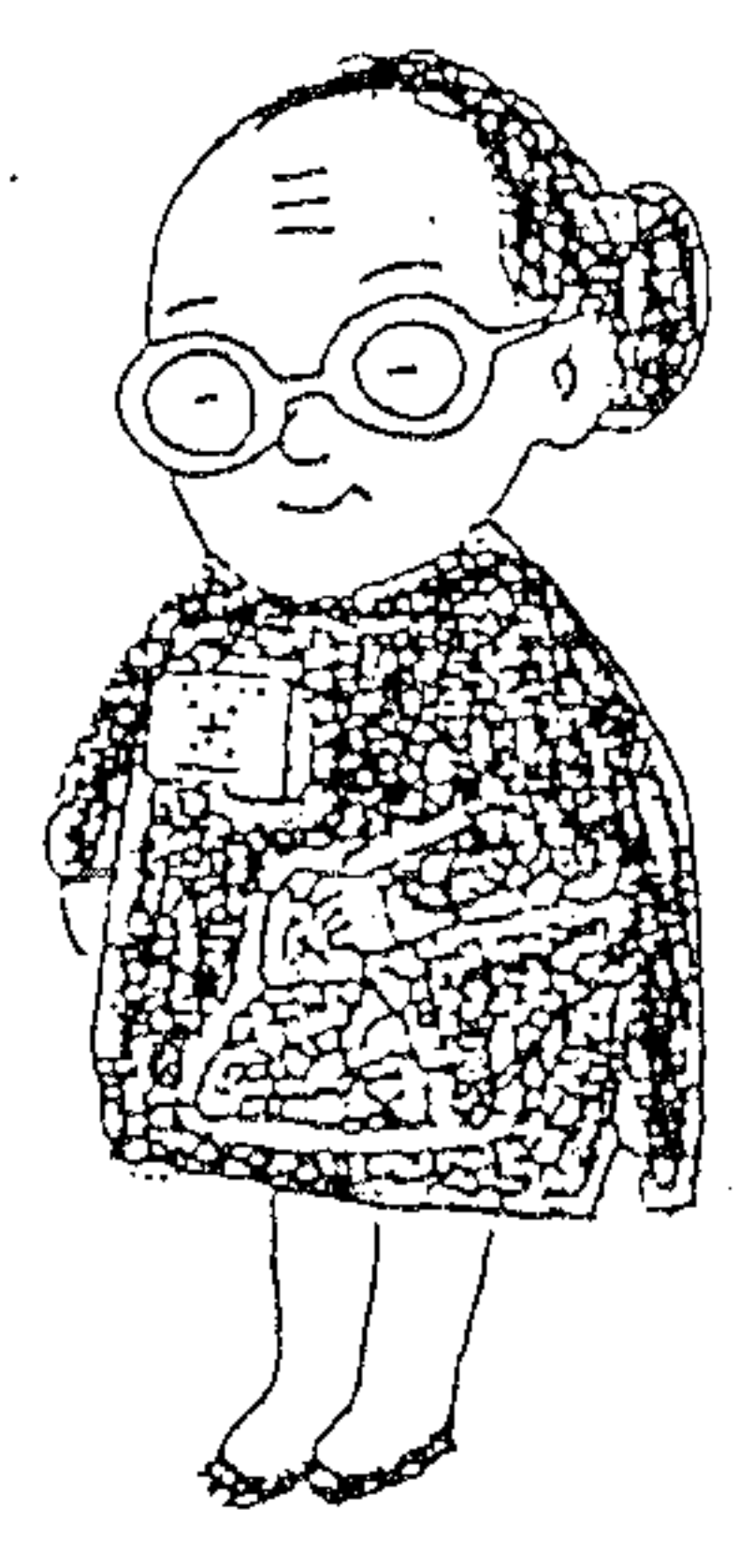
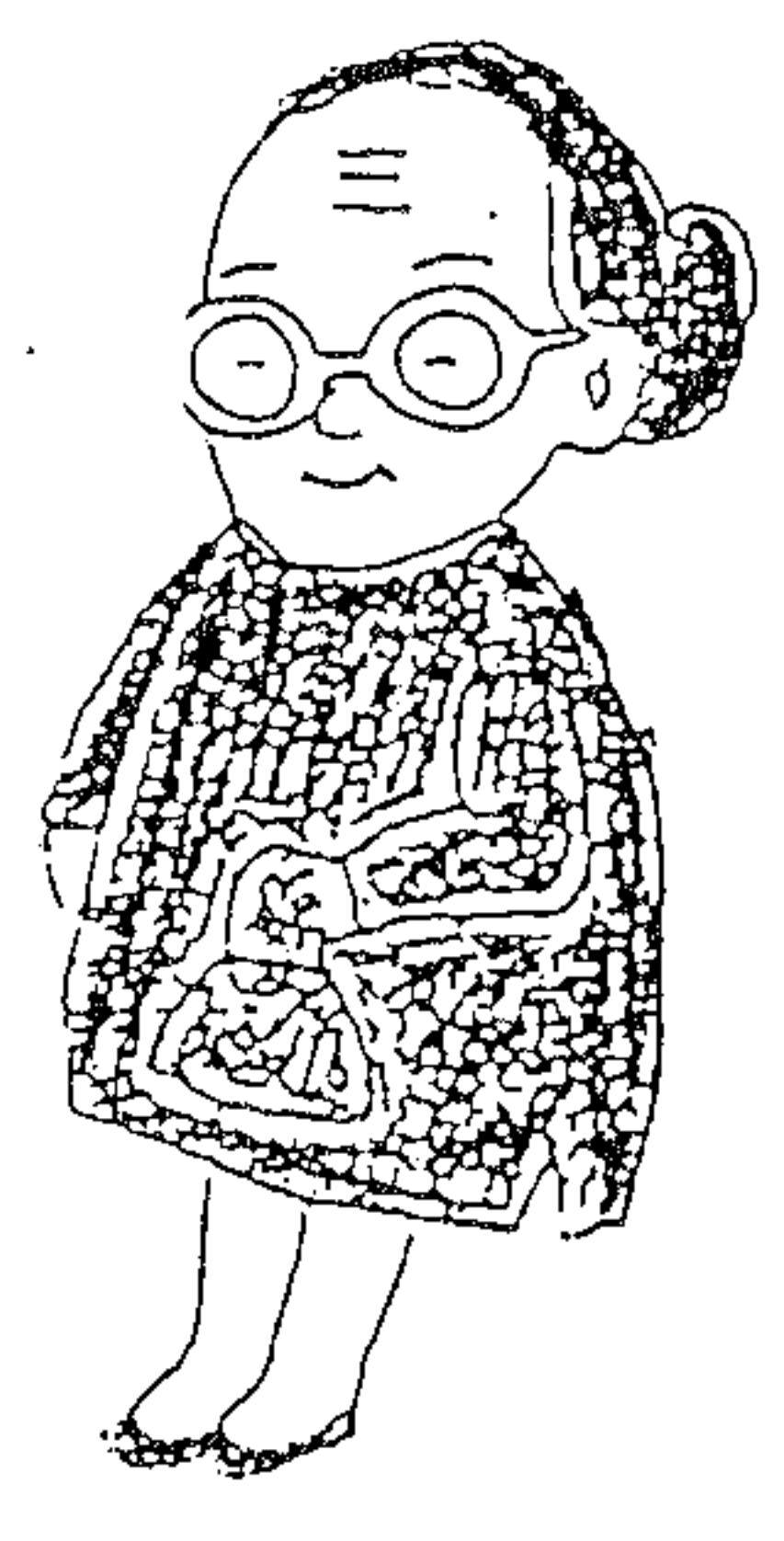
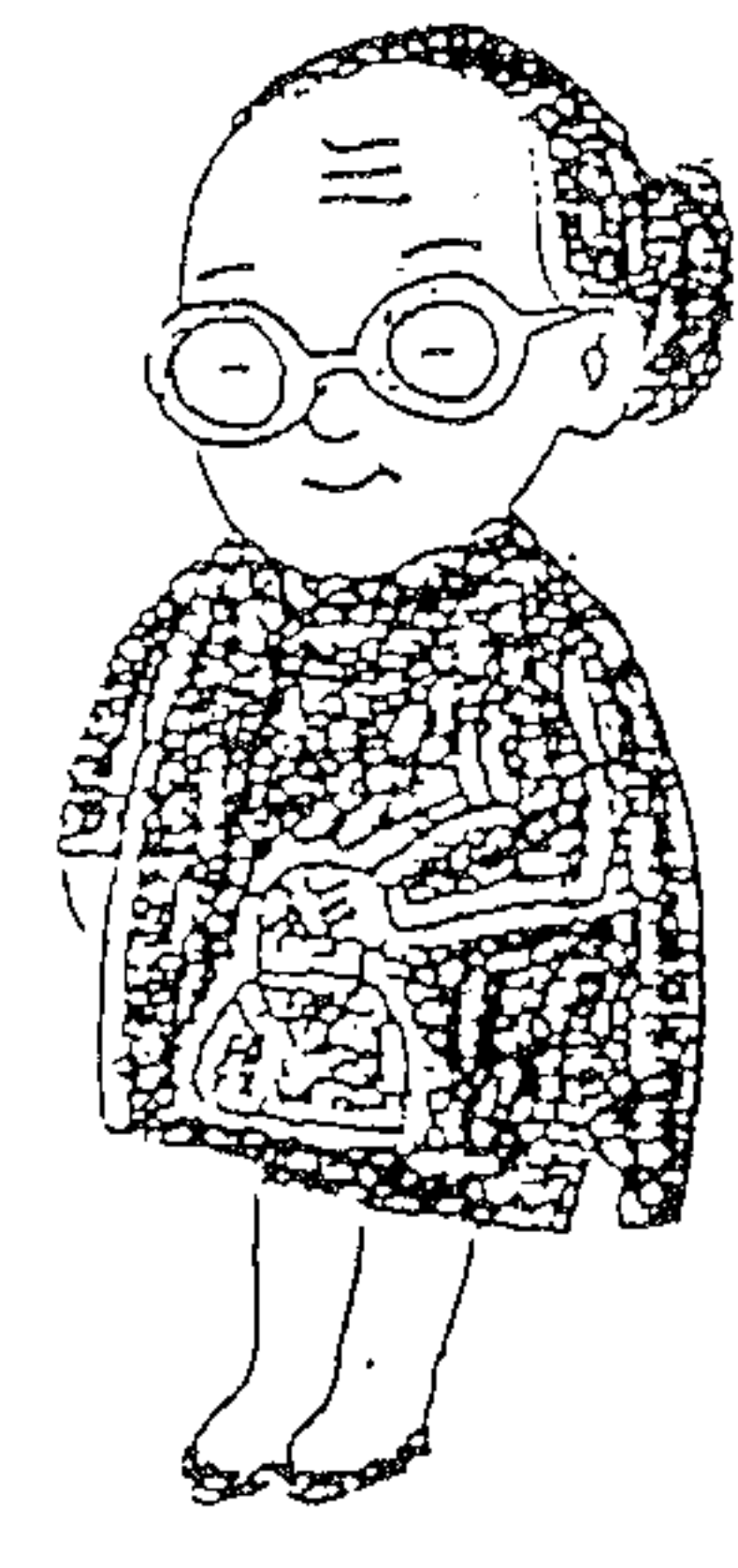
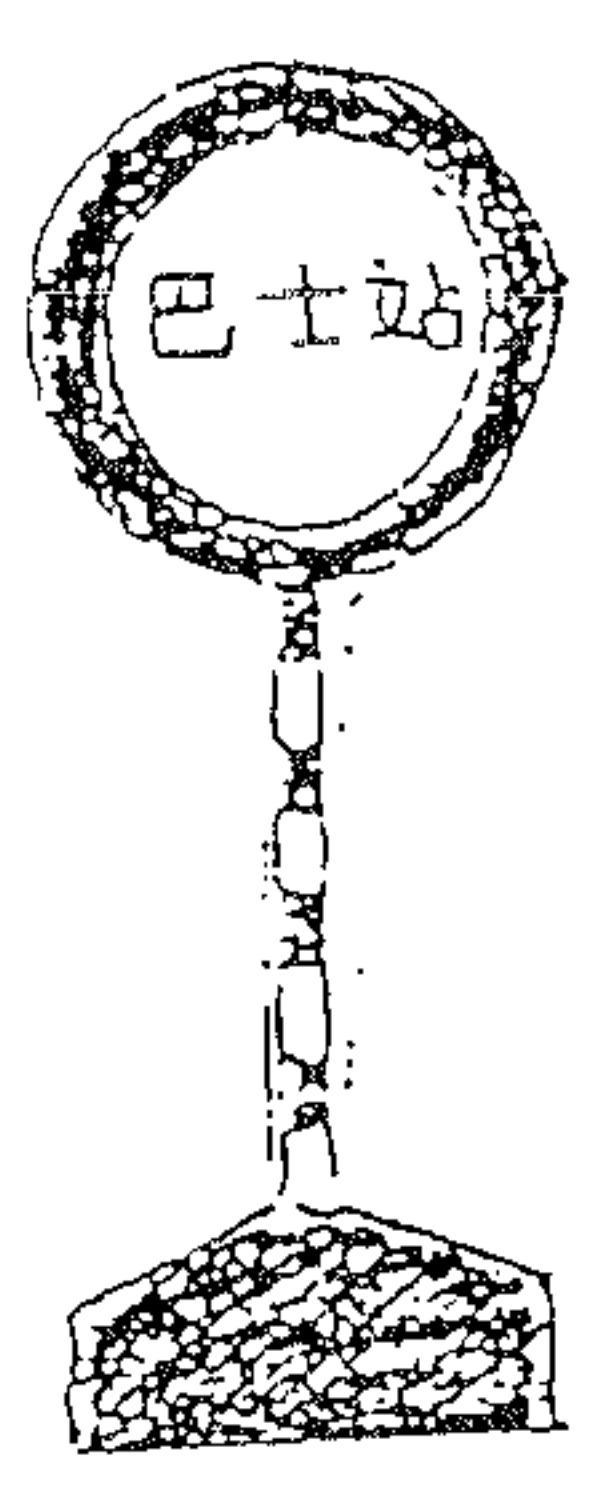
Siuming maai kei bei keijung-yatgo popo
sell flag to one-of granny

