The acquisition of nominal structure, word order and referentiality in Chinese: Corpus and experimental findings on the numeral phrase

Thomas Hun-tak Lee ¹ and Zhuang Wu ^{1, 2}

¹The Chinese University of Hong Kong ²Xiangtan University huntaklee@cuhk.edu.hk, woodblcu@yahoo.com.cn

In this paper we study two phenomena related to numeral phrases in Mandarin Chinese from the acquisition perspective: (a) First, we will clarify the nature of the so-called subject specificity constraint, in particular the types of numeral phrases that fail to occur in subject position; and (b) second, we examine the specificity difference between what has been called IMN (inner modifier nominals) and OMN (outer modifier nominals), each instantiating a different positioning of the prenominal modifier.

The two phenomena are complex and have been subject to different analyses in the literature. The subject specificity constraint says that a sentence with a numeral phrase subject, such as *yige tongxue dao le* "A classmate has arrived" in (1) is ungrammatical, but then in many contexts such sentences sound quite natural, and in fact corpus data show that native speakers use numeral phrase subjects quite a lot. The distinction between IMN and OMN, as illustrated by the meaning difference between the sentences in (2) and (3), is said to be one of specificity. Scholars (notably Zhang 2006) argue that nominals like *dai yanjing de san ge xuesheng* (OMN) in (3), with the modifier in outer position preceding the numeral-classifier, are specific but not definite, whereas nominals like *san ge dai yanjing de xuesheng* (IMN) in (2), with the modifier in inner position, following the numeral-classifier, can be specific or non-specific.

- (1) ??yi ge tongxue dao le.
 one CL classmate arrive asp
 "A classmate has arrived"
- (2) san ge dai yanjing de tongxue (IMN; specific/ non-specific) three CL wear glasses DE classmate
- (3) dai yanjing de san ge tongxue (**OMN**; specific) wear glasses DE three CL classmate

First we examined numeral phrases in the naturalistic production of two Mandarin-speaking children from the Beijing Child Early Language Acquisition (BJCELA) corpus. For both the children's production and the child-directed adult speech, all the numeral phrase tokens were

extracted and classified according to their linguistic form. Then the different types of numeral phrases were divided into categories according to the syntactic positions in which they occurred. The results show that, (a) The early use of numeral phrases by the two children was adult-like. The most productive forms were Numeral-Classifier, and then Numeral-Classifier-Noun and Classifier-Noun; (b) similar to adults, children generally favored the object position for numeral phrases to occur. Although the subject numeral phrases in the children's naturalistic production were exclusively definite or quantity-denoting. No non-specific subject numeral phrases were found in the children's language, suggesting that the onset of subject specificity constraint is very early. Children were also sensitive to the prohibition against Classifier-Noun nominals in subject position, using these structures only as objects; (c) numeral phrases with inner modifiers (IMN) were rare in children's naturalistic speech. Those with outer modifiers (OMN) were not found at all in child speech. The rare occurrence of both IMN and OMN in child language leads us to hypothesize that the mapping between the internal order of numeral phrase and specificity may be late in acquisition.

Next we tested this hypothesis through an experiment. Our experimental design made use of the distinction between IMN and OMN in relation to focus: while the numeral in an IMN can be stressed and thus focused by a restrictive operator (4a), the numeral in an OMN cannot (4b). In other words, while (4)a is ambiguous between a quantity reading and an individual reading, (4)b only has the individual reading.

(4) a. Xiaohouzi zhi nazhe <u>san ge bole pi de xiangjiao.</u> 小猴子 只 拿着 三个 剥了 皮 的 香蕉 (IMN)

monkey only hold-asp three CL peel-asp skin NOM banana

"The monkey is holding only three peeled bananas (and nothing else)."

"The monkey is holding only THREE peeled bananas (not other quantities of peeled bananas)."

b. Xiaohouzi zhi nazhe <u>bole pi de san ge xiangjiao.</u>

小猴子 只 拿着 剥了皮的 三个香蕉 (OMN)

The monkey only hold-asp peel-ASP skin NOM three CL banana

"The monkey is holding only three peeled bananas (specific ones) (and nothing else)."

