Lexical Gaps and Lexicalization: Implications for Word Segmentation Systems for Chinese NLP

Chan-Chia Hsu
Graduate Institute of Linguistics, National Taiwan University
No. 1, Sec. 4, Roosevelt Road, Taipei, 10617 Taiwan
chanchiah@gmail.com

Abstract

This paper is motivated by the observation that not all adjectives in Chinese have a canonical antonym. For example, most Chinese speakers choose to translate the English word *dishonest* into a word string *bu chengshi* 'not honest' instead of any antonym candidates of *chengshi* suggested in antonym dictionaries. Our discourse evidence from corpus data suggests that *bu chengshi* is evolving into a word in discourse at a faster pace than some other '*bu* + adjective' strings, and this may result from the lexical gap for a canonical antonym of *chengshi* and the communicative need for such a word. As a consequence, it is proposed that if the lexicalization process of *bu chengshi* continues in the future, the string may need to be considered a single word in a segmentation system (i.e., *buchengshi* 'dishonest'). For a segmentation system to distinguish between words and phrases, discourse factors should be taken into consideration.

1 Introduction

The issue of antonym canonicity has been empirically investigated in English and other languages (Paradis et al., 2009; Willners and Paradis, 2010). This paper is motivated by the observation that not all adjectives in Chinese have a generally accepted antonym. For example, although the antonym of *chengshi* 'honest', according to antonym dictionaries (e.g., Han and Song, 2001; Xu, 2000), can be *xuwei* 'hypocritical', *xujia* 'unreal', and *jiaohua* 'cunning', none of them are canonical for most native speakers. Intriguingly, in translating *dishonest* into Chinese, most Chinese speakers also choose to translate the English word into a word string *bu chengshi* 'not honest' instead of any antonym candidates of *chengshi*. The aim of this paper is thus to address the question: Will the lexical gap for a canonical antonym of *chengshi* enable the string *bu chengshi* to evolve into a word?

To answer the question, we adopt a corpus-based approach and see how *bu chengshi* behaves in discourse. The results will have implications for word segmentation systems for Chinese NLP in that if *bu chengshi* functions like a word both linguistically and conceptually, then the string may not need to be further segmented into '*bu* + *chengshi*' in a segmentation system. This line of study can shed light on the segmentation task from a discourse perspective.

This paper is organized as follows. Section 2 reviews different views of what a word is. Section 3 introduces the data examined in the present study, and Section 4 presents the results. Section 5 discusses the implications of the results for Chinese wordhood and the segmentation task in Chinese. Section 6 offers the conclusion and some suggestions for future research.

2 What Is a Word?

Generally, a word is defined as "a unit which has universal intuitive recognition by native-speakers, in both spoken and written language" (Crystal, 2008:521). Though most native speakers intuitively know what a word is, there exists no definition of the concept 'word' that is universally applicable. (Crystal, 1991; Dai, 1998; Packard, 2000). This has complicated the segmentation task in natural language processing.

26th Pacific Asia Conference on Language,Information and Computation pages 191–198

Packard (2000) provides a comprehensive review of what a word is, and Packard's review suggests that a word can be defined from various perspectives. A common, straightforward way to define words is based on writing conventions, i.e., orthographic words. In many languages, words are separated by spaces. However, words in Mandarin cannot be discerned orthographically since they are not physically separated. Sociologically, words are forms intermediate between phonemes and sentences, forms the general public are conscious of and find relevant in many ways (Chao, 1968:136-138). Most speakers may regard Chinese characters as sociological words. Semantically, a word is seen as a form with a semantic value. Words can be combined to form complex expressions, but they may not be further decomposed into smaller units (e.g., Baxter and Sagart, 1997; Dowty et al., 1981).

Dai (1998), also reviewed in Packard (2000), argues that words can function in different domains and that words, as a consequence, can be defined in phonological, morphological, and syntactic terms (Dai, 1998:104-105):

> A syntactic word is a minimal constituent to which syntactic rules may refer; a phonological word is a certain prosodic domain in which internal phonological rules may apply (as opposed to external or phrasal sandhi rules); and a morphological word is a maximal domain in which morphological rules may apply.

Dai's conception of the word works in Chinese, as in many other languages.

