Available online at www.sciencedirect.com

ScienceDirect

Lingua 164 (2015) 189–214

Lingua

www.elsevier.com/locate/lingua

Complex numerals in Mandarin Chinese are constituents

Chuansheng He *

Department of English/Institute of Language and Cognition, College of Foreign Languages, Hunan University, Changsha City, Hunan Province 410082, China

Received 7 June 2014; received in revised form 15 June 2015; accepted 23 June 2015
Available online

Abstract

In this article, three types of data from syntax, semantics, and morpho-phonology are presented to argue that Ionin and Matushansky's (IM, 2006) non-constituency analysis of complex numerals may not work in Mandarin Chinese and that the traditional constituency analysis is more plausible. In order to address these data under the constituency analysis, we propose and justify a fine-grained phrase structure for complex numerals in Mandarin Chinese, which can accommodate all these data satisfactorily. We also investigate the cross-linguistic implications of our argumentation by showing that similar morpho-phonological data in many minority languages of China also support the constituency analysis for complex numerals.

Keywords: Complex numerals; Constituenthood; Phrase structure; Phrasehood; Mandarin Chinese

1. Introduction

Since Hurford's (1975) seminal work on the linguistic theory of numerals, it has been generally agreed that complex numerals like *five hundred* (multiplicative) and *fifty two* (additive) are phrasal constituents (the constituency analysis). This view has been widely accepted and defended, implicitly or explicitly, as a null hypothesis for the syntactic studies of complex numerals (Jackendoff, 1977; Selkirk, 1977; Borer, 2005; Corver and Zwarts, 2006; Kayne, 2010, among others), until it was challenged by some other studies. For example, Ionin and Matushansky (IM, 2006) proposed a novel syntactic analysis toward complex numerals, mainly based on the following case marking data in Russian (and some other languages).

(1) a. četyre šagá — IM's (30a)
four step$_{GEN.SG}$
'four steps'

b. šest' šagov — IM's (30b)
six step$_{GEN.PL}$
'six steps'

* Tel.: +86 18374825097.
E-mail address: hcsmail@163.com.

http://dx.doi.org/10.1016/j.lingua.2015.06.014
0024-3841/

(2) a. četyre tysjači šagov IM's (32a)
four thousand$_{GEN.SG}$ step$_{GEN.PL}$
'four thousand steps'

b. pjat' tysjač šagov IM's (32b)
five thousand$_{GEN.PL}$ step$_{GEN.PL}$
'five thousand steps'

In the above examples, the morphological case on the lexical nouns and numerical bases seems to depend on the numerals. Following the standard case theory that only a head can assign case to a nominal element in a c-commanding relation, IM argued that numerals like *two*, *hundred*, or *thousand* are nominal heads selecting lexical nouns or other numeral-noun combinations as complements. Therefore, a complex numeral expression like *two hundred thousand students* projects a complementative structure, as bracketed in (3), in which the complex numeral does not form a constituent (the non-constituency analysis).[1] According to this structural analysis, the expression (3) should be interpreted as, roughly, two groups of one hundred groups of one thousand students, a total of 200,000 students.

(3) [two [hundred [thousand [students]]]]

Consequently, an additive numeral expression like *fifty two students* should be transformed from a full NP coordination in which the head noun is either right-node-raised or PF-deleted as illustrated in (4), and should be interpreted as a sum of fifty students and two students, a total of 52 students.

(4) a. [[[fifty t_i] (and) [two t_i]] students$_i$]
b. [fifty ~~students~~] (and) [two students]

Thus, IM developed a structure-interpretation mapping theory for complex numerals by arguing that complex numerals are derived in syntax and compositional in semantics. It is possible, however, to favor the view that complex numerals are derived in syntax and compositional in semantics (see Section 6 for independent argumentation for the phrasehood of complex numerals in Mandarin Chinese) while remaining unconvinced by IM's suggestion that the non-constituency analysis for complex numerals is cross-linguistically applicable. It is generally acknowledged that languages differ a great deal in how complex numerals are formed (e.g., Hurford, 2003). In order to generalize some universal rules on the formation of complex numerals, it is necessary to analyze as many different languages as possible. It may be that there are no universal rules on the formation of complex numerals; it is possible that some languages have a constituency structure for complex numerals while others do not, or both structures are available even in the same language, as Danon (2011) proposed.

As is well known, Mandarin Chinese is a classifier language that does not have overt case-marking. Therefore data directly comparable with that used by IM is not available to determine how complex numerals are organized. Fortunately Mandarin Chinese exhibits some other special properties in the syntax, semantics, morphology and phonology of its numeral system, which may be useful in helping us understand the syntax of complex numerals. In this paper, we present three types of evidence to argue that complex numerals in this language are constituents, as bracketed below.[2]

[1] But see Meinunger (2015) for a recent treatment of the Russian case data under the constituency analysis.

[2] We assume that classifiers do not form constituents with head nouns (Li, 1998; Cheng and Sybesma, 1999; Zhang, 2013); instead they form constituents with numerals (Tang, 1990; Gao, 1994; Krifka, 1995). This choice does not affect the current topic. If we adopt the first option, the classifier-noun should undergo deletion or raising for additive numerals according to the non-constituency analysis. And in a more complex additive (i), *wàn gè xuéshēng* should undergo deletion or raising, as shown in (ii).

(i) èr shí sān wàn gè xuéshēng
two ten three 10,000 Cl student
'230,000 students'

(ii) a. [[èr shí t_i] [sān t_i] wàn gè xuéshēng$_i$]
b. [èr shí ~~wàn gè xuéshēng~~] [sān wàn gè xuéshēng]
If we adopt the second option, the classifier should undergo deletion or raising for additive numerals. And in (iii), *wàn gè* should undergo deletion or raising.

(iii) a. [[[èr shí t_i] [sān t_i] wàn gè$_i$] xuéshēng]
b. [[èr shí ~~wàn gè~~] [sān wàn gè] xuéshēng]]

The arguments to be presented in this paper apply to both structural analyses. However, in our argumentation, we sometimes only focus on one analysis, believing that readers can easily see the same point in the other analysis. Another reminder is that some arguments apply to both the deletion approach and raising approach; while others are mainly targeted at the deletion approach and may not be a problem to the raising approach if additional assumptions are made.