jeung ji kei tip hai kui gin saam-dou
take CL flag attach on her CL clothing-LOC

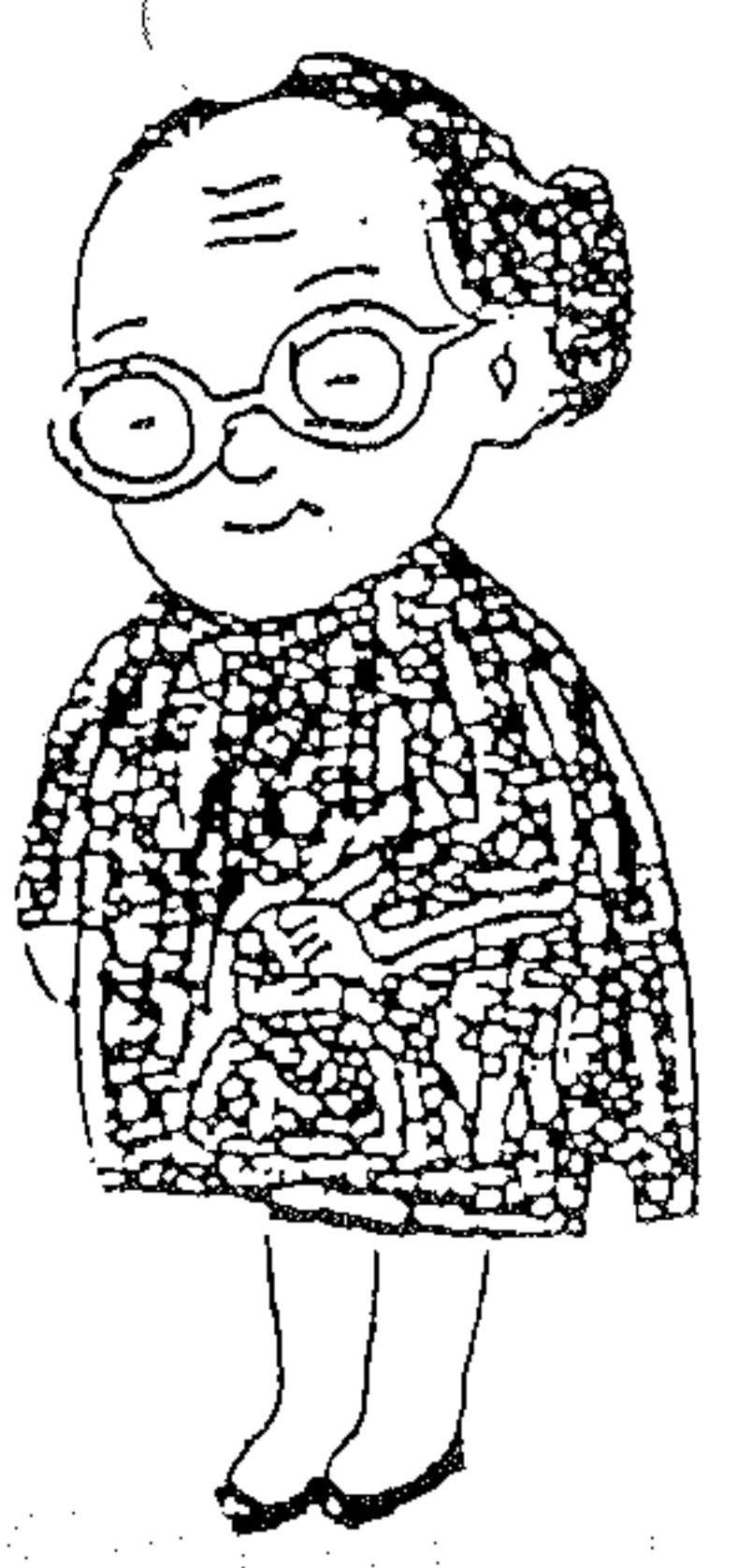
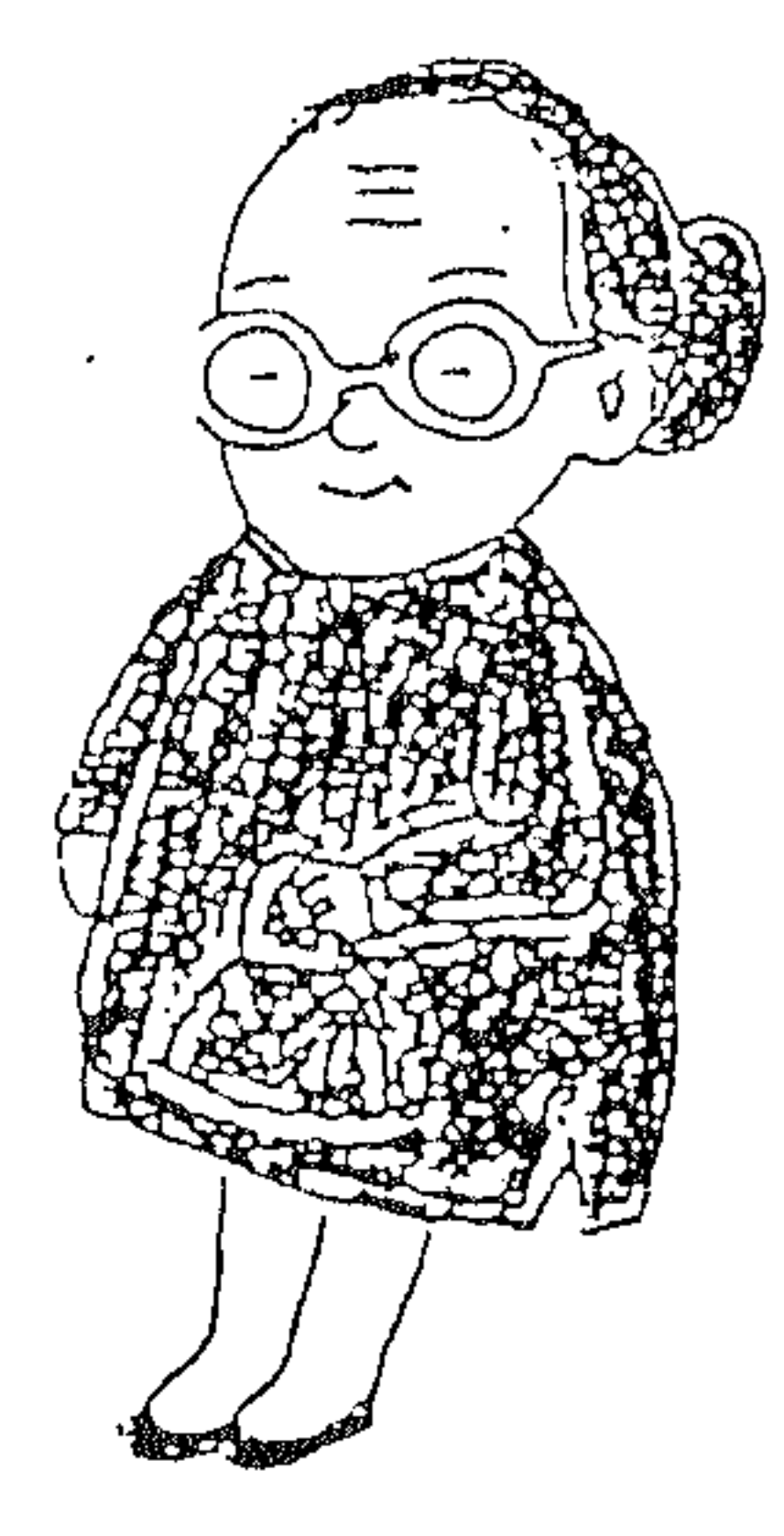
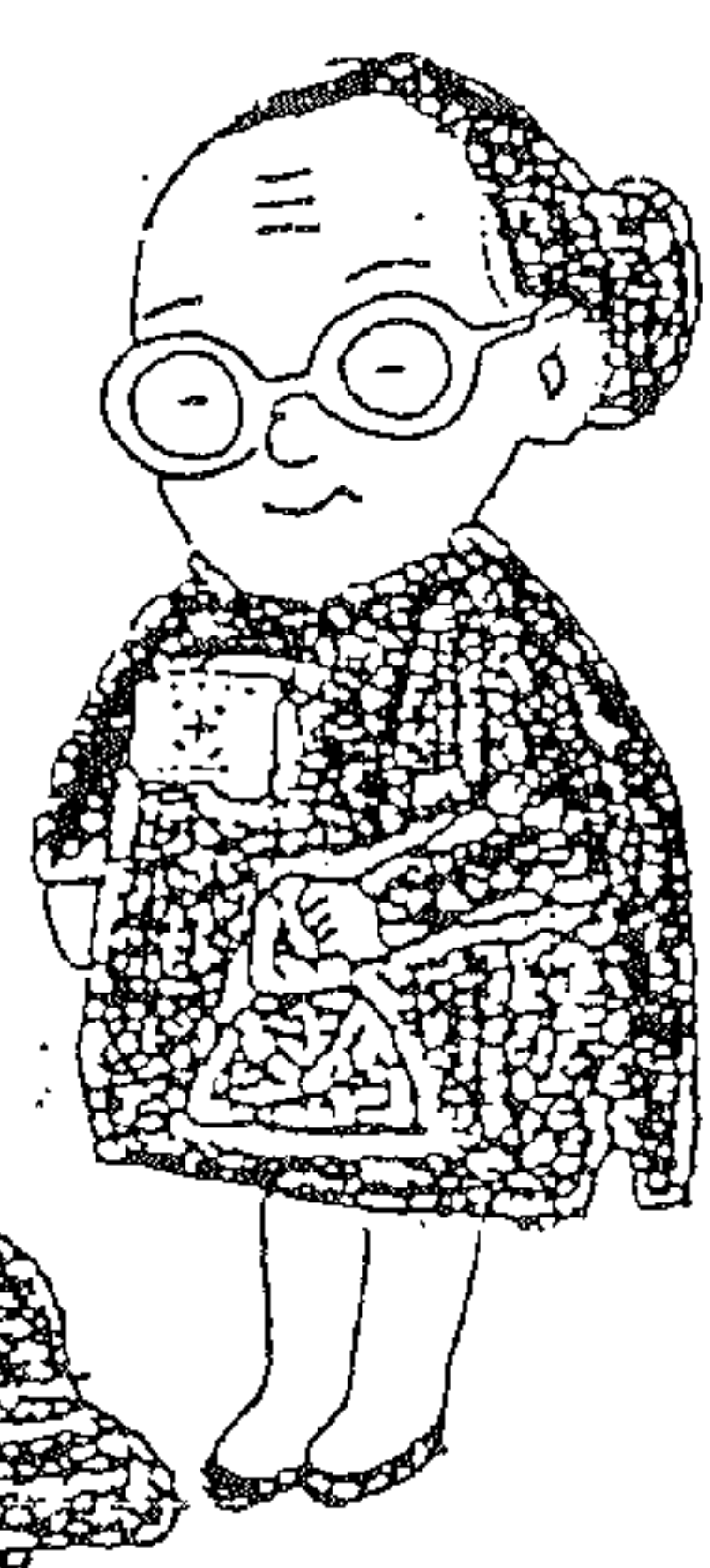
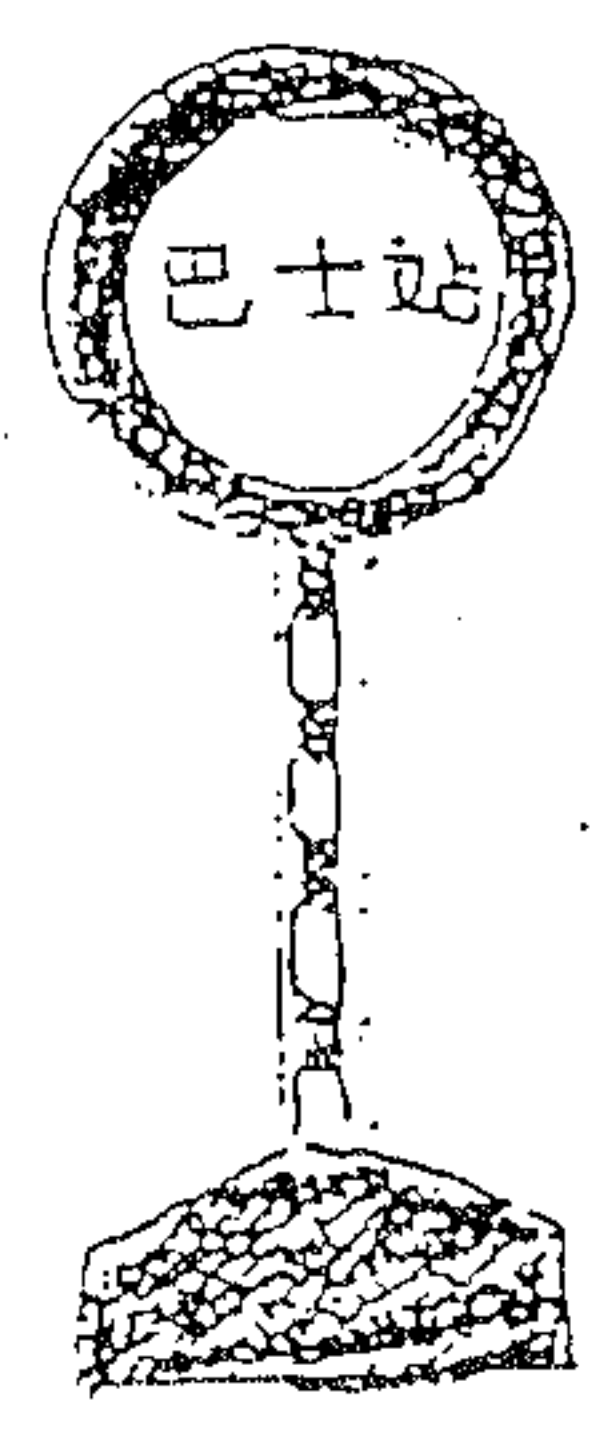
"Siuming sells a flag to one of the elderly women, and pins a flag on her dress."



godzo hounoi, jung meiyau che
 after long-while, still not-have car
 "After a while, there is still not a car"
 go popo jau dzo
 CL granny leave ASP "The elderly woman left."

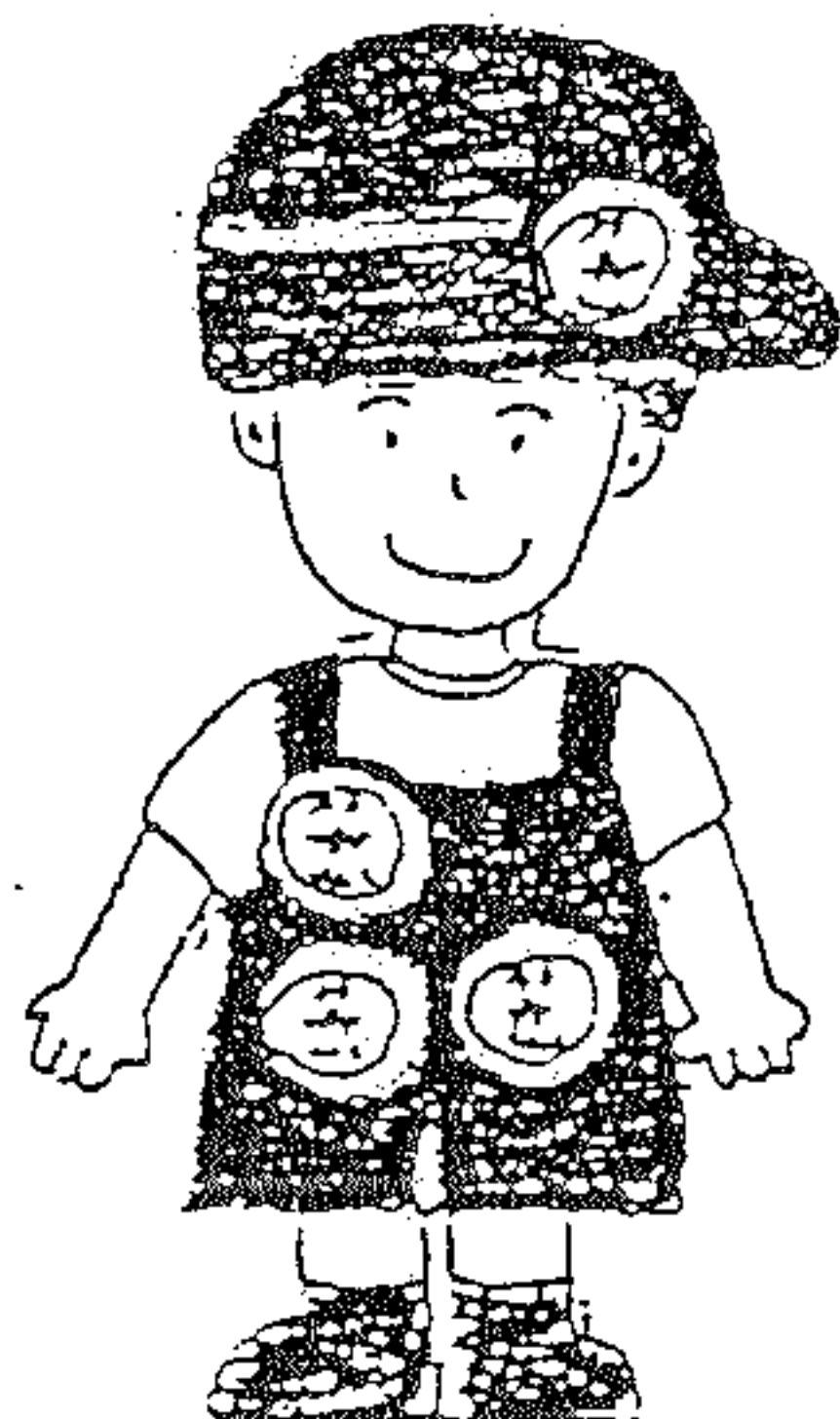
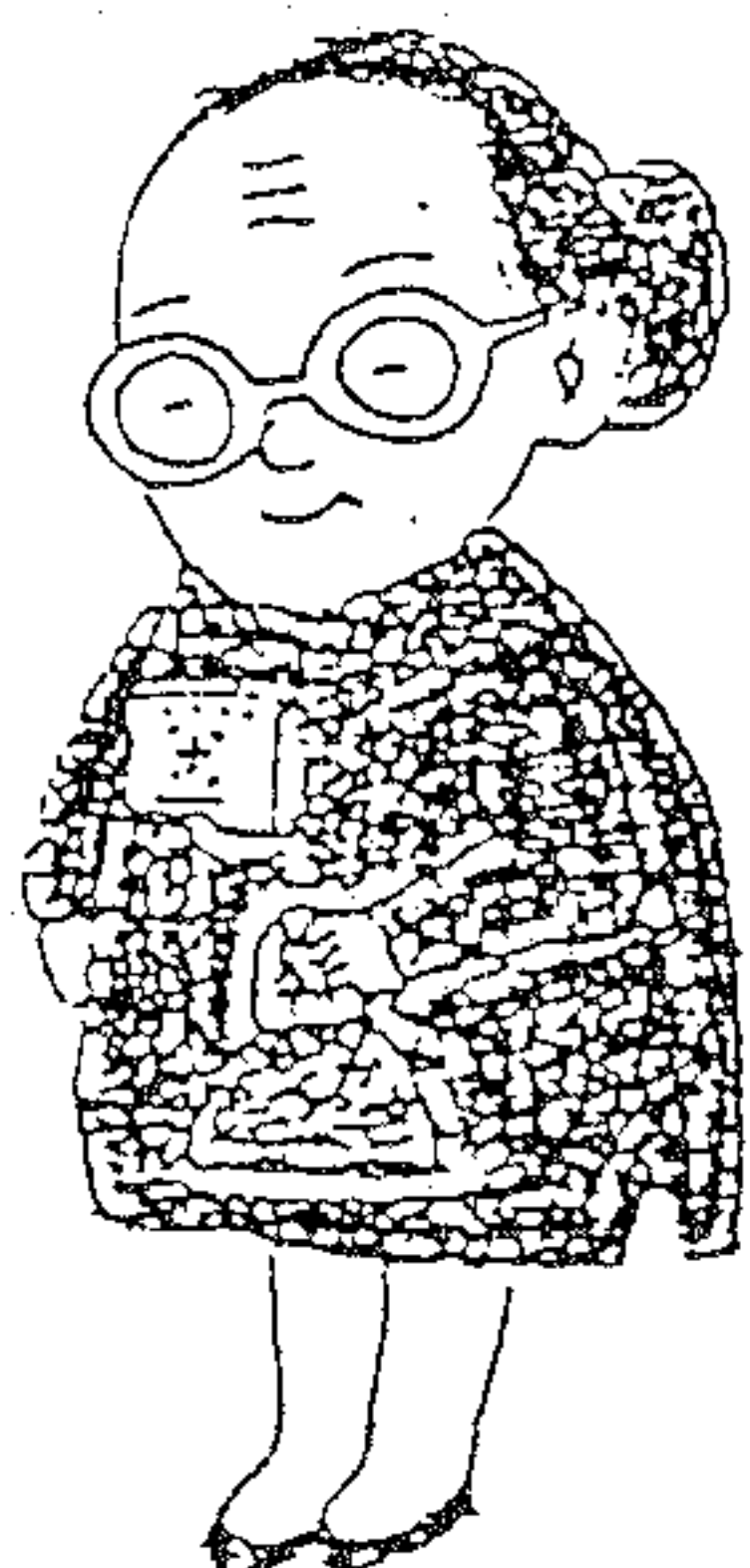
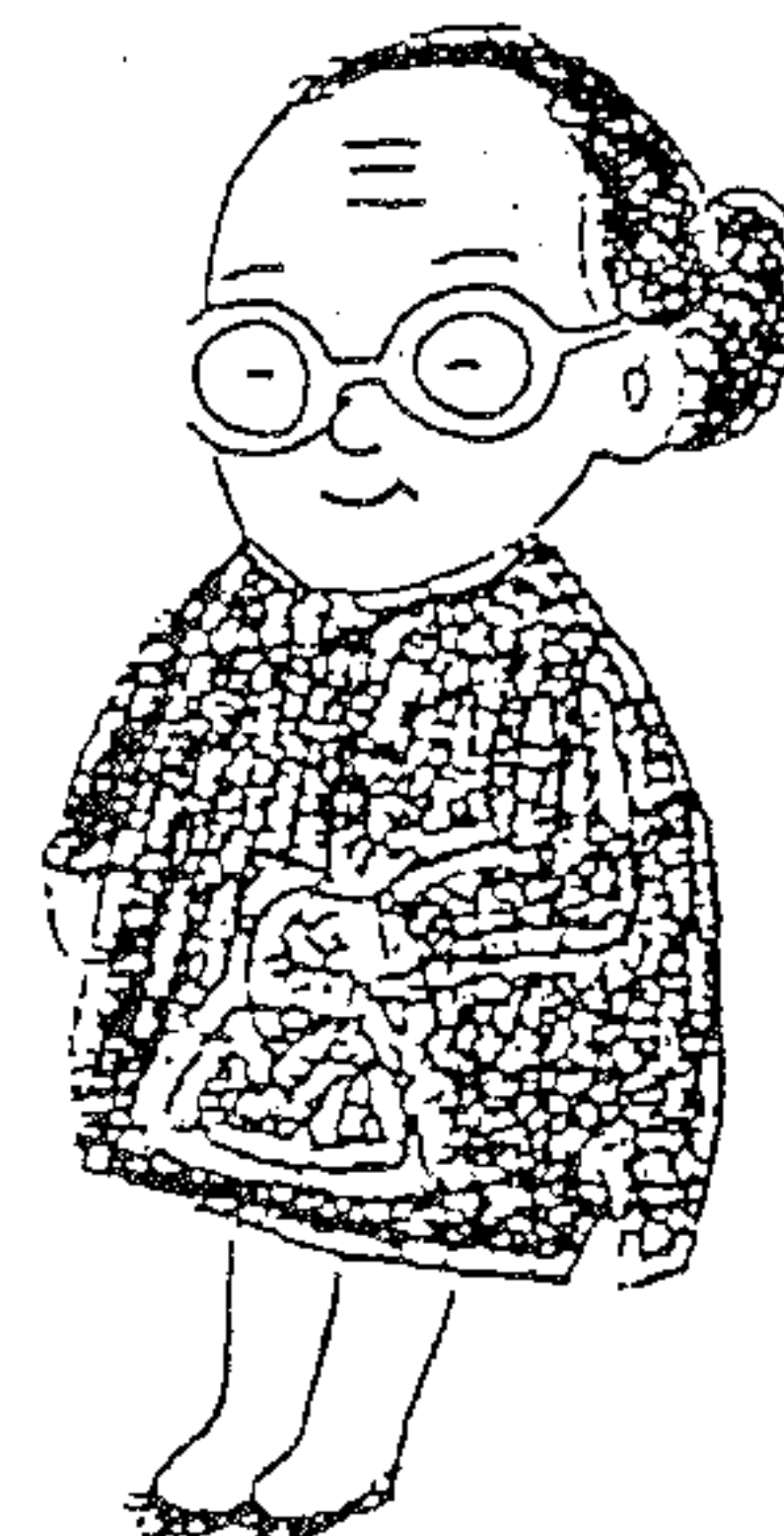
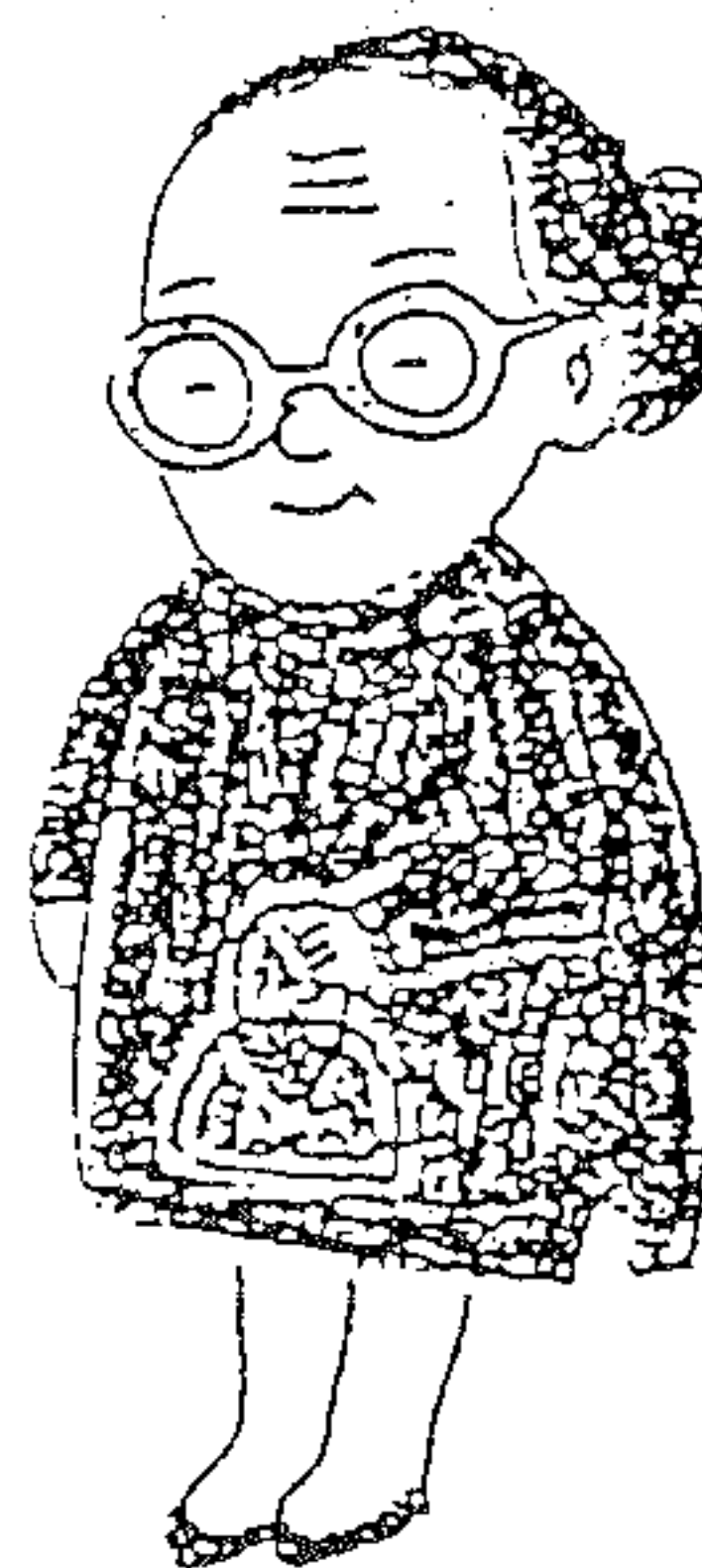
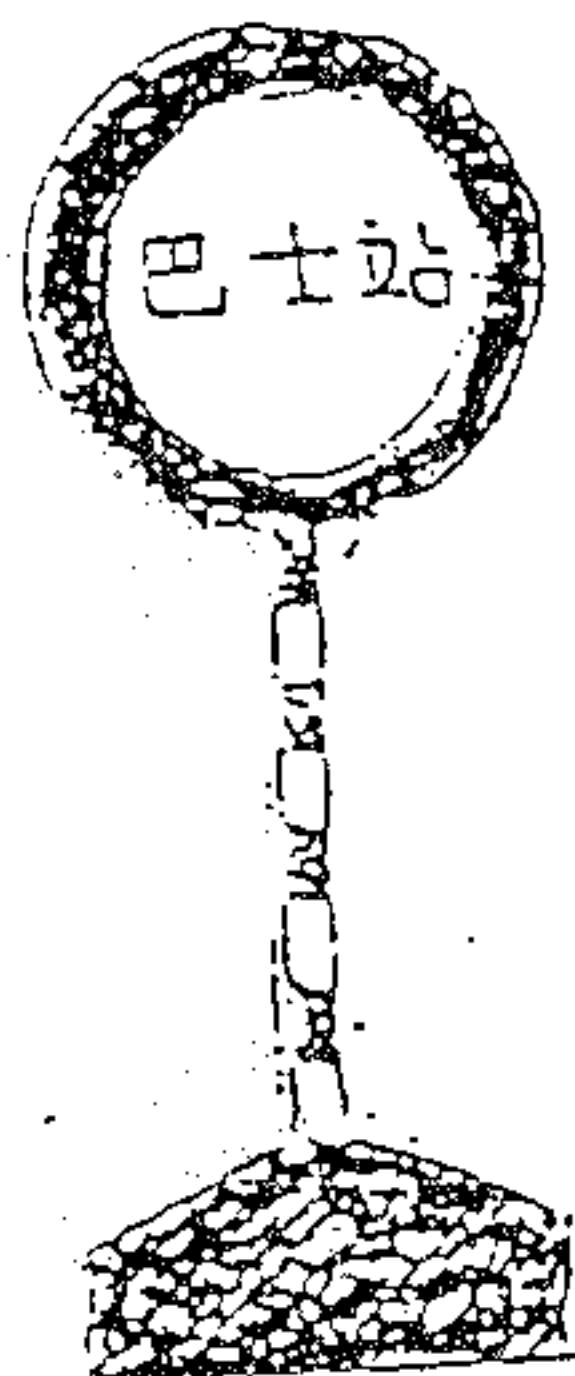


