

## Productivity in the early word combinations of Cantonese-speaking children\*

Thomas Hun-tak Lee  
Institute of Cognitive Science, Hunan University

**Abstract:** The current paper examines the early syntax of Cantonese-speaking children with respect to the issue of productivity of word combinations. It addresses the question whether the early multi-word utterances of children support a continuity theory that presupposes knowledge of syntactic categories or a theory that assumes that children learn by induction entirely on the basis of lexically-based positional patterns. The results show that frozen phrases account for less than one-quarter of children's multi-word combinations, disconfirming the lexically-based accounts of syntactic development.

**Key words:** syntactic acquisition, continuity theory, Cantonese child language

In this paper, we examine the early syntax of Cantonese-speaking children with respect to the issue of productivity of word combinations. We address the question whether the early multiword utterances of children support a continuity theory that presupposes knowledge of syntactic categories or a theory that assumes that children learn entirely on the basis of lexically based positional patterns. In particular, we evaluate some recent proposals that purport to demonstrate that a large proportion of children's early word combinations are not productive syntactic patterns (Lieven, Pine and Baldwin 1997).

### 1. The issue of continuity in language development

The classic position on continuity is stated in clear and unambiguous terms in Chomsky (1965), in which he assumes there is content knowledge specific to the language faculty, which the child is born with. The units and the general form of linguistic representations constitute part of the child's biological endowment. Chomsky assumes that the child comes to the language learning task with the linguistic knowledge detailed in (1):

- (1) (i) a technique for representing input signals
- (ii) a way of representing structural information about these signals
- (iii) some initial delimitation of a class of possible hypotheses about language structure
- (iv) a method for determining what each such hypothesis implies with respect to each sentence
- (v) a method for selecting one of the (presumably, infinitely many) hypotheses that are allowed by (iii) and are compatible with the given primary linguistic data" (Chomsky 1965: 30)

---

\* I wish to thank Stephen Crain, Paul Fletcher, Brian MacWhinney, James Myers, James Tai, Jane Tsay and William Wang for helpful discussions of various points raised in the paper. Versions of this paper were presented at the Workshop on Language Acquisition and Language Change held at City University of HK on May 31-June 1, 2001, the Symposium on Studying Grammatical Development through Computational Analyses of Large Corpora held at the Ninth International Congress for the Study of Child Language held in Madison, Wisconsin on July 17, 2002, and the Language Acquisition Workshop of Tsinghua University, Taiwan, held on March 16, 2004. The faults that remain are mine. The support of City University of Hong Kong (faculty grant # 9380025) is hereby acknowledged. Email address: huntaklee@yahoo.com.

This position of strong continuity is further developed and strengthened in the works of Wexler and Culicover (1980) and Pinker (1984). The latter even maintains continuity of language processing mechanisms as well as linguistic representations. As he puts it, "in the absence of compelling evidence to the contrary, the child's grammatical rules should be drawn from the same basic rule types, and be composed of primitive symbols from the same class, as the grammatical rules attributed to adults in standard linguistic investigations." In addition, the continuity assumption should be applied to accounts of children's language not only in the qualitative nature of the child's abilities and in the formal nature of the child's grammatical rules, but also in the way that those rules are realized in comprehension and production (Pinker 1984: 6-8).

The strongest argument for continuity comes from the poverty of stimulus argument, which observes the existence of linguistic knowledge which cannot be traced to any conceivable positive data. Examples of complex linguistic knowledge which cannot be traced to an inductive learning source include the structure-dependence of transformations, the subadjacency condition on syntactic movement, and the Binding Principles (Chomsky 1971, 1972, 1981). By the structure-dependence requirement, syntactic rules must be stated in terms of phrasal categories and phrasal structure, and not simply as a linear arrangement of words. Thus, faced with data such as (2a,b), the learner will opt for a structure-dependent rule such as subject-auxiliary inversion rather than a rule that moves the first verb to the initial position of the sentence. Such a learner will be able to give the correct question form for a sentence such as (3a), accepting (3b) derivable by the subject-auxiliary rule, and rejecting (3c), derivable from a rule that fronts the first verb. The subadjacency condition on movement forbids movement of a constituent across more than one bounding node. By this condition, fronting a noun phrase from a coordinate structure would be prohibited, since such movement would cross a noun phrase node and a clausal node. Thus, the sentence (4b) is ungrammatical. Binding Principle A requires an anaphor such as a reflexive to be co-indexed with an antecedent that c-commands it in a local domain such as the minimal clause containing the anaphor. In (5) the reflexive *herself* should be co-indexed with an antecedent that c-commands it in the minimal clause containing it. In (5a), *John's sister* c-commands *herself*. However, in (5b), the only potential antecedent of the reflexive, i.e. *Mary*, does not c-command the reflexive; and in (5c), the only potential antecedent of *herself* does not fall within the latter's minimal clause. Thus the sentences (5b,c) are ill-formed.

