

# AN OT ANALYSIS OF CHINESE ABBREVIATORY COMPOUNDS

Man Gao

Department of Linguistics  
University of Victoria

## 1. INTRODUCTION

The abbreviatory compound is a very common type of word formation in Chinese. Like abbreviations or acronyms in other languages, Chinese abbreviatory compounds are the abbreviations or shortened forms of long words or phrases; unlike those abbreviations and acronyms in languages like English, they are compounds in Chinese, and they are not formed in the way of first letter concatenation like *UN* for *United Nations*, or the cropping of long words like *Mass.* for *Massachusetts*. Chinese abbreviations are composed of the individual morpheme from the original form, and have essentially the same meaning of the original form but with shorter form, e.g.

- |     |  |  |     |   |
|-----|--|--|-----|---|
| (1) | <i>bian-fang</i><br>frontier-protect<br>'frontier' | <i>jian-cha</i><br>inspect-examine<br>'inspect(ion)' | --> | <i>bian-jian</i><br>frontier-inspect<br>'frontier-inspect(ion)' |
|-----|--|--|-----|---|

The chosen morphemes are in bold in the input (same for all the following examples). In this example, *bian-jian* is the abbreviatory compound. It is formed by two morphemes from the original form. It is claimed that the abbreviations are predominantly formed by taking the first morpheme of each word in the original form (Packard, 2000), like in (1). But it is common to form the abbreviatory compounds in Chinese like (2) and (3) too.

- |     |   |  |     |   |
|-----|---|--|-----|---|
| (2) | <i>chang-tu</i><br>long-path<br>'long distance' | <i>dian-hua</i><br>electricity-speech<br>'telephone' | --> | <i>chang-hua</i><br>long-speech<br>'long-distance call'     |
| (3) | <i>lao-shi</i><br>old-teacher<br>'teacher'      | <i>xue-sheng</i><br>study-person<br>'student'        | --> | <i>shi-sheng</i><br>teacher-person<br>'teacher and student' |

The abbreviation in example (2) is formed by taking the first morpheme of the first compound *chang* of *chang-tu*, and the second morpheme of the second compound *hua* of *dian-hua*; while in (3) it is formed by taking the second morpheme of each compound in the original form. All possible combinations are found, and there are six different ways of abbreviation according to Jin (1999), including truncating one of the compounds in the original form like *Qing-hua* for *Qing-hua da-xue* 'Qing-hua University'.

However, there are few attempts to look for the rules of how these abbreviatory compounds are formed since it seems that there is no pattern at all. In this paper, I am going to explore the formation of the Chinese abbreviatory compounds in the framework of Optimality Theory (Prince & Smolensky 1993, OT hereafter).

The structure of the paper is as follows: part 2 talks about the assumptions and theoretical background on which the analysis is based; part 3 analyzes the data by introducing relevant constraints in OT and modifying them on the basis of some morphological principles for this analysis; part 4 summarizes this paper.

## 2. THE ASSUMPTIONS AND THEORETICAL BACKGROUND

Most phrases in Chinese can be abbreviated in some way, and so can some sentences. The length of these abbreviations is from 2 to 7 or more morphemes. However, the focus of this paper is on the abbreviatory compounds with two morphemes which makes up 70.9% of the total 1066 abbreviatory compounds found in Chinese (Jin, 1999). In addition, most of the two-morpheme abbreviatory compounds are originating from four-morpheme phrases (consisting of two compounds with two morphemes each). Considering that this is a small-scale study, my data contains 169 abbreviatory compounds that all originate from four-morpheme phrases.

Of all the compounds in the original forms in my corpus, there are three types of compounds: the endocentric, exocentric and co-ordinate compounds. The endocentric compounds are the compounds which have a head (Fabb 1998). In Chinese, the head of a compound may be defined canonically, by its position within the compound (Packard 2000). The canonical head in Chinese is defined by Packard (2000) as follows: the function of the form class of the word, following the headedness Principle: verbs have their canonical head on the left and nouns have their canonical head on the right. For example, the head of *xue-xiao* study-school 'school' is *xiao*, which is on the right of the noun; while the head of the verb *tiao-wu* jump-dance 'dance' is *tiao*, on the left side. This is a straightforward way to determine the head of nominal and verbal compounds in Chinese.

