

# Truncation and headedness in Chinese compounding: A dictionary-based study

Michael Oppen 欧迈珂    Alexander Sugar 唐山龙  
*University of Michigan*    *University of Washington*

Morphological headedness in Chinese compounding has been subject to theoretical debate (Huang 1998, Packard 2000, Ceccagno 2009 and many others). This study intends to contribute to this debate with insight from recent research on the linguistic phenomenon of “Word Length Flexibility” in Chinese (Duanmu 2012, Huang & Duanmu 2013). The main hypothesis of this study is that the preserved syllables in truncation are either the morphological heads of “flexible” words or the morphological non-heads of “inflexible” words. This hypothesis is empirically tested on a 2,922 item corpus of Chinese abbreviations (commonly referred to as *suōlüèyǔ* 缩略语 in Chinese) extracted from a specialized dictionary – *Xiàndài Hànyǔ Suōlüèyǔ Cídiǎn* 现代汉语缩略语词典 [Dictionary of Modern Chinese Abbreviations] (Yuan & Ruan 2002).

## 0. Introduction

The status of morphological headedness in Chinese compounding has been subject to debate amongst several scholars. Starosta et al (1998) advocate for the right-headedness of Chinese compounds. Huang (1998) concludes that Chinese compounds are headless as the morphological categories of compounds cannot be generalized based on their internal constituents. Packard (2000) takes a more nuanced view, suggesting that verbs are by nature left-headed, while nouns are right-headed. He refers to this division as the “Headedness Principle”. A recently-developed framework referred to as “Metacompounding” (Ceccagno & Scalise 2006, Ceccagno & Basciano 2007, Ceccagno 2009) has refined the general understanding of headedness in Chinese compounding. This approach is highly-relevant for our analysis, and therefore a brief outline is provided in the next paragraph. The proposal in this paper enhances the Metacompounding approach with the incorporation of the independent linguistic phenomenon referred to as “Word Length Flexibility” (Duanmu 2007, Duanmu 2011, Huang & Duanmu 2013).

Unlike the earlier studies on morphological headedness, Ceccagno & Basciano (2007), alternatively, see the determination of headedness as a function of the semantic relationship between the constituents of disyllabic compounds. They propose that there are three types of relations between constituents in disyllabic compounds – subordinate,

attributive, and coordinate. In the case of subordinate compounds, one constituent acts as an argument of the other. For instance, consider the verb-object ([V O]) compound, *zūnzhòng lǎorén* 尊重老人 ‘respect the elderly’. In this is verbal compound, the noun *lǎorén* ‘the elderly’ is assigned the theme role by the verbal head *zūnzhòng* ‘to respect’. In contrast, attributive compounds are defined as compounds in which one constituent acts as modifier of the other. For example, the compound *shòushāng zhànfú* 受伤战俘 ‘wounded prisoner of war’ is a typical example of an attributive compound, since the nominal head *zhànfú* ‘prisoner of war’ is modified by the adjective *shòushāng* ‘wounded’. While the classic case of an attributive compound is the adjective-noun ([A N]) type like the example above, either constituent can be an A, N or V. Finally, there are coordinate compounds, in which the two constituents are the same part of speech, resulting in a serial verb, synonym, antonym or pair reading. We see an antonymous relationship between the two constituents of *bǎocún fèichú* 保存废除 ‘preserve or eliminate’. The location of a head—the most semantically prominent constituent—in a Chinese compound, then, will depend on the relationship between its constituents. If the compound is subordinate, then the head is the constituent that requires an argument, usually a V. If the compound is attributive, on the other hand, then its head is the constituent that is being modified. While left-headedness is more common in subordinate compounds and right-headedness is more common in attributive compounds, both compound types can have either left or right heads. In the case of coordinate compounds neither constituent can be considered the head since the two constituents behave as semantic equals. Ceccagno & Basciano refer to the collective possibility of left-headedness, right-headedness and two-headedness as the “Cerberus” head.

The remainder of this paper is structured as follows. Section 1 reviews the existing literature on Word Length Flexibility in Chinese. Section 2 introduces the Chinese “abbreviation” word-formation process (commonly referred to as *suōlǚèyǔ* in Chinese) and discusses the data selection used in this paper. Section 3 provides an analysis of the data introduced in the previous section. Section 4 concludes this study with directions for further research.

## 1. Word Length Flexibility in Chinese

The phenomenon of “Word Length Flexibility” (hereafter WLF) in Chinese refers to the property of a word to have two semantically-equivalent variants differing only in prosodic size (Guo 1938, Pan 1997, Feng 2002, Duanmu 2007, Duanmu 2012, Ke 2012, Huang & Duanmu 2013).<sup>1</sup> Typically, the two variants of “flexible words” are a monosyllabic form and a disyllabic form; consider the flexible pair *shī-lǎoshī* 师—老师 ‘teacher’, for instance. If these forms are analyzed syllable-by-syllable, the monosyllabic

<sup>1</sup> Although this phenomenon is discussed in all the sources above, only the research by Duanmu and Guo refers to this phenomenon as “Word Length Flexibility”.

form means ‘teacher’, while the disyllabic form means ‘old-teacher’. However, both of these variants have the exact same meaning of ‘teacher’ because even though the first syllable – *lǎo* – independently means ‘old’, the disyllabic form can refer to a teacher of any age. The remainder of this section discusses the existing literature on WLF in Chinese and provides further analysis.