- (2a) The man is tall.
- (2b) Is the man tall?
- (3a) The man who is tall is running.
- (3b) Is the man who is tall \_\_ running?
- (3c) \*Is the man who \_\_ tall is running?
- (4a) Mary knew [Paul and John].
- (4b) \*John, [Mary knew [Paul and \_\_]].
- (5a) John's sister<sub>i</sub> touched herself<sub>i</sub>.
- (5b) \*Mary<sub>i</sub>'s brother touched herself<sub>i</sub>.
- (5c) \*Mary<sub>i</sub> said that John's brother touched herself<sub>i</sub>.

If knowledge of the grammaticality of the above sentences cannot be derived from linguistic experience, a reasonable hypothesis would be to posit such knowledge as part of Universal Grammar or the initial state of the language faculty of the mind. Besides

receiving support from the poverty of stimulus argument based on adult linguistic knowledge, the continuity position is also supported by empirical data from child language. We would mention three kinds of data that are uncontroversial in demonstrating young children have access to syntactic categories and structure. First, there is strong evidence from Crain and Nakayama (1987)'s study of children's yes-no question formation that two- and three-year-olds pay attention to constituent structure, and Otsu (1981) provides experimental evidence for young children's sensitivity to the Complex Noun Phrase Constraint (cf. Crain 1991, Crain and Thornton 2001). Second, the series of experiments by Chien and Wexler on a range of languages, including English and Chinese, have shown that young children by four years old make use of c-command in determining binding relations for anaphors and pronouns (Wexler and Chien 1985, Chien and Wexler 1986, 1990). At the same time, a number of studies have shown that children must have access to functional categories before they are two years old. Deprez and Pierce (1993) demonstrated that French children's positioning of the negative *pas* and the verb depends crucially on whether the verb is finite, so that they place the finite verb before the negator and the infinitive form of the verb after it. Likewise, Poeppel and Wexler (1994) show convincingly that their German child subject placed the verb in second (non-final) position when it is finite, and in clause final position when it is non-finite. These data suggest that the functional category of inflection must be present for young children to be sensitive to the finite/nonfinite distinction.

## **2. Recent challenges to the continuity position**

In the last decade, there has been a return to a position that is even more conservative than the early semantic-based grammar of Bowerman (1973) in the analysis of children's early linguistic categories. More recent approaches, exemplified by the works of Tomasello and Lieven, are claiming that children may not be even organizing their utterances around semantic categories, let alone syntactic categories. Instead, they may be just organizing their sentence patterns around specific lexical items, in particular verbs. After acquiring a lot of these patterns pivoted on particular lexical items, they generalize across these patterns and break into categorial syntax. These researchers report experimental as well as naturalistic findings showing that children are conservative learners that generalize from lexically-based constructions instead of applying syntactic rules based on general syntactic or semantic categories.

In Tomasello (1992, 2000, 2003), the nativist position of continuity is rejected outright. He argues that there is no strong evidence from child language for the nativist position, and that syntactic learning can be achieved through the use of general cognitive principles of induction and generalization. In his view,

"children's early language is organized and structured totally around individual verbs and other predicative terms; that is, the 2-year-old child's syntactic competence is comprised totally of verb-specific constructions with open nominal slots. Other than the categorization of nominals, nascent language learners possess no other linguistic abstractions or forms of syntactic organization. This means that the syntagmatic categories with which children are working are not such verb-general things as 'subject' and 'object', or even 'agent' and 'patient', but rather such verb-specific things as 'hitter' and 'hittee', 'sitter' and 'things sat upon'." (Tomasello 2000:214)

Another group of researchers who advocate a similar point of view in rejecting both the syntactic-based and the semantic-based theories of early grammatical development is Eleanor Lieven and her associates. They developed an elaborate, explicit coding scheme for measuring productivity in terms of three kinds of multiword utterances: frozen phrases, intermediate utterances and constructed utterances. Using this taxonomy, they were able to show that individual variation among children previously characterized in cognitive terms as referential vs expressive style of language can be shown to be related to the degree to which children use frozen phrases (Lieven, Pine and Barnes 1992).<sup>1</sup> These researchers then argued that in fact acquiring a stock of frozen phrases may be a typical route for syntactic acquisition, so that the process of syntactic acquisition is one of initially learning a stock of unanalyzed phrases and then breaking them down into component parts (Pine and Lieven 1993). They claim that the frozen phrases have a facilitating role in the learning of syntax. Using longitudinal data from a dozen children, they demonstrated that lexically-based constructions can account for about 60% of children's early multi-word utterances, with the remainder consisting primarily of frozen phrases (Lieven, Pine and Baldwin 1997).