There are also some compounds that do not have a head. The exocentric compounds in Chinese are composed of morphemes that cannot be the head of the compound. For example, with *wei-sheng* protect-live 'hygiene', both morphemes *wei* and *sheng* have their own meanings, but neither of them is directly related to the meaning 'hygiene' as a word.

The co-ordinate compounds consist of components that both share head-like characteristics (Fabb 1998). For example, *jian-cha*, check-examine 'check up' is a word with both morphemes as the head, since both components have a similar meaning to the compound.

With all the above three types of compounds found in the original form, the puzzle we want to solve in this analysis is if there are any rules to determine which morpheme is preserved in the abbreviatory compounds, and whether they have any interactions if there are more than one rules. I am going to analyze it under the framework of OT, which I anticipate would provide a satisfying answer.

Optimality Theory (Prince & Smolensky 1993) proposes an input and an output and a formal relation between the two. OT's viewpoint of language, and in fact every grammar, is a system of conflicting forces (Kager 1999). Markedness and Faithfulness are the two major forces engaged in a fundamental conflict in every grammar. These 'forces' are embodied by constraints (Kager 1999). Thus OT is a theory of constraint interaction. Constraints in OT are violable and they are ranked differently to form the conflict-regulating mechanism specific to every grammar. The formal model of OT consists of three parts. Gen (the Generator) first creates a number of candidates; then Eval (the Evaluator) selects the optimal candidate from all these candidates by the ranking of constraints (Universal Con). The best candidate of a grammar is the least costly violation of the constraints (Kager 1999). A conflict shows in the formation of Chinese abbreviatory compounds since, for some compounds, the first morpheme is chosen in the abbreviation, while the last is chosen for other compounds. Thus OT holds the promise of regulating this conflict with a ranking of a set of constraints.

### 3. MORPHOLOGICAL CONSTRAINTS IN OT

In analyzing the truncative plurals in Koasati, Horwood (2001) advocates another type of constraint other than Faith and Markedness we discussed above; the anti-faithfulness constraints were added to the inventory of CON. An anti-faithfulness constraint proposed by Alderete (1999) is defined as the logical negation of a faithfulness constraint. Taking; for example, a constraint immediately applicable to the problem at hand, consider  $\neg$ Max, negatively quantified from Max (McCarthy & Prince 1999): (Horwood 2001)

①  $\neg$ Max-Cat: ('Delete at least one *Cat*')

It is not the case that every element of type *Cat* in S1 has a correspondent element of type *Cat* in S2.

(Horwood 2001)

This anti-faithfulness constraint is also applicable to the Chinese abbreviatory compounds, with the more specific element type- morpheme in this case:

①'  $\neg$ Max-Morpheme: It is not the case that every morpheme in the input has a corresponding morpheme in the output. ('Delete at least one morpheme.')

$\neg$ Max-Morpheme captures the characteristics of Chinese abbreviatory compounds that they are shorter than the original form.  $\neg$ Max will penalize any candidate whose output is identical to input; if a morpheme of the

corresponding input (the original form) is not present in the output (the abbreviatory compound), the constraint will be satisfied. According to Horwood (2001), where  $\neg$ Max dominates all related Max constraints in a grammar, subtraction will occur. Therefore, this constraint should outrank any Max constraints in this analysis of abbreviation in Chinese.

To illustrate the  $\neg$ Max-Morpheme constraint for Chinese abbreviatory compounds, let's look at tableau 1:

	<i>bian-fang jian-cha</i>	$\neg$ Max-Morpheme
a.	<i>bian-fang jian-cha</i>	*!*
b.	$\varphi$ <i>bian-jian</i>	

Table 1.

Candidate (a) is exactly the same as the input *bian-fang jian-cha*, thus it violates  $\neg$ Max-Morpheme. Since there are two compounds in the input, and both of them remain unchanged in candidate (a), the  $\neg$ Max-Morpheme constraint is violated twice, each by one compound. For candidate (b), on the other hand, the morphemes *fang* and *cha* are missing, it is the optimal candidate as it does not incur any violation of  $\neg$ Max-Morpheme.