Di Sciullo and Williams (1987) define words from a cognitive view. They suggest that words are *psychologically real* in our language use. Words are listed units in the lexicon and have idiosyncratic properties which are not governed by rules but must be memorized by speakers. Packard (2000) critically pinpoints the weaknesses of the previous approaches and also offers new insights into the cognitive nature of Chinese words (e.g., Chinese X-bar morphology).

However, while there have been various theories about how to define a word, few have taken discourse factors into account. This motivates us to examine corpus data and see how words emerge and function in our actual communication (cf. Sun, 2006).

3 Methodology

The database for the present study is the Academia Sinica Balanced Corpus of Modern Chinese.[1] In the Sinica Corpus, every text is segmented, and every word is tagged with its part-of-speech. There are 489,2324 words in total.[2] The Chinese Gigaword Corpus (the second edition), which is much larger than the Sinica Corpus, is an alternative, yet it only collects newswire texts. The present study prefers not to consider genre factors, so the data in the Chinese Gigaword Corpus were not used.

When the data were collected, it was specified that *bu* should occur within five words to the left of the target.[3] In addition to *bu chengshi*, *bu zhijie* 'not direct' and *bu hefa* 'not legal' were also examined for comparison. In Chinese, the canonical antonym of *zhijie* is *jianjie*, and that of *hefa* is *feifa*. Therefore, it was predicted that *bu chengshi*, which lacks a lexical counterpart against *chengshi*, would behave more like a word than *bu zhijie* and *bu hefa*.

Here are the criteria for selecting *bu zhijie* and *bu hefa* for analysis in the present study. First, gradable antonyms such as *kuaile*/*beishang* 'happy/sad' were not considered, for the negation of a gradable antonym does not entail its counterpart (i.e., *bu kuaile* 'not happy' does not necessarily mean *beishang* 'sad'). Thus, the present study selected complementary antonyms, which are mutually exclusive (Cruse, 1986); the negation of one can be regarded as near-synonymous with the other. Next, Jones' (2002) list is adopted as a point of departure. The list includes 56 antonym pairs in English, which are considered to be an effort to approximate a

[1] It is open to the public at http://db1x.sinica.edu.tw/kiwi/mkiwi/.

[2] For more information about the Sinica Corpus, refer to http://db1x.sinica.edu.tw/kiwi/mkiwi/98-04.pdf.

[3] The collocation span usually ranges from three words to five words (Sinclair, 1991), and the present study follows the tradition.

representative set for a study of antonymy. The scope of the present study was limited to adjectives, and the complementary pairs were singled out and then translated into Chinese. There are two problems, though. First, it can be impossible to find a Chinese equivalent for some antonyms in Jones' (2002) list. Second, after translated into Chinese, the negation of an antonym may sound inappropriate or be semantically different from the counterpart of that antonym (e.g., *si*/?*bu huo* 'dead/not alive'). With these problems, the present study finally selected two strings only, i.e., *bu zhijie* 'not direct' and *bu hefa* 'not legal', for a comparison with *bu chengshi*.[4]

In our analysis, accidental co-occurrences were excluded. The following is an example:

(1) 工作上遭遇**不**平等的待遇，**直接**與上司溝通多次仍屬無效。
Gongzuo shang zaoyu **bu** pingdeng de daiyu, **zhijie** yu shangshi goutong duo ci reng shu wuxiao.
'(Someone) was not fairly treated in the workplace, and it was useless to communicate with the supervisor many times.'

In (1), *bu* works with *pingdeng* 'equal' rather than *zhijie*. Therefore, the sentence in (1) was excluded.

The following table summarizes the numbers of the tokens analyzed in the present study.

Strings	Tokens
bu zhijie	9 (56.3%)
bu…zhijie	**7 (43.7%)**
bu hefa	**9 (100%)**
bu…hefa	0 (0%)
bu chengshi	**7 (87.5%)**
bu…chengshi	1 (12.5%)

Table 1: The numbers of the tokens analyzed in the present study

The analysis of how the tokens are used in Chinese will be presented in the following section.

4 Results

As Table 1 shows, the '*bu* X' strings do not occur frequently in the corpus. However, when the expected value was calculated, it was found that the token numbers were larger than expected by chance.