Attempts to show that syntactic categories may be inductively acquired by children have had a long history in child language study, going back to the writings of Braine (1976) and MacWhinney (1982). Braine (1976) argues that pivot grammar does not characterize all children. Perhaps what children are learning are simply "positional productive patterns" or "positional associative patterns", which are learned individually. Braine argues that each positional productive pattern is semantically consistent. For example, 'more X' means the thing that recurs or whose recurrence is desired; 'other X' means indication of or request for some object distinct from the one at hand.

"We can infer from the productivity and semantic consistency of the patterns that they represent formulae for expressing the meaning associated with the pattern, formulae that map semantic elements into phrase positions, that is, they define where the words expressing component concepts go in the utterance-- for example, 'more +X' is a formula for expressing recurrence of X" (Braine 1976: 9).

MacWhinney (1982) is an early non-nativist account that describes three different kinds of strategies relevant to syntactic development: rote learning, analogy and combination. He goes through different kinds of evidence that support rote learning and analogy and at the same time observe inadequacies of these processes for grammar learning. He discusses item-based pattern learning, citing the range of evidence for this mode of learning as well as its limitations. The more recent studies of Tomasello and Lieven have further developed the early views of Braine and MacWhinney, drawing empirical evidence from extensive corpora data and experimental studies.

### **3. Multi-word combinations in early child Cantonese**

To see whether the early word combinations of children reflect frozen units or productive patterns, we used data from the two youngest children in the Hong Kong Cantonese Child Language Corpus: MHZ and CKT (Lee and Wong 1998). The first child MHZ was observed from 1;6 to 2;8 in 26 sessions, while the second child CKT was recorded from 1;5 to 2;7 in 25 sessions. To observe the early word combinations in child Cantonese in closer detail, we adopt the explicit coding scheme of Lieven et al (1992, 1997) with some minor adaptations. We would like to address the issue whether the

majority of children's early combinations are frozen phrases or lexically-based combinations. In our analysis, we point out the severe rigidity of Lieven's methodology and its inadequacies in capturing data like Cantonese. It will be argued that its central weaknesses stem from a refusal to acknowledge categorial information, and an over-reliance on word-hood properties of English.

### 3.1 The Lieven et al (1992, 1997) Coding scheme

The coding scheme of Lieven et al is given in Table 1, with examples of coding classifications given in (6). Three kinds of multi-word combinations are identified: frozen phrases, intermediate combinations and constructed combinations.

**Table 1. The Coding Scheme of Lieven et al (1992, 1997) and Pine and Lieven (1993)**

<p><u>Multi-word utterances</u>: utterances which consist of more than one word (as defined by the adult language) including forms which are frozen in adult speech or in the child's, but excluding word compounds and reduplications.</p>
<p><u>Frozen phrases</u>: utterances which contain two or more words which have not previously occurred alone in the child's vocabulary or which contain one such word, provided it has not occurred in the same position in a previous multi-word utterance.</p>
<p><u>Intermediate utterances</u>: utterances in which            (i) all the words or phrases have previously occurred independently in the child's vocabulary, provided none of them has occurred in the same position in two previous multi-word utterances; and/or            (ii) one or more of the words or phrases has occurred in the same position as in one but only one other previous previous multi-word utterance, provided the word or phrase which makes up the remainder of the utterance has already occurred independently in the child's vocabulary.</p>
<p><u>Constructed utterances</u>: utterances which            (i) contain one or more words or phrases which have occurred independently in the child's vocabulary together with a word or phrase which has occurred in the same position in at least two other previous multi-word utterances, or            (ii) conform to a positional pattern as in (i) which is already established in the child's vocabulary, regardless of whether the variable word has occurred independently.</p>

(6) Examples of classification according to Lieven et al's coding scheme:

Example 1:

Words in child's single-word vocabulary:

Daddy

Cat

Dog

Multiword utterances:

Daddy gone                      frozen phrase

Cat gone                            intermediate

Dog gone                            constructed

Example 2:

Words in child's single-word vocabulary:

Gone  
Mummy  
Car  
Plane

Multi-word utterances:

Mummy gone	intermediate
Car gone	intermediate
Plane gone	constructed

Example 3:

Words in child's single-word vocabulary:

Gone  
ball  
dolly  
birdie  
dinner

Multi-word utterances:

ball gone	intermediate
book gone	frozen phrase
dolly gone	constructed
birdie gone	constructed
dinner gone	constructed

To understand how this taxonomy works, let us consider the simple case of a combination of two words X+ Y. To qualify as a frozen phrase, at least one of them has to be a bound form. ie either X or Y has not occurred as an independent utterance in previous discourse. One can still classify an 'X Y' combination as frozen if one of them has occurred alone previously, but neither has occurred in the same position in a previous utterance. Thus in Example 1, the words *daddy*, *cat* and *dog* have occurred alone previously. The combination *Daddy gone* is frozen, because while *Daddy* has occurred independently, *gone* has not. In Example 3, *book gone* counts as a frozen phrase, since only *gone* has occurred alone previously, and *book* has not occurred alone or in the same position in any previous combination.