Huang & Duanmu 2013 sample a large portion of the Chinese lexicon – one-tenth of the monosyllabic entries in the *Xiàndài Hànyǔ Cídiǎn* 现代汉语词典 [Modern Chinese Dictionary] (2016 word senses in total). The results of their sampling process indicate that roughly 65% of non-compound words in Chinese have flexible length. Examples of flexible words are provided in (1). Since adjectives, nouns, and verbs constitute 92% of the sample analyzed in Huang & Duanmu 2013 and the truncation patterns of adjective, noun, and verb phrases are the focus of this study, we only discuss the flexible nature of words belonging to these three parts of speech.

(1) Flexible Words in Chinese (monosyllabic forms in bold)

Adjectives	Nouns	Verbs
<b>wěnjiàn</b> 稳—稳健 ‘calm’	<b>lǎoshǔ</b> 鼠—老鼠 ‘rat’	<b>fēnkāi</b> 分—分开 ‘to split’
<b>kuàilè</b> 乐—快乐 ‘happy’	<b>yāzi</b> 鸭—鸭子 ‘duck’	<b>zhòngzhi</b> 种—种植 ‘to plant’
<b>měilì</b> 美—美丽 ‘beautiful’	<b>shāngdiàn</b> 店—商店 ‘store’	<b>shībài</b> 败—失败 ‘to lose’
<b>qīngchū</b> 清—清楚 ‘clear’	<b>dàsuàn</b> 蒜—大蒜 ‘garlic’	<b>zhīdào</b> 知—知道 ‘to know’
<b>pínqióng</b> 贫/穷—贫穷 ‘poor’	<b>xuéwèn</b> 学—学问 ‘knowledge’	<b>xuéxí</b> 学—学习 ‘to study’

There are several noteworthy observations about the flexible words introduced in (1) and a summary of these observations is provided in (2). First, the monosyllabic variant of a flexible word can be the left syllable, the right syllable, or either syllable of the disyllabic variant; this is explained in detail in the next paragraph. Second, flexible disyllabic words differing in one syllable can share the same monosyllabic form. For instance, the words *xuéwèn* ‘knowledge’ and *xuéxí* ‘to study’ share the same monosyllabic form – *xué* – which has two meaning senses differing in part of speech; a nominal form meaning ‘knowledge’ and a verbal form meaning ‘to study’. Third, for many flexible words, the disyllabic form appears to be a derived word by affixation. The disyllabic forms of the first two noun examples provided in (1) appear to be derived words; analyzed syllable-by-syllable the literal meaning of *lǎoshǔ* is ‘old-rat’ and the literal meaning of *yāzi* is ‘duck-child’. However, regardless of age, rats and ducks can be referenced by their disyllabic forms in Chinese; as these affixes do not contribute an additional meaning to their roots, these classes of words can be referred to as “Prefixed Flexible: 0X” or “Suffixed Flexible: X0”.

The nature of flexible words that look like true compounds, which we refer to as “pseudo-compounds”, differs from the aforementioned “affixed flexible words”. True compounds such as *hēibǎn* 黑板 ‘blackboard’ generally consist of two constituents, each of which contributes a semantic component to the composite meaning of the compound;

in this example, the first syllable *hēi* contributes the meaning ‘black’ and the second syllable *bǎn* contributes the meaning ‘board’ to derive the compositional meaning ‘blackboard’. Contrastively, in the case of “pseudo-compounds”, one constituent determines the composite meaning of the disyllabic form despite the fact that the disyllabic form appears to be a compound. Consider the “pseudo-compound” *wěnjiàn* ‘calm’, for instance. The first syllable *wěn* also means ‘calm’ in isolation; furthermore, the monosyllabic form of *wěnjiàn* cannot be *jiàn* since *jiàn* has an entirely different meaning in isolation – ‘strong (adj), strengthen (v)’. Pseudo-compounds can be “Left Flexible: XY – X” (the left syllable of the disyllabic form is the monosyllabic variant) or “Right Flexible: XY – Y” (the right syllable of the disyllabic form is the monosyllabic variant). In contrast with the preceding example, the monosyllabic form of several flexible words can be either the left or right syllable of the disyllabic form. Take the disyllabic form *pínqióng* ‘poor’ as provided in (1), for instance. Either syllable can contribute the meaning of ‘poor’ in compound word formation – both *qiónggrén* and *pínmín* mean ‘the poor’ or ‘poor person’. In this study, this class of flexible words is referred to as “Fully Flexible: XX”. In addition to pseudo-compounds, other classes of flexible words include proper nouns (PN) and polysyllabic truncations (XYZ).<sup>2</sup>