The formula for the expected value of the surface co-occurrence is as follows (cf. Event, 2008):[5]

(2) $f_1 \times f_2 / N$
(f_1: the token number of *bu*; f_2: the token number of the adjective; N: the token number of the corpus)

The trouble was that the online version of the Sinica Corpus does not provide the token number if a word occurs more than 5,000 times in the corpus. The negative marker *bu* is such a frequently-occurring word. We could estimate the token number of *bu* by referring to Xiao and McEnery (2008). On average, *bu* occurs approximately 800 times per 100,000 words (Xiao and McEnery, 2008:290). Based on their calculation, we estimated that *bu* occurs approximately 39,000 times in the Sinica Corpus.[6] Based on the formula in (2), Table 2 presents the expected values and the observed values of *bu zhijie*, *bu hefa*, and *bu chengshi*.[7]

Strings	Values
bu zhijie (ex.)	8.68
bu zhijie (ob.)	9
bu hefa (ex.)	1.32
bu hefa (ob.)	9
bu chengshi (ex.)	0.84
bu chengshi (ob.)	7

Table 2: The expected values and the observed values of *bu zhijie*, *bu hefa*, and *bu chengshi*

[4] The behavior of *jianjie* and that of *bu zhijie* were compared, and the same analysis was done for *hefa* and *bu hefa*. However, the analyses are beyond the scope of the present study, and the results will not presented here.

[5] The present study adopted the most common approach and calculated surface co-occurrences (Event, 2008) rather than textual co-occurrences and syntactic co-occurrences.

[6] There are 489,2324 words in the Sinica Corpus.

[7] In the Sinica Corpus, there are 1,089 tokens of *zhijie*, 166 tokens of *feifa*, and 106 tokens of *chengshi*.

As Table 2 shows, the observed values of *bu zhijie*, *bu hefa*, and *bu chengshi* are higher than their expected values. This justifies the analyses in the following, for the co-occurrences of *bu* and *zhijie*, *hefa*, and *chengshi* are not haphazard.

To address the issue of how close *bu* and its following adjective are, the present study analyzed how often *bu* and the adjective are interrupted and how often a modifier is used to modify *bu* rather than the whole construction '*bu* + adjective'. The first question has been answered in Table 1.

As shown in Table 1, *bu* and *zhijie* is interrupted by a modifier more often than the other two. Here is an example:

(3) 活動之間的關係錯綜複雜，並**不那麼直接**而明顯。
Huodong zhijian de guanxi cuozongfuza, **bing bu name zhijie** er mingxian.
'The relationships between activities are complex, not very direct and obvious.'

In (3), *name* 'so' is inserted between *bu* and *zhijie*. There are some other patterns, including *bu shi zhijie* 'not be direct', *bu hen zhijie* 'not very direct', and *bu shi name zhijie* 'not be that direct'.

As for the modifiers of *bu*, the following are two examples:

(4) 使用者**並不直接**指定所要的字。
Shiyongzhe **bing bu zhijie** zhiding suo yao de zi.
'The user does not directly specify words they want.'

(5) 儀式**根本不合法**。
Yishi **genben bu hefa**.
'The ceremony is not legal at all.'

In (4), *bing*, which serves as an intensifier, cannot modify *zhijie* if *bu* is not present. Therefore, *bing* is analyzed as modifying *bu* rather than *bu zhijie*. With a modifier attached to *bu*, the relationship between *bu* and *zhijie* seems to become weaker. The same analysis is true of the sentence in (5). Table 3 summarizes the modifying patterns:

Strings	Tokens
zero + *bu zhijie*	6 (66.7%)
modifier + *bu zhijie* (Attested pattern: *bing bu zhijie* 'entirely not direct')	3 (33.3%)
zero + *bu hefa*	6 (66.7%)
modifier + *bu hefa* (Attested patterns: *bing bu hefa* 'entirely not legal', *jibenbu hefa* 'basically not legal', *genben bu hefa* 'fundamentally not legal')	3 (33.3%)

Table 3: Modifiers of '*bu*'

As Table 3 shows, *bu* in *bu zhijie* and *bu hefa* can be modified. However, in the Sinica Corpus, *bu* in *bu chengshi* is not found to be modified.