To qualify as a constructed combination, both X and Y have to show some productivity in terms of either their ability to occur alone or their tendency to occupy the same slot in two previous combinations. Either both X and Y are free forms and one of them has occurred in the same slot in two previous combinations, or one of them is free and the other has occurred in the same slot in two previous combinations. In Example 1, *cat gone* is not a constructed combination, since while *cat* has occurred alone, *gone* has not; neither has *gone* occurred in the same slot in two previous combinations. On the other hand, *dog gone* will be a constructed utterance, since *dog* has occurred alone, and *gone* has occurred in the same position in *daddy gone* and *cat gone*. In Example 2, both *mommy* and *gone* are free forms, having occurred alone in previous utterances. However, *mommy gone* still fails to be classified as constructed, since *gone* has not occurred in the same slot in two previous utterances. In contrast, *plane gone* is constructed because both are free forms, and *gone* has occurred in final position in *mommy gone* and *car gone*.

Intermediate utterances are combinations that are not frozen, but whose components do not show as much productivity as those of constructed utterances. If X and Y are free

forms, and neither has occurred in the same position in more than one previous combination, 'X Y' would be regarded as an intermediate form. Or if one of them is a free form, and the other component has occurred in the same position in exactly one previous utterance. In Example 1, *daddy gone* is classified as intermediate; since *daddy* is a free form, and *gone* has occurred in the same position in one and only one previous utterance. In Example 2, *car gone* is also categorized as intermediate for the same reason. Both *car* and *gone* are free forms, but the only component that has recurred in the same position in a previous utterance is *gone*, and it has only appeared thus once.

In applying Lieven's coding scheme to the Cantonese data, we had to make some modifications with respect to functional categories such as the negator *m4*, determiners, aspect markers, and final particles, as these never occur as free forms. Thus productivity for these items will be defined according to their occurrence in the same position in two previous utterances. We examined all the combinations of MHZ and CKT before 2;0, and ascertained whether the component morphemes have occurred alone in previous utterance and in how many of previous combinations they have occurred in the same position.

### 3.2 Analysis of all the spontaneous verb-containing word combinations of MHZ and CKT before 2;0 (adopting the coding scheme of Lieven et al)<sup>2</sup>

Table 2A gives a breakdown of the verb-containing word combinations of MHZ before 2;0. There were 241 such combinations of which about only one quarter (55) can be classified as frozen phrases. Table 2B gives a breakdown of the verb-containing word combinations of CKT before 2;0. There were 364 such combinations of which about less than one quarter (77) can be classified as frozen phrases. These are much lower percentage than what one might expect given the implications of the conservative claims of Lieven et al.

**Table 2A. Types of combinations in the multi-word combinations of MHZ before 2;0**

Age	frozen phrases	intermediate utterances	constructed utterances	total number of multi-word utterances
1;7-2;0	55 (23%)	109 (45%)	77 (32%)	241

**Table 2B. Types of combinations in the multi-word combinations of CKT before 2;0**

Age	frozen phrases	intermediate utterances	constructed utterances	total number of multi-word utterances
1;5-2;0	77 (21%)	69 (19%)	218 (60%)	364

To illustrate how the Lieven et al methodology is adapted for the analysis of Cantonese, Table 3A gives an exhaustive listing of the first 45 verb-containing word combinations of MHZ before 2;0, classified according to the three categories of Lieven et al. Table 3B gives a corresponding list of the first 45 such word combinations of CKT before 2;0. An 'I' following a constituent indicates that it has occurred alone in previous utterances. We counted identical combinations in successive conversational turns as one since we are interested in combination types and not tokens.

To see how the scheme is applied to the Cantonese data, let us consider the first combination *daap ce* ("take car"). Both have occurred in prior discourse as independent utterances, but neither has occurred in the same slot in any previous combination, this

being the very first combination produced by the child. Thus the utterance is identified as intermediate. On the other hand, the third and fourth combinations produced in 1;8 are frozen phrases. The combination *daai ze* ("bring umbrella") consist of two components neither of which has occurred alone or in the same slot in a previous utterance. By the same token, in the combination *paa tai* ("climb ladder"), while *paa* ("climb") has appeared independently, *tai* ("ladder") has not; nor has it appeared in the same slot in a previous utterance. Therefore this utterance is also classified as intermediate.