However, in order to limit the distance between input and output, we need a faithfulness constraint to 'restrict the shape variability of lexical items' (Kager 1999). The second constraint is:

② Max-IO: Every segment of the input has a correspondent in the output. (McCarthy & Prince 1995, 1999)

The segment in the Max-IO constraint proposed by McCarthy and Prince originally refers to the phonological segment. However, since this is an analysis of morphological phenomenon and, like the  $\neg$ Max-Morpheme constraint, the segment in this analysis refers to the morphological segment, i.e. the morpheme.

②' Max-IO: Every morpheme of the input has a correspondent in the output.

This constraint keeps the abbreviation of the input to two morphemes in the output. Thus, input with only one morpheme or none at all is ruled out by this constraint.

	<i>chen-shi jian-she</i>	$\neg$ Max-Morpheme	Max-IO
a.	<i>chen-shi jian-she</i>	*!*	
b.	$\varphi$ <i>chen-jian</i>		**
c.	<i>chen</i>		***!

Table 2.

The deletion of one morpheme in the output incurs one violation of Max-IO. Because the abbreviatory compounds under analysis are all composed of two morphemes, the optimal candidate should violate this constraint twice. Candidate (a) in the tableau above does not incur any violation of Max-IO at all, but it violates  $\neg$ Max-Morpheme which is ranked higher than the Max constraint, so it is ruled out. Although candidate (c) does not incur any violation of  $\neg$ Max-Morpheme, it violates Max three times by deleting three morphemes. Candidate *chen-jian* is then the optimal output with the fewest violations of Max.

According to the classifications of the compounds I discussed earlier, the two compounds in the input *chen-shi jian-she* are co-ordinate ones:

(4) *chen-shi* *jian-she* --> *chen-jian*  
 town-city build-establish  
 'city' 'construction' 'city construction'

The first morpheme in these two co-ordinate compounds is chosen for the abbreviated form in this example, and this is the case for most co-ordinate words in Chinese. Statistics based on my database show that 90.0% of the co-ordinate compounds are abbreviated in this way (103 out of 115). Thus we may come up with another constraint based on the fact that most co-ordinate compounds keep their left-most morpheme in the abbreviated form:





- |      |  |  |     |   |
|------|--|--|-----|---|
| (10) | <u><i>nong-ye</i></u><br>agriculture-industry<br>'agriculture' | <u><i>xue-xiao</i></u><br>science-study<br>'science' | --> | <i>nong-xiao</i><br><br>'school of agriculture' |
| (11) | <u><i>shang-ye</i></u><br>business-industry<br>'business'      | <u><i>xue-xiao</i></u><br>science-study<br>'science' | --> | <i>shang-xiao</i><br><br>'school of business'   |

All the compounds in (10) and (11) are headed nouns. According to the definition of canonical head, *ye* and *xiao* are the head in the corresponding compounds since they are at the rightmost position of the nouns. The head *xiao* is chosen according to the Faith-head constraint, but the head *ye* in *nong-ye* and *shang-ye* is not chosen. There must be some other constraint on the selection for the headed compounds, and it must be ranked higher than the Faith-head constraint to override it. Let's first assume that the head *ye* is chosen in both examples. The result would be that the same abbreviation is derived from two different original forms, and of course with different meaning. If *\*ye-xiao* is used for both phrases, ambiguity will arise as the reader or hearer would not know to which original form it corresponds. In other words, this does not provide any information about the specific industry that the school is involved in at all.

Similarly for the headed verbs, the choice of the head (which is the leftmost morpheme for the verb) will also result in ambiguity.

- |      |   |   |     |   |
|------|---|---|-----|---|
| (12) | <u><i>jie-yue</i></u><br>save-restrict<br>'economize' | <u><i>yong-dian</i></u><br>use-electricity<br>'electricity' | --> | <i>jie-dian</i><br>save-electricity<br>'economize on electricity' |
| (13) | <u><i>jie-yue</i></u><br>save-restrict<br>'economize' | <u><i>yong-shui</i></u><br>use-water<br>'water'             | --> | <i>jie-dian</i><br>save-water<br>'economize on water'             |

If the head *yong* of both verbs in (12) and (13) is chosen, we will not know what we are economizing on (*jie-yue*), so ambiguity arises. Therefore, we need another constraint to ban the outputs that will cause ambiguity.

⊙ \*Ambiguity: Avoid ambiguity.

