

## Boundary Issues of Chinese Phrases, Words and Morphemes

Sun Yan

Beijing University of Technology  
No.100, Pingleyuan, Chaoyang District, Beijing, China  
Yan802@bjut.edu.cn

Selected paper from CLSW2019

**ABSTRACT.** *From the perspective of Prototype Theories, there are no obvious boundaries among lexicon, morphology and syntax. This applies to Chinese too. However, it is impossible to define the members at the boundaries of different categories with the conventional dichotomy approach. In this paper, four criteria were raised to test the solidification of the compound, according to which the ones that meet all the 4 criteria are considered as typical words (grade 4), while the ones that meet 0 criterion are considered as typical phrases (grade 0). By analyzing the constituents of compound continuum, we further put forward the continuum from bound morpheme to word.*

**Keywords:** prototypical category, continuum; word, phrase; morpheme

**1. Introduction.** One of the basic principles in cognitive linguistics is the concept of continuum, which is a sub-concept of the prototype category. Specifically there are no obvious boundaries among lexicon, morphology and syntax, which constitute a continuum of symbolic units. Therefore, it is unreasonable to discuss grammar without meaning, and it is unacceptable to divide grammatical structure into parts. Although the lexicon, morpheme and syntax are different in many parameters, the classification on them is only a subjective act- ion [1].

Opposite to the concept of continuum, is the view on grammatical structure that emphasizes the separation of elements, which is a sub-concept under the theoretical background of classical category view. Because of the emphasis on dispersion, researchers mostly focused on representative instances at both ends of a continuum, ignoring situations in the middle, so that discrete, distinctive categories are easily observed. Under the guidance of this concept, researchers generally adopt strict dichotomize concepts, such as synchronic and diachronic framework, language competence and language use,

grammar and vocabulary, morphology and syntax, semantics and pragmatics.

**2. There is no clear distinction between lexicon and syntax.** Under the concept of continuum, there is no clear distinction between lexical and grammatical structure, both of which are symbolic units. The difference among symbolic units are not in essence, but in concrete degree [2-4]. In the study of Chinese language, the structural consistency between lexicon and syntax has been recognized by many scholars [3] [5]. But because the formation of compound words and the syntactic structure are basically the same, it also challenges us to distinguish words and phrases.

By definition, word group (phrase) is a combination of words, and word is a combination of morphemes [6]. But the generative nature and the universal nature of rules in lexicon cannot be compared with that in syntax, and not all types of semantic relations that can be expressed by syntax can be expressed by lexicon [2]. Besides, Chinese syntax is considered to be descriptive, while the lexical meaning was suggestive. Some meanings of compound words were the addition of morphological meaning, while some meanings could not be obtained from the addition of morphological meaning, which were restricted by many factors [7].

**3. The problem of dividing standard of dichotomy.** As for the distinction among words and phrases, we take the standards put forward by Lü Shuxiang and Zhu Dexi as an example.

Lü proposed five perspectives: (1) whether the structure can be used alone, whether the constituents of the structure can be used alone. (2) Whether the combination of structure can be separated. (3) Whether the structure can be extended. (4) Whether the structural meaning is equal to the additive meaning of constituents. (5) How many morphemes does this structure contain [8].

But Lü also pointed out that in practice these five perspectives are often contradictory. For example, according to (1), if constituents of a structure cannot be used alone, then the structure should be considered as word. Like ‘tuó(camel)’ and ‘yā(duck)’, which cannot be used alone, thus ‘tuómáo(camel hair)’ and ‘yādàn(duck eggs)’ should be regarded as words. While ‘yáng(sheep)’ and ‘jī(chicken)’ can be used alone, then ‘yángmáo(wool)’ and ‘jīdàn(egg)’ should be regarded as phrases. But Lü believed that ‘yángmáo(wool)’ and ‘jīdàn(egg)’ are common words.

Zhu Dexi raised three criteria for judging words: (1) whether sentences can be formed separately. (2) Whether constituents can be replaced by other constituents. (3) Whether the structure can be extended [6]. But Zhu also said that it was difficult to identify Chinese words according to any of the three criteria alone, especially when some criteria were contradictory to each other.

