Allomorphy of 'One' and 'Two' in Mandarin Chinese

Danfeng Wu (University of Oxford)

1. Introduction. In Mandarin, numerals '1' and '2' appear in different forms depending on the context. For example, in a multi-digit cardinal number, the multipliers of the singles and tens bases appear in the *absolute form* (1a), but multipliers of higher bases (e.g. of hundreds and thousands) can appear in the *absolute* or *contextual* form (1b-c). While the two forms of '2' have distinct lexical forms (*liang* and *er*), the forms of '1' are solely distinguished by tone: while its absolute form *yī* has high tone, its contextual form is subject to tone sandhi: it has falling tone *yì*, unless when followed by another falling tone (*wàn* in (1c)), in which case it surfaces as a rising tone *yí*.

(1)a. er shí {yī / er} b. {yì / liang / yī / er} bǎi
2.ABS ten {1.ABS / 2.ABS} '21 / 22' {1.CONT / 2.CONT / 1.ABS / 2.ABS} hundred '100/ 200'
c. {yí / liang / yī / er} wàn
{1.CONT / 2.CONT / 1.ABS / 2.ABS} ten.thousand '10000 / 20000'

***Analysis*.** Despite its apparently impoverished morphology, Mandarin actually distinguishes between bound and free morphemes: the contextual form is free like English *two*, while the absolute form is bound like English *twen-*. The cardinal phrase (CardP) must be spelled out as a free-standing word or phrase. A multi-digit cardinal involves additive conjunction of cardinals (contra Ionin & Matushansky 2018). This complex cardinal then combines with a right-branching classifier-noun structure (2) (following Cheng & Sybesma 1999). The contextual form of '1' undergoes (3).

(2)*Our analysis of additive complex cardinals + classifier + noun:*
[$_{ClP}$ [$_{CardP/\&P}$ [Multiplier Base] [$_{\&'}$ &^{0} [Multiplier Base]]] [$_{Cl'}$ Clas^{0} NP]]

(3)*Lexical Tone Sandhi Rule:* yì (CONTEXTUAL) → yí/__ falling tone

***Implications*.** Our analysis has consequences in two domains. First, we make a cross-linguistic generalization about the morphological form of multipliers in multi-digit cardinals that may challenge the containment analysis (e.g. Bobaljik 2012) but not the monotonicity-based approach (Graf 2019). Second, we observe that in Mandarin multiplicative bases cannot be elided, but classifiers can. This challenges Ionin & Matushansky's ellipsis-based approach to multi-digit cardinals, and suggests that multi-digit cardinals are conjunction of smaller cardinals.

2. Multi-digit cardinals involve addition of multiplier-base combinations, which we call *additive complex cardinals* (*ACC*). The multipliers of the singles and tens bases must appear in the absolute form, while multipliers of higher bases have an implicational distribution: when a multiplier is in the contextual form, all the multipliers above it must also be in the contextual form (4b); but when a multiplier is in the absolute form, the higher multipliers can be either absolute or contextual (4a). This holds for '1' and bases above the thousands as well.

(4)a. {liang /**er**} qiān **er** bǎi **er** shí *1000: CONT/**ABS**; 100: **ABS***
{**2.CONT**/**2.ABS**} thousand **2.ABS** hundred **2.ABS** ten '2220'
b. {liang /***er**} qiān liang bǎi **er** shí *1000: CONT/***ABS**; 100: CONT*
{**2.CONT**/**2.ABS**} thousand **2.CONT** hundred **2.ABS** ten '2220'

We argue that the multiplier of the tens base is a bound morpheme, while the multiplier of higher bases can be bound or free, in parallel to the contrast between English *twen+ty* and *two#hundred*. Our analysis only applies to cardinal numbers, and has nothing to say about digits (e.g. in phone numbers and math problems), which we take to be free-standing homophones to *yī* and *er*.

