The Syntax and Semantics of Chinese Counting Numerals

Prof. Brendan Gillon, Yingrui He
McGill University

Background

There have been debates on whether or not counting numerals are constituents. Hurford (1987) provides a constituency analysis for numerals applicable universally. Ionin and Matushansky (2006) argue that numerals are not constituents. He (2015) shows that Mandarin Chinese numerals are constituents and applies Hurford's analysis to them. He's analysis, like Hurford's, overgenerates. We propose an amended version of He's analysis, which we formalize and for which we provide a semantics. To see better the advantages of our analysis, we illustrate its advantages by applying it first to three simpler sets of Chinese counting numerals, used in the pre-Tang era, known as the lower scale (*xià děng*), the middle scale (*zhōng děng*) and the upper scale (*shàng děng*). We then apply the analysis to the more complex contemporary counting numerals of Mandarin.

Topics we investigated

Pre-Tang Chinese counting numerals
Phrase structure rules and the packing strategy
Our alternative proposal
Contemporary Mandarin counting numerals

Acknowledgements

Many thanks to Prof. Brendan Gillon for his guidance throughout this project. Thanks also to the Arts Internship Office and to Mr. Harry Samuel, whose funding made this project possible.

References

Hurford, James R. 1987 *Language and Number: The Emergence of a Cognitive System*.

Zhao, Yuan-ren 1968 *A Grammar of Spoken Chinese*. Berkeley, California: University of California Press.

Qian Baocong 1932 *Zhōnggúo Suànxúe Shǐ*. Beijing, Republic of China: Guólì Zhōngyāngyánjiūyuànlìshǐyǔyányánjiūsǔo.

Data Analysis & Examples

The examples below illustrate portions of our work on three topics we investigated.

Hurford's phrase structure rules

Hurford uses the phrase structures rule below to generate counting numerals.

(1.1) Numeral → Digit
(1.2) Numeral → Phrase
(1.3) Numeral → Phrase Numeral
(1.4) Phrase → Numeral Measure

Hurford's packing strategy

However, as Hurford points out, his rules overgenerate. To exclude numerical expressions which are not numerical expressions of that language, Hurford proposes the Packing Strategy. It states:

> the sister constituent of a Number must have the highest possible value, that is, the highest value that a constituent of its category can have less than or equal to the value of the immediately dominating node (p. 246; cited from LTN 67–68).

Results

We set out three analyses of counting numerals. One is a formalization of He's analysis, a second is an equivalent formalization of He's analysis, which we think is easier to grasp, and a third is a new formal analysis, which does not overgenerate.

We investigate the adequacy of the three formalizations with respect to four systems of counting numerals. Three are pre-Tang, one is the contemporary modern Mandarin counting numerals.

Future work

We shall investigate the applicability of the analysis proposed here for Mandarin counting numerals to the counting numerals of other Chinese languages, building on other work by He.

Lower scale

We illustrate the main ideas of the paper with a sketch of their application to lower scale counting numerals.

There are 9 simple words for unit numbers,

yī (1), èr (2), sān (3), sì (4), wǔ (5), liu (6), qī (7), bā (8), jiu (9)

whose syntactic category we shall call 'Digit'.
There are 14 simple words for measure numbers,

shí (10^1), bǎi (10^2), qiān (10^3), wàn (10^4), yì (10^5), zhào (10^6), jīng (10^7), gāi (10^8), zǐ (10^9), ráng (10^{10}), gōu (10^{11}), jiàn (10^{12}), zhèng (10^{13}), zǎi (10^{14})

whose syntactic category we shall call 'Measure'.

The counting numerals of the lower scale can be characterized by the following schema:

$(U\ M_{14})\ (U\ M_{12})\ \ldots\ (U\ M_2)\ (U\ M_1)\ (U)$

where M_i stands for the Chinese word in M whose value is 10^i. The counting numerals are instances of the schema where any of the parenthetical elements may be omitted but clearly not all of them.

A special case of the rules in (1), where `Numeral' in the last two rules is replaced by `Digit' and the rules in (2) below are equivalent.

(2.1) Numeral → Digit;
(2.2) Numeral → Digit Measure;
(2.3) Numeral → Digit Measure Digit.

Moreover, the special case of the rules in (1), constrained by the packing strategy, is equivalent to the rules in (2) constrained by a syntactic version of the packing strategy, which uses a ranking of the measure words. We say that M_i outranks M_k iff $i > k$.

Our syntactic rules, which use this ranking, are these.

(3.1.1) If $x \in U, y \in M$, then $xy \in D$;
(3.1.2) If (1) $x, y \in D$ and (2) l(x) outranks m(y), then $xy \in D$;
(3.1.3) nothing else is in D;
(3.2.1) If $x \in U$, then $x \in Nu$;
(3.2.2) If $x \in D$, then $x \in Nu$;
(3.2.3) If $x \in D$ and $y \in U$, then $xy \in Nu$;
(3.2.4) nothing else is in Nu.

Modifications of the rules in (3) properly analyze the counting numerals of the upper scale, the middle scale and Modern Mandarin. The rules in (1) and the rules in (2), even if supplemented with a packing strategy, progressively overgenerate. We prove these results.