This constraint is tricky in regard to how we determine which output violates \*Ambiguity. If the output containing the head of the compound violates \*Ambiguity as discussed in (10) to (13):

- |      |  |  |     |   |
|------|--|--|-----|---|
| (10) | <u><i>nong-ye</i></u><br>agriculture-industry<br>'agriculture' | <u><i>xue-xiao</i></u><br>science-study<br>'science' | --> | <i>nong-xiao</i> / <i>*ye-xiao</i><br><br>'school of agriculture' |
|------|--|--|-----|---|

What about those optimal output containing the head of compound discussed in (8) and (9)?

- |     |  |  |     |  |
|-----|--|--|-----|--|
| (8) | <u><i>xue-yuan</i></u><br>study-college<br>'college' | <u><i>xue-xiao</i></u><br>study-school<br>'school' | --> | <i>yuan-xiao</i> / <i>*xue-xue</i><br><br>'college and school' |
|-----|--|--|-----|--|

The choice of the head *ye* in *\*ye-xiao* violates \*Ambiguity so it is not optimal, but the choice of the head *yuan* and *xiao* in *yuan-xiao* does not, so it is optimal. How can we know whether the choice of the head violates \*Ambiguity?

One way of doing this is to see whether the canonical head matches the semantic head. If yes, the choice of the head will cause ambiguity. According to Packard (2000), the semantic head is that part of the word which is a more general instance of what the entire word means, often defined in terms of the 'IS A' relation, e.g., *ye* is both the semantic head and the canonical head for *nong-ye* agriculture-industry 'agriculture', which is a kind of industry. In other words, if the canonical head is a more general term, it is also the semantic head, and the choice of the head

will result in ambiguity so it violates \*Ambiguity; whereas if it is not, the choice of canonical head will not result in ambiguity thus preserved in the abbreviation. Therefore, any output with the head *ye* 'industry' will cause ambiguity as it is a more general term.

- (10) *nong-ye*                      *xue-xiao*                      -->                      *nong-xiao*  
 agriculture-industry      science-study                      'school of agriculture'  
 'agriculture'                      'science'

<i>nong-ye xue-xiao</i>	—Max-Morpheme	*Ambiguity	Faith-head	Anchor-L	Max-IO
a. <i>nong-ye xue-xiao</i>	*!*				
b. <i>nong-xue</i>			**!		**
c. <i>nong-xiao</i>			*	*	**
d. <i>ye-xue</i>		*!	*	*	**
e. <i>ye-xiao</i>		*!		**	**

Table 7.

Candidate (b) violates Faith-head twice and the second violation makes it fatal. In spite of the fact that candidate (c) violates Faith-head once, it is the optimal output given that the \*Ambiguity constraint is ranked higher than the Faith-head constraint. And the other two candidates incur violations of \*Ambiguity with the head *ye* preserved.

The fundamental difference between the canonical and semantic head of a headed compound is that the former has a stable position (right for nouns and left for verbs), while the latter does not have a fixed position. According to Packard (2000), the structural head (the canonical head) often matches with the semantic head, but there are cases in which they differ. One example is the word *xue-xiao* study-school 'school', in which the canonical head is on the right, but the semantic head is on the left (since a school is more generally understood as a kind of place to 'study' compared with *xue-yuan* study-college 'college'). This can be further proved by the statistics based on my data that 70% of the headed nouns have a match between the canonical head and the semantic head as the non-head morpheme chosen in the abbreviation, e.g. (the morpheme in black is both the canonical and semantic head):

- (14) *wen-xue*                      /                      *guang-xue*                      /                      *ke-xue*                      ...  
 writing-study                      light-study                      science-study  
 'study of literature'                      'optics'                      'study of science'

Furthermore, the reason that ambiguity will arise with the head preserved in the abbreviatory compound is because there is more than one word with the same head.

- (15) *da-xue*                      /                      *zhong-xue*                      /                      *xiao-xue*  
 big-study                      middle-study                      small-study  
 'university'                      'middle school'                      'elementary school'

The three compounds in (15) have the same structure and have the same head *xue*, thus, the choice of *xue* in abbreviation is very likely to cause ambiguity. It is also the case for headed verbs like in example (12) and (13), *yong-dian* use-electricity, *yong-shui* use-water, *yong-you* use-oil, etc. Because they are words with the same structure, the choice of the head *yong* that they all have will be ambiguous in abbreviations.