(2) Categories of Flexible Words in Chinese

Category	Example
0X: Prefixed Flexible Word	<i>lǎoshǔ</i> – <i>shǔ</i> ‘rat’
X0: Suffixed Flexible Word	<i>yāzi</i> – <i>yā</i> ‘duck’
XY – X: Left Flexible Pseudo-Compound	<i>fēnkāi</i> – <i>fēn</i> ‘to split’
XY – Y: Right Flexible Pseudo-Compound	<i>shāngdiàn</i> – <i>diàn</i> ‘store’
XX: Fully Flexible Pseudo-Compound	<i>pínqióng</i> – <i>pín</i> OR <i>qióng</i> ‘poor’
PN: Proper Nouns	<i>xiānggǎng</i> – <i>gǎng</i> ‘Hong Kong’
XYZ: Polysyllabic Truncations	<i>jiānádà</i> – <i>jiā</i> ‘Canada’

The primary object of inquiry for this paper is the nature of the truncation of four syllable compound phrases into disyllabic forms. More specifically, this study explores which syllables of the four syllable long forms are preserved in their disyllabic truncated variants. This pattern of truncation is incredibly productive in Chinese – the dataset analyzed in this study includes nearly 3000 items. On a theoretical note, the productivity of this process is not surprising given the overwhelming evidence for syllabic foot structure in Chinese. Evidence for syllabic foot structure has been exhibited in studies of the phonology of English-based loanwords (Lin 2007), the disyllabic realization of tone three in isolation (Chao 1968, Duanmu 2007), Tone Three Sandhi (Chen 2000, Duanmu 2007), Prosodic Syntax (Feng 2002) and length preferences in compounding (Duanmu 2012). Via the Strict Layering Hypothesis (Selkirk 1982, Ito & Mester 1992/2003), a consequence of having syllabic feet is that the minimal

<sup>2</sup> Polysyllabic forms are typically loanwords, such as the example ‘Canada’ provided in (2).

prosodic word should be disyllabic (a binary foot with no additional material) and therefore a minimal prosodic phrase should consist of two prosodic words with no additional material. By nature, truncation processes should minimize production efforts by speakers (Zipf 1935). At the same time, truncation processes need to generate output which is prosodically well-formed (McCarthy & Prince 1986/1993, Benua 1997). In sum, disyllabic truncation of four-syllable phrases is natural from a prosodic morphology account of word formation.

In this section, the authors have reviewed the literature on the phenomenon of WLF observed in Chinese, thereby establishing the preliminary basis necessary to explore the nature of Chinese abbreviations. This section concludes with a statement on the synchronic and diachronic nature of these phenomena in Chinese. As the authors hope to demonstrate in this paper, strong synchronic generalizations can be made with regard to the nature of abbreviation in Chinese. However, there are diachronic factors which affect this process; that is, some items have lexicalized truncations. The diachronic nature of WLF has not been explored in detail and we choose not to delve into this topic in this paper.

## 2. Chinese abbreviation patterns

As mentioned earlier, the object of inquiry for this paper is Chinese abbreviation (or *suōlüèyǔ* 缩略语 in Chinese). The data analyzed in this paper comes from the *Xiàndài Hànyǔ Suōlüèyǔ Cídiǎn* (Yuan & Ruan 2002), hereafter *XDHYSLYCD*. The *XDHYSLYCD* is a monolingual Chinese specialized dictionary of abbreviations in the Modern Chinese language. Entries in this dictionary are ordered alphabetically based on Hànyǔ Pīnyīn Romanization and contain detailed information including: part of speech category (either noun, verb, or adjective), lexicalization (indicated by the phrase *yǐcíhuà* 已词化, meaning “lexicalized”), earliest source of attestation, and definition. Of the 8400 entries in this dictionary, the most productive pattern – a total of 2951 entries – is that between a given four-syllable phrase and a disyllabic truncation of that phrase. Of this dataset, 1687 entries are noun phrases, 1203 entries are verb phrases, and 61 entries are adjective phrases. In this study, four syllable compounds are expressed formally as ABCD with each letter representing a syllable. Abbreviation patterns are expressed by the corresponding letters which are preserved in truncation; therefore, the most productive pattern of truncation which takes the first syllable of each constituent can be formally expressed with the rewrite rule  $ABCD \rightarrow AC$  or simply with the shorthand AC. Examples of noun abbreviation patterns along with the type counts of these patterns in our dataset, regardless of part of speech, are provided in (3).

## (3) Noun Abbreviation Patterns

Pattern	Type Count	Example
AC	1704 (57.7%)	<i>gōnggòng jiāotōng</i> 公共交通 → 公交 ‘public transportation’
AD	663 (22.4%)	<i>mínzú yīnyuè</i> 民族音乐 → 民乐 ‘ethnic music’
BC	290 (9.8%)	<i>lǚkè yùnshū</i> 旅客运输 → 客运 ‘passenger transport’
BD	229 (7.7%)	<i>kuàisù jìsuàn</i> 快速计算 → 速算 ‘short-cut’
AB	33 (1.1%)	<i>rèxiàn diànhuà</i> 热线电话 → 热线 ‘hotline’ (telephone)
CD	21 (0.6%)	<i>guǎngbō diàntái</i> 广播电台 → 电台 ‘broadcasting station’
Others	12 (0.4%)	<i>lìrùn liúchéng</i> 利润留成 → 留利 ‘retained profit’

Interestingly, the distributions between abbreviation patterns do not differ very much when separated according to the part of speech of the full form listed in the dataset. Type counts and their percentages are provided for each productive abbreviation pattern (AC, AD, BC, and BD) divided by phrase type (noun, verb, adjective) in (4).