It has been found that canonical antonyms co-occur far more frequently than expected by chance (e.g., Charles and Miller, 1989; Fellbaum, 1995; Jones, 2002, 2006, 2007; Jones and Murphy, 2005; Justeson and Katz, 1991). In the Sinica Corpus, *zhijie* and *jianjie* are found to co-occur twenty times, which confirms the previous observations.[8] Since *jianjie* is near-synonymous to *bu zhijie*, *zhijie* and *bu zhijie* may co-occur in the corpus. When the window size was specified as within four words, *zhijie* and *bu zhijie* were not found to co-occur in the Sinica Corpus. The same search was conducted for *hefa*/*bu hefa* and *chengshi*/*bu chengshi*, and no co-occurrences were attested when the span was specified as within four words. However, when the span was extended, *chengshi* and *bu chengshi* were found to co-occur. Of the seven tokens of *bu chengshi* in the Sinica Corpus, three (42.9%) co-occur with *chengshi* and serve some discourse functions. Here is an example:

(6) 你**誠實**，我就喜歡跟你交朋友。你**不誠實**，我就不喜歡跟你交朋友。
Ni **chengshi**, wo jiu xihuan gen ni jiao pengyou. Ni **bu chengshi**, wo jiu bu xihuan gen ni jiao pengyou.

[8] In the Sinica Corpus, there are 1089 tokens of *zhijie* and 134 tokens of *jianjie*. Based on the formula in (2), the two words are expected to co-occur fewer than once!

'If you are honest, I like to make friends with you. If you are not honest, I do not like to make friends with you.'

In (6), *chengshi* and *bu chengshi* co-occur in an ancillary manner (cf. Jones, 2002, 2006; Jones and Murphy, 2005). That is, the co-occurrence of *chengshi* and *bu chengshi* helps to signal another contrast in the context, i.e., whether the speaker wants to make friends with someone. Such a co-occurrence is not accidental, but helpful in organizing the discourse. Unlike *chengshi*/*bu chengshi*, *zhijie*/*bu zhijie* and *hefa*/*bu hefa* were not found to co-occur and serve discourse functions even though the span was extended.

5 Discussion

This study examined the behavior of *bu chengshi*, *bu zhijie* and *bu hefa* to deal with Chinese wordhood from a discourse perspective. Though *bu zhijie*, *bu hefa*, and *bu chengshi* have not been regarded as words by native speakers, they may be on the way of lexicalization at different paces. Table 4 demonstrates their distributional differences observed in the corpus (cf. Table 1, Table 2, Table 3).

Strings	Distributional differences
bu chengshi	1. **O/E: 8.33** 2. *bu…chengshi*: 1 (12.5%) 3. **modifier + *bu*: 0 (0.0%)** 4. **co-occurs with *chengshi* when the span is extended**
bu hefa	1. O/E: 6.82 2. *bu…hefa*: 0 (0.0%) 3. modifier + *bu*: 3 (33.3%) 4. never co-occurs with *hefa*
bu zhijie	1. O/E:[9] 1.04 2. *bu…zhijie*: 7 (43.7%) 3. modifier + *bu*: 3 (33.3%) 4. never co-occurs with *zhijie*

Table 4: Evidence for the lexicalization of *bu zhijie*, *bu hefa*, and *bu chengshi*

As shown in Table 4, the O/E ratio of *bu chengshi* is the highest of the three, which suggests that the combination of *bu* and *chengshi* is far from accidental. Second, *bu* and *chengshi* is interrupted only once in the Sinica Corpus, and *bu* in *bu chengshi* is never modified in the corpus. The two facts may indicate that the boundary between *bu* and *chengshi* may be breaking down. Third, only *bu chengshi* is found to be used as a whole contrastively with *chengshi*. The discourse evidence in Table 4 supports that *bu chengshi* is being lexicalized at a faster pace than the other two.

The strings *bu zhijie*, *bu hefa*, and *bu chengshi* share an identical structure, i.e., '*bu* + adjective', but they are functionally different from a

[9] In Table 4, O/E stands for the ratio between the observed value between the expected value. See Section 4 for more details.

communicative perspective. As mentioned earlier, the antonym of *chengshi* can be *xuwei*, *xujia*, and *jiaohua*, but such pairs are regarded as non-canonical by most native speakers. The three words are much more pejorative than *bu chengshi*, and they represent different ways of being dishonest. Therefore, a word is needed in Chinese to neutrally represent the concept of being dishonest. Moreover, such a word can serve as a canonical antonym for *chengshi*. These communicative needs may contribute to the relatively high O/E ratio of *bu chengshi* and help it to fulfill its potential to evolve into a single word (cf. Kjellmer, 2003, 2005). On the other hand, the semantic difference between *bu hefa* and *feifa* is not so substantial as that between *bu chengshi* and *xuwei*/*xujia*/*jiaohua*, and neither is that between *bu zhijie* and *jianjie*. In other words, *bu chengshi* is more useful than *bu zhijie* and *bu hefa* in communicative terms. This results in different lexicalization processes, and the functional differences have been formally reflected. By analyzing discourse data from the corpus, we can advance our understanding of Chinese wordhood.

