

Jiajin Xu, Lancaster
Corpus informed approach to Canonical Typology
With special reference to splitable compounds in Mandarin

Canonical typology has been recognised as a strategy of studying fuzzy and cross-border grammatical categories (Bond 2007; Corbett 2007; Nikolaeva 2008). In this paper, we propose a corpus-informed approach to canonical typology. Corpus linguistics and canonical typology can be employed as two methodological approaches to capturing typological generalisations, with the former being bottom-up and inductive, and the latter top-down and deductive. This paper will address the disputed splitable compound issue in Mandarin, in which corpus method will be used to inform the generalisations about verb compound canons.

In Mandarin, the dual (word/phrase) status of splitable verb noun/complement compounds (SCs henceforth) remains to be a matter of controversy (Chao 1968; Lu 1957; Lü 1959; Lü 1979; Packard 2003; Yu 2003; Zhou 2006). The two parts of the compound, the SC Head (SC_H) and the SC Tail (SC_T), can be joined to form one composite word, as in *dan1xin1* (to worry), or separated by intervening grammatical elements, as in *dan1-le yi1shang4wu3 xin1* (to be worried the whole morning), without jeopardizing its lexical semantics. This poses a challenge to the general typology of verb compounds.

We, however, categorise some key canonical properties of verb compounds in terms of grammatical and phonological criteria (c.f. Dixon & Aikhenvald 2001).

1. The key grammatical/structural criteria

Host dependency: head dependence > tail dependence

Host dependency criterion of the canonical approach perceives: 1) SCs with a clitic-like aspect marker (like perfective marker LE) as compounds instead of phrases; 2) SCs with resultative verb complements attached to the main verb; 3) and other modifiers (classifiers, modifiers etc) attached to SC_T, represented typically by a noun or complement.

The structural canon is also profiled by the lexical and grammatical patterning of SCs based on two corpora, the *Lancaster Corpus of Mandarin Chinese* (LCMC) and the *Lancaster Los Angeles Corpus of Spoken Chinese* (LLSCC), which are composed of one million words each, representing written and spoken Chinese respectively.

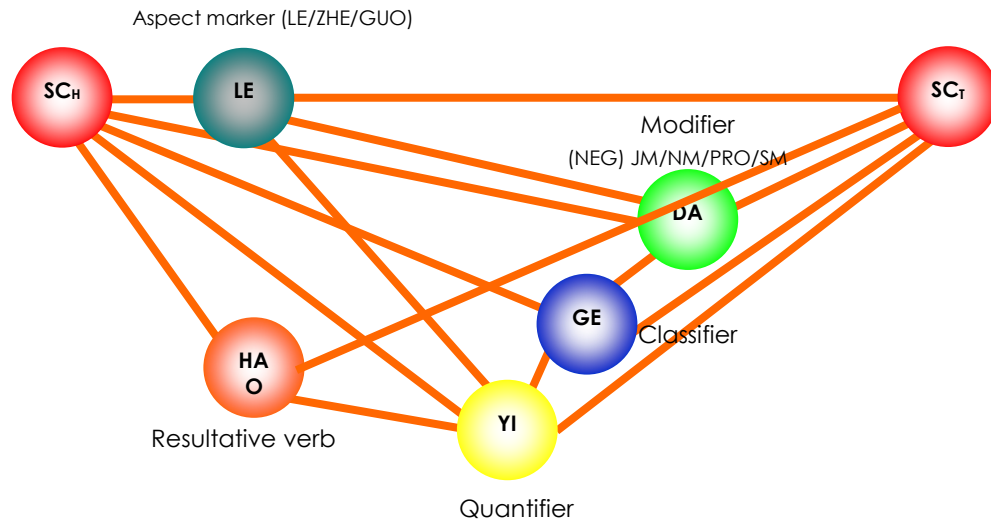


Figure 1: Lexical and grammatical patterning of SCs based on corpus analysis

2. Phonological criteria

We propose, based on corpus analysis, the following continuum criteria for SCs ($a > b > c$):

- a) Combined use of SC_H and SC_T will be argued to be compound;
- b) SC_H and SC_T separated by one single aspect marker will be considered as a compound, given that an aspect marker may be viewed as the near-inflectional suffix to the preceding verb; while
- c) SC_H and SC_T separated by other grammatical categories like quantifiers, adjectival modifiers etc. are seen as phrases.

In conclusion, this study starts with 2-million-word authentic language data and establishes the quantitatively informed canons on two interactive planes, grammar and phonology, to differentiate Chinese SCs. To make typological generalisations, computer corpora of different language families shall be collected. Against the criterial properties of verb compounds, we can check their actual linguistic behaviours, or their instantiations, to see whether they conform to the criteria. In this fashion, the conflation of the two methods will help to make quantitatively and qualitatively sound typological generalisations.

References

- Bond, O. (2007). Towards a canon for negation. In P. Austin, O. Bond, & D. Nathan (eds). *Proceedings of Conference on Language Documentation & Linguistic Theory*. London: SOAS.
- Chao, Y. (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Corbett, G. (2007). Canonical typology, suppletion, and possible words. *Language*, 83: 8-42.
- Feng, S. (2002). *The Prosodic Syntax of Chinese*. Munich: Lincom Europa.
- Lü, S. (1979). Problems in grammatical analysis in Chinese. Reprinted in S. Lü. (ed.). (1984). *Anthology of Chinese Grammar*. Beijing: The Commercial Press.

- Lu, Z. (1957). *The Word-formation of Chinese Language*. Beijing: The Scientific Press.
- McCarthy, J. & A. Prince. (1993). *Prosodic Morphology*. Ms. University of Massachusetts, Amherst and Rutgers University.
- Nikolaeva, I. (2008). Towards a typology of finiteness: A canonical approach. Paper presented at Universitaet Konstanz, January 24th, 2008.
- Packard, J. (2003). *The Morphology of Chinese*. Cambridge: Cambridge University Press.
- Yang, Q. (Ed.). (1995). *A Dictionary of Split Word Usage in Modern Chinese*. Beijing: Beijing Normal University Press.
- Yu, S. (2003). *Discontinuous Verb-object Compounds in Cantonese and Mandarin*. Unpublished MA dissertation. University of Hong Kong, Hong Kong.
- Zhou, S. (2006). *A Study on Chinese Split Words: The Peculiarities of Chinese Morphemes, Words, and Phrases*. Shanghai: Shanghai Foreign Language Education Press.