## (4) Abbreviation patterns by phrase type

	Noun Phrase		Verb Phrase		Adjective Phrase	
Pattern	Count	Percentage	Count	Percentage	Count	Percentage
AC	982	58	685	56.7	37	60.7
AD	361	21.7	289	24.4	13	23
BC	160	9.7	126	10.6	4	6.6
BD	130	7.7	93	8	6	9.8

The nature of the abbreviations with regard to the aforementioned Metacompounding framework (Ceccagno & Scalise 2006, Ceccagno & Basciano 2007, Ceccagno 2009) is discussed for the remainder of this section. As mentioned in the introduction, there are three semantic-relationships between constituents in Chinese compounds – attributive, coordinate, subordinate – in the Metacompounding framework. Furthermore, this framework allows for three possible locations of the morphological head: left, right, two-headed. In order to analyze this dataset in the Metacompounding framework, three additional annotations needed to be added to each entry. First, each constituent of the full form needed to be assigned to a part of speech category. This decision was automated based on part of speech listed in the *Xiàndài Hànyǔ Cídiǎn* (2005) for the first meaning of the given constituent. Second and third, the relationship between the constituents and the location of the morphological head for each entry needed to be added. The second author (a non-native fluent speaker of Chinese) went through the dataset and made these decisions.

In our dataset, we observed several important generalizations. Not including two-headed forms, verb phrases are slightly more likely to be left-headed (59.1%) than right headed (40.9%). Noun and adjective phrases are generally right headed; based on token count, right-headed noun and adjective phrases account for 97.4% and 83.3% of the

data points in their respective part of speech categories within the dataset we consider in this study. Virtually all two-headed phrases are coordinate in our dataset; a potential contradictory example is *zēngjiā biānzhi* 增加编制, which we consider subordinate since *zēngjiā* ‘to add’ modifies *biānzhi* ‘to edit’ to derive the meaning of the full form – ‘to add contents (to a piece of writing)’. Not including coordinate phrases, most verb phrases are subordinate (46%) while most noun and adjective phrases are attributive (93.4% and 73% respectively).

Ceccagno & Basciano (2007:225) define Metacompounding as “a mode of compound formation in which at least one of the constituents refers to an underlying compound that does not appear on the surface. The constituent on the surface can therefore be regarded as a truncated form of the underlying compound.” The thesis of this study does not depart far from this definition. In fact, the authors intend to demonstrate that this is in fact true through the formal addition of Word Length Flexibility.

The shortcoming of the Metacompounding analysis is simply that the proposed semantic relationships between constituents, internal structure of constituents, and location of morphological heads within the full-forms of compounds cannot predict which syllables will survive truncation. Three chi-square tests of independence were performed on the subset of productive truncations (truncation patterns AC, AD, BC, BD) in R. First, the result of the chi-square test for semantic relationship suggests independence between semantic relationship and truncation patterning –  $X^2 = 17.36$ ,  $df = 6$ ,  $p = 0.008$ . Second, the result of the chi-square test for internal structure of constituents ([A A], [A N], [A V], [N N], [N V], [V N], [V V]) suggests independence between internal structure of constituents and truncation patterning –  $X^2 = 48.37$ ,  $df = 18$ ,  $p < 0.001$ . Third, the result of the chi-square test for the location of the morphological head suggests independence between the location of the morphological head and truncation patterning –  $X^2 = 22.71$ ,  $df = 6$ ,  $p < 0.001$ .

This section has introduced the attested patterns of Chinese abbreviation. While the metacompounding analysis is a significant enhancement on the theories of compound formation in Chinese, it falls short when applied to the prediction of truncation patterning. In the next section, we account for these truncation patterns with the formal addition of the morphological property of word length flexibility.

### 3. Analysis

The trends observed in the abbreviation data presented in (3) above can be summarized in the following generalization. For a given four-syllable compound word ABCD, its probable truncation falls in the order:  $AC > AD > BC > BD$ . In this section we assert that the syllables which are preserved in truncation are either the syntactic non-heads or the monosyllabic flexible variants. Both kinds of truncation, along with examples and theoretical explanation are discussed. By the end of this section, we demonstrate that this hypothesis regarding truncation patterns in Chinese accounts for

roughly 98% of the truncation patterns in our dataset. We explore the exceptional cases in the conclusion.

As the morphological divisions of compound phrases have proven ineffective for predicting truncation patterning, we choose to analyze the truncation patterns of each disyllabic constituent of the productive truncation patterns in our dataset – a total of 5772 tokens. In the analysis of the truncation of disyllabic constituents, we follow a theoretical assumption summarized as the “Information-Stress Principle” as outlined in Duanmu 2010. The basic tenet of this principle is that morphological non-heads contribute more information than morphological heads in compounding and hence receive stress. Furthermore, cross-linguistic studies have demonstrated that truncatory processes are sensitive to stress (Lappe 2007) and may even preserve stress, if possible (Thomason & Thomason 2004). In sum, Chinese abbreviations of true compounds preserve their stressed constituents. This is expanded upon in the next paragraph.