- (12) *jie-yue*                      *yong-dian*                      -->                      *jie-dian*  
 save-restrict                      use-electricity                      save-electricity  
 'economize'                      'electricity'                      'economize on electricity'

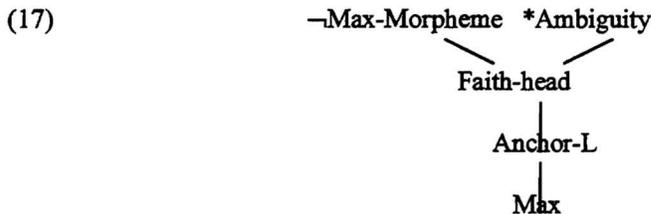
<i>jie-yue yong-dian</i>	¬Max-Morpheme	*Ambiguity	Faith-head	Anchor-L	Max-IO
a. <i>jie-yue yong-dian</i>	*!*				
b. <i>jie-yong</i>		*!	*		**
c. <i>jie-dian</i>			*	*	**
d. <i>yue-yong</i>		*!	*	*	**
e. <i>yue-dian</i>			*	**!	**

Table 8.

Although *yong* is the canonical head on the left, *dian* is chosen in the abbreviation. Candidate (b) and (d) incur a violation of \*Ambiguity since the head *yong* that will cause ambiguity is preserved. Candidate (c) and (e) incur one violation of the Faith-head constraint ranking lower than the \*Ambiguity constraint. Since candidate (e) violates Anchor-L twice, while (c) only violates it once, (e) is not the optimal candidate.

In summary, we have accounted for how the Chinese abbreviatory compounds are formed in the framework of OT by using five constraints based on morphological principles. The ranking of all these five constraints are:

(16) ¬Max-Morpheme, \*Ambiguity >> Faith-head >> Anchor-L, Max



Though there are some exceptions that our OT analysis is unable to account for, this analysis provides a straightforward explanation for the majority of the abbreviatory compounds in my corpus.

#### 4. CONCLUSION

This paper is a study of the formation of Chinese abbreviatory compounds based on the framework of Optimality Theory. A number of constraints in morphology play a very important role in determining the choice of the morphemes kept in the abbreviation. The ranking of those constraints can explain the majority of the data from the database of 169 abbreviatory compounds, which is indeed a clear analysis for this morphological phenomenon in the Chinese language.

This analysis successfully combines the framework of OT with morphological notions (e.g. morpheme and head), and principles like the Headedness Principle and Avoid ambiguity, together with a set of established constraints and the rankings among them, which provides evidence that OT can extend to areas other than phonology and syntax.

#### REFERENCES

- Alderete, John (1999) *Morphologically governed accent in Optimality Theory*. Doctoral dissertation, UMASS, Amherst. ROA-309.
- Fabb, Nigel (1998) Compounding. In Spencer, Andrew and Arnold Zwicky (eds), *The Handbook of Morphology*. Blackwell Publishers.
- Horwood, Graham (2001) *Anti-faithfulness and Subtractive Morphology*. UBC. ROA-466.

- Jin, Zhenguo (1999) *Dangdai Hanyu Suolueyu de Shehui Wenhua Jiazhi*. M.A. Thesis, Beijing Language and Culture University.
- Kager, Rene (1999) *Optimality Theory*. Cambridge University Press.
- McCarthy, John J. & Alan S. Prince (1995) Faithfulness and Reduplicative Identity. In Jill N. Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.), *Papers in Optimality Theory*. 249-384. University of Massachusetts, Amherst.
- McCarthy, John J. & Alan S. Prince (1999) Faithfulness and Identity in Prosodic Morphology. In Rene Kager, Harry van der Hulst and Wim Zonneveld (eds.), *The Prosody Morphology Interface*. 218-309. Cambridge University Press.
- Packard, Jerome L. (2000) *The Morphology of Chinese*. Cambridge University Press.
- Tang, Ting-chi (1994) More on the Relation Between Word-Syntax and Sentence-Syntax in Chinese Case Study in Compound nouns. In Jose Camacho and Lina Choueiri (eds.), *NACCL 6*, Vol1. 195-248. University of Southern California.
- Wei, Dongya, etc. (1997) *Chinese-English Dictionary (Revised Edition)*. Foreign Language Teaching and Research Press.