godzo hounoi, jung meiyau che
 after long-while, still not-have car
 "After a while, there is still not a car"
 yau go popo jau dzo
 have CL granny leave ASP "An elderly woman left."

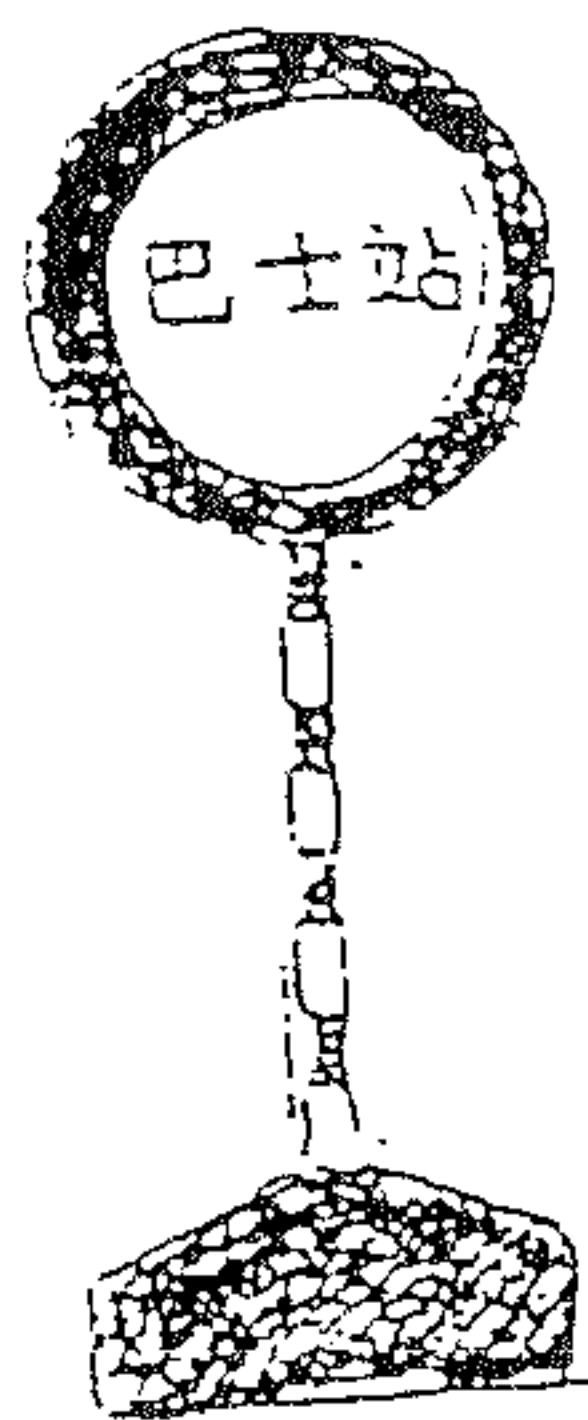


sample episode 2

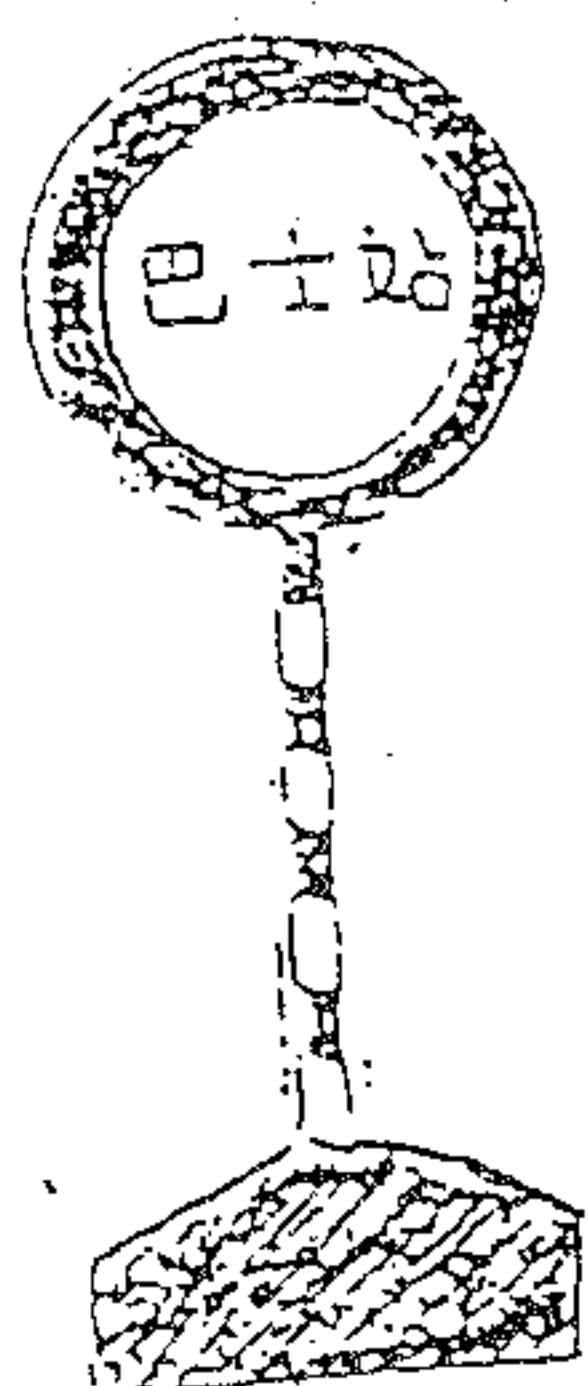
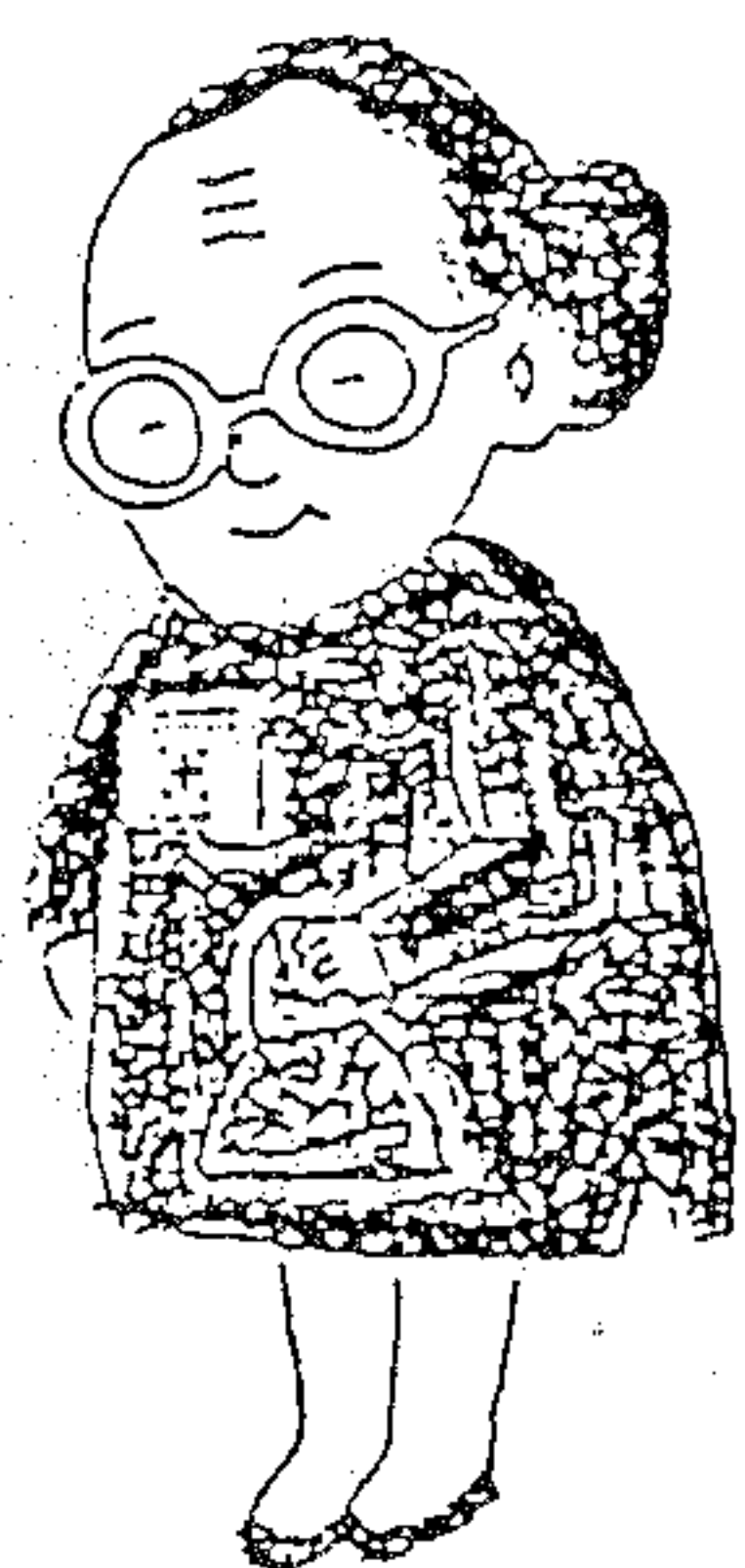
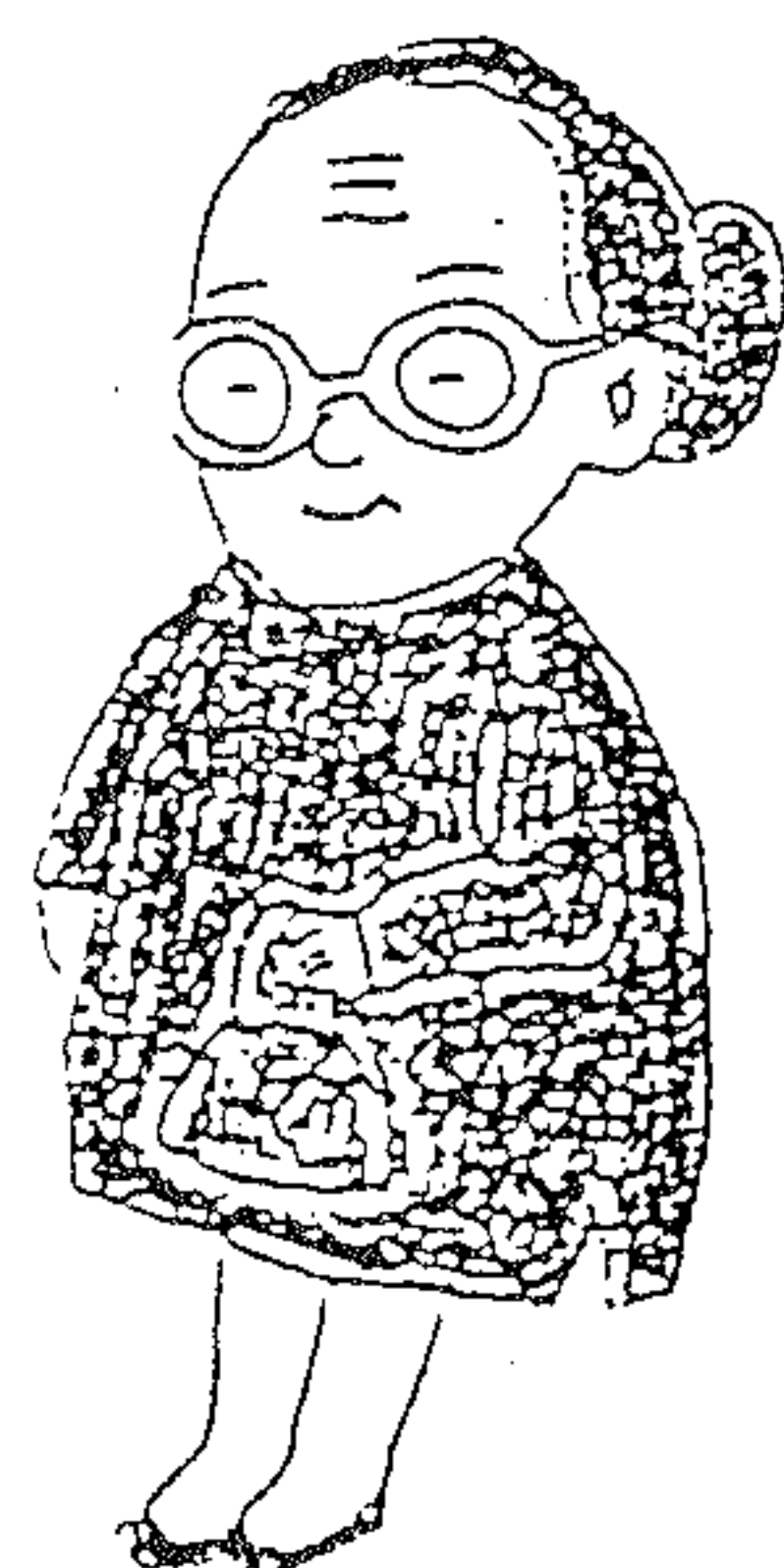
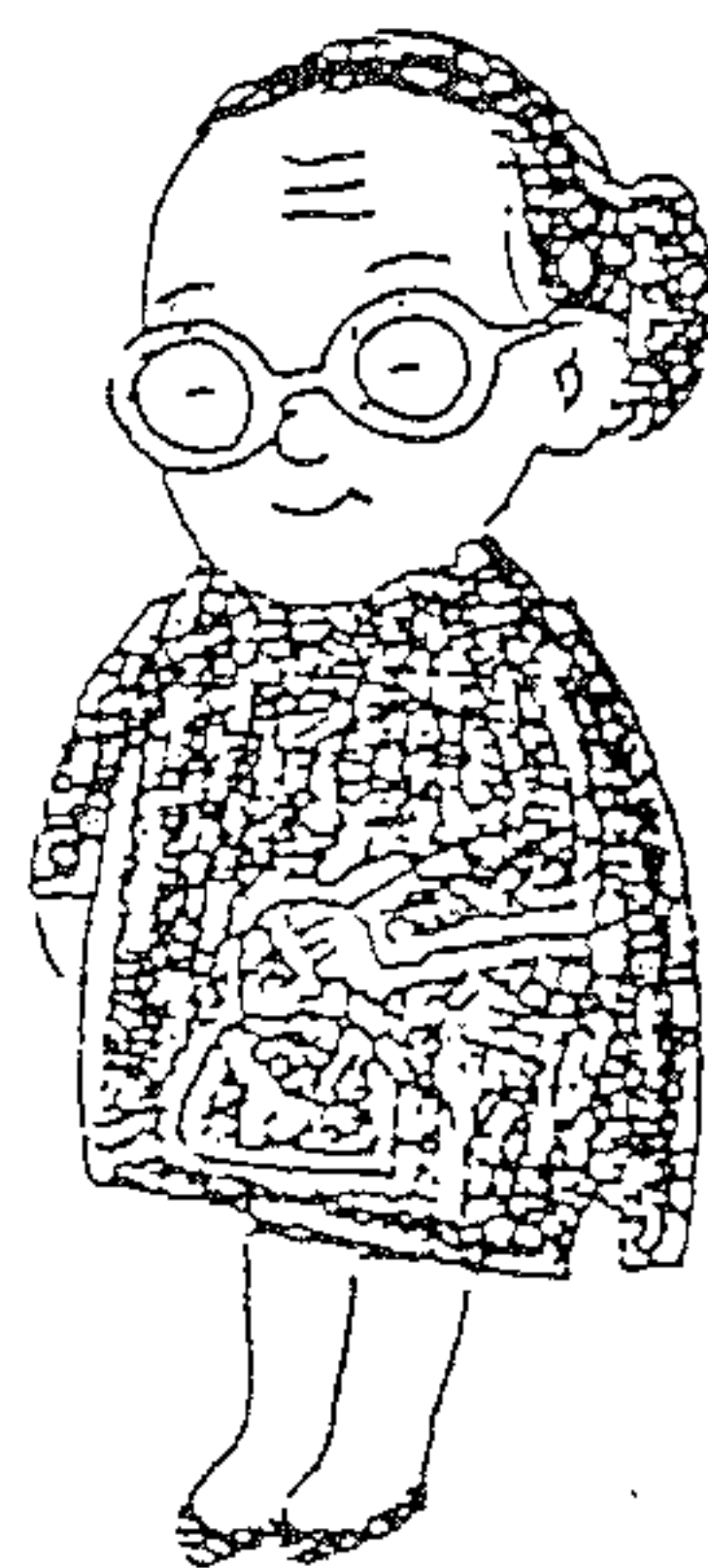
Siuming gamyat jeukdzo gin san saam chutgai
 today wear CL new clothing go-street
 gin saam soengbin yau sei go jeung
 CL clothing on-top have four CL button
 "Siuming is wearing a new outfit today and goes out.
 On his outfit are four buttons."