In the case of noun compounds, the modifier component (typically an adjective or a noun) would receive stress placement at the phrase level. For example, according to the Information-Stress Principle, phrasal stress in the word *dàxué* ‘university (lit. big-study)’ should fall on the left syllable. Furthermore, the left syllable of ‘university’, *dà*, is generally the syllable which survives truncation; consider the abbreviations of *Běijīng Dàxué* ‘Peking University’ and *Yúnnán Dàxué* ‘Yunnan University’ – *BěiDà* and *YúnDà* respectively, for instance. Likewise, in verb-object compounds such as *yòngshuǐ* 用水 ‘use water’, the object (in this case, the right syllable *shuǐ* ‘water’) would receive stress placement at the phrase level. The right syllable of this compound, *shuǐ*, is in fact, the syllable that survives truncation of the four syllable phrase *jiéyuē yòngshuǐ* 节约用水 ‘conserve water’; the disyllabic truncation of this phrase is *jiéshuǐ* 节水.

Word-stress judgments in Chinese have been subject to a good deal of debate in academic circles.<sup>3</sup> This is not surprising given that Chinese is a tone language which uses pitch to distinguish minimal pairs and most theories require stress to have phonetic correlates which are generally limited to pitch, duration, intensity, or some combination of these three qualities. Since the data analyzed in this study are text-based, we do not intend to contribute to the phonetic realizations of stress in Chinese for this study. However, recent phonetic-based work provides support for our claims regarding phrasal stress. Using a speech data corpus, Lai, Sui & Yuan 2010 found that the duration and  $F_0$  differences are significantly longer and greater, respectively, for the left syllable than the right syllable in disyllabic words that were not verb-object compounds. Shen, Vaissière & Isel 2013 conducted an experiment demonstrating that the object in verb-object compounds generally has a longer duration than the verb and that verb-object compounds generally have a larger  $F_0$  range than modifier-noun compounds.

For the remainder of this section, we analyze the truncation patterns of each disyllabic constituent in our dataset. As mentioned in the first paragraph of this section,

<sup>3</sup> Refer to Lai, Sui & Yuan 2010:1 for references regarding this debate



there are a total of 5772 disyllabic constituents in our dataset. Of these 5772 items, there are many more tokens that preserve the left syllable – 4361 – than tokens that preserve the right syllable – 1411. This is not surprising given that AC is the most productive truncation pattern in the XDHYSLYCD, that there are more noun phrase abbreviations in our dataset, and that the monosyllabic variant of disyllabic flexible words can be either the left or the right syllable of the disyllabic form. The first author (a non-native fluent speaker of Chinese) went through a large sample of the truncations of the disyllabic constituents and labeled types as either non-heads, flexible-primary, flexible-sense, prefixed, suffixed, place name, and exceptions. Each of these types is explained in the next paragraph; the sampling procedure is discussed in the paragraph after next.