We also observe that not only Mandarin and English, but perhaps all languages follow a ***monotonicity*** generalization, echoing Greenberg (1978): multipliers of lower number bases are less regular (e.g. suppletive, re-adjusted) than those of higher number bases (e.g. analytic). For example, French single digits are suppletive (e.g. *cinq* '5' vs. *un* '1'), multipliers of 10 are re-adjusted (e.g. *cinqu+ante* '50') and multipliers of 20, 100 and higher bases are analytic (e.g. cinq#cents). Our generalization only applies to multipliers of bases (e.g. 10, 100) but not cardinals themselves. It does not apply to borrowed words, either (e.g. Russian 40 and Belgian French 80).

We can state this generalization as *ABA, where B is less regular than A. But it cannot be explained by the containment analysis (e.g. Bobaljik 2012), the standard analysis of *ABA. The containment analysis would require that in English, *thir+teen* contain *twelve* structurally, and in French, the twenties base (e.g. *quatre#vingts*) contain the tens (*cinqu+ante*), which is not plausible. In contrast, a monotonicity-based approach like Graf (2019) can account for this generalization.

Hurford (2003) observed that in ACCs that lack an overt conjunction, the high cardinal almost always precedes the low cardinal (e.g. *twenty-two* vs. **two-twenty*). Assuming our monotonicity generalization, we propose a modification to Hurford's observation–it is not about high vs. low cardinals, but about morphological regularity: in these additive cardinals, regular cardinals precede the less regular ones. This accounts for the implicational distribution in (4).

3. Cardinals + classifier. As He (2015) observed, before a classifier, a single-digit cardinal is in the contextual form (5a), but the last digit of a multi-digit cardinal is in the absolute form (5b).

(5)a. **liang** gè li 'pear'
2.CONT CLAS pear '2 pears'

b. liang bǎi er shí **er** gè li 'pear'
2.CONT hundred 2.ABS ten **2.ABS** CLAS pear '222 pears'

Under our analysis in (2), the classifier and the NP form a constituent, and the base does not form a constituent with them. Assuming that only constituents can be elided, we predict that the NP (6b) or classifier + NP (6c) can be elided, but not base + classifier + NP (6d). This is borne out.

(6)Zhangsan sold 30000 pears, and Lisi sold …

a. liang wàn gè li. '20000 pears'
2.CONT ten.thousand CLAS pear

b. liang wàn gè ~~li~~.

c. liang wàn ~~gè li~~.

d. *liang / er ~~wàn gè li~~.

4. Approximate Expressions involve two numerals in sequence (7), which both occur in the contextual form. Our analysis accounts for this straightforwardly because we assume it involves disjunction of two CardPs. (7) cannot involve backward ellipsis of *wan ge li* (yì ~~wàn gè li~~ liǎng wàn gè li) because (6d) showed this is not possible.

(7)yì liǎng wàn gè li '10000 to 20000 pears'
1.CONT 2.CONT ten.thousand CLAS pear

5. Alternative proposals and challenges. He (2015) proposed that the contextual form appears only when the numeral is the sister of the following morpheme. This analysis is challenged by modified classifiers (8) because according to Cheng & Sybesma (1999), the numeral is not the sister of the following modifier, and yet the numeral is in the contextual form.

(8)a. [$_{ClP}$ liang[$_{Cl'}$ [$_{Cl'}$ dà wan][$_{NP}$ tang]]]
2.CONT big CLAS soup

b. shí er dà wan tang '2/12 big bowls of soup'
ten 2.ABS big CLAS soup

Ionin & Matushansky (2018) analyzed ACC as additive coordination of right-branching multiplier-base-NPs plus backward ellipsis. They would analyze (9) as [200 ~~wàn ge li~~] & [20 ~~wàn ge li~~] & [2 wàn ge li], but (6d) showed that the base cannot undergo ellipsis along with the classifier + NP.

(9)	liang	bǎi	er	shí	**er**	wàn	gè	li	
	2.CONT	hundred	2.ABS	ten	**2.ABS**	ten.thousand	CLAS	pear	‘2220000 pears’