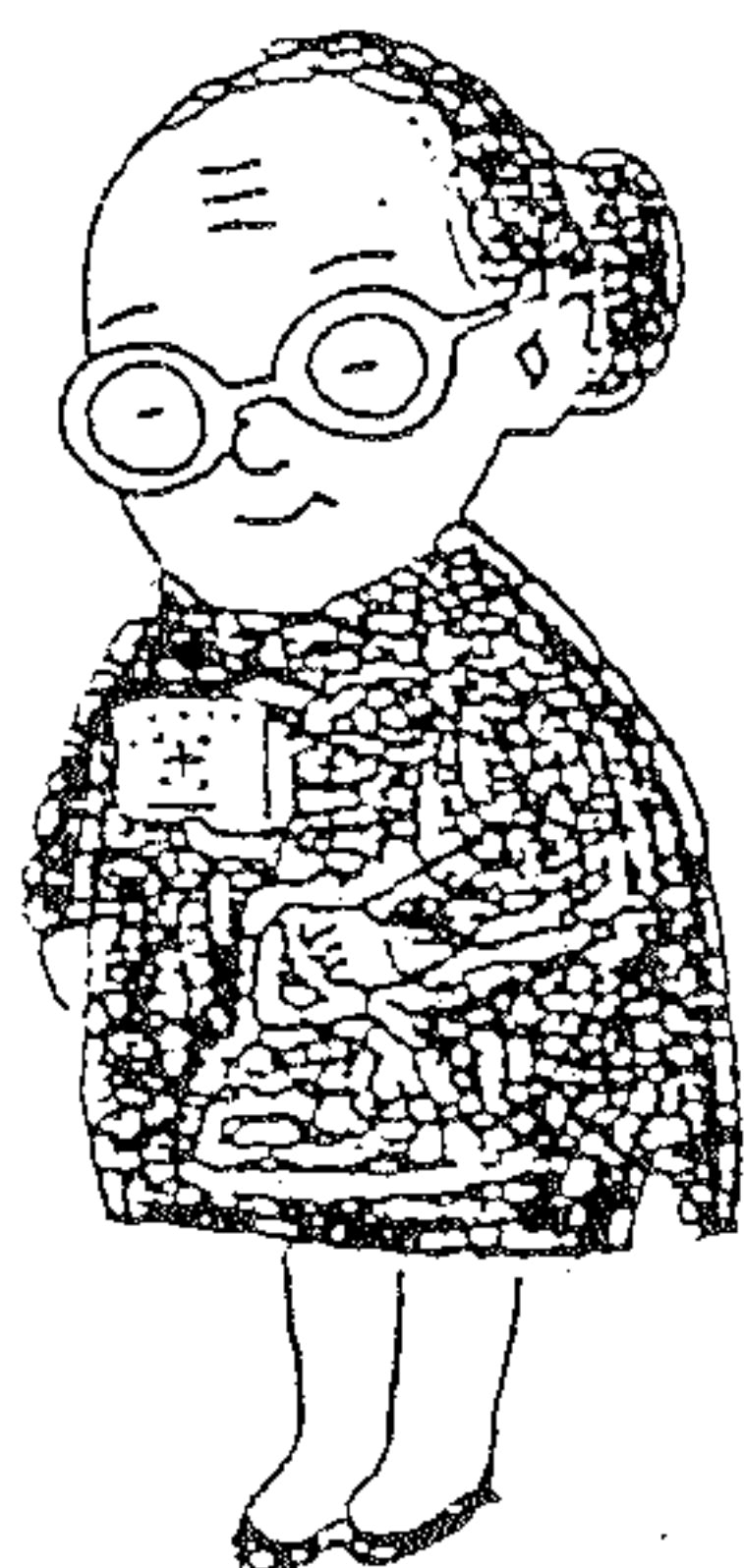
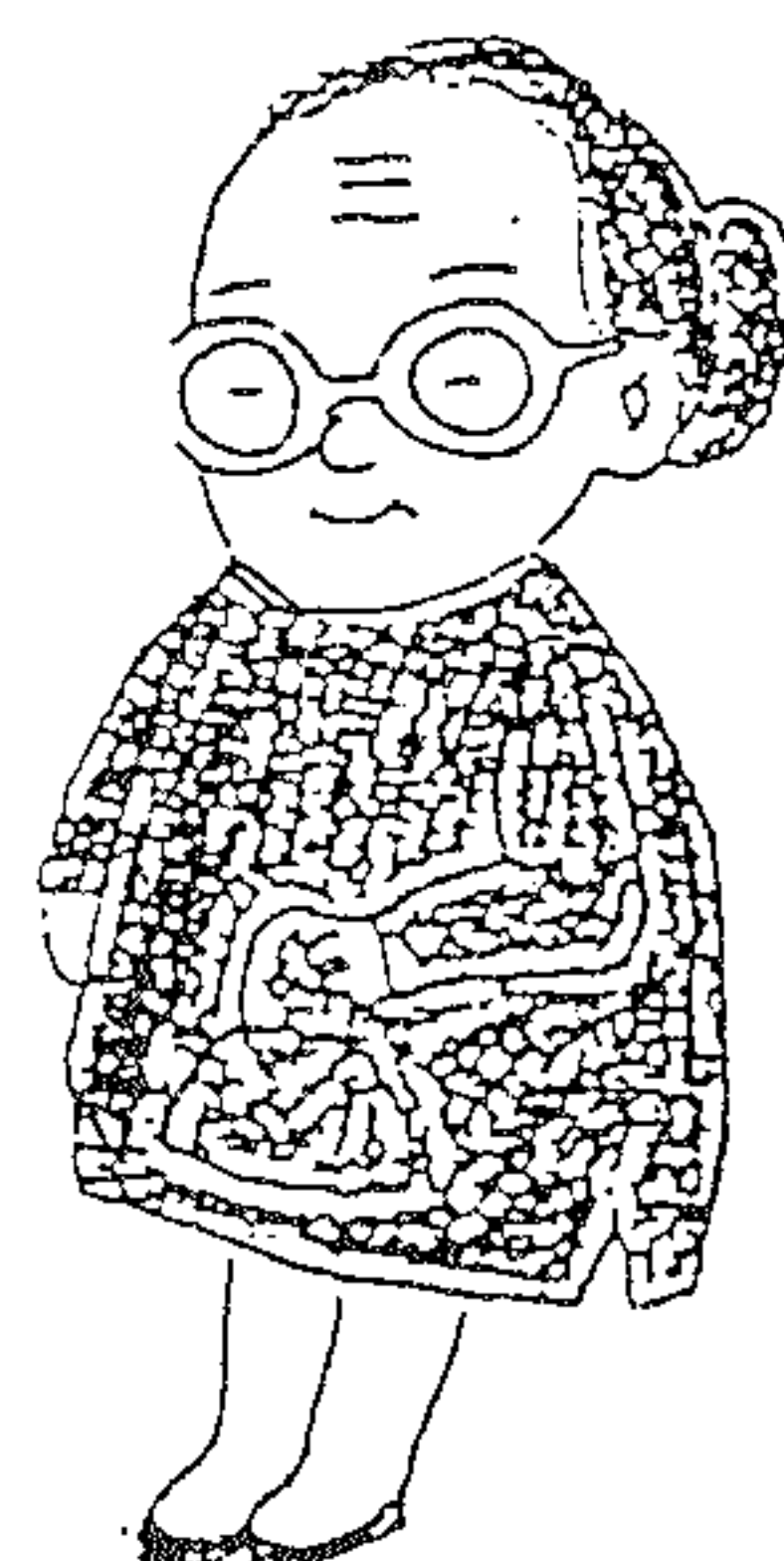
kui jeung keijung-yatgo jeung kau haai deng mou-dou
 he take one-of button pin on CL hat-LOC
 Siuming jau haang fan ukkei la,
 then walk back home SFP
 "He pins one of the buttons on top of his hat. Then he walks back home."



haang haang ha, go jeung lat dzo
walk walk ASP, CL button drop ASP
"As he is walking, the button drops".



haang haang ha, lat dzo go jeung
walk walk ASP, drop ASP CL button "the button drops"
"As he is walking, a button drops."



The stories were presented to the child subjects in two test sessions each containing 12 episodes. The members of a contrasting sentence pair do not co-occur in the same test session: if a member of a test pair appeared in one test session, the other member would appear in the other session. The children were tested individually. The ordering of the stories was the same for each child. In the first test session, four practice episodes were given to the child subjects to familiarize them with the procedures. Since the understanding of the singular cardinality of [CL-N] was considered essential, 6 pre-test items (3 in each test session) were given to the child subjects to see whether they would understand [CL-N]s indicate singular referents. Child subjects taking the test would have to pass all 6 pre-test items if their responses were to be included in the analysis.

4. Results on the comprehension experiment

The results of the comprehension experiment are shown in Table 3, classified according to whether the subject picked a prior-mention object or a first-mention object. Bearing in mind that young children may interpret definite NPs deictically, they may appear to be correct in giving a large number of prior-mention responses for NPs, irrespective of syntactic position. To see if they are sensitive to word order, one needs to check whether they respond differently to NPs in pre- and post-verbal positions.

(a) As can be seen from Table 3, children of all age groups on average gave more prior-mention responses to NPs not marked by the existential verb *jau*, than to NPs marked by *jau*. Table 3 shows that the difference was significant for the 5-year-olds and for the children as a whole. This difference was relatively independent of animacy. For animate NPs, the tendency for NPs not marked by the existential verb *jau* to receive more prior-mention responses than NPs marked by the existential verb reached significance by 5;1. For inanimate NPs it reached significance somewhat later, by 5;7.

(b) Table 3 also shows that with respect to ergative verb sentences, children of every age group gave more prior-mention responses on NP-Verg sentences than on Verg-NP sentences (8a vs 8b). This tendency began to be significant by 5;7 and was significant for the child subjects as a whole. This sensitivity was again true of both animate and inanimate referents.

(c) The figures indicate that the younger children, 3-year-olds in particular, did not sufficiently differentiate between preverbal and postverbal positions, since they gave a high proportion of prior-mention responses to both preverbal and postverbal NPs, on both existential-verb and ergative verb sentences.

(d) While children began to make systematic discriminations of word order in relation to definiteness by five and a half years old, their accuracy rates were surprisingly low, when compared with the figures in Maratsos' comprehension experiment on which this experiment was modelled (cf. Tables 2). In Maratsos' experiment, 3- and 4-year-olds were able to refer to a prior-mention object around 75% of the time; with the Cantonese children, the figure was only around 60% for the five and a half year olds (cf. 58% for existential verb sentences and 44% for ergative verb sentences).

Table 3: Cantonese children's interpretation of [CL-N] NPs in pre- and post-verbal positions.

Mean percentage of prior-mention choices on existential-verb sentences and sentences with ergative verbs

Age	jau-NP V (animate)	[]-NP V (animate)	jau-NP V (inanimate)	[]-NP V (inanimate)	jau-NP V	[]-NP V
3;7-4;0	63	75	75	71	69	73
4;1-4;6	33	46	42	63	38	54
4;7-5;0	4	21	50	50	27	35
5;1-5;6	29	58 **	58	75	44	67 *
5;7-6;0	25	54 *	25	63 *	25	58 **
adult	13	79 **	50	88 *	31	83 *
all children	31	51 **	50	64 *	40	58 **

Age	V-erg NP (animate)	NP-Verg (animate)	Verg-NP (inanimate)	NP-Verg (inanimate)	Verg-NP	NP-Verg
3;7-4;0	54	63	63	67	58	65
4;1-4;6	33	54	33	46	33	50
4;7-5;0	33	33	21	33	27	33
5;1-5;6	42	58	50	58	46	58
5;7-6;0	17	50 **	25	38	21	44 *
adult	29	63 ?*	58	92 *	44	77 **
all children	36	52 *	38	48	37	50 *

(* indicates that the difference represented by the adjacent pairs of means is significant at $p < .05$ level;
** indicates significance at $p < .01$ level; ?* indicates significance at $p < .07$ level)

Table 4. Interpretation of [CL-N] NPs in preverbal and postverbal positions. Number of correct choices on existential-verb sentences and sentences with ergative verbs (max=6)

Age	Subject	jau-NP V	[]-NP V	Verg NP	NP Verg
3;7-4;0	#19	2	4	2	5
	#17	0	5	2	3
	#18*	4	4	2	3
	#37	0	5	2	3
	#39	5	2	4	2
	#20	0	6	0	6
	#38	0	5	0	6
	#40*	3	4	5	3
4;1-4;6	#15*	1	4	3	5
	#13	6	1	5	1
	#14	2	5	2	3
	#16*	4	4	4	0
	#36	6	1	4	2
	#33	1	6	1	6
	#35*	5	2	5	4
	#34*	3	3	4	3

4;7-5;0	#11	4	1	6	0
	#10	4	2	3	1
	#29	4	2	6	1
	#30*	4	3	4	2
	#31*	3	3	2	4
	#9	4	2	5	2
	#12	5	1	4	0
	#32*	5	3	3	6
5;1-5;6	#7*	3	4	5	1
	#25	1	5	2	4
	#6*	4	5	2	5
	#5	2	5	0	4
	#27*	5	1	4	3
	#8*	4	4	4	5
	#26	1	6	1	5
	#28	5	2	5	1
5;7-6;0	#2*	5	3	6	0
	#3*	4	3	4	4
	#4*	4	5	5	3
	#21*	0	1	6	3
	#22*	4	3	2	2
	#23*	0	6	5	5
	#1	1	4	3	2
	#24*	5	3	5	2
adults	#51*	4	4	3	4
	#57*	5	6	5	4
	#58*	4	5	5	6
	#53	5	2	5	2
	#52*	4	5	4	4
	#55	2	6	0	6
	#56*	6	6	3	6
	#54*	3	6	2	5

* indicates individuals who obtained at least 3 correct choices on both existential-verb sentences or on both ergative-verb sentences.

(e) If one looks at individual performance what one sees is that the children's grasp of the definiteness/indefiniteness distinction is not straightforwardly correlated with age. Table 4 gives the number of correct choices for all the subjects. For the preverbal condition, the child was considered correct only if a prior-mention referent was picked. For the postverbal condition, the expected response was a different referent than the one that had been mentioned. If one adopts a lax criterion and considers individuals who have obtained at least half of the choices correct on either both existential-verb sentences or on both ergative verb sentences to have acquired the distinction. Then the ones marked with asterisks were the children and adults who have acquired the distinction. We can see that very few children before five and a half satisfy this criterion, and the development was not a linear increase.

(f) The last finding is concerned with the interaction between animacy and definiteness. Table 3 shows that for the existential verb sentences, both children and adults tended to give more first-mention responses when the referent was animate than when it was inanimate. This difference was significant for the four-and-a-half year-olds and the 5;1 year-olds and for the children as a whole. It was also marginally significant for adults. However, no such animacy effect was observed for the ergative verb sentences for the child subjects [cf. the difference was however significant for adults]. The percentage of first-mention responses for ergative verbs was more or less the same for animate as well as inanimate referents. Why should an animacy effect show up only for *jau*-sentences? It turned out that the animate referents in the *jau*-sentences are all human referents, while the animate referents in the ergative verb sentences are all nonhuman. This may reflect a sensitivity to the role of humanness in Chinese grammar. Faced with a sentence like "John walked near one of the boys. CL-boy left". there is a stronger probability of the child picking another referent, given that 'boy' is human. The child may have been operating under some Gricean principle. Normally, since there are pronominal forms for human objects but generally not for inanimate objects, one could have used a pronoun to refer back to the antecedent. The fact that a [CL-N] NP is used suggests therefore that the reference may not be the same. [cf. child language evidence for children's use of conversational maxims, e.g. the maxim of quantity, Surian and Job 1987]

Discussion

Recall that in terms of the rules of file change semantics given in (3) and (4), we should predict the comprehension act out task to be easy. All we need to do is translate 'definite NP' into preverbal position and 'indefinite NP' into postverbal position in the rules. If one hears a preverbal NP, one updates an old file card; if one hears a postverbal NP, one starts a new file card. So why is it that the children did not do well on this task? I'd like to suggest two factors that may account for this difference between English and Cantonese. First, unlike the contrast between indefinite and definite articles in English, the word order difference in Chinese only applies to restricted verb types, namely the existential verb and ergative verbs. For other verbs one will not get this minimal contrast, in that it is not generally the case that postverbal arguments have to be indefinite. In fact it has been shown that the correlation between postverbal position and indefiniteness is not particularly strong in discourse data (Sun and Givon 1985). They report that a high percentage of definite objects occur in postverbal position [cf. Wang Mingquan 1988 for a discussion of Sun and Givon.

Wang shows that 74% of the VO objects are indefinite, and 26% of the VO objects are definite]. Thus while it is true that preverbal NPs are subject to the definiteness restriction, there is only a tendency for postverbal objects to be indefinite. The fact that the contrast applies to restricted contexts and represents only a tendency in other contexts may make it harder to learn. The other factor derives from the use of word order instead of a determiner system. The range of meanings conveyed by the article system is much broader than that conveyed by word order. For instance, it seems to me that preverbal position does not have a deictic function, unlike the definite articles in English, which has weak deictic functions (cf. Lyons 1975, 1978). Suppose one wants to express the meaning 'pick up the ball' referring to a ball in the immediate context, one would use either the preverbal order or the postverbal order: *go bo lik heilei* 'cl-ball pick up' vs *lik hei go bo* 'pick up the ball'. Determiners can pick out individuals with descriptions as their complements; there is no comparable device for modification in word order, since word order does not represent a syntactic constituent. Determiners also denote genericity in complex ways, but genericity is clearly not part of the meaning of word order. So word order may be a pure device for representing familiarity vs novelty, without the kinds of additional semantic dimensions that one finds in English. But perhaps for this reason, the positive evidence for the word order distinction is more impoverished than that for the article contrast, which delays the learning process.