However, in distinguishing between words and phrases, few studies have considered discourse data (Packard, 2000:15). This is reflected in most segmentation systems for Chinese. Since *bu zhijie*, *bu hefa*, and *bu chengshi* share an identical structure, the algorithm segments the three strings in a similar manner. The following results come from the segmentation system of Academia Sinica:[10]

(7) 不(D) 誠實(VH)
bu chengshi
'dishonest'

(8) 不(D) 直接(VH)
bu zhijie
'indirect'

(9) 不(D) 合法(VH)
bu hefa
'illegal'

In the present study, it is suggested that the string *bu chengshi* is more likely to further develop into a word than *bu zhijie* and *bu hefa* even though they share the same structure. If the evolution of *bu chengshi* continues, the string may need to be considered to be a single word in a segmentation system in the future (i.e., *buchengshi* 'dishonest'). Our results can serve as a reference for segmentation systems. With ample evidence from discourse data, strings with a similar, or even identical, structure can be segmented differently.

Then, how can we compromise with the conflict that after all, *buchengshi* is inarguably formed from the concatenation of the negation marker *bu* and an adjective? The speaker's communicative need may have helped *bu* + *chengshi* to achieve a relatively high frequency (in terms of its O/E value), and frequent repetitions may enable the whole string to gain an independent representation in the lexicon (cf. Bybee, 2000, 2006). Consequently, from a cognitive perspective, *buchengshi* may be processed as a whole rather than in a compositional manner. As manifested in discourse data, the boundary between *bu* and *chengshi* is less clear than that between *bu* and *zhijie* and that between *bu* and *hefa*. That is, *buchengshi* may become psychologically real if its lexicalization process continues, and it may eventually become a listed unit memorized by speakers in their lexicon (Di Sciullo and Williams, 1987; Hoosain, 1992). It appears that morphological rules alone cannot explain why strings with the same structure can have different representations in the grammar of Mandarin. The morphological boundary is fluid (Hoosain, 1992:118-120), and communicative needs and discourse factors should be taken into account in a theory about Chinese wordhood.

6 Conclusion

By analyzing corpus data, this study suggests that *bu zhijie*, *bu hefa*, and *bu chengshi* may be on the way of lexicalization at different paces. Of the three, *bu chengshi* is the most useful in communicative terms and is evolving the fastest in discourse. The boundary between *bu* and *chengshi* may gradually become blurred. In fact, words and phrases in Chinese are so closely connected that "one must investigate and study the links between speech sounds, syntax, semantics, and discourse

[10] The segmentation system of Academia Sinica is available at http://ckipsvr.iis.sinica.edu.tw/.

factors in forming Chinese words in actual communication" (Sun, 2006:75).

The findings of this study can have computational, lexicographical, and pedagogical applications. First, the results provide some feedback for segmentation programs for Chinese. In designing a segmentation system or a computer algorithm to parse texts, computational linguists need to take discourse factors into consideration. Second, if *bu chengshi* eventually evolves into a single word and gains a representation in the speaker's lexicon, a lexicographer may need to consider listing the word in the dictionary. Third, teaching materials may need to be revised according to corpus data so that language learners can learn to speak and write natural-sounding Chinese. For example, students should be taught to distinguish between *bu chengshi* and *xuwei* though both can serve as antonyms of *chengshi*.

Further studies are still needed to explore Chinese wordhood from a usage-based perspective. First, the present study decides to ignore genre factors at the cost of the sample size (cf. Section 3), yet more quantitative data are needed to make the calculations more reliable. Second, the present study focuses on three adjectival strings, yet more need to be analyzed. For example, the scope can extend to verbal ones (e.g., *fandui*/*bu tongyi* 'disagree/not agree') so that the generalizations made in the present study will be more valuable. Third, psycholinguistic experiments (e.g., self-paced reading tasks) can de conducted to investigate how the speaker processes the morphological boundary online. To develop a fuller understanding of Chinese grammar and provide more feedback on the performance of a segmentation system, converging evidence from different fields is encouraged.

References

徐安崇 [Anchong Xu]. 2000. 反義詞應用辭典 [Antonym Application Dictionary]. Language and Culture Press, Beijing.