In 1;09;18, we observe the occurrence of the sequence *sik1 aa3* ("know-sfp"). The word *sik1* has appeared alone, and the final particle has occurred in final position in two previous utterances, ie *co ce aa3* ("ride car sfp") and *hai5 dou aa3* ("exist here sfp"). Thus we label this utterance as constructed. A neighboring combination *bei zeze* ("give sister") is also considered constructed, since both *bei* and *zeze* have occurred independently. Further, *bei* has appeared in initial position in two previous positions: *bei aa3* ("give sfp") at 1;08;14, and *bei Houzeon* ("give MHZ") at 1;09;04.

**Table 3A. The first 45 verb-containing word combinations of MHZ and their coding.**

**(An 'I' means that the form has previously occurred as an independent utterance in the child's speech. In the case of forms that only occur bound in the language, they will be considered as independent if they have occurred in the same position in two or more previous utterances. Identical combinations in successive conversational turns are counted as one).**

Age	Utterance	Type
1;08;00	daap [I] ce [I] ride car	intermediate
	m sik not know	frozen
	nei bong ngo aak1 you help me sfp	frozen
	daai ze bring umbrella	frozen
	paa [I] tai climb ladder	frozen
1;08;14	bei [I] aa3 [I] give sfp	intermediate
1;08;28	hai caak [I] ngaa be brush tooth	frozen
1;09;04	hoi dang turn-on light	frozen
	zeze [I] mit sister tear-off	frozen
	m [I] hai not be	frozen
	bei [I] Houzeon [I] give MHZ	intermediate
	ke maamaa [I] ride horse	frozen

	jam [I] milk-milk drink milk	frozen
	sik [I] caangcaang [I] eat orange	intermediate
	sai tau wash head	frozen
	m [I] hai not be	frozen
1;09;18	heoi co cece [I] go ride car	frozen
	co cece [I] aa3 [I] ride car sfp	intermediate
	ke [I] maamaa [I] ride horse	intermediate
	jau aa3 [I] have sfp	frozen
	zong maamaa [I] bump horse	intermediate
	daap [I] cece [I] take car	intermediate
	baai [I] nei [I]-di put this-CL (=these)	intermediate
	baai [I] aa3 [I] put sfp	intermediate
	daap [I] cece [I] take car	intermediate
	daap [I] go-gaa take that-CL	intermediate
	daap [I] nei [I]-gaa take this-CL	intermediate
	ke [I] maamaa [I] ride horse	intermediate
	fangaa u aa4 sleep q-sfp	frozen
	ke [I] maamaa [I] ride horse	intermediate
	co ce[I] sit/ride car	frozen
	co ce[I] aa3 [I] sit/ride car sfp	intermediate

hai5 dou aa3 [I] exist here sfp	frozen
sik [I] aa3 [I] know sfp	constructed
laan zo5 aa3 [I] break asp sfp	intermediate
bei [I] zeze [I] give sister	constructed
ngo [I] sik [I] aa3 [I] I know sfp	constructed
laan zo5 aa3 [I] break asp sfp	intermediate
ngo [I] cai [I] I pile-up	intermediate
co ce[I] aa3 [I] sit/ride car sfp	intermediate
laan zo5 aa3 [I] break asp sfp	intermediate
caak aa1 [I] break-open sfp	intermediate
baai [I] haphap [I] put box	constructed
baai [I] geigei [I] put plane	constructed
maat gei [I] wipe plane	intermediate

**Table 3B. The first 45 verb-containing word combinations of CKT and their coding. (An 'I' means that the form has previously occurred as an independent utterance in the child's speech. In the case of forms that only occur bound in the language, they will be considered as independent if they have occurred in the same position in two or more previous utterances. Identical combinations in successive conversational turns are counted as one).**

Age	Utterance	Type
1;05;22	bei ngo aa3/ bei ngo give me sfp/ give me	frozen
1;07;03	geng aa3 [I] fear sfp <10522: biubiu aa3 watch sfp 1;05;22: gaaigaai aa3 street sfp>	frozen
	tai aa3 [I] look sfp	frozen