5. Conclusions

To conclude, I would like to reiterate that file change semantics has allowed us to formulate the rules for the use of indefinite and definite NPs more explicitly. I hope that I have shown this can help clarify the confusing acquisition literature on definiteness (cf. earlier review in Cziko 1986). It now emerges that instead of asking when definiteness is acquired, it may be more fruitful to ask when particular pragmatic rules for definite and indefinite reference are acquired. The findings from Cantonese indicate that the linguistic encodings for (in)definiteness is not acquired before five years of age. I have attributed the relative lateness of emergence of this distinction in Chinese to the way the distinction is marked in the language.

References

- Avrutin, S. (1994) Psycholinguistic Investigations in the Theory of Reference. Ph.D. dissertation. MIT.
- Bamberg, M. (1986) "A functional approach to the acquisition of anaphoric relationships", *Linguistics* 24: 227-284.
- Brown, R. (1973) *A First Language*. Cambridge, Mass: Harvard University Press.
- Burton-Roberts, N. (1981) "Review of J. Hawkins: Definiteness and Indefiniteness" in *Language* 57 (1): 191-196.
- Chang Hsin-wu and Tseng Ling-Yi (1988) Guoyu cixue dui lijie de yingxiang. [Word order and sentence comprehension in Mandarin]. Paper presented at the 1988 Conference on Chinese Linguistics, Taipei.
- Chao, Yuen-ren (1968) *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

- Chen Ping (1987) Xi Hanyu zhong yu mingci chengfen xiangguan de sizu gainian [On four sets of concepts related to nominal constituents in Chinese] *Zhongguo Yuwen* [Chinese Language] 197:81-92.
- Chesterman, A. (1991) On Definiteness. London: Cambridge University Press.
- Cheung, Hung-nin. (1972) *Xianggang Yueyu Yufa de Yanjiu* [Cantonese as spoken in Hong Kong]. Hong Kong: Chinese University of Hong Kong Press.
- Chien, Yu-chin and K. Wexler (1990) "Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics", *Language Acquisition* 1.3:225-295.
- Clark, H. and S. Haviland (1977) Comprehension and the given-new contract. *Discourse Production and Comprehension*. Norwood, NJ: Ablex.
- Clark, H. and C. Marshall (1981) Definite reference and mutual knowledge. *Elements of Discourse Understanding*. ed. by A. Joshi, B. Webber, and I. Sag. London: Cambridge University Press.
- Cziko, G. (1986) "Testing the language bioprogram hypothesis: a review of children's acquisition of articles", *Language* 62.4: 878-898.
- Declerck, R. (1986) "Two notes on the theory of definiteness", *Journal of linguistics* 22 (1): 25-39.
- Declerck, R. (1987) "Definiteness and inclusive reference" *Journal of Literary Semantics* 16:12-29.
- Emslie, H. and R. Stevenson (1981) Preschool children's use of the articles in definite and indefinite referring expressions. *Journal of Child Language* 8: 313-328.
- Erbaugh, Mary. 1982. Coming to Order: natural selection and the origin of syntax in the Mandarin-speaking child. Doctoral dissertation, University of California at Berkeley.
- Garton, A. (1983) "An approach to the study of determiners in early language development", *Journal of Psycholinguistic Research* 12.5: 513-525.
- Grodzinsky, Y. and T. Reinhart (1993) "The innateness of binding and coreference", *Linguistic Inquiry* 24.1:69-101.
- Hawkins, J. (1978) Definiteness and indefiniteness. London: Croom Helm.
- Hawkins, J. (1991) "On (in)definite articles: implicatures and (un)grammaticality prediction" *Journal of Linguistics* 27:402-442.
- Heim, I.R. (1982) The semantics of definite and indefinite noun phrases. PhD dissertation, University of Massachusetts, Amherst.
- Heim, I. R. (1983) "File change semantics and the familiarity theory of definiteness", in R. Bauerle, C. Schwarze and A. von Stechow (eds) *Meaning, Use and Interpretation of Language*. New York: Walter de Gruyter.
- Hickmann, M. (1980) "Creating referents in discourse: a developmental analysis of linguistic cohesion", *Papers from the Parasession on Pronouns and Anaphora*. eds. by J. Kreiman and A. Ojeda. Chicago Linguistic Society.
- Hickmann, M. (1987) "The pragmatics of reference in child language: some issues in developmental theory", in M. Hickmann (ed) *Social and Functional Approaches to Language and Thought*. New York: Academic Press.
- Hickmann, M. and J. Liang (1990) Clause-structure variation in Chinese narrative discourse: a developmental analysis. *Linguistics* 28: 1167-1200.
- Huang, C-T J. (1987) Existential sentences in Chinese and (in)definiteness. *The Representation of (In)definiteness*. ed. by E. Reuland and A. ter Meulen. Cambridge, Mass: MIT Press.

- Hurford, J. and B. Heasley (1983). *Semantics: a coursebook*. London: Cambridge University Press.
- Kadmon, N. (1985) "The discourse representation of noun phrases with numeral determiners", *Proceedings of NELS* 15.
- Kadmon, N. (1990) "Uniqueness", *Linguistics and Philosophy* 13:273-324.
- Kail, M. and M. Hickmann (1992) "French children's ability to introduce referents in narratives as a function of mutual knowledge", *First Language* 12:73-94.
- Kamp, H. and U. Reyle (1993) *From Discourse to Logic*. Dordrecht: Kluwer. Student edition.
- Karmiloff-Smith, A. (1979) *A Functional Approach to Child Language*. London: Cambridge University Press.
- Karmiloff-Smith, A. (1981) The grammatical marking of thematic structure in the development of language production. *The Child's Construction of Language*. ed. by W. Deutsch. New York: Academic Press.
- Karmiloff-Smith, A. (1985) "Language and cognitive processes from a developmental perspective", *Language and Cognitive Processes* 1.1:61-85.
- Karttunen, L. (1976) "Discourse Referents" in J. McCawley (ed.), *Syntax and Semantics*, Vol. VII: 365-385. New York: Academic Press.
- Keenan, E. (1974) "Conversational competence in children" *Journal of Child Language* 1: 163-183.
- Keenan, E. and E. Klein (1975) "Coherency in children's discourse", *Journal of Psycholinguistic Research* 4.4:365-380.
- Kramsky, J. (1972) *The Article and the Concept of Definiteness in Language*. The Hague: Mouton.
- Lewis, D. (1979/1983) "Scorekeeping in a language game", in *Philosophical Papers*. vol. 1, London: Oxford University Press, pp. 233-249.
- Li, Charles and Sandra Thompson. 1981. *Mandarin-Chinese: a functional reference grammar*. Los Angeles: University of California Press.
- Lyons, J. (1975) "Deixis as the source of reference", in E. Keenan (ed) *Formal Semantics of Natural Language*. Cambridge University Press, reprinted in J. Lyons (1991). Lyons, J. (1978) "Deixis and anaphora" in T. Myers (ed) *The Development of Conversation and Discourse*. Edinburgh University Press. reprinted in J Lyons (1991).
- Lyons, J. (1991) *Natural Language and Universal Grammar*. Cambridge University Press.
- MacNamara, J. (1982) *Names for Things*. Cambridge, Mass: MIT Press.
- Maratsos, M. (1973) Non-egocentric communication abilities in preschool children. *Child Development* 44: 697-700.
- Maratsos, M. (1974) Preschool children's use of definite and indefinite articles. *Child Development* 45: 446-455.
- Maratsos, M. (1976) *The Use of Definite and Indefinite Reference in Young Children*. London: Cambridge University Press.
- Markman, E. (1989) *Categorization and Naming in Children*. Cambridge, Mass: MIT Press.
- Menig-Peterson, C. (1975) The modification of communicative behavior in preschool-aged children as a function of the listener's perspective. *Child Development* 46: 1015-18.
- Min, Rui-fang (1994) *The Acquisition of Referring Expressions by Young Chinese Children*. Ph.D. dissertation. The Catholic University of Nijmegen.

- Power, R. and M. Dal Martello (1986) The use of the definite and indefinite articles by Italian preschool children: *Journal of Child Language* 13: 145-154.
- Quine, W. (1960) *Word and Object*. Cambridge, Mass: MIT Press.
- Quine, W. (1973) *The Roots of Reference*. La Salle, Illinois: Open Court.
- Searle, J. (1969) *Speech Act*. London: Cambridge University Press.
- Sun, C-F and T. Givon (1985) On the so-called SOV word order in Mandarin Chinese: a quantified text study and its implications. *Language* 61.2: 329-351.
- Surian, L. and R. Job (1987) Children's use of conversational rules in a referential communication task. *Journal of Psycholinguistic Research* 16.4: 369-382.
- Thorne, J. P. (1987) "The indefinite article and the numeral one", in R. Steele and T. Threadgold (eds) *Language Topics: essays in honour of Michael Halliday*, vol. 2. Amsterdam: John Benjamins.
- Warden, D. (1976) "the influence of context on children's use of identifying expressions and references" *British Journal of Psychology* 67: 101-112.
- Warden, D. (1981) "Learning to identify referents", *British Journal of Psychology* 72:93-99.
- Wigglesworth, G. (1990) "Children's narrative acquisition: a study of some aspects of reference and anaphora", *First Language* 10: 105-125.
- Wang, Mingquan (1988) Comments on Sun and Givon's study of the OV constructions in Mandarin. *Journal of the Chinese Language Teachers' Association* 23.2: 33-53.
- Warden, D. (1976) The influence of context on children's use of identifying expressions and references. *British Journal of Psychology* 67: 101-112.
- Zhou Xinping (1990) *Aspects of Chinese Syntax: ergativity and phrase structure*. Doctoral dissertation, University of Illinois, Urbana-Champaign.

Cantonese-speaking Children's Comprehension of Wh-Questions*

Alice Shuk-ye Cheung and Thomas Hun-tak Lee

1. Introduction

Three issues have figured prominently in the study of interrogative acquisition. One issue that has been looked at in almost every study is the order of acquisition of argument and adjunct questions. The production data from Bloom, Merkin and Wootten (1982) and Stromswold (1988) indicate that *what*- and *who*-questions appear earlier than *how*- and *when*-questions. Stromswold further shows that for the same question type, the *wh*-phrase would occur in argument position earlier than in adjunct position. For example, *where*-argument questions such as *where is the book?* would occur earlier than *where*-adjunct questions like *where did Mary meet John?*. The relative superiority of argument questions over adjunct questions is also supported by comprehension data. Ervin-Tripp (1970), Tyack and Ingram (1977), and Cairns and Hsu (1978) all indicate that *what*- and/or *who*-questions are understood earlier than *when*- and *how*- questions. The literature also suggests that *where*-argument questions are among the earliest acquired.¹

The early acquisition of *what*-, *who*-, and *where*-argument questions in comparison to *when*- and *how*-questions is supported to varying degrees by naturalistic findings from other languages than English, in Tamil (Vaidyanathan 1988), Korean (Clancy 1989) and Chinese (Erbaugh 1982, Lee 1989 for Mandarin, and Cheung 1992 for Cantonese).

Secondly, previous studies have identified the syntactic position of the *wh*-phrase as a relevant factor in interrogative acquisition. Although claims have been made (Stromswold 1988) that object-questions are acquired significantly earlier than subject-questions, the data suggest a more controversial picture. Whether object-questions are easier than subject-questions depends on the question type. Both Stromswold (1988)

* This study was supported by UPGC research grant #CUHK/2/91. We wish to thank Colleen Wong, Sam Leung and participants of the forum for comments on earlier versions of the work. Thanks are also due to Ms. Leung Lai Ying of Lap Yan Kindergarten for her support for this project. Please address correspondence to T. Lee, Dept of English, Chinese University of Hong Kong, Shatin, Hong Kong.

¹ There is considerable difference between production and comprehension data with respect to *why*- questions. The production data from Stromswold (1988) and Bloom et al (1982) indicate that *why*-questions appeared after *what*- and *who*- questions. However, in both Ervin-Tripp (1970) and Tyack and Ingram (1977), *why*-questions were understood at the same time as *who*- subject questions. The latter report that *why*-questions were easier than *what*-questions.

and Tyack and Ingram (1977) agree that *what*-object questions are acquired earlier than *what*-subject questions. On the other hand, these studies as well as Ervin-Tripp (1970) show that *who*-subject questions are acquired at the same time as or earlier than *who*-object questions.

Thirdly, verb type has been shown to be relevant to the acquisition of interrogatives. Tyack and Ingram (1977) observed that the transitivity of the verb affects children's performance. It was found, for instance, that *where*-questions were understood better with intransitive verbs than with transitive verbs, and that certain error patterns were contingent on verb transitivity. Cairns & Hsu (1978) found that *who*-object questions with auxiliary *do* were easier than those with auxiliary *be*. Bloom et al (1982) discovered that while argument questions tended to cooccur with the copula and pro-verbs like *do*, adjunct questions tended to accompany verbs with more specific semantic content, e.g., *sing* or *draw*.

Different types of explanations have been proposed to account for the acquisition order observed. Cognitive proposals attribute the relative difficulty of adjunct questions to the complexity of concepts of cause, manner and time in comparison to concepts underlying names of people or things (cf. Tyack and Ingram 1977). As pointed out by Bloom et al (1982), children encode causation and temporal relations through other linguistic devices at a very early stage (e.g., in their use of conjunctions like *because* and *and then*). Thus the key to the order of difficulty of wh-questions must lie not in cognitive factors but in linguistic factors. They suggest that since adjunct questions go with semantically more specific verbs than argument questions, the former are linguistically more complex, and are therefore acquired later. As regards the difference between subject- and object- questions, both functional and formal accounts have been suggested. Ervin-Tripp (1970) argues that *who*-subject questions and *what*-object questions are easier, because subjects tend to be animate and objects tend to be inanimate. On the other hand, Stromswold (1988) appeals to the difference between lexical and antecedent government to explain the superiority of object-questions.

This cross-sectional study explores the comprehension of wh-questions by Cantonese-speaking children. It aims to find out (a) when Cantonese-speaking children comprehend the different types of wh-questions: *bin1go3* 'who'-; *mat1ye5* 'what'-; *bin1dou6* 'where'-; *dim2yeung2* 'how'-; *dim2gaai2* 'why'-; and *gei2si4* 'when'-questions;² (b) to what extent children's understanding of wh-questions depends on the actionality and transitivity of the predicate; (c) whether, in a language without syntactic wh-movement such as Cantonese, there would be evidence for a subject/object asymmetry in children's performance; (d) whether there would be an argument/adjunct difference for questions that allow the wh- phrase to be expressed as an argument or adjunct (viz. *bin1dou6* 'where'-questions).

2 The numeral following the romanized form of a Cantonese morpheme indicates the tone borne by the morpheme. Cantonese has six contrastive tones.

2. Materials and Procedure

The subjects were 30 Cantonese-speaking children, aged between 2;6 and 5;0, evenly distributed into 5 age groups: I 2;6-3;0 (mean=2;9), II 3;0-3;6 (mean=3;4), III 3;6-4;0 (mean=3;9) IV 4;0-4;6 (mean=4;3) and V 4;6-5;0 (mean=4;8). Each group consisted of 3 boys and 3 girls.

The experiment included 4 verb types, with 3 verbs chosen for each type, giving a total of 12 verbs. They were transitive actional verbs, intransitive actional verbs, transitive stative verbs, and functional verbs including the copula *hai6* 'be', the existential verb *yau5* 'exist/have', and the locative verb *hai2* 'be-in/at'. Intransitive verbs are defined as verbs that do not subcategorize for patient NPs.³ Transitive stative verbs are distinguished from transitive actional verbs in that the former cannot occur in imperative sentences, while the latter can. The test questions were based on 12 short stories, each centered on one verb. Because of the verb types chosen, it was not possible to ask the full range of argument and adjunct questions for each verb type. The stories were therefore not uniform in length, ranging from 31 to 60 morphemes, and with a mean length of 48 morphemes. Each story was depicted with two pictures which illustrated clearly the action or spatial dimension encoded by the verb. The stories were audiotaped. The test contained 73 questions in total, with a mean of 18 questions for each verb type. Wh-phrases were included in subject and object positions for 'who'- and 'what'-questions, and in subject, post-verbal argument, and adjunct positions for 'where'-questions.⁴

Each subject was tested individually by two researchers in three separate sessions. The child was first presented a set of two pictures depicting a story, and was asked to identify the participants, location and actions in the pictures. Then the child listened to a story played from a tape recorder. As s/he listened to the story, a researcher pointed to relevant parts of the pictures. At the end of the story, the pictures were put away and the child was asked a set of questions related to the story. The answers were recorded by hand and also audiotaped. The subject was then tested with the next story. The stories were arranged so that stories drawn from the same verb type were not adjacent to each other. The sequence of the types of questions asked within a story was randomized. All the subjects listened to the stories and answered the questions in the same sequence.

3. Results

³ Intransitive verbs in Chinese are defined with restrictive criteria, since most verbs can be followed by locative and cognate objects. Transitivity is generally based on the type of object that can immediately follow a verb (cf. Chao 1968).

⁴ Examples of the three types of 'where'-questions are: *bin1dou6 yau5 mou4mou1gau2 a?* 'where-exist/have-puppy?' ('where'-subject); *jeuk3jai2 fei1 heui3 bin1dou6 a?* 'bird-fly-to-where?' ('where'-postverbal argument); *jeuk3jai2 hai2 bin1dou6 fei1 a?* 'bird-at-where-fly' ('where'-adjunct).

Order of acquisition. Table 1 shows the percentage of correct responses given by the different age groups to various types of questions. Responses on the same type of wh-phrase were pooled together irrespective of the verb type contained in the question. The point of acquisition is taken as the point at which 70% accuracy is consistently maintained from there on. As can be seen from the figures, *bin1go3* 'who'- and *mat1ye5* 'what'-questions were understood by Age Group II, and *dim2gaai2* 'why'- and *bin1dou6* 'where'-questions by Age Group III. *Dim2yeung2* 'how'-questions were understood by Age Group V. *Gei2si4* 'when'-questions were not yet understood even by the oldest children. The order of acquisition is 'who-/what'-<'why-/where'-<'how'-<'when'-questions.

Table 2 gives the breakdown of the percentage of correct responses according to the four categories of verbs. As can be seen, some question types which were acquired at a certain point if all verb types were pooled together were understood earlier for some of the verb types. For example, 'what'-questions were acquired as early as Group I and 'where'-questions by Group II for transitive actional verbs. 'How'-questions with intransitive actional verbs were understood by Group IV, and 'when'-questions with functional verbs by Group V.

