

The Application of Contemporary Chinese Verbal Model Markers and Their Construction Corpus in an Electronic Dictionary of New Words

Ming Zhao^{1,*}

¹ School of Literature and Journalism Sichuan University, Chengdu, Sichuan 610207, China

*Corresponding author. Email: zhaoming528@126.com

ABSTRACT

Corpus can provide more powerful support and convenience for linguistic information processing, and the information obtained through corpus is richer and more complete, and the relevant research results are more objective and credible. In this paper, on the basis of statistical screening of more than 100,000 new words in 51 dictionaries of Contemporary Chinese, nearly 800 monosyllabic verbal model markers are compiled to build a corpus of Contemporary Chinese verbal model markers and their constructions. The article presents a multidimensional study of the verbal model markers and their word formation corpus, and points out its application prospect in the compilation and revision of electronic dictionary of new words.

Keywords: Contemporary Chinese, Verbal model markers, Corpus, Electronic dictionary of new words.

1. INTRODUCTION

It is generally considered that Contemporary Chinese is part of modern Chinese, the Chinese language since China's reform and opening up. During this period, social development was rapid, and new words emerged constantly. In order to make the object of study more focused and quantifiable, this article focuses on words newly created in the 40 years since China's reform and opening up in 1978, words borrowed from other ethnic languages, words borrowed from dialects, industrial words and domain vocabularies of our own language, words that have produced new meanings and uses or old words or phrases that have been revived after a dormant period of more than 70 years [1]. The term "new" is a relative term for a word or phrase that has come back into use after a dormant period of more than 70 years. The term "new" is relative, i.e. the year 1978 is used as a point of reference for words that are new or old, and no strict distinction is made between "words" and "phrases", which are collectively referred to as "new words".

In the Contemporary Chinese language, the family-like, explosive production of new words has

attracted the attention of many people. Li Yuming called this pattern of word production "words model". According to Li Yuming (1999:146), "words models are various frames that function as new word production. Such frames are composed of two parts: the 'model marker' and the 'model slot'. The 'model markers' refer to the unchanging words in the words model and the 'model slots' refer to the empty spaces in the words model" [2]. The disadvantage is that Li Yuming did not limit the minimum number of words in this framework, that is, how many new words should be produced at least to be counted as a words model. On the basis of the previous studies, in order to account for a greater number of Contemporary Chinese verbal model markers, this paper limits the words model to a framework consisting of two parts, the model marker and the model slot, which can construct three or more words. The model marker is a constant component of the words model, consisting of affixes, affix-like words, morphemes and compounds, and the model slot can be a morpheme, word or phrase, excluding clauses.

There are three main types of model markers: noun model markers, verbal model markers and adjective model markers. At present, there are more

studies on noun model markers in academic circles, but there is a lack of systematic studies on verbal model markers. In order to make the object of study more focused and pure, the verbal model markers in this paper refer to single-syllable verbs or verbal morphemes acting as model markers alone. In the words model, verbal model markers are used as morphemic constituents to participate in word formation. Therefore, the verbal model markers in this paper contain affixes, affix-like affixes, and verbal morphemes, and do not contain synthetic words or verbal phrases. Verbosity means that the grammatical character of the model marker is verbal.

In today's society, a systematic and in-depth study of the monosyllabic verbal model markers of Contemporary Chinese that can produce new words in batches and its word formation corpus is necessary and will be of some value for the compilation and revision of electronic dictionaries of new words.

2. AN OVERVIEW OF CONTEMPORARY CHINESE VERBAL MODEL MARKERS AND THEIR CONSTRUCTION CORPUS

2.1 Sources of New Words Corpus

Su Xinchun (2013: 11) pointed out that "dictionaries have also become the most effective, useful and informative type of special corpus for lexicometric studies" [3]. Therefore, the sources of the corpus in this paper are mainly from tools such as new words dictionaries [4]. In order to make the statistics more comprehensive and accurate, this paper follows the principle of well-known editors and well-known publishers, and carefully screens and examines 51 corpora published after 1978, with a total of more than 103,315 new words, and extracts 793 verbal model markers, which can be said to cover the situation in the more than 40 years of reform and opening up. To facilitate verification, a source index was adopted to mark the specific source of each new word in the constructed corpus in the tool book.