1;07;10	tai aa3 [I] look sfp  sau aa3 [I] collect sfp  zai aa3 [I] put sfp  bei ngo give me	frozen  frozen  frozen  frozen
1;08;00	pou pou [I] aa3 [I] carry-in-arm sfp  fing aa3 [I] sling sfp  co ce [I] sit/take car  zaa zyu grasp DUR-asp  bei [I] aa3 [I] give sfp  co aa3 [I] sit sfp  co cece [I] sit/take car  deng aa3 [I] fling sfp  zing laan make break  pou [I] aa4 [I] carry-in-arm sfp <1;07;13: mou aa4 neg sfp 1;08;00 kam4 aa4 piano sfp>  lo aa4 [I] get sfp  zaa zyu grasp DUR-asp  geng aa3 [I] fear sfp  m jam laa3 neg drink sfp  co [I] aa3 [I] sit sfp	constructed  frozen  frozen  frozen  constructed  intermediate  constructed  frozen  frozen  constructed  frozen  frozen  intermediate  frozen  constructed

	tai [I] aa3 [I] look sfp  zeze [I] zai1 [I] sister put  baai [I] aa3 [I] put sfp  geng aa3 [I] fear sfp	constructed  intermediate  constructed  constructed
1;08;07	baai baai baai [I] aa3 [I] put put put sfp  zeze [I] sik [I] sister eat  baaibaai aa3 [I] put sfp  co [I] cece [I] aa3 [I] sit car sfp  bei [I] ngo give me  bei [I] aa3 [I] give sfp  bei [I] ngo [I] give me  zi [I] aa1 [I] know sfp <1;08;07: faaidi aa1 quicker sfp hou aa1 good sfp>  zaa1 zyu grasp DUR-asp  bei [I] aa4 [I] give sfp  baaibaibaai [I] aa3 [I] put sfp  taai [I] aa3 [I] look sfp  paak [I] aa3 [I] hit sfp  zaizai [I] aa4 [I] put sfp	constructed  intermediate  constructed  constructed  intermediate  constructed  constructed  constructed  frozen  constructed  constructed  constructed  constructed  constructed
1;08;21	geng aa3 [I] fear sfp	constructed

zaa zyu grasp DUR-asp	frozen
gaau co get wrong	frozen
hoi ce [I] aa3 [I] start car sfp	frozen
gam6 [I] aa3 [I] press sfp	constructed
hoi no3 open sfp	frozen
ngo [I] daa [I] I hit	intermediate
ngo [I] pou [I] aa3 [I] I carry-in-arm sfp	constructed
ngo [I] jiu [I] I want	constructed

Table 4 gives examples of the constructed utterances of MHZ before 2;0, which were ascertained according to the coding scheme. In some of these cases, one could identify the pattern without much difficulty, as in the pattern '*bei*+X' ("give X") or the pattern '*baai* + X' ("put X"), where X would be a variable word that can be filled in. In other cases, however, one is hard put as to what the central pivoting element is in the combination. For example, in the combination *jam naainai* ("drink milk") produced at 1;09;25, which is the pivoting element and which the variable? Since Lieven's methodology makes no reference to categories, sometimes it may be some instances of X that establish the pattern as constructed though syntactically it is the other element that forms the pivoting element. In this particular example, while *naainai* ("milk") has recurred in previous combinations, it is *jam* ("drink") that is central to the construction rather than the word for milk. This kind of important distributional fact will be missed by accounts that do not make reference to categories. Another kind of difficulty one may observe is that final particles are extremely versatile forms that combine with many other combinations to form an utterance. If one defines distribution entirely in terms of linear position, one can see that 'X +sfp' will be an easily established productive pattern. But the kinds of combinations that come under 'X +sfp' will be extremely varied. In that case, Lieven's distributional methodology is too weak in picking out the syntactic characteristics of early word combinations.

**Table 4. Examples of the constructed utterances of MHZ before 2;0 (based on Lieven-style criteria)**

Age	Utterance	Lexically-based pattern
1;09;18	sik [I] aa3 [I] know sfp	X+ aa3 sfp
	bei [I] zeze [I] give sister	bei + X give
	ngo [I] sik [I] aa3 [I] I know sfp	X+ aa3

	baai [I] haphap [I] put box	baai + X put
	baai [I] geigei [I] put plane	baai + X
1;09;25	teoi [I] cece [I] push car	X + cece
	bei [I] zeze [I] give sister	bei + X give
	jam [I] naainaai [I] drink milk	??
	bei [I] zeze [I] give sister	bei + X give
	ngo [I] jam [I] naainaai [I] I drink milk	?
	ngo [I] cai [I] I pile-up	X + cai pile-up
	nei bong ngo [I] cai [I] you help me pile-up	X + cai pile-up
	ngo [I] bei [I] nei I give you	ngo + X I
	ngo [I] caak [I] ngaa I brush tooth	ngo + X I
	jam [I] naai [I] drink milk	?
	daap [I] cece [I] take car	X + ce ? car
	gogo [I] zaa ce [I] boy drive car	?

#### 4. Problems with the surface distributionist account

To sum up, in our detailed analysis of the first multi-word utterances of MHZ and CKT, not only have we identified more syntactically productive patterns than frozen phrases (77 to 55 out of a total of 241 for MHZ, and 287 to 77 out of a total of 364 utterances for CKT), we have also experienced inherent difficulties of the methodology that arise from its excluding categorial information from consideration. An approach like Lieven's is bound to be beset with a number of problems.