Table 1: Cantonese-speaking children's comprehension of different question types (% of correct responses)

Question Type	Gp I	Gp II	Gp III	Gp IV	Gp V
<i>bin1go3</i> 'who'-	60	81	84	87	90
<i>mat1ye5</i> 'what'-	42	71	81	76	85
<i>bin1dou6</i> 'where'-	42	58	73	92	86
<i>dim2gaai2</i> 'why'-	15	39	79	89	94
<i>dim2yeung2</i> 'how'-	52	60	71	67	81
<i>gei2si4</i> 'when'-	4	18	29	28	58

(GpI: 2;6-3;0 GpII: 3;0-3;6 GpIII: 3;6-4;0 Gp IV: 4;0-4;6 Gp V: 4;6-5;0)

Table 2: Cantonese-speaking children's comprehension of questions containing different verb types (% of correct responses)

Q.Type	Transitive actional verbs						Transitive stative verbs					
	Gp I	Gp II	Gp III	Gp IV	Gp V	Mean	Gp I	Gp II	Gp III	Gp IV	Gp V	Mean
Wha	75	83	75	75	75	77	28	39	67	67	72	54
t-												
Who	67	77	83	87	93	81	67	83	100	96	96	88
-												
Whe	50	89	94	100	100	86	50	42	63	100	88	68
re-												
Why	22	50	100	94	94	72	5	22	67	89	94	56
-												
Ho	67	67	83	56	83	71	17	17	50	33	33	30
w-												
Whe	0	11	39	28	61	28	5	28	17	22	44	23
n-												

Q.Type	Intransitive actional verbs						Functional verbs					
	Gp I	Gp II	Gp III	Gp IV	Gp V	Mean	Gp I	Gp II	Gp III	Gp IV	Gp V	Mean
Wha	50	83	100	75	92	80	33	80	83	83	93	75
t-												
Who	58	92	58	83	92	77	46	79	83	79	79	73
-												
Whe	39	67	83	92	81	72	28	33	44	72	83	52
re-												
Why	11	44	78	94	94	64	25	42	67	75	92	60
-												
Ho	50	67	67	72	94	70	-	-	-	-	-	-
w-												
Whe	5	17	28	22	50	24	6	17	33	39	78	34
n-												

(Q.Type: Question Type GpI 2;6-3;0 GpII: 3;0-3;6 GpIII 3;6-4;0 GpIV 4;0-4;6 GpV 4;6-5;0; '-' marks categories for which there were no test items)

Thus, if children are considered to have understood a question type so long as they reach criterion on any verb type, the order of acquisition is 'what'-'who'-'where'-'why'-'how'-'when'-questions. This is similar to the order found in comprehension data in English in that 'what/who' questions are acquired before 'how/when' questions.

Verb type. Children's performance on the same question type differs according to the nature of the verb used. As can be seen from Table 2, questions containing

transitive actional verbs were among the earliest to be comprehended by the children. 'What'- and 'where'-questions both first reached criterion in transitive actional verbs. 'Why'-questions first reached criterion in two categories, one of which was transitive actional verbs.

A one-way ANOVA was carried out with each subject's percentage-correct score as the dependent variable and verb type as the repeated-measures independent variable. It was found that verb type was a significant factor ($F(3, 596)=3.2, p<0.05$). The tukey test revealed that across the sample as a whole, performance on transitive actional verbs was significantly better than on transitive stative verbs (mean 70.0 vs 58.0 ; $p<0.05$). To find out how the relative difficulty of question types depends on the verb contained in the question, a one-way ANOVA was carried out for each question type. Verb type was a significant factor for 'what'-, 'how'-, and 'where'-questions ($p<0.005$) but not for 'who'-, 'why'- or 'when'-questions. For 'what'-questions, performance on transitive stative verbs was significantly worse than that on the other verb types. Similarly, the results on 'how'-questions reveal that transitive stative verbs were significantly more difficult than transitive actional and intransitive actional verbs. As for 'where'-questions, comprehension was significantly poorer on functional verbs than on transitive actional and intransitive actional verbs. The effect of verb type on children's comprehension of questions across the various age groups is illustrated by Figures 1-3.

Subject-questions vs object-questions. Apart from the variation in performance due to verb type, children's comprehension also differed according to the syntactic positions of the wh-phrases. A significant difference was observed with respect to overall results on 'who'-subject-questions and 'who'-object-questions (76 vs 89; $p<0.05$). The superiority of 'who'-object questions over 'who'-subject questions was significant for transitive actional verbs (90 vs 76; $p<0.05$), and marginally significant for transitive stative verbs (93 vs 83; $p=0.056$). The subject/object distinction can also be observed in 'what'-questions: children understood 'what'-object questions significantly better than 'what'-subject questions for functional verbs (83 vs 69; $p<0.01$).

Argument vs adjunct questions. As reported above, adjunct questions such as *dim2yeung2* 'how'- and *gei2si4* 'when'- questions were in general comprehended later than argument questions such as *bin1go3* 'who'- or *mat1ye5* 'what'- questions. The argument/adjunct distinction came out differently when the question type was controlled. As can be seen from Table 3, children's overall comprehension of 'where'-adjunct questions was significantly better than that of 'where'-postverbal-argument questions (83 vs 57). To see if this is a genuine difference, one would need to focus on the intransitive actional verb category, in which a minimal contrast can be found between adjunct and postverbal-argument positions for each verb.

Table 3: Cantonese children's comprehension of *bin1go3* 'who'-, *mat1ye5* 'what'-, and *bin1dou6* 'where'-questions according to verb type and syntactic position (mean percentage of correct responses for each category)

Type Verb	Question	'Who'		'What'		'Where'	
		S	O	S O		S Ad	Pv
trans. actional		76		-		-	-
		90*		77		87	
<i>tek3</i> 'kick'		73		-		-	-
		93		57		80	
<i>da2</i> 'hit'		63		-		-	-
		87		-		90	
<i>chit3</i> 'cut'		90		-		-	-
		-		97		90	
intr. actional		77		80		-	68
		-		-		77	
<i>tiu3</i> 'jump'		80		-		-	77
		-		-		67	
<i>pa4</i> 'crawl'		73		70		-	70
		-		-		87	
<i>feil</i> 'fly'		-		90		-	57
		-		-		77	
trans. stative		83		-		50	53
		93*?		54		85	
<i>nam2jy6</i> 'have-in-mind'		80		-		-	-
		90		40		80	
<i>geng1</i> 'fear'		87		-		-	-
		97		33		90	
<i>joi3jy6</i> 'hold'		-	-	-		50	53
				90		-	
functional		71		69		65	27
		80		83*		-	
<i>yau5</i> 'exist/have'		93		80		73	-
		-		87		-	
<i>hai2</i> 'be-at/in'		50		37		-	27
		-		-		-	
<i>hai6</i> 'be'		70		90		57	-
		80		80		-	
All verbs		76		73		60	57
		89*		69		83*	

(trans.=transitive; intr.=intransitive; S=subject; O=object; Pv=postverbal argument; Ad=adjunct; '-' marks categories for which there were no test items; '*' indicates the adjacent pair of means represents a statistically significant difference, $p<0.05$; '*?' represents a difference level of $p<0.06$).

The table shows that children understood 'where'-postverbal argument questions better than 'where'-adjunct questions for the verb *tiu3* 'jump'. However, the tendency was in the other direction for the verbs *pa4* 'crawl' and *feil* 'fly'. Overall results for intransitive actional verbs were higher on 'where'-adjunct questions than on 'where'-postverbal-argument-questions (77 vs 68). This difference was, however, only marginally significant ($p=0.073$), and was not significant for any of the age groups. Our findings should therefore be taken to reflect the absence of a systematic argument/adjunct distinction.

4. Discussion

As reported above, in Cantonese as in English, 'what'- and 'who'-questions are acquired earlier than 'how'- and 'when'-questions. This result holds independent of verb type, since 'who'- and 'what'-questions were comprehended better than 'when'-questions for each verb type and at every age level (cf. Table 2). Further, 'who'- and 'what'-questions were better understood than 'how' questions for transitive stative verbs across all ages. Thus, at least within certain verb types, performance on 'who' and 'what' questions generally surpassed that on 'how' and 'when' questions. This means that the relative difficulty of adjunct questions should not be attributed to their tendency to cooccur with more descriptive verbs than argument questions, contrary to the proposal of Bloom et al (1982). The extremely poor performance on 'when'-questions points to the complexity of specifying temporal locations for 3-to-5-year-olds.

With respect to the influence of verb type on the comprehension of wh-questions, two major findings deserve attention. First, it is generally not the case that questions containing transitive actional verbs were significantly easier than those containing the other verb types. Results on transitive actional verbs were only significantly better than those on transitive stative verbs. As will be seen, this relative superiority of transitive actional verbs could be ascribed to the children's generally poor performance on stative verbs. Such an advantage was not observed when comparison was made with intransitive actional verbs or functional verbs. The high prototypicality of transitive actional predicates does not always make it the most accessible category for children in interrogative acquisition.

Secondly, the relative difficulty of a verb type depends crucially on the question type involved. Stative verbs were significantly more difficult than some or all of the other verb types for 'what'-and 'how'-questions. In addition, functional verbs were significantly more difficult than actional verbs for 'where'-questions. This pattern was found for all but the oldest age group (see Figures 1-3). 'How'-questions containing stative verbs were difficult because they asked for manner, which is a property typical of actions and not states. These questions had an average accuracy rate of only 30% (cf. Table 2). 'What'-questions containing stative verbs posed a problem because 2 out of the 3 items in this verb category were object-questions in which a clausal complement was questioned (cf. Table 3). For example, *Siu2ming4 geng1 mat1ye5 a?*

'Siuming fear what?' is based on a sentence in the story: 'Siu2ming4 fear [teacher punish]'. Less than half of the subjects' responses were correct on these two test items. 'Where'-questions containing functional verbs showed a low overall accuracy rate, because of the poor scores on the locative verb *hai2* 'be-at/in' and the copula *hai6* 'be'. Less than one-third of the responses to *gung1yn2 hai2 bin1dou6 a* 'park be-at/in where' (where is the park located?), and only slightly over half of the answers to *bin1dou6 hai6 seung2tong4 ge3 dei6fong1 a?* 'where be hold-class nominalizer place' (where is the place to hold classes?) were correct. The difficulty of these questions may arise from their querying the location of a location (i.e. the park or the place to hold classes) rather than that of a theme-object.

The data have established firmly that 'who'-object questions were significantly easier than 'who'-subject questions. Likewise, 'what'-object questions were significantly easier than 'what'-subject questions for functional verbs. Table 3 shows clearly that with virtually all verbs that had both subject and object test items, object questions were better comprehended than subject questions. That this psycholinguistic asymmetry was observed in a language without wh-movement suggests that it is not related to wh-movement at the syntactic level.⁵ Since both 'who'- and 'what' questions queried the same types of entities (human entities in the former and inanimate entities in the latter) in subject and object positions, it would also be implausible to link the subject/object difference to the tendency for subjects to be animate and objects to be inanimate (cf. Ervin-Tripp 1970).

Why should 'what'-object questions be more difficult than 'what'-subject questions for functional verbs (cf. Table 3)? This difficulty should be understood in terms of the presuppositional complexity of wh-subject questions containing the locative verb. There are two ways to ask about the location of a theme-object in Chinese. One is to use a presentative order, 'location-existential verb-theme' (e.g. *hap2yap6min6 yau2 mat1ye5?* 'box-in exist/have what?'). The alternative structure is 'theme-locative verb-location' (e.g. *mat1ye5 hai2 hap2yap6min6?* 'what-be at/in-location'). The two questions differ in their presuppositional status. In declarative sentences containing the existential verb *yau2*, the NPs following the verb must be indefinite; a question such as the above example querying the object of the verb does not presuppose the existence of any entities at the location specified. On the other hand, since declarative sentences containing the locative verb have the theme-NP in subject position, the theme-subject must be definite following a general constraint on subjects in Chinese (cf. Chao 1968, Li and Thompson 1981). A subject-question corresponding to this type of sentence must presuppose the existence of entities that satisfy the description of the predicate. In other words, it would be odd for speakers to use the locative-verb structure if they do not already assume that some things are at the location. The question merely served as a request to identify the referents. The presuppositional complexity of wh-subject

⁵ One may also argue that the subject/object asymmetry stems from differences in how the subject and object traces are governed in Logical Form representations. The data from this study are not rich enough to confirm or refute this hypothesis.

questions containing the locative verb is reflected by the fact that for this verb, only 50% or less of the responses to 'who' and 'what' questions were correct.

It is unclear why 'who'-object questions should be generally easier than 'who'-subject questions. If one conceives of the task of answering a 'who'-question as identifying a set of referents that satisfy criteria mentioned in the rest of the sentence, the difference between subject and object questions in Chinese reduces to whether the criteria for identification of referents are mentioned before or after the query. It might be suggested that the processing load is lighter if the criteria for selecting referents are mentioned before the query is made. This type of explanation, however, has serious limitations. It will not be able to explain the subject/object asymmetry observed in languages with wh-movement, since in these languages all wh-questions have the wh-phrase situated in sentence-initial position, before the criteria for identifying referents are given.

With regard to the argument/adjunct distinction, our findings do not reveal a significant difference between 'where'-argument and 'where'-adjunct questions. All other things being equal, Cantonese children did not find adjunct questions more difficult than argument questions. This runs counter to Stromswold's finding and earlier reports of Cheung (1990) suggesting that Cantonese two-year-olds actually used postverbal argument NPs to signal static location normally expressed by preverbal adjuncts. It also deviates from the allative bias reported for young children (cf. Freeman, Sinha and Stedmon 1981). One factor that may account for the difference between our study and earlier studies is that our subjects, primarily in the 3- and 4-year-old range, had become sensitive to the fact that Cantonese is a head-final language, in which the principal branching direction is leftward, with adjuncts in preverbal position.

To summarize, this study has highlighted the important role of verb type in interrogative acquisition. The relative accessibility of a question depends on the nature of the verb it contains. Questions with transitive actional verbs were not generally easier than those with other verbs. Questions containing stative verbs were difficult because of the semantic properties of specific sentence types. Object questions were comprehended better than subject questions. This asymmetry was partially explained by the presuppositional differences between questions involving the existential and locative verbs. No systematic argument/adjunct difference was found for the same question type.

References

- Bloom, Lois, Susan Merkin and Janet Wootten. 1982. Wh-Questions: linguistic factors that contribute to the sequence of acquisition. *Child Development* 53:1084-92.
- Cairns, Helen and Jennifer Hsu. 1978. Who, Why, When and How: a developmental study. *Journal of Child Language* 5:477-488.

- Chao, Yuen-Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Cheung, Shuk-yee A. 1992. The Acquisition of Interrogatives by Cantonese-speaking Children. Paper presented at the 1992 Linguistic Society of Hong Kong Annual Research Forum.
- Cheung, Sik-lee. 1990. The Acquisition of Locative Constructions in Cantonese Children. *Papers and Reports in Child Language Development* 29:20-27.
- Clancy, Patricia. 1989. Form and Function in the Acquisition of Korean Wh-Questions. *Journal of Child Language* 16:323-347.
- Erbaugh, Mary. 1982. Coming to Order: natural selection and the origin of syntax in the Mandarin-speaking child. Doctoral dissertation, University of California at Berkeley.
- Ervin-Tripp, Susan. 1970. Discourse Agreement: how children answer questions. *Cognition and the Development of Language*. ed. J.R. Hayes. New York: Wiley.
- Freeman, N., C. Sinha and J. Stedmon. 1981. The Allative Bias in Three-year-olds Is Almost Proof Against Task Naturalness. *Journal of Child Language* 8: 283-296.
- Lee, Hun-tak T. 1989. Development of a Mandarin-speaking Child's Comprehension of Wh-Questions. *Cahiers de Linguistique Asie Orientale* 18.1: 29-62.
- Li, Charles and Sandra Thompson. 1981. *Mandarin-Chinese: a functional reference grammar*. Los Angeles: University of California Press.
- Stromswold, Karin 1988. Linguistic Representations of Children's Wh-Questions. *Papers and Reports in Child Language Development* 27:107-114.
- Tyack, Dorothy and David Ingram. 1977. Children's Production and Comprehension of Wh-Questions. *Journal of Child Language* 4:211-224.
- Vaidyanathan, R. 1988. Development of Forms and Functions of Interrogatives in Children: a longitudinal study in Tamil. *Journal of Child Language* 15:533-549.

**Part C: Commentaries on the Hong Kong Cantonese
Child Language Corpus**

A Cantonese Child Language Corpus

Thomas Hun-tak Lee

Cantonese differs from familiar European languages such as English or French in interesting ways. It is well-known that the grammar of Cantonese permits the dropping of subjects and objects in both main and subordinate clauses; such omission is severely restricted in English or French. It is also common knowledge that Cantonese is a classifier language, in which the enumeration of objects would require a particular class of words called classifiers. Unlike English or French, Cantonese encodes the indefinite/definite distinction by means of word order instead of a system of articles. The way nouns, verbs and adjectives are patterned is also different in these languages.

How do children acquire the grammar of Cantonese? What are the characteristics of their early language? For the past two years, a team of researchers from several tertiary institutions have been looking at these questions.¹ A major portion of the work consists in observing the longitudinal development of children from around one-and-a-half to approximately three years of age. The naturalistic data collected for analysis contain two hundred hours of audio recording of child-adult conversations. In addition, cross-sectional experiments were carried out with three-to-five year-olds to supplement the longitudinal work.

Two broad themes underlie the goal of the research. We are interested in when children become sensitive to the salient properties of Cantonese grammar. At what point are they aware that Cantonese allows zero pronouns in subject and object position? When do children begin to understand that the classifier is an integral part of the noun phrase? At what age do they signal definiteness using the position before the verb in the sentence? Answers to these questions will inform us about the stages of normal language development. Further, we wish to see how children's utterances in Cantonese are similar to and different from those in other languages. Does the language of Cantonese two-to-three year-olds resemble that of English-speaking children of the same age range? Do they differ in systematic ways that reflect the divergent properties of Cantonese and English? Results on the latter type of issue are relevant to considerations of the role of innate universals and the language environment in shaping the grammar in the child's mind.

In the course of our investigation, we have found that, for instance, Cantonese-speaking children drop subjects just as their English counterparts do. However, unlike English-speaking children, the rate of subject drop does not decline as the children grow older (cf. Man 1992). In relation to classifiers, we have observed that Cantonese-speaking children by three years of age have developed a basic grasp of the syntax of classifiers. They realize that if a numeral occurs, then a classifier must be used before a noun. While they may be unsure about which classifier to use with which noun, the obligatory presence of the classifier is clear to them. They also know that not more than one classifier should be used. Thus errors such as **saam1 jan4* 'three person' or **saam1 go3 zek3 jan4* 'three classifier classifier person' are extremely rare.² Our study has also found that the pragmatic ability to distinguish between definite and indefinite

reference in terms of word order is a late acquisition: it is not firmly established until after five years of age. These findings have given us an initial understanding of the acquisition of Cantonese grammar.

The generalizations one can glean from child language data are of interest to both the field worker and the theorist. Observing patterns emerge from the data and seeing one's hypotheses confirmed is doubtless a rewarding experience. Before one arrives at this stage, however, one must go through an extremely laborious process of data collection, transcription and input. First, the researcher must interact with the child at regular intervals. Getting to know the child and the parents, especially if they are not one's relatives, requires patience and persistence. Eliciting speech from the two-year-old in their one-word stage, depending on the temperament of the child, could also be taxing. The audio recordings of the conversational interaction need to be transcribed, which may take up to 15 hours per hour of tape recording. Only then will one have the raw data for preliminary analysis. The written transcripts do not lend themselves to easy tabulation or statistical calculations. Traditionally, one would have to make a large number of note cards to record occurrences of a particular linguistic feature. The note cards used by one researcher for a certain purpose may not be relevant to, or may not be available to, another researcher working on a different problem.

With the widespread use of personal computers, data analysis can be facilitated if the transcriptions are typed in computer files according to some kind of coding convention. Once such a corpus is created, text-search software can be designed to do string extraction or statistical calculation. If these computerized data can be accessed freely by scholars of the international community, it will mean a rapid multiplication of the database available for theorization. The use of a commonly agreed coding convention will also make it possible for scholars to refer to data from languages other than the one they are working on for comparative purposes.

Child language corpora from different languages have been made available since the mid eighties due to the effort of the Child Language Data Exchange System (CHILDES) project headed by Brian MacWhinney and Catherine Snow at Carnegie Mellon and Harvard respectively (cf. MacWhinney and Snow 1985, MacWhinney 1991). Currently the data one could access through file transfer protocol at Carnegie Mellon University include 14 languages ranging from Indo-European languages such as English, French, German, Swedish to non-Indo-European languages such as Hebrew, Turkish and Hungarian. English is the most richly represented language in the database, with around 30 corpora contributed by various researchers. (1) shows the beginning of a text file recording the speech of one of the celebrated children studied by Roger Brown and his colleagues at Harvard (cf. Brown 1973):

Begin

Coding: September 1987

Participants: EVE Eve Child, MOT Sue Mother, COL Colin Investigator, RIC
Richard Investigator

Sex of EVE: Female

Age of EVE: 1;6.0

Date: 15-OCT-1962

Time Duration: 10:00-11:00

*EVE: more cookie.