Basing on those facts, we can see that the traditional method of dividing word and phrase has some inevitable limitations. On one hand it was believed that all the structures can be classified into different categories with dichotomy method, on the other hand, atypical members at the boundary always bring controversies. In this paper, we try to describe the

relationship among phrase, word and morpheme under the concept of continuum, so as to make our cognition closer to the actual situation.

#### 4. Solidification of disyllable compound structure.

4.1. **Definition of language units.** We adopt Zhu Dexi's definition on language unit. A *Morpheme* is the smallest meaningful linguistic element. Morpheme that can be used alone are *compound-word morpheme* or *free morpheme*, such as *hǎo*(good), *lái*(come), *kàn*(look), etc. The morpheme that cannot be used alone are called *bound morpheme* or *non-word morpheme*, such as: *rì*(day), *mín*(people), *yī*(clothing).

The unit larger than *morpheme* is *word*. A *word* is the smallest meaningful language unit capable of acting independently. A *phrase*, referred to as the *word group* or *syntactic structure* by Zhu, is the combination of words. *Morpheme* and *word* are always discussed within the lexicon system, while *word* and *sentence* within the syntax system. (See Figure 1 below):

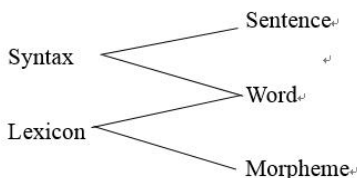


FIGURE 1. CORRESPONDENCE BETWEEN LANGUAGE UNITS AND THE RULE SYSTEM

4.2. **Criteria of solidification of disyllable compound structure.** As the traditional criteria have obvious shortcomings in classification. We will try to classify words and phrases according to the degree of typicality. In this case, prototype, that is, the strongest example of typicality, should be identified first, and then structures will be graded respectively according to their typicality.

We have synthesized explanations on distinguishing word and phrase in many textbooks [6], [8-10], and summarized a number of criteria for classifying words and phrases. In general, four criteria are formed:

(1) *It tends to be a word that has bound morphemes.* If two morphemes in a compound are both bound morphemes, such as *zǎotáng*(bathroom), *zhíwù*(plant), or one of them is bound morpheme, such as *jìyú* (crucian), *diànshì* (television), then it tends to be a word. Of course, the combination of two free morphemes are not necessarily phrase, such as *báicài*(cabbage) and *niúròu*(beef).

(2) *Words usually cannot be extended.* Because a word always has solidified and specific meanings, normally it cannot be extended; while the meaning of phrase is analytic, so it can usually be extended. Of course, a structure that can be extended may not necessarily be a phrase, because some words can also be extended, such as, *dàchē*(big cart → large cart), in which sense ‘*dàchē*’ is a phrase. But ‘*dàchē*’ also has a special meaning——a cart pulled by a horse or a mule, in which sense, it is inappropriate to classify it as a word.

(3) *If the meaning of an extended structure changed, it tends to be a word.* The meaning of a word is generally solidified, and often not a simple combination of the constituents, so

the meaning of a word will always change after extension. While for phrases, as the meaning of a phrase is analytical, and the meaning is normally the combination of the constituents, thus the meaning will not be changed after extension. But this is not always the case. For some words, the meaning will not change even after the extension because the meanings of these structures are combinations of constituents, such as: kuòdà(extension), yādàn( duck egg). And for some idiomatic syntactic structures that have conventional meanings, like ‘bànpíngcù(half bottle of vinegar, a dabbler), yǒu-yìsi(interesting, be attracted to)’, the meaning will change completely after extension.

(4) *In phonetic aspect, a word normally do not allow a pause within the structure.* For example, ‘huā’ and ‘hóng’ respectively mean ‘flower’ and ‘red’. There can be no pause between ‘huāhóng (name of tree) ’ which is a small deciduous tree, ‘huāhóng’ in this sense cannot contain any pause, but in the sense of ‘huā hóng le(the flowers are turning red)’, there can be obvious pause between ‘huā(flower)’ and ‘hóng(red)’, similar to ‘cǎolǜ(grass green)’, ‘tiānlán( sky blue) ‘and so on.