As underscored by Wexler and Culicover (1980) and Wexler (1982), the question of learnability and continuity with the adult language should be tackled right from the start. If children do not have any categories to begin with, how are they to progress to acquire the complex syntactic categories and principles that underlie structure-dependence of transformations, subadjacency and binding?

Second, the coding scheme of Lieven et al is far too restrictive in excluding first time occurrences as indicative of productive patterns. For example, if a child says *Daddy drink* and *drink milk*, and later says *Daddy drink milk*, the latter utterance will not be classified as constructed according to Lieven's criterion, since one can argue that *drink* has not occurred in the same slot in two previous utterances. But one could see that the child may be generalizing from its knowledge of subject-predicate and verb-object structures (cf. McNeill 1966).<sup>3</sup> For the combinations in MHZ at 1;10;23, *soe waat tai* ("slide-down slide") and *toutou soe waat tai* ("rabbit slide-down slide"), the classification is 'frozen,' even though we see that the child is actively tagging a subject to a predicate.

The methodology is overly dependent on the distribution of words rather than groupings of words. As a result, we run into problems of classification of the pivot of lexically based combinations, in the case of utterances in which do not contain a clear functional categorial element (e.g. a sequence of two nouns or two verbs). In other cases, the pivot may not correspond to what we intuitively think of as the most important element in the phrase.

The methodology may give arbitrary results. Two combinations of apparently the same type may be grouped as constructed in one case but intermediate as another, because of idiosyncrasies of the methodology. In 1;09;25, the utterance *jam naainaai* ("drink milk" ) is regarded as constructed since *jam* is a free form, but *sik naai* ("eat milk") is classified only as intermediate, since *sik* ("eat") has not occurred alone in previous utterances at this point. These counterintuitive features of the methodology reflect the unreasonable implicit denial of the possibility that some first-time combinations could be instances of productive syntactic patterns.

Lexically-based accounts of early syntactic development fail to capture the fact that the evolution from frozen and intermediate combinations to constructed utterances is a very rapid process. In our data, we see constructed occurrences occurring at the same time as frozen and intermediate phrases in a span of two or three months. If one limits one's conception of early syntax to inductive generalization from frozen phrases, one still needs to explain why that transition is often a rapid process (cf. Lee and Tsay 2001, Zhang and Fang 2003, Xiao, Yang and Yang 2003). Further, since this is an account based purely on language production, it may underestimate children's syntactic competence, which is reflected more fully in comprehension (cf. the studies on Principal branching direction by Lust and Chien 1984).

## 5. Conclusions

Early child Cantonese shows that frozen phrases account for less than one-quarter of children's multi-word combinations, contrary to what the accounts of Lieven et al would lead us to believe. The claims of Lieven et al (1992, 1993, 1997) seem to be heavily influenced by the particular way in which coding of combinations is carried out. Serious problems remain with lexically-based accounts of syntactic acquisition with respect to how the categorial distinctions of constructions evolve, the relatively rapid spread of constructions and continuity with adult syntax.