%spa: \$IMP
 %int: distinctive, loud
 *MOT: you xxx more cookie?
 *MOT: how about another graham cracker?
 *MOT: would that do just as well?
 *MOT: here.
 *MOT: here you go.
 *EVE: more cookie.
 %spa: \$IMP
 %int: distinctive, loud
 *MOT: you have another cookie right on the table.
 *EVE: more juice?
 *MOT: more juice?
 *MOT: would you like more grape juice?
 *MOT: where's your cup?
 *MOT: oh # I took it.
 %com: Fraser mentioned twice in conversation between Colin and Mother
 *EVE: Fraser.
 %spa: \$IMT

The file header gives information about the child, such as the name, sex, and age of the child, the time and date of the recording, as well as the identity of the participants. Each line beginning with an asterisk indicates an utterance by either the child or the adult. In addition, there are lines that describe special features of an utterance, such as whether it is an imperative speech act, or an imitation, or whether the intonation is distinctive. The Harvard (Brown) corpus comprises a set of around 200 files coded in such a format.

Accompanying the CHILDES data is a set of software one could retrieve from the archive for data analysis. Depending on the level of detail the data are coded, the programs will allow one to calculate frequencies of words and word classes, type/token ratios, and mean length of utterance. With these programs, one could also do keyword-in-context search, detect repetition of words in successive utterances, and track the longest word or utterance in a set of files. The implications of the computerization of transcripts of child language are evident. A task that used to take hours to complete can now be accomplished with the typing of a single command, and with greater accuracy and reliability. It also means that the raw data of other researchers are readily at one's disposal if one wishes to check research hypotheses against other data sets.

Our task in creating a Cantonese child language corpus has been to transcribe the audio- data in the internationally accepted CHILDES format, and to convert the transcripts into text files.³ We had two objectives in mind when developing this corpus. First, it should be easy to use by anyone who finds the data relevant to their purpose, whether they be professional researchers or practitioners in language-related fields. This means that the data should definitely have Chinese display and not be simply entered in romanized forms or in formats only familiar to the professional researcher. Second, the data should be usable for child language researchers who may not be able to read Chinese characters. To achieve the second objective, some kind of conversion from Chinese characters into romanized forms would be needed. It would

also be desirable to have word-for-word English glosses for each romanized line in the text.

To this end, we have been entering our data in plain Eten text files following the CHILDES convention. The corpus at the present stage consists of approximately a hundred text files, with each file based on a session of recording. The average file size is approximately 30 kilobytes. Upon completion, the corpus will consist of around 150 files reflecting the longitudinal language development of eight children over the span of a year. A typical file in the corpus begins as in (2):

```

Begin
Name of CHI:  Cheung Chun Him
Birth of CHI:  28-Oct-1989
Sex of CHI:    Male
Tape location: 1
Date:          07-MAR-1991
Time duration: 1100-1300
Age of CHI:    2;04;08
Participants:  CHI child INV investigator MOT mother FAT father SIS sister
Situation:     At child's home. The child's family is in the sitting room.
*MOT:          謙謙.
*MOT:          同姐姐講嘢啦.
*INV:          喂.
*MOT:          姐姐咁 &a1.
*CHI:          姐姐.
*MOT:          頭先有冇踩單車 &a3?
*CHI:          0 [=! nods].
*MOT:          有 &a4.
*MOT:          叫姐姐睇電視咁 &a1.
%exp:          child does not act according to mother's words.
*CHI:          0 [=! looking at investigator's recorder].
*INV:          你睇吓.
*INV:          嘩, 佢呢就會 &dzaam2 吓 &dzaam2 吓咁㗎.
*MOT:          咩嘢㗎?
*CHI:          &tsa3
*MOT:          &tsa3
*MOT:          錄音機係咪 &a3?
*CHI:          〈機〉 [ ] 機.
*MOT:          做乜嘢?
*MOT:          叫姐姐做乜嘢?
*CHI:          睇電視.
*MOT:          睇電視.
*INV:          睇電視.

```

As can be seen, the header follows the CHILDES format. The asterisked lines are the utterances of the participants of the conversation, with the identity of the speaker indicated. In addition to the asterisked lines, there are lines prefixed by a percentage sign and bracketed comments which give information about the situation or the child's non-linguistic behavior.

Once the files are in place, data analysis can be performed. For example, one could look at the entire files of a particular child, and extract all the utterances said by

the child containing a particular character, with the number of occurrences of the character indicated. The search can be easily extended to adult utterances or child and adult utterances combined. If one is only extracting strings, the speed at which the search could be done will be a matter of seconds. For example, the entire set of strings produced by one of the children containing the adverb zau6 'then' adds up to only eight utterances, as shown in (3):

- *CHI: 就返囉。
 *CHI: 就嚟冇晒啦3 [=! points at the tape-recorder].
 *CHI: 就嚟冇帶。
 *CHI: 我一陣就開俾妳。
 *CHI: 陣間就開俾妳。
 *CHI: 好似我本書嗰個就係〈紅〉[/] 紅蘿蔔 # 嗰個 &mit1 &si4 &a3 # 紅一：蘿蔔。
 *CHI: 〈嗱〉[<]。呢個就係 &jyn6 餅嘞 [=! holding a piece of play-dough].
 *CHI: 有啲唔咬就唔擺。

The surprisingly small number of tokens for this adverb as compared with other morphemes (e.g. sin1 'first') is thus ascertained by executing a simple command. The figure suggests that the adverb zau6 'then' is probably not as productive as some of the other adverbs. If this pattern is replicated with other children, the discrepancy will call for an explanation.

The child's utterances in isolation do not, of course, tell us how they understand the adverb, or whether these occurrences are spontaneous. Context is essential in determining whether the child is simply imitating adult speech or is using the words productively. Context is also crucial for interpreting the meanings intended by the child. It would therefore be desirable to have text search software that will include not only the single lines that contain a particular morpheme, but also a window that covers the utterances preceding and following those lines that are of interest. The size of the window should of course be adjustable.

As mentioned above, the CHILDES programs have a keyword-in-context search function. Nevertheless, they cannot be directly applied to Chinese data: the software identifies words by means of spaces, but Chinese characters and words are not generally separated by spaces. The programs will therefore fail to segment the Chinese strings properly. To resolve this problem, we have developed a program for use on an Eten platform. It can extract from any number of files in a directory lines containing a particular character, with a search window of specified size (0–9 lines). It can also do a Boolean search, for example extracting lines with a particular character that are preceded or followed by lines containing another specified character. The result of that search will be directed to a separate file. (4) gives an example of the main menu for this program.

FILE

INPUT FILE(s): C:\LONGDATA\cgk1*.*
 OUTPUT FILE: c:\lee\cgk1zo.chi

PHRASE

THE PHRASE TO BE SEARCHED: 㗎

LINES

NUMBER OF LINES OUTPUT: 3

BOOLEAN CONDITIONS

SPEAKER(s) SELECTED: 1. CHI

STRING CONTAINED: 未

The screen display indicates that the search is performed on all files beginning with the prefix 'cgk1', which means all files for the child 'cgk' before two years of age. The phrase searched is the Cantonese perfective aspect marker zo2. The window size of the search is one line preceding and following that carrying the aspect marker, giving a total of three lines. Only the lines whose surrounding context also contains the negative perfective morpheme mei6 were included. The output of this search will be stored in a file called 'cgk1zo.chi', which will list every extracted string together with its context, including information about the line number of the string, as well as the name of the file from which it is extracted. A segment of the file is illustrated in (5).

Input file:	C:\LONTDATA\CGK11108
The phrase to be searched is:	咗
The 3 line(s) including the phrase:	
*INV:	係啦，妳食咗嘢未呀？
*CHI:	食咗。
*INV:	食咗呀？
The speaker is:	CHI
Line number:	362
The number of occurrence(s) of the phrase:	1

In addition to the Chinese keyword in context search program, we have also developed a program which will romanize Chinese texts according to any user-defined romanization convention. Given a particular Chinese child language text file, one could produce a romanized text that preserves the format of the original. One problem encountered in such conversion is the problem of homonymy: many characters have more than one pronunciation and hence more than one romanization. At present, the program will list the various romanizations corresponding to a character, and manual editing would be necessary to determine the appropriate choice for the word in context.

Currently, the Cantonese files do not include information about parts of speech of the various morphemes. Addition of such information on a separate tier under each string would enable the researcher to ask further questions about grammatical categories, and to identify particular phrasal structures. With part-of-speech tags, one could, for instance, extract all utterances that begin, not with a noun phrase, but with a verb; or one can collect all utterances that contain classifiers. The tagging of part-of-speech labels is one of our major concerns in the final phase of the project.

In the process of designing text search software with an aim to developing future exchanges with scholars in other countries, we have encountered other problems. For example, how can the data, which are encoded in Big-5 codes, be used by scholars who use Guo-biao codes, the system used in PRC, and to a certain extent in North America? Our corpus makes use of a long list of Cantonese words which are coded in terms of the user-fonts we designed. The codes for these dialect characters are non-standard, and

may not be shared by others who do not have the user-font information. How can one process the data so that it can also be used by researchers using Macintosh systems? These are questions we need to address in the long run.

While we fully recognize the value of creating such a corpus, we are also aware that such computerization efforts are no substitute for careful qualitative analysis. To understand the dynamics of language acquisition, one would have to reflect on conceptual premises and examine the data with meticulous care. What the computer can do is speed up the initial task of data processing. We have also learned that despite its limitations, the trend toward computerization carries an important message for researchers in the humanities. It is true that the use of computer techniques does not guarantee the quality of the research. It is also undeniable that in certain areas of linguistic research one cannot afford to be uninformed about these techniques. To do so would put one in a highly disadvantaged position, especially if most other researchers in the field have been able to adopt more sophisticated methods of empirical analysis.

Notes

- 1 The research is being funded by a UPGC grant to Thomas Lee (CUHK), principal investigator, Colleen Wong (HKP), and Samuel Leung (HKU). Support from a Faculty of Arts direct grant and a grant from the Freemason's Fund for East Asian Studies is hereby acknowledged. Other members of the research team include Patricia Man, Alice Cheung, Kitty Szeto, and Cathy Wong. Lawrence Cheung and Joseph Kwok helped us with software development.
- 2 The Cantonese romanization convention used in this text is that adopted by the Linguistic Society of Hong Kong.
- 3 There have been other efforts to create a Chinese child language corpus, notably the Mandarin corpus at the Dept of English, National Taiwan University. Their computer texts are also coded in Eten files, but they include only the children's utterances, with information about adults excluded.

References

- Brown, R. (1973), *A First Language: The Early Stages*. Cambridge, Massachusetts: Harvard University Press.
- MacWhinney, B. (1991), *The CHILDES Project: Tools for Analyzing Talk*. Hillsdale, New Jersey: Lawrence Erlbaum.
- MacWhinney, B. and C. Snow (1985), "The Child Language Data Exchange System," *Journal of Child Language*, 12, pp. 271-296.
- Man, Yuk-hing (1993), "Subject-object Distinctions and null NPs in Child Cantonese," M.Phil. thesis, Department of English, Hong Kong Polytechnic.

THE CHALLENGE IN CREATING A HONGKONG CANTONESE CHILD-LANGUAGE CORPUS

Colleen Wong and Patricia Man

§1 Introduction

We are going to describe to you what a team of researchers in language acquisition, among whom the writers of this report are members, have undertaken: to create a computerized Hongkong Cantonese child-language corpus. This is to add to the computerized data that is freely accessible to colleagues of the international community. In our description of the creation process, we will highlight the problems that we have faced, how we have tackled them (i.e., the decisions that we have made), and what we think can further be done when the world becomes a bit more ideal. All this we now have the privilege to call *a challenge*; you can let us know how well we have met the challenge September this year when we shall be making the corpus accessible to you.

The creation of the corpus follows the formula: data collection, data transcription, computerized input and, voila! output, with software development along the way between input and output. Decisions have to be made from day one to the day we call it a day. The decisions are theoretical and methodological in nature, with theory hopefully in the driver's seat most of the time. There are of course other factors that affect our decision making. But they are technical in nature, such as a lack of certain ideal equipment, or not being able to get the total amount of funding we applied for. We know that they are not crucial. What is crucial is a vision, or call it a dream if you like. But we were indeed endowed with institutional monies to start with, and by the time they more or less ran out, we had already acquired knowledge about potential funding sources and accumulated evidence enough to give us and the funders confidence that we

would put their money to good use and deliver the goods.¹ And while we are working, the dream takes care of itself and begins to take the form of a vision; and the vision helps us grow and become sanguine researchers.

§2 The data

The *HongKong Cantonese child language corpus* consists of 171 Eten text files put into the internationally accepted CHILDES format (MacWhinney and Snow 1985).² Each file consists of an average size of approximately 1500 lines (30 Kilobytes). These text files are based on longitudinal observations and audio-recordings of adults' conversations with eight children, aged from one-and-a-half to three-and-a-half. Each individual child was audio-recorded at home fortnightly for a year. The youngest child was 1-year-5-months old (1;5) and the oldest was 2-year-8-months old (2;8) when recording began. So as a whole our data consist of spontaneous language of eight Cantonese-speaking children, aged from 1;5 to 3;8, recorded in a total of 171 sessions.³ These are very young children because we want to look at the emergence of syntax in the Cantonese of a child.

We'll now look at the kind of questions that our longitudinal data can and cannot answer.

§2.1 *What is there in longitudinal data?*

A first decision we had to make is related to data collection. What data to collect? How much? In fact such questions don't quite arise on their own, because we had already decided to do a longitudinal study on the grammatical competence of Cantonese-speaking children. The questions we ask and the problems we anticipate are those that hinge on what longitudinal studies entail: what information we can get out of longitudinal data and hence what questions we can ask; what information we cannot get out of longitudinal data, and hence what questions are irrelevant; and if there are certain questions that ought to be asked (certainly there are!), what other kind of data should be

¹ See Acknowledgements, at the end of the paper.

² CHILDES is an acronym for *Child Language Data Exchange Service*.

³ See Appendix:: Child Subjects, Periods of Recording and Number of Recording Sessions.

collected. (Answer: cross-sectional, experimental, ... data -- cf. Lust, Chien & Flynn 1987)

Data from longitudinal studies can provide evidence of first occurrences of morphemes and are useful for studies of say, the emergence of categories. While we can say things like that, broadly and fairly confidently, with reference to our corpus, we want to reiterate the worries that researchers have when using longitudinal data in general. Here they are: first the methodological limitations, and then the limits in aiding theoretical formulations.

The first recording of a certain linguistic form may not necessarily be the first occurrence in the child's speech. In other words, there is a fairly good chance that the first occurrence is uttered when the researcher is not around. That can explain why data in the form of diaries, collected by parent-investigators, have an edge over those collected by visiting investigators -- assuming all other things are equal.

Limitations on the investigator's abilities to record adult speech preceding the specific child utterance, requires the researcher to be particularly careful in distinguishing exemplary first utterances from repetitions of adult speech by the child.

And confidence, or reservations if you like, will come in useful when one infers the order of acquisition of linguistic elements or rules as reflected in the longitudinal data or decides upon the emergence or absence of a form based on the data. The child may not have acquired certain rules. Or the child, or any learner for that matter, may have acquired one thing but doesn't use it right away and may have learned another and uses it immediately. There is differential opportunity to display knowledge. Take our case for instance. Based on the absence of the form, we tried to claim (in Man 1993) that reflexives are absent in the postverbal position during a particular stage of language development. But one of our readers pointed out that the claim is invalid based on longitudinal data alone. Another instance: although our corpus is of considerable size, containing language from a total of 171 sessions of eight children, we could only find a total of thirty-seven utterances containing wh-words (interrogative words) in

indirect questions.⁴ The absence of certain forms in children's speech may indicate that the child's rules do not generate those particular forms or that communicative intentions do not require those rules to be exercised (Pinker 1984:127-8, 147).

There are straightforward and consistent data in our corpus that can be used to make arguments. For example, after examining the language of two children (aged 1;11 to 2;9 and 2;3 to 3;4), Man (1993) uses the following findings in her thesis, *Do children have subject-object distinction around the age of two to three?*: (a) the correct response rate to object questions is far higher than the correct response rate to subject questions (when the mean-length-of-utterance is about 2.3 and the age of the children is between 2 and 2 1/2 – cf. Man 1993: 7-8); (b) reflexives only occur in the preverbal position; (c) omission of preverbal arguments (subjects) is far higher than that of postverbal arguments (objects).⁵

§1.2 How much longitudinal data is enough for researchers to make statements with confidence?

In the subject-object distinction study (Man 1993), which represents the earliest stage of the research project on the grammatical competence of Cantonese-speaking children, we had only two children's data ready for analysis. But this body of data not only shows consistent language patterns in both children as mentioned in (a), (b) and (c) in the previous paragraph, it also shows similar patterns in other areas. For example, the data show that both children start using reflexives with agent noun phrases and then extend to other thematic roles. Also the frequency of dropping preverbal arguments is in an inverse proportion to age for both children. The investigator is confident in making statements in some areas with the data from two children. For example, on the discourse factors which affect the use of various structures; on how commonly a form is produced in this age

⁴ This is a finding in a current M. Phil. study (A. Cheung, *Acquisition of wh-words by Cantonese-speaking children*, Department of English, Hong Kong Polytechnic University).

⁵ Man uses these findings as support for her argument for subject-object distinction in early child language. She is aware that there can be alternative arguments using the same findings – e.g. pre-/post- verbal distinction. She thinks that experimental studies will be useful to clarify the issues.

range, namely frequency; and the presence of idiosyncrasies (or errors) in language development.⁶

But there is the problem of differential frequency. Some structures are found only in one child, like the use of connectives like *jin4-hau6 then* and *jan1-wai6 because* to form compound sentences. One child uses reflexives quite a lot whereas the other uses few of them. One child doesn't show much response to questioning subjects and objects whereas the other does. We then have to collapse the data of the child with few responses in two investigative periods in order to compare the response patterns with the other child in one period.

Now that we have the data of eight children aged 1;5 to 3;8 ready for analysis, we naturally have more data to look at. For example, the study on the acquisition of classifiers conducted by a team member reports that a total of sixty-nine classifiers have been produced by the children. This is in contrast with experimental studies on classifiers in Cantonese and Mandarin, in which far fewer classifiers have been studied.⁷

A large body of data supported by good analysis can not only give us a more comprehensive understanding of a certain question but also prompt rethinking of existing statements in a particular language and possibly other languages as well.

But the point we want to make is not a simplistic "the more the merrier". Recall what we have cited earlier about the inadequacy of longitudinal data: the absence of form (reflexives) and the small sample size (wh-words in indirect questions) in our corpus.

6 Examples of idiosyncrasies or errors in child language are:

- (1) syntactic forms - *hissell*, and *self* [Brown data]; [cgk11101],
 - (2) semantic roles - *dang3 dang3 zau2 zo2* (chair-go-ASP);
(*zau2* selects an animate NP in the non-figurative sense); [cgk20008].
 - (3) position of a category - *maau1 maau1 sik6 ping4 guo2 jau5*
(kitten-eat-apple-HAVE)
- As an aspect marker, *jau5* should occur in the preverbal position;
m4 (not) occupying the utterance initial position can precede different syntactic category - i.e., V, N, A etc.;
m4 ji1 waa2 a3 (NEG-auntie-play-prt) [ckt20215].

⁷ This is K. Szeto's current M. Phil. study on the acquisition of classifiers by Cantonese-speaking children (Dept of Speech & Hearing Sciences, University of Hong Kong). Szeto points out that, e.g. Erbaugh (1986 - longitudinal, 3 children) reports on 17 Mandarin classifiers; Loke & Harrison (1986 - experimental) work with 10 Cantonese and 8 Mandarin classifiers; Mak (1991 - experimental) 10 Cantonese classifiers, Poon (1980 - experimental) 27 Cantonese classifiers; and Ying et al (1987 - experimental) 56 Mandarin classifiers.