According to the concept of prototype category, typical words meet the most criteria, while typical phrases violate the most criteria. In the middle of the continuum are atypical words and phrases. We take the criteria above to test the typicality of a structure and can roughly divide the disyllable compound structures into several levels.

**5. Solidification grade of disyllable compound structure.** The disyllable compound structures that meet all the four criteria above belong to grade 4. On the contrary, the structures that meet no criterion belong to grade 0. According to the criteria (1) there are bound morphemes. Bound morphemes cannot be used as words alone, but must adhere to other morphemes, so the presence of bound morphemes indicates no possibility of extension, which means there is no need to question about whether the meaning changes after extension, or whether it can contain any pause. A structure that meets the criterion (1) is bound to meet the criteria (2) (3) and (4). All structures that with bound morpheme belong to grade 4, with the highest degree of solidification and can be regarded as the most typical words. For example, words containing two bound morphemes, like zǎotáng(bathroom), zhíwù(plant), or one bound morpheme, like diànshì(television), shūcài(vegetables), etc.

If criterion (1) is not met, then the criterion (2) should be considered. The absence of bound morphemes means that both constituents are free morphemes and can be used separately. But compound structures with strong solidification cannot be extended. For instance, ‘huǒchē(train)’, ‘shǒubiǎo(watch)’ cannot be extended to ‘huǒ de chē(fire car)’, ‘shǒu de biǎo(watch of hands)’. The non-extension indicates that it doesn't matter if the meaning changes after the extension (criterion 3) or whether it can contain pause (criterion 4). So if a structure meets criterion (2), it must meet the criterion (3) and (4), then it belongs to the grade 3. Examples are huāchá(flower tea), diànhuà( telephone), etc.

If criteria (2) is not met, then look at criterion (3). Compared with phrase, the meaning of a word is usually solidified, and not a simple addition of constituents. Although the two morphemes in some structures are compound-word morphemes and can be extended, the

extended meaning is different from the original meaning. For example, ‘càiyóu(rapeseed oil)’ can be extended to ‘cài de yóu(cooked food oil)’, and ‘lǜdòu(mung bean)’ can be extended to ‘lǜ de dòu(green beans)’. But ‘càiyóu (rapeseed oil)’ is not ‘cài de yóu (cooked food oil)’, but ‘oil pressed with rapeseed’, and ‘cài de yóu’ can refer to the oil in dishes; ‘lǜdòu (mung bean)’ is not ‘lǜ de dòu (green bean)’, but the seed of an annual herb plant. All in all, although for these structures, no bound morphemes exists and extension is possible, their meanings will change after extension, thus meeting the criterion (3). While the existence of solidified meaning indicates that there should be no pause among morphemes because pauses cause structural separation. Therefore, as long as they meets (3), they must meet (4) too, and their grade is 2.

If the criteria (1) (2) and (3) are not met, then we have to consider the criterion (4) in that a pause is not allowed. If both morphemes of a structure are compound-word morphemes and can be extended, and the extended meaning is similar to that of the original structure, then there are basically no conditions for typical words, but some of these structures are usually considered as words, such as yángmáo (wool of sheep), niúròu (meat of cattle), etc. We believe that criterion (4) can only be used to test whether the constituents can have a pause. It can be regarded as the last ‘line of defense’ between ‘word’ and ‘phrase’. If this criterion is not met, then there is no reason to identify a structure as a word rather than a phrase. A structure which only meets this criterion belong to grade 1.

Significantly, if a structure does not meet the above four criteria, it means that there is no ‘word’ characteristics at all, such as: ‘zuǐchán(greediness)’, ‘zuǐ(mouth)’ and ‘chán(greediness)’ are free morphemes. The structure can be expanded to ‘zuǐ hěn chán(very greedy)’, and the extended meaning is basically unchanged. There can be a pause in the middle, such as ‘(tā de zuǐ hěn chán)his mouth is very greedy’. ‘Tiānrè(the day is hot)’, ‘rè(hot)’ and ‘tiān(day)’ are both free morphemes, which can be extended to ‘tiān hěn rè (the day is very hot)’. The meaning after extension is basically the same, and there can be a pause in the middle, such as ‘tiān, rè a (the day, so hot!)’ The grade of these structures is 0. Thus we can establish a continuum of solidification between phrase and word, as shown in Figure 2:

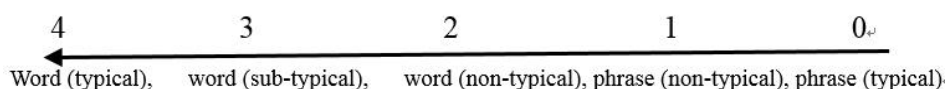


FIGURE 2. GRADE OF SOLIDIFICATION BETWEEN PHRASE AND WORD

This continuum reflects the gradual transition from a typical word to a typical phrase. If we must use a dichotomy to classify, we presume that grade 4 and 3 can be regarded as typical and sub-typical words, because neither of them can be extended. In contrast, for grade 2, the meaning changes after extension, thus should also be considered as word, but not a typical one. For the two grades on the right. Grade 1 can be regarded as the transition between typical phrase and non-typical word, and its solidification is a little stronger than

typical phrase (no pause in the middle), and has no conventional meaning compared with non-typical word. Grade 0 is typical phrase (syntactic structure), with the lowest solidification. Identification method and examples are offered in Table 1:

TABLE 1. IDENTIFICATION METHOD AND EXAMPLES.

Criteria Grade	(1)	(2)	(3)	(4)	Example
4	√	√	√	√	diànshì(television), shūcài(vegetables)
3	×	√	√	√	huāchá(flower tea), diànhuà( telephone)
2	×	×	√	√	càiyóu(rapeseed oil), lǚdòu (mung bean)
1	×	×	×	√	yángmáo (wool of sheep),niúròu (meat of cattle)
0	×	×	×	×	zuǐchán(greediness) ,tiānrè(the day is hot)

6. **The continuum of words and morphemes.** Under the concept of prototype category, since there is no obvious boundary between the word and phrase itself, the solidification is constantly declining from the typical word to the typical phrase, then a clear-cut boundary between morpheme and word (both taken as the constitutes of a compound) loses its factual base. So we end up representing the continuum as figure 3:

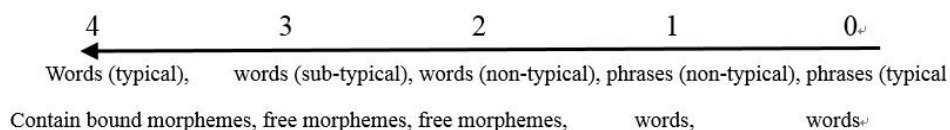


FIGURE 3. GRADE OF SOLIDIFICATION BETWEEN PHRASE AND WORD

Although the concept of continuum is more in line with the facts of the language, the solution we adopted in this paper is still segmentation of the ‘continuum’, which may be regarded as a compromise when there is no better way to describe language units.

## REFERENCES

- [1] Langacker, R.W., *Foundations of cognitive grammar, Vol. I, II*. Stanford University Press, Stanford, 1987/1991.
- [2] Dong Xiufang, *Chinese Lexicon and Morphology*, Peking University Press, Beijing, 2004.
- [3] Shen Jiakuan, Morphological Typology and Syntactic Typology, *Minority Languages of China*, no.6, pp.3-9, 2006.
- [4] Geert Booij, *The grammar of words: An introduction to linguistic morphology*, Oxford University Press, Oxford, 2005.
- [5] Shen Jiakuan, *Asymmetry and Markedness Theory*, the Commercial Press, Beijing, 2005.

- [6] Zhu Dexi, *Lectures on Grammar*, the Commercial Press, Beijing, 1982.
- [7] Zhang Yichun, An Analysis of Different Morpheme Sequences of Compound Modifier Structures, *Journal of Nanjing Normal University (Social Science)*, no.4,pp157-160,2002.
- [8] Lü Shuxiang, Issues on Chinese Grammatical Analyses (Hanyu Yufa Fenxi Wenti), *the Commercial Press*, Beijing 2005.
- [9] Shao Jingmin, *General Theory of Modern Chinese (xiandai hanyu tonglu)*, Shanghai Educational Publishing House, Shanghai, 2007.
- [10] Xing Fuyi, Wang Guosheng, *Modern Chinese(Xiandai Hanyu)*, Central China Normal University Press, Wuhan,2011.