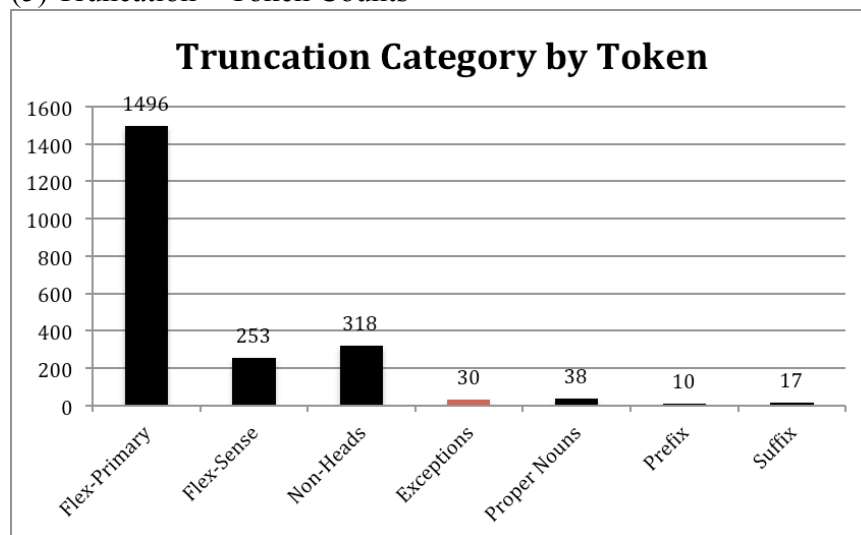
Truncations were categorized as “non-heads” if the morphological non-head is the monosyllabic variant of the disyllabic full form. For instance, the flexible word pair *piào-tóupiào* 投票 was categorized as “non-head” since the full form is a verb-object compound meaning ‘to vote (lit. cast-ticket)’ and the object of this compound – *piào* ‘ticket’ – is the syllable which survives truncation. The designation “flexible-primary” was given to flexible word pairs in which the monosyllabic variant has no difference in meaning from the disyllabic variant; this includes pairs belonging to flexible word categories 0X, X0, XY-X, XY-Y, XX as listed in (2). The most common type of this category, which has a total of 44 tokens, is *xiào-xuéxiào* 学校 ‘school (lit. study-school)’. The category “flexible-sense” was given to pairs in which the monosyllabic variant does not generally have the primary meaning of the disyllabic variant, but the *Xīnhuà Zìdiǎn* 新华字典 [Xinhua Dictionary] (2004) lists the disyllabic form as a “semantic-sense” of the monosyllabic form. Consider the flexible pair, *yǐng-diànyǐng* 电影 ‘movie’ for instance. The monosyllabic form generally means ‘shadow’ and the disyllabic form is a true compound with the literal interpretation ‘electric-shadow’. This pair, in particular, is well attested in the dataset analyzed in the study with 11 tokens. The high token count of these types can imply some extent of lexicalization. Additional support for the lexicalization argument can be provided with reference to the semantic-senses provided for monosyllabic morphemes in a comprehensive Chinese-language reference dictionary such as the *Xīnhuà Zìdiǎn*. Similarly, if the *Xīnhuà Zìdiǎn* does not list the disyllabic form of a given pair in the dataset as a “semantic-sense” of its monosyllabic variant and that monosyllabic variant is not a non-head, the first author categorized that given pair as “non-flexible”. In cases which could either be classified as “flexible sense” or “non-head”, such as *dà-dàxué* ‘college (lit. big-study)’, the first author chose to classify such cases as “non-head”. Prefixed and suffixed forms such as *gǎnshàng* 赶上 ‘immediately upon’ and *měiyuè* 每月 ‘each month’ are morphologically complex in composite meaning; the data demonstrate that theses affixes do not survive truncation. In the aforementioned cases, the surviving syllables after truncation are *gǎn* ‘quickly’ and *yuè* ‘month’, respectively. The classification “Proper Nouns” has already been discussed in the first section of this paper. All remaining short-long pairs that did not fulfill the criteria

of the aforementioned designations were considered “exceptions”. This section concludes with a quantitative analysis of the aforementioned truncation patterning of the disyllabic constituents.

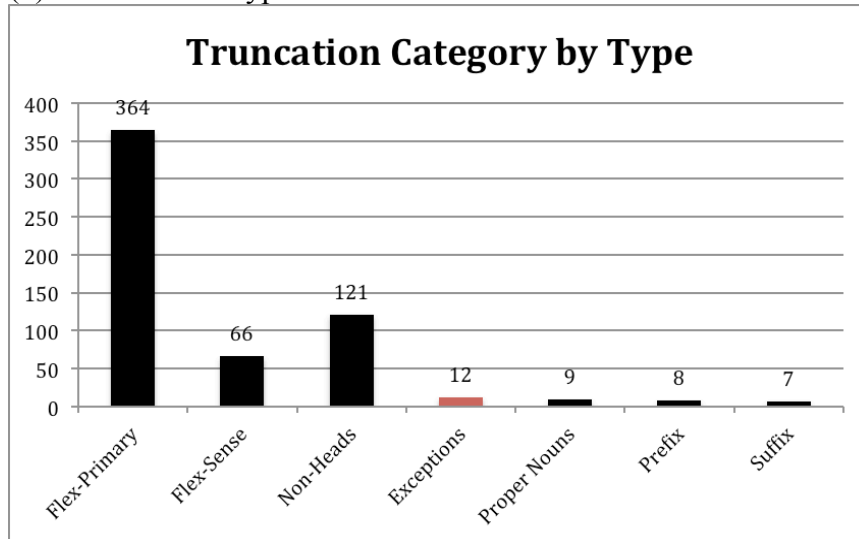
The subset of disyllabic constituents that preserve the left syllable contains 4361 tokens of 2014 types. By part of speech, nouns make up 1080 types, verbs make up 748 types, and adjectives make up 186 types. The first author categorized the truncations of roughly 30% of this subset; this 30% included of all types with a token count of 10 or greater (946 tokens of 62 types) and a randomly selected one-ninth of the remaining types (376 tokens of 216 types). The subset of disyllabic constituents that preserve the right syllable contains 1411 tokens of 877 types. By part of speech, nouns make up 459 types, verbs make up 363 types, and adjectives make up 55 types. The first author categorized the truncations of roughly 60% of this subset; this 60% included of all types with a token count of 2 or greater (754 tokens of 221 types) and a randomly selected one-seventh of the remaining types (93 tokens/types).

Combining the left syllable truncation subset with the right syllable truncation subset, we have a total of 2169 tokens and 592 types. The systematic sampling procedure outlined above provides empirical coverage for roughly 50% of the tokens and 25% of the types. Bar graphs are provided for the tokens (5) and types (6). Overall, we have demonstrated that WLF alone (Flex-Primary, Flex-Sense) can account for 83.6% of the token truncation and 76.6% of the type truncation. Including WLF, truncation patterns of morphological non-heads, proper nouns, and affixed forms, the analysis introduced in this paper can account for 98.6% of the token truncation and 98% of the type truncation. The truncation patterning of exceptional cases, some of which are coordinate compounds such as *qīnpéng* 亲朋 ‘family and friends’, is still not understood. All exceptional cases are provided in the appendix.