In the study on the acquisition of wh-words conducted by a team member,⁸ an asymmetry between subject and object questions is observed using the corpus. She decided to add a cross-sectional experimental study to see whether the asymmetry still holds and to explore the effects that verbs might have on wh-questions.

So how does one deal with the inherent limits and limitations of longitudinal data? Researchers (e.g., Pinker 1984: 56) adopt the assumptions that (i) in general what is first recorded is what is first uttered; (ii) in general the opportunity is frequent for the child to use what is acquired; (iii) in general the child will not suppress the linguistic elements that s/he has learned; and (iv) in general linguistic knowledge is not dormant. Yes, after all, this is a sort of empiricist methodological consideration!

Methodological considerations, however, can make up for human limitations to some extent. For example, adult speech is recorded and transcribed together with the child's in our corpus. That way we at least have the more immediate antecedent adult speech to help decide whether a certain utterance of the child is an exemplar occurrence or an imitation.⁹ Also reading a larger body of data can give confidence in saying whether a certain entity is used productively (i.e., *spontaneously*), or whether the utterance is an un-analysable (i.e., *formulaic*) entity, or whether it is simply some isolated case that is there only to baffle the investigators. Including contextual information in the corpus (as will be discussed later under *running commentary*) can help us decide whether the meaning is acquired or not.

Longitudinal data is basically language recorded over a period of time. The information one can cull from the data is plentiful. We will, in the next section, give you an example of how reading a large body of data helps us arrive at an answer with more confidence. Context is useful for investigators to make an informed guess at what the child means, when a guess or an inference has to be made. In this regard we have developed a text-search program that gives us lines of utterances preceeding and following a particular line uttered by the child. We'll talk about that too. But what we have been talking about so far are

8 This is A. Cheung's study, cf. Note 1 above.

9 The Mandarin corpus at the Department of English, National Taiwan University, is a notable effort to create a Chinese child language corpus. Their computer texts are also coded in Elen files, but they include only the children's utterances.

matters related to ascertaining exemplary occurrences of linguistic forms in children's language development. And having said all this, somehow a decision will have to be made.

In an investigation on the emergence of categories,¹⁰ we count an entity as having been acquired by the child when it has occurred in three consecutive recording sessions. This is a fairly conservative criterion for emergent syntax. A conservative criterion does not allow one to hypothesize much, and a corpus of considerable size is often needed to generate the statements made. The trade-off is that one might be able to talk with a bit more confidence and conviction.

But in a study on classifiers,¹¹ the investigator counts the first occurrence of a classifier. We appreciate that acquisition of a linguistic form might be easier to ascertain by means of experimental data than longitudinal data. But the investigator thinks that *classifier*, although a *closed class*, is lexicalized by many different forms in Cantonese (69 classifiers have been observed in the corpus), and it can be said that each classifier has its own unique meaning. The investigator hopes to be able to describe the development of classifier meaning as well. That's why *emergence* is also looked at and first occurrences are counted.

§3 Audio-recording

There are altogether seven members in our team. Three members, who were graduate students then, each followed the individual child at home to audio-record its language during the fortnightly visits, for a year.¹² Each recording session lasted for about sixty minutes. So each investigator took charge of two to three children's language collection for the whole year. They were also responsible for transcription. But data collection and transcription were not the only tasks in that year. There were fortnightly seminars that the team organized to keep abreast of

¹⁰ This refers to an on-going investigation by members on the acquisition of categories, including: post-verbal auxiliaries, classifiers, aspect and complementizers (sentence final particles, relative clause markers, & resultative clausal complement markers).

¹¹ This is a current M. Phil. study on the acquisition of classifiers by Cantonese-speaking children (K. Szeto).

¹² Audio-recording and data transcription were carried out by the following team members (the children are identified by codes): P. Man (*cgg, wbi*); A. Cheung (*ckt, ltf, mhz*) and K. Szeto (*ccc, lthc, lly*).

issues related to linguistics, especially syntax, language acquisition and methodology and to report on progress. So data collection was the predominant task in the first year or so; transcription work and seminars were concurrent activities and continued beyond the data collection phase.

§3.1 *Equipment*

We used Sony TCS2000 recorders to audio-record the children's language when we visited them at their homes. This model doesn't have an external microphone but it offers quality recording. We had considered external microphones because they can pick up children's speech even when they are moving about. But recorders with external microphones that we came across did not meet the quality requirement that we desired. They would pick up too much noise when children moved about. Clip-on microphones attracted children's attention; they looked on them as toys and wanted to play with them. At the end we decided to use the present recorder and put it on the table while we engaged the child with toys and pictures and interacted with him/her.

The toys we brought along included play dough, lego, playhouse pieces and picture books. Thanks to the toys (and the people of course), after one or two visits, the children were looking forward to our coming to play with them. To induce variations in the content of conversations, we varied the toys from time to time. Care-takers such as the mothers were interested in participating in the conversations. But most of the time they were preoccupied with household chores and left the children with the investigators.

For transcribers, we use Sony BM77, which is a standard cassette system. The transcriber can be used together with a foot control unit (FS75) and an earphone (DE35).

§3.2 *Running commentary*

Let's back track a bit to audio recording. We could only manage one investigator per visit so she had to primarily interact with the child and try to supply contextual information at the same time by commenting on the situation so that the comments would be recorded and transcribed.

Recording contextual information is important. For example, the child might be pointing rather than making an utterance. Contextual information fills the omission of deictic referents in the audio-recording of child language. One of our children at one stage uttered negative forms but nodded his head simultaneously, and uttered positive forms but shook his head. This kind of behaviour needs commenting on; otherwise we would have taken the acquisition of a form as necessarily equivalent to the acquisition of meaning.

It is useful to have two investigators to work in each visit, with one interacting with the child and the other acting as an observer and to giving direct commentary on another tape-recorder, especially in cases where no visual-recording is taken. When there are two investigators, the work includes the synchronization of information on separate tapes for transcription purposes.

Now we are going to show you what the data looks like after being computerized and coded in the CHILDES format. Then we will return to transcription to show you some problems and working solutions.

§4 A file in the CHILDES format

We understand that it is only when our data has been coded in an internationally acceptable format and computerized, and text-search software designed, that it can be freely accessible to colleagues in the international community. We want to contribute to the growth of child language databases freely available for comparative and theory purposes. This means we have to transcribe the audio-data into an internationally accepted format, and to convert the transcripts into text files. We entered our data in plain Eten text files following the CHILDES format. A typical file in our corpus is given on the following page:

You can see that the file header contains information of the name, birth date, sex and age of the child, the date, time and location of the recording, and the identity of the participants. The lines prefixed by the * sign indicate the utterances of the individual participants. The % lines and bracketed comments give situational or behavioral information accompanying the utterances.

Example of a transcript by hand

@Begin
 @Name of CHI: CHEUNG-CHUN-HIM
 @Birth of CHI: 28-Oct-1989
 @Sex of CHI: Male
 @Tape Location: 15
 @Date: 6-Oct-1992
 @Time duration: 1730-1830
 @Age of CHI: 2;11;08
 @Participants: CHI child INV investigator SER maid
 SIS sister MOT mother
 @Situation: Inside the bathroom in the child's home. The child finished bathing and the maid is helping the child put on his clothes.
 #INV: 沖完涼 &la3 係咪?
 #CHI: 係 &a3.
 #CHI: 擺上去, 一二.
 %exp: The maid helps the child out from the tub.
 #INV: 哎咗, 要人抱 &a4 咁大個都?
 #INV: &fing6 乾未?
 #CHI: 0.
 #INV: 0 [=! giggles].
 #INV: 著衫未?
 #INV: 你啲衫咁邊度 &a3?
 #CHI: &eik3, 喺度 [=! points].
 #INV: 自己識唔識擺?
 #CHI: 0 [=! nods].
 #INV: &ap6 頭即係識 &lu3 &wo3.
 #INV: 有冇著衫 &a3 你睇清楚自己?
 #CHI: 0.
 #INV: 有冇?
 #CHI: 有.

Example of CHILDES format

A: 點解呀? 你吓吓, 試吓你吓!
 (A opens the McDonald's)
 A: 麥當勞早啲囉!
 A: 你呀, 吓吓仲有冇得開嘅?
 有隻手囉!
 A: 機械人囉!
 A: 一九九三年二月一日
 李天仁
 我開唔到²呀, 點解開唔到?
 [o²], [o²] /
 咁呀
 你吓吓啲囉呢個?
 (Tiffany plays with the robot)
 [eu²]! (Tiffany laughs)

While we were transcribing the data by hand, we started to employ help to input the transcripts into Chinese text files. And we switched our format to that of CHILDES even in the hand-transcription. This was to facilitate the data input process.

S6 Problems in transcription

We would like to talk about two problems we have encountered in the transcription work. They are: demarcating pauses and the level of transcription in unclear cases.

In accordance with the CHILDES format, a line is an utterance. An utterance is the basic unit for analysis. And we have to decide on utterance boundaries in transcribing. Utterance boundaries have a tremendous effect on linguistic analysis, and we'll show you what we mean shortly. A pause is a natural demarcation of an utterance boundary, you might say. And what an innocently

S5 A transcript by hand

The three investigators who collected the data also transcribed it before it was put into the computer.¹³ We followed the conventions set out in Bloom & Lahey (1978: see Appendix A, *Conventions for Transcription of Child Language Recordings*). As you can see from the sample below, a page is divided into two halves, with all the adult speech on one side and the child's on the other.

¹³ Audio-recording and data transcription were carried out by the following team members (the children are identified by codes): P. Man (cgk, wbb); A. Cheung (ckt, llf, mhz) and K. Szeto (ccc, lhc, lly).

simple answer *pause* appears to be! Consider the following example please (child at 1;9):

- (1) *CHI: ping4-gwo2 # ngau5. (# indicates a pause)
(apple bite)

Example (1) presents the particular token of speech as one utterance, and therefore makes the utterance subject to a topic-structure analysis, which is common in adult Cantonese. But, depending on the transcriber's (to mean the person who does the transcribing, not the machine) perception of the length of the pause, the same instance can be interpreted to contain two utterances, and transcribed as follows:

- (2) *CHI: ping4-gwo2.
(apple)
*CHI: ngau5.
(bite)

Extra care is needed when one argues for a topic-structure analysis with (2).¹⁴

A pause marks an utterance boundary. The way how pauses are assigned affects the measurement of "MLU" (mean length of utterance), which is used as an indicator of children's linguistic maturity and from which developmental stages can be inferred (cf. Klee & Fitzgerald 1985).

The actual length of silence in real time is but one factor which helps determine a pause. It is the human being who perceives and interprets. We think it is useful for the transcriber first to identify the ways that the indication of a pause can affect the linguistic analysis that is to be applied on the data. Then s/he will (not) arrive at an arbitrary decision, and be as consistent as possible when transcribing the data.

Level of transcription is another decision the transcriber has to make. Morphemic or phonemic transcription for unclear cases?

- (3) (English) [ə want] / &a want/ [I want]
(Cantonese) [ə jiu:3] / &a jiu3/ [ŋo5 jiu3]

¹⁴ We think that if a researcher wants to argue for topic-structure analysis, s/he would already have observed some occurrences of topic-structure in the child's language. Then seeing (2) being two consecutive utterances of the child, s/he interprets the utterance boundary as a long pause.

The transcription (between slashes '/') is marked by the & sign, is the phonemic transcription of the phonological string; and the other (between square brackets '[]') is the morphemic transcription of the string. An investigator holding a non-deterministic view of language development will favour phonemic transcription in unclear cases, such as (3) above. Other users of the corpus are thus given space to determine for themselves the meaning of the phonological strings. Another example:

- (4) /mou (rising tone)/

Cantonese has two rising tones -- tone 2 and tone 5. The two tones are difficult to distinguish if a syllable is pronounced in isolation with regular speed. In adult Cantonese, /mou5/ is a single morpheme meaning 'no' in the linguistic context [__NP]; /mou2/ is a contracted form for /m4 hou2/ meaning 'don't' in the context [__VP]. So both /mou5/ and /mou2/ express negation. A non-deterministic transcriber will put down (4) to indicate that it is a phonological string and leave the analysis to the users of the data.

§7 Tagging

Remember we want ultimately to be able to account for the acquisition of certain structures. Tagging is assigning labels to structures. But the data at this stage do not even tell us what a structure is. So we have to delineate structure, or structural units, just as in the process of transcribing the data when linguistic analysis is called for. We can't say what we mean by *structure* without committing ourselves to a theory of what that structure is. A theoretical stand is therefore fundamental.

Take for example the structural unit of what we call a category. A complementizer in generative grammar is considered a functional category. This particular grammar says that every clause is headed by a complementizer whose function it is to characterize the clause; that is, a complementizer tells us what sort of a clause it is. In English, the complementizers are *that* which introduces a tensed declarative clause; *if*, which introduces a tensed interrogative clause; *whether*, which introduces a tensed or un-tensed interrogative clause; and *for ... to*, which introduces an un-tensed noninterrogative clause (cf. Radford 1988: 303).

Before Tagging vs. After Tagging

We have developed a tagging program to parse the main line (i.e., the transcription) and identify the Chinese characters, and then for each character create a corresponding entry on the %mor: line with information on part-of-speech. The tagging program feeds on a lexicon. We applied for two separate small funds for creating the lexicon and designing the tagging program. In the following paragraphs we'll describe what we did in creating the lexicon to provide crucial linguistic input to the designing of the tagging program.

CHKWIC program has been developed such that it can extract, from any number of Eten files in a directory, lines containing a particular character or category, with a search window of specified size from zero to nine lines. A window specifies the lines preceding and following the line in which the morpheme of interest is found. CHKWIC can also do a Boolean search, for example, extracting lines with a particular character that are preceded or followed by lines containing another specified character. The result of that search will be directed to a separate file. An example of the main menu for this program is given in Appendix B.¹⁵

A program has been designed to count the mean-length-of-utterance. Another program we call DICTON is designed to compute frequencies and related statistics such as type and token of the morphemes and categories in the corpus. A sample of the output is given in Appendix C.

§9 Ending and Beginning

Our corpus will be put into the CHILDES database, but in a server.¹⁶ The computational tools in their present shape cannot accommodate Chinese characters. Of course we could romanize the characters in the files and gloss every word for the benefit of those researchers who don't read Chinese. But the team members have decided not to pick up the romanization and glossing tasks. In the five years researching the development of grammatical competence of Cantonese-speaking children and creating the corpus along the way, we have always been in the mainstream of life — i.e., teaching and research for some and studying for higher degrees for others. But a community has been created. And the members are now dispersing for other focussed pursuit in Cantonese linguistics and language acquisition or some other concerns in life.

As for the eight children, we wish them continued growth in age, grace and wisdom with God and man.

¹⁵ The example is the same as the one used in Lee (1994: 66-67).

¹⁶ Personal communication with B. MacWhinney.

Acknowledgments:

On behalf of the research team, we wish to thank the following bodies for funding various portions of the research: UPGC: Earmarked Research Grant for 1991-94, *The development of grammatical competence in Cantonese-speaking children*, to Thomas Lee (CUHK) principal investigator, Colleen Wong (HKPU) and Samuel Leung (HKU), together with Patricia Y.-H. Man, Alice Cheung, Kitty Szeto and Cathy S.-P. Wong. Lawrence Cheung and Joseph Kwok helped with software development.

CUHK: Faculty of Arts, for a direct grant; Department of English HKPU: Research Committee, for supporting 3 related projects; Staff Development Programme, for 2 programmes; Department of English, for 3 grants, supervision of 2 related M.Phil projects, and the Working Papers HKU: Department of Speech & Hearing Sciences, for supervision of a related M.Phil project.

In particular we are grateful to colleagues, local and overseas, for collegial support and to the Linguistic Society of Hong Kong for giving us opportunities to report our findings at various stages of our research in child

REFERENCES:

- Bloom, L. & M. Lahey. 1978. *Language Development and Language Disorders*. New York: John Wiley.
- Erbaugh, M. 1986. Taking stock: the development of Chinese classifiers historically and in young children. In, C. Craig (Ed.), *Noun Classes and Categorization*. Amsterdam: John Benjamins.
- Klee, T. and M.D. Fitzgerald. 1985. "The relation between grammatical development and mean length of utterance in morphemes." *Journal of Child Language*, 12, 251-269.
- Lee, T. H.-T. 1994. "A Cantonese Child Language Corpus." *The Humanities Bulletin*, Chinese University of Hong Kong, 3, June 1994, 62-68.
- Loke, K.K. & G. Harrison. 1986. Young children's use of Chinese (Cantonese and Mandarin) sortal classifiers. In, H. Kao & R. Hossain (Eds.), *Linguistics, Psychology, and the Chinese Language*. University of Hong Kong. pp. 125-146.
- Lust, B. (Ed.). 1987. *Studies in the Acquisition of Anaphora*. IL D. Reidel.
- Lust, B., Y.-C. Chien & S. Flynn. 1987. "What children know: Methods for the study of first language acquisition." In B. Lust (Ed.), 1987: 271-356.
- MacWhinney, B. & C. Snow. 1985. "The Child Language Data Exchange System." *Journal of Child Language*, 12, 271-296.

- Mak, D. 1991. *The acquisition of Cantonese classifiers*. Unpublished D. Phil dissertation. University of Reading.
- Man, Y.-H. 1993. *Subject-object distinction and null NPs in Child Cantonese*. M.Phil. dissertation, Department of English, Hong Kong Polytechnic University.
- Pinker, S. 1984. *Language Learnability and language Development*. Cambridge, MA: Harvard University Press.
- Poon, E. 1980. *Some Aspects of the Ontological Development of Nominal Classifiers in Cantonese*. Unpublished M.A. dissertation. University of Hong Kong.
- Radford, A. 1988. *Transformational Grammar: A First Course*. New York: CUP.
- Ying, H. 1987. Si zhi gi shui er tong zhang wo lian ci de te din [*The characteristics of children's mastery of classifiers from 4 to 7 years old*]. Child Language Developmental Research. East China Normal University press. pp.91-103.

APPENDIX A

Child Subjects, Periods of Recording and Number of Sessions:

Child	Periods of Recording	Number of Sessions
CKT	1;05;22-2;07:02	25
MHZ	1;07;00-2;08;06	26
CCC	1;10;08-2;10;27	22
CGK	1;11;01-2;09;09	19
LTF	2;02;10-3;02;18	16
WBH	2;03;23-3;04;08	27
HHC	2;04;08-3;04;14	16
LLY	2;08;10-3;08;09	20

APPENDIX B

Example of CHKWIC format

```

FILE
INPUT FILE(s):      C:\LONGDATA\cgk1*.*
OUTPUT FILE:        c:\lee\cgk1zo.chi
PHRASE
THE PHRASE TO BE SEARCHED:  咗
LINES
NUMBER OF LINES OUTPUT:    3
BOOLEAN CONDITIONS
SPEAKER(s) SELECTED:      1.CHI
STRING CONTAINED:         未

```

APPENDIX C

Example of DICHION format

Input File(s): C:\TAGFILE\CCC21027.TAG

Speaker(s) Selected:

CHI
INV
NOT

Total number of tokens = 8656

Category	Types in Tokens in p.c. of Category	Morpheme No. of Tokens	Filename	Line no.
----------	-------------------------------------	------------------------	----------	----------

adj	73	445	5.14094% 咗 一樣	2	CCC21027	4744,4748
				12	CCC21027	1360,1366
					CCC21027	1915,2020
					CCC21027	2693,2696
					CCC21027	2700,2702
					CCC21027	2704,2709
					CCC21027	2717,2721
det	6	430	4.96765% 咗	76	CCC21027	144,158
wh	14	296	3.41959% 咗	59	CCC21027	24,42,136

total number of types = 726
type/token ratio = 0.08387