Anna-Maria Di Sciullo and Edwin Williams. 1987. On the Definition of Word. MIT Press, Cambridge.

Carita Paradis, Caroline Willners, and Steven Jones. 2009. Good and bad opposites: using textual and experimental techniques to measure antonym canonicity. The Mental Lexicon 4:380-429.

Caroline Willners and Carita Paradis. 2010. Swedish opposites: A multi-method approach to 'goodness of antonymy'. In Petra Storjohann (ed.), Lexical-Semantic Relations: Theoretical and practical perspectives (pp. 15-48). John Benjamins, Amsterdam.

Chaofen Sun. 2006. Chinese: A Linguistic Introduction. Cambridge University Press, Leiden.

Christiane Fellbaum. 1995. Co-occurrence and antonymy. International Journal of Lexicography 8:281-303.

David Crystal. 2008. A Dictionary of Linguistics and Phonetics (6th edition). Blackwell, Oxford.

David R. Dowty, Robert E. Wall, and Stanley Peters. 1981. Introduction to Montague Semantics. D. Reidel, Dordrecht.

Göran Kjellmer. 2003. Lexical gaps. In Sylviane Granger and Stephanie Petch-Tyson (eds.), Extending the Scope of Corpus-based Research (pp. 149-158). Rodopi, Amsterdam.

Göran Kjellmer. 2005. Negated adjectives in modern English. Studia Neophilologica 77:156-170.

Jerome L. Packard. 2000. The Morphology of Chinese: A Linguistic and Cognitive Approach. Cambridge University Press, Cambridge.

韓敬體, 宋惠德 [Jingti Han and Dehui Song]. 2001. 反義詞辭典 [Antonym Dictionary]. Sichuan People's Publishing House, Chengdu.

Joan L. Bybee. 2000. The phonology of the lexicon: Evidence from lexical diffusion. In Michael Barlow and Suzanne Kemmer (eds.), Usage-based Models of Language (pp. 65-85). CSLI Publications, Center for the Study of Language and Information, Stanford.

Joan L. Bybee. 2006. From usage to grammar: The mind's response to repetition. Language 82:711-733.

John S. Justeson and Slava M. Katz. 1991. Co-occurrence of antonymous adjectives and their contexts. Computational Linguistics 17:1-19.

John Sinclair. 1991. Corpus, Concordance, Collocation. Oxford University Press, Oxford.

John Xiang-Ling Dai. 1998. Syntactic, phonological, morphological words in Chinese. In Jerome L. Packard (ed.), New Approaches to Chinese Word Formation: Morphology, Phonology and the Lexicon in Modern and Ancient Chinese (pp. 103-134). Mouton de Gruyter, Berlin.

Richard Xiao and Tony McEnery. 2008. Negation in Chinese: A corpus-based study. Journal of Chinese Linguistics 36:274-330.

Rumjahn Hoosain. 1992. Psychological reality of the word in Chinese. In Hsuan-Chih Chen and Ovid J. L. Tzeng (eds.), Language Processing in Chinese (pp. 111-130). North-Holland and Elsevier, Amsterdam.

Stefan Evert. 2008. Corpora and collocations. In Anke Lüdeling and Merja Kytö (eds.), Corpus Linguistics: An International Handbook (pp. 1212-1248). Mouton de Gruyter, Berlin.

Steven Jones and M. Lynne Murphy. 2005. Using corpora to investigate antonym acquisition. International Journal of Corpus Linguistics 10:401-422.

Steven Jones. 2002. Antonymy: A Corpus-based Perspective. Routledge, New York.

Steven Jones. 2006. A lexico-syntactic analysis of antonym co-occurrence in spoken English. Text & Talk 26:191-216.

Steven Jones. 2007. 'Opposites' in discourse: A comparison of antonym use across four domains. Journal of Pragmatics 39:1105-1119.

Walter G. Charles and George A. Miller. 1989. Contexts of antonymous adjectives. Applied Psycholinguistics 10:357-375.

William H. Baxter and Laurent Sagart. 1997. Word formation in Old Chinese. In Jerome L. Packard (ed.), New Approaches to Chinese Word Formation: Morphology, Phonology and the Lexicon in Modern and Ancient Chinese (pp. 103-134). Mouton de Gruyter, Berlin.

Yuen Ren Chao. 1968. A Grammar of Spoken Chinese. University of California Press, Berkeley.