*"The monkey is holding only THREE peeled bananas (not other quantities of peeled bananas)."

We used a picture verification task with 4- to 5-year-old Mandarin-speaking kindergarteners in Shenzhen (N=30), as well as a control group of adult subjects (N=24). The following 3 types of sentences were included, with 4 sentences for each type, each sentence paired with an affirming situation and a falsifying situation. Altogether there were 24 core test sentences (3 types x 4

tokens x 2 situations). For each sentence type, two of the four test tokens had agentive subjects and theme objects, while the other two had locative subjects and theme objects. The experimental design made use of the difference between IMN and OMN in their interactions with restrictive focus (cf. 4 above), to create situations in which the two structures will yield a truth conditional difference. The only difference between Type A and Type B items is that the former type has the numeral stressed, making the quantity reading salient, while the latter does not have the numeral stressed. These two types were tested on two groups of subjects, which means a group of subjects received Type A and Type C sentences, while another group received Type B and Type C sentences, with each subject receiving 16 items performed in two separate sessions: the core test items (2 sentence types x 4 tokens x 2 situations), 4 warm-up trials, 8 pretest items examining their understanding of *zhi* "only", and 46 fillers.

(5) <u>Types of IMN/OMN test items</u>

Type A. (IMN, numeral stressed)

Xiaohouzi **zhi** nazhe **SAN GE** bole pi de xiangjiao. monkey only hold-asp three CL peel-asp skin NOM banana

"The monkey is holding only THREE peeled bananas (not other quantities of peeled bananas)."

Type B. (IMN, numeral not stressed)

Xiaohouzi zhi nazhe san ge bole pi de xiangjiao. monkey only hold-asp three CL peel-asp skin NOM banana

"The monkey is holding only THREE peeled bananas (not other quantities of peeled bananas)."

"The monkey is holding only three peeled bananas (and nothing else)"

Type C. (OMN)

Xiaohouzi zhi nazhe bole pi de san ge xiangjiao. monkey only hold-asp peel-asp skin NOM three CL banana

"The monkey is holding only the three peeled bananas (specific ones) (and nothing else)."

Picture 1. Critical situation Picture 2 Counterbalancing situation 1





Picture 3 Counterbalancing situation 2



A sharp contrast between children and adults was observed in the findings. Adults differentiated OMN and IMN sharply, with 100% of the adults consistently rejecting Type C OMN sentences in a falsifying situation (i.e. an extra object with different attribute situation) and only 58.3% of them rejecting Type B IMN sentences for the same situation, as the IMN sentences are ambiguous and may be true or false depending on the interpretation. Only a low percentage of adults (8.3%) rejected the test sentences when the numeral is stressed in the falsifying situation when the numeral is stressed (Type A). If the numeral is not stressed, the percentage of subjects who consistently rejected the sentences rises to 58.3% (Type B). Around 57% of the children consistently rejected the Type B IMN sentences (see Figure 2), and around 60% of them rejected the OMN sentences, suggesting that they were treating IMN and OMN on a par with each other. Around 57% of the children consistently rejected the IMN sentences, whether the numeral is stressed or not. Our findings clearly point to a lack of sensitivity to the OMN/IMN distinction, and delayed mastery of properties on the syntax-semantics interface in language development.

To conclude, our detailed analysis of the early language of two Beijing children indicate that Mandarin-speaking children are sensitive to the Subject Specificity Constraint from the very earliest stage of grammatical development, using numeral phrases primarily in object position for both specific and non-specific reference, and using them in subject position only for definite reference or for denoting quantities.

Children produce IMN sparsely, and do not use OMN at all, in their naturalistic speech. While IMNs appear to some extent in the adult input to children, OMNs are virtually absent. Thus the distinction between IMN and OMN is not visibly present in the positive evidence received by the children. Our experimental study of the difference between IMN and OMN, as reflected in their distinct interactions with stress and focus, has confirmed that late development of this interface property in child grammar.