(5) Truncation – Token Counts



## (6) Truncation – Type Counts



In this section, we have provided empirical evidence that truncation patterns in Chinese can be attributed to either WLF in which the monosyllabic variant survives truncation or to headedness in which the non-head survives truncation. The productive abbreviation patterns AC, AD, BC, and BD account for approximately 98% of the data extracted from the XDHYSLYCD and of that subset, approximately 98% of truncations have been accounted for as examples of word length flexibility and non-headedness. This study concludes with brief discussion of some counterexamples and directions for further inquiry.

#### 4. Conclusion

In this paper, the authors have proposed an enhancement to the Metacompounding theory as applied to Chinese compounding (Ceccagno & Basciano 2007). In the Metacompounding theory, the morphological structure of a given compound is in part explained by the syllables that appear in the abbreviation of that compound. That is, the syllables which survive abbreviation are regarded as truncations of underlying compounds. This theory, however, does not provide any means to explain which syllable of a disyllabic compound will appear in abbreviation. Through the addition of Word Length Flexibility (Duanmu 2012, Huang & Duanmu 2013), the authors have demonstrated that approximately 98% of monosyllabic truncations of disyllabic constituents in Chinese can be accounted for as either monosyllabic forms of flexible words or morpho-syntactic non-heads. In the next paragraph we account for some of the remaining abbreviations.

In the dataset considered in this study, there are 33 types of compounds with abbreviation pattern AB (approximately 1.1% of the dataset). These compounds are all noun phrases and therefore truncation occurs at the level of the entire phrase, that is, the

morpho-syntactic non-head of the entire phrase is preserved in truncation. These are unlike the productive abbreviation patterns AC, AD, BC, BD in which truncation is applied constituent-by-constituent. In the case of abbreviation pattern CD, which represents 21 types and 0.6% of this dataset, roughly two-thirds (14 types) are labeled as lexicalized. If the non-lexicalized CD types are combined with the other types (CA, CB, DA, and DB) there are 19 truncations (0.006% of the dataset) remaining. We consider these patterns to be non-productive and do not provide any additional explanation.

Within the productive patterns (AC, AD, BC, BD) from the subset analyzed in section 3, a small number of constituents did not truncate according to the principles outlined in this study (1.3% of the tokens and 2.02% of the types). We propose that disyllabic truncation of four-syllable compounds is preferred over trisyllabic truncation and that an arbitrary choice regarding which syllable survives abbreviation must be made to satisfy this preference. Perhaps an information-based approach to Chinese abbreviation – such as the approach explored in Shaw, Han & Ma (to appear) in which informative syllables are theorized to survive truncation in order to optimize recoverability – can shed light on these examples. This conflict between information recoverability and prosodic preferences is left open for future research.

## REFERENCES

- BENUA, LAURA. 1997. Transderivational identity. PhD Dissertation. University of Massachusetts.
- CECCAGNO, ANTONELLA. 2009. Metacompounds in Chinese. *Lingue e linguaggio*, 8(2). 195-212.
- CECCAGNO, ANTONELLA; and BASCIANO, BIANCA. 2007. Compound headedness in Chinese: an analysis of neologisms. *Morphology*, 17(2). 207-232.
- CECCAGNO, ANTONELLA; and SCALISE, SERGIO. 2006. Classification, structure, and headedness of Chinese compounds. *Lingue e linguaggio*, 2. 233-260.
- CHEN, MATTHEW. 2000. *Tone sandhi: patterns across Chinese dialects*. Cambridge: Cambridge University Press.
- DUANMU, SAN. 2007. *The phonology of Chinese*. Second Edition. Oxford: Oxford University Press.
- DUANMU, SAN. 2010. Stress, information, and language typology. In 《语言学文选》 [Papers in Linguistics], edited by Feng Shi and Hongming Zhang. 3-21. Beijing: 世界图书出版社 [World Book Press].
- DUANMU, SAN. 2012. Word-length preferences in Chinese: a corpus study. *Journal of East Asian Linguistics*, 21(1). 89-114.

- FENG, SHENGLI. 2002. *Prosodic Syntax of Chinese*. LINCOM Studies in Asian Linguistics. Muenchen: Lincom Europa.
- GUO, SHAOYU (郭绍虞). 1938. 中国语词之弹性作用 [Flexible usage of words in the Chinese language]. 《燕京学报》 24.
- HUANG, LIJUN (黄丽君); and DUANMU, SAN (端木三). 2013. 现代汉语词长弹性的量化研究 [A quantitative study of elastic word length in Modern Chinese]. 《语言科学》 [Linguistic Sciences], 12(1). 8-16.
- HUANG, SHUANFAN. 1998. Chinese as a headless language in compounding morphology. In *New approaches to Chinese word formation: morphology, phonology, and the lexicon in Modern and Ancient Chinese*, edited by Jerome Packard. 261-283. Berlin: Mouton de Gruyter.
- ITÔ, JUNKO; and MESTER, ARMIN. 1992/2003. Weak layering and word binarity. Ms. *University of California, Santa Cruz*.
- KE, HANG (柯航). 2012. 现代汉语单双音节搭配研究 [Modern Chinese mono/di-syllabic combination research] Shanghai: 商务印书馆 [Commercial Publishing Press].
- LAI, CATHERINE; SUI YANYUAN; and YUAN, JIAHONG. 2010. A corpus study of the prosody of polysyllabic words in Mandarin Chinese. In *Proceedings of Speech Prosody*.
- LAPPE, SABINE. 2007. *English prosodic morphology*. Springer.
- LIN, YEN-HWEI. 2007. *The sounds of Chinese*. Cambridge: Cambridge University Press.
- MCCARTHY, JOHN; and PRINCE, ALAN. 1986/1993. Prosodic morphology I: constraint interaction and satisfaction. Ms. *University of Massachusetts and Rutgers University*.
- PACKARD, JEROME. 2000. *The morphology of Chinese*. Cambridge: Cambridge University Press.
- PAN, WENGUO (潘文国). 1997. 《汉英语对比纲要》 [A comparative outline of Chinese and English]. Beijing: 北京语言文化大学出版社 [Beijing Culture and Language University Publishing Press].
- SELKIRK, ELISABETH. 1982. *The syntax of words*. Cambridge, MA.: Linguistic Inquiry Monographs.
- SHAW, JASON; HAN CHONG; and MA, YUAN. To appear. Surviving truncation: informativity at the interface of morphology and phonology. *Proceedings of Morphology and its Interfaces*, 13 November 2013, Lille, France.
- SHEN, WEILIN; VAISSIÈRE, JACQUELINE; and ISEL, FRÉDÉRIC. 2013. Acoustic correlates of contrastive stress in compound words versus verbal phrase in Mandarin Chinese. *Computational linguistics and Chinese language processing*. 18(3). 45-58.
- STAROSTA, STANLEY; KUIPER KOENRAAD; NG, SIEW-AI; and WU, ZHI-QUIAN. 1998. On defining the Chinese compound word: Headedness in Chinese compounding and Chinese VR compounds. In *New approaches to Chinese word formation*:

- morphology, phonology, and the lexicon in Modern and Ancient Chinese*, edited by Jerome Packard. 347-370. Berlin: Mouton de Gruyter.
- THOMASON, LUCY; and THOMASON, SARAH. 2004. Truncation in Montana Salish. *Studies in Salish linguistics in honor of M. Dale Kinkade*, 354-376.
- XIANDAI HANYU CIDIAN [Dictionary of Modern Chinese]. 2005. Beijing: 中国文史出版社 [China Written History Press].
- XINHUA ZIDIAN [Xinhua Dictionary]. 2004. Shanghai: 商务印书馆 [Commercial Publishing Press].
- YUAN, HUI (袁晖); and RUAN, XIANZHONG (阮显忠). 2002. 《现代汉语缩略语词典》 [Dictionary of Modern Chinese Abbreviations]. Beijing: 语文出版社 [Language and Writing Press].
- ZIPF, GEORGE KINGSLEY. 1935. The psycho-biology of language. Cambridge: Cambridge University Press.

## APPENDIX: EXCEPTIONAL CASES

Compound	Part of Speech	Truncation	Token Count
yǐngpiàn 影片 ‘movie’ (lit. shadow-slice)	Noun	piàn 片 ‘slice’(mw)	7
zhùmíng 著名 ‘famous’ (lit. marked-name)	Adjective	míng 名 ‘name’	6
huáqiáo 华侨 ‘overseas Chinese’ (lit. China-bridge)	Noun	qiáo 桥 ‘bridge’	3
diànhuà 电话 ‘telephone’ (lit. electric-speech)	Noun	huà 话 ‘speech’	3
zhāoshēng 招生 ‘recruit students’	Verb	zhāo 招 ‘to recruit’	3
wénmáng 文盲 ‘illiterate’ (lit. character-blind)	Noun	máng 盲 ‘blind’	2
jiàoyuán 教员 ‘teacher’ (lit. teach-person)	Noun	yuán 员 ‘person’	1
shuífáng 睡房 ‘sleep-place’	Noun	fáng 房 ‘place’	1
dǎnshí 胆识 ‘courage and insight’	Noun	shí 识 ‘insight’	1
qīnpéng 亲朋 ‘family and friends’	Noun	qīn 亲 ‘relative’	1
zhuāndì 专递 ‘special delivery’	Noun	dì 递 ‘to deliver, pass’	1
shōuyīn 收音 ‘to receive sound’	Verb	shōu 收 ‘to receive’	1

ACKNOWLEDGEMENTS

The authors would like to thank San Duanmu, Andries Coetzee, Bill Baxter, Richard V. Simmons, and Jason Shaw for discussions related to this research. Comments offered by audience members at NACCL-25, particularly by Zheng Xu, have been helpful to the development of this paper. The first author would like to acknowledge financial support from NSF GRFP grant 1000117925.