2.2 Principles for the Selection of New Words

Different dictionaries have different criteria for determining whether a word is new, and the criteria for inclusion are not always the same. After comparative checks and verification of the BCC

Chinese ephemeral corpus at the Beijing Language University, those that were clearly misclassified were removed. In the new word examples created by the verbal model markers, the conjunctive phrases (including abbreviated or shortened forms) were deleted. Different verbal senses under the same verbal model marker are marked as different model markers.

2.3 New Word Deletion and Reweighting Criteria

Different dictionaries of new words generally contain duplicates. When aggregating the total number of new words in a constructed corpus, censoring is required. The criteria for censoring are as follows: if the same word in different dictionaries has different meanings or different lexical properties, it is mainly based on whether there is a derivational relationship, and if there is a derivational relationship, it is regarded as one word; if there is no derivational relationship, but only the same form or the same sound, it is regarded as a different word. Specifically, words with different connotations are included together; those with the same connotation, or those with little difference in interpretation between dictionaries, are only selected for notation, with the main reference being the retention of the New Words Dictionary (1978-2018) [5].

3. APPLICATION IN THE ELECTRONIC DICTIONARY OF NEW WORDS

The research and development of electronic dictionaries is an important aspect of modernizing dictionary compilation [6] and automating language information processing. Using a corpus linguistic approach, Kang Shiyong et al. (2008:28-29) developed and established an electronic dictionary of information on new words in modern Chinese [7]. There are 36271 new words describing grammatical information in the electronic dictionary. The relevant research results of this paper can be applied to the compilation and revision of new words electronic dictionary in the following aspects.

3.1 Study on the Classification of Verbal Model Markers Sources

Classified by source, the 793 monosyllabic verbal model markers in the corpus can be divided into endogenous and exogenous types (see the

Table 1 below for details). Peng Xiao (2015:199-203) classified the autogenic model markers in Contemporary Chinese into four categories: firstly, those directly based on morphemes, secondly, those produced through abbreviation, thirdly, those produced through semantic specificity, and fourthly, those inherited from ancient times [8]. In this paper, the endogenous model markers are divided into five categories based on Peng Xiao's classification: firstly, those generated by abbreviation, secondly, those generated by common

semantic items, thirdly, those generated by semantic specificity, fourthly, those adopted from ancient times, and fifthly, those absorbed from dialects (i.e., dialectal model markers). Model markers from dialect refer to the model markers of Chinese dialect from outside Mandarin. The exogenous model markers are divided into English and Japanese model markers. The results of this study can help to better identify the morphemic origin of some new words in the electronic dictionary of new words.

Table 1. Contemporary Chinese verbal model markers by source

By origin	Secondary classification	Number (pcs)	Total number (pcs)
Endogenous model markers	Model markers from abbreviation	69	773
	Model markers from the common meanings of the morphemes	638	
	Model markers from Semantic specificity of the morphemes	23	
	Model markers from the past	19	
	Model markers from dialect	24	
Exogenous model markers	Model markers from English	13	20
	Model markers from Japanese	7	

3.2 Study on the Nature of Verbal Model Markers Constructions

In terms of the nature of the constructions, of the 793 verbal model markers in the corpus, 599 prepositional words are mostly verbal in nature, while 194 postpositional words are divided into three types of nature: still verbal, being transformed from verb to noun, and already transformed into noun. This indicates that a large number of verbs are being transformed into nouns due to the postpositioning of the verbal model markers. With the evolutionary development of the language, noun usage of monosyllabic verbs has become very common in Contemporary Chinese. As for the motivation of this phenomenon, Liu Danqing (2010) pointed out that "the use of verbs as nouns is an ontological metaphor, and any action, event and situation has the need to be referred to. Therefore, it is extremely common for verbs to be used as self-referential nouns, and is a very direct and easy to understand transformation" [9]. The results of this study can help to better determine and correct the part-of-speech problem of some new words in the new word electronic dictionary.

3.3 Study on the Grading of Verbal Model Markers Constructions

The word formation number of verbal model markers in the corpus has been set as a numerical column, and marked as low, medium, high and super high - those who can make 3-5 new words are marked as low; those who can make 6-10 new

words are marked as medium; those who can make 11-20 new words are marked as high; and those who can make 21 or more new words are marked as super high. As a result, a Contemporary Chinese