## References

- Bowerman, M. 1973. "Structural relationships in children's utterances: syntactic or semantic?", in T. Moore. ed. *Cognitive Development and the Acquisition of Language*. New York: Academic Press. Reprinted in L. Bloom ed. 1978. *Readings in Language Development*. New York: John Wiley.
- Braine, M. 1963. "The ontogeny of English phrase structure: the first phase", *Language* 39: 1-13.
- Braine, M. 1976. *Children's first word combinations*. Monographs of the Society for research in child development vol. 41, no.1.
- Chien, Yu-chin and K. Wexler 1987. "A comparison between Chinese-speaking and English-speaking children's acquisition of reflexives and pronouns", paper presented at the 1987 Boston University Conference on Language Development.
- 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.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. 1971. *Problems of Knowledge and Freedom*. New York: Vintage.
- Chomsky, N. 1972. *Language and Mind*. Enlarged Edition. New York: Harcourt, Brace and Jovanovich.
- Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Crain, S. 1991. "Language acquisition in the absence of experience", *Behavioral and Brain Sciences* 14, reprinted in P. Bloom ed. 1994. *Language Acquisition: core readings*. Cambridge, MA: MIT Press.
- Crain, S. and M. Nakayama 1987. "Structure dependence in grammar formation", *Language* 63: 522-43.
- Crain, S. and R. Thornton 1998. *Investigations in Universal Grammar*. Cambridge, MA: MIT Press.
- Deprez, V. and A. Pierce 1993. "Negation and functional projections in early grammar", *Linguistic Inquiry* 24.1: 47-85.
- Lee, T. and C. Wong 1998. "Cancorp-The Hong Kong Cantonese Child Language Corpus", *Cahiers de Linguistique Asie Orientale* vol. 27, no. 2: 211-228.
- Lee, T. and J. Tsay 2001. "Argument structure in the early speech of Cantonese-speaking and Taiwanese-speaking children", paper presented at the Tenth Annual Conference of the International Association of Chinese Linguistics, June 22-24, 2001, UC Irvine
- Lieven, E., J. Pine and H. Barnes 1992. "Individual differences in early vocabulary development: redefining the referential-expressive distinction", *Journal of Child Language* 19: 287-310.
- Lieven, E., J. Pine and G. Baldwin 1997. "Lexically-based learning and early grammatical development", *Journal of Child Language* 24: 187-219.
- Linguistic Society of Hong Kong Jyutping Editorial Group 1997. *Guide to LSHK Cantonese Romanization of Chinese Characters*. Hong Kong: Linguistic Society of Hong Kong.
- Lust, B. and Yu-chin Chien (1984) "The structure of coordination in first language acquisition of Mandarin Chinese: evidence for a universal", *Cognition* 17: 49-83.
- MacWhinney, B. 1982. "Basic syntactic processes", in S. Kuczaj III ed. *Language Development Vol. 1: syntax and semantics*. Hillsdale, NJ: Lawrence Erlbaum.
- Nelson, K. 1975. "The nominal shift in semantic-syntactic development," *Cognitive Psychology* 7: 461-479.
- Otsu, Y. 1981. *Universal Grammar and Syntactic Development in Children: toward a*

- theory of syntactic development. PhD Dissertation. MIT.
- Pine, J. and E. Lieven 1993. "Reanalyzing rote-learned phrases: individual differences in the transition to multi-word speech", *Journal of Child Language* 20: 551-571.
- Pinker, S. 1984. *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.
- Pinker, S. 1987. "The bootstrapping problem in language acquisition", in B. MacWhinney (ed) *Mechanisms of Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum.
- Poeppel, D. and K. Wexler 1993. "The full competence hypothesis of clause structure in early German", *Language* 69.1: 365-424.
- Tomasello, M. 1992. *First Verbs*. London: Cambridge University Press.
- Tomasello, M. 2000a. "Do young children have adult syntactic competence?" *Cognition* 74: 209-253.
- Tomasello, M. 2000b. *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.
- Tomasello, M. 2003. *Constructing a Language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Wexler, K. 1982. "A principle theory for language acquisition", in E. Wanner and L. Gleitman eds. *Language Acquisition: the state of the art*. Cambridge University Press.
- Wexler, K. and Yu-chin Chien 1985. "The development of lexical anaphors and pronouns", *Papers and Reports on Child Language Development* 24: 138-149.
- Wexler, K. and P. Culicover 1980. *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.
- Xiao, Dan, Yang Xiaolu and Yang Li 2003. "The BA construction in Child Mandarin," paper presented at the First Yuelu Language Acquisition Workshop, December 12-14, Hunan University, China.
- Zhang Qingwen and Fang Li 2003. "The acquisition of sentence final particles in Mandarin Chinese," paper presented at the First Yuelu Language Acquisition Workshop, December 12-14, Hunan University, China.

### 粤语儿童早期多词联组的句法能产性

李行德

湖南大学认知科学研究所

摘要： 本文考察粤语儿童多词联组的句法能产性，探讨早期儿童语句是反映语言发展连续性理论所假设的先天语法范畴，还是反映以词汇的位置分布为基础的归纳习得模式。结果表明固定词组只占粤语儿童多词联组不到四分之一，并不支持仅以词汇位置分布为获得机制的理论。

关键词： 句法获得，连续性理论，粤语儿童语言

---

### Notes

<sup>1</sup> Nelson (1975) observed that children differed in the amount of referential nominals vs social/personal words they used, and described this difference in terms of the referential/ expressive dichotomy.

<sup>2</sup> Romanization of Cantonese morphemes is based on the Linguistic Society of Hong Kong Jyutping Romanization system (1997). The numbers at the end of romanizations indicate the tones of the morphemes concerned. Except for the focus particles, sentence final particles and interjections, the tones of

---

Cantonese morphemes have been left out in the example sentences. The following abbreviations and conventions are used in this paper: 'm;n;p'= year;month;day; 'sfp'= sentence final particle; 'q-sfp'=question final particle; 'asp'=aspect marker.

<sup>3</sup> McNeill (1966) entertains the possibility that on the basis of a rule that expands a sentence into NP-V, and another rule that expands the sentence into V-NP, the child will be able to arrive at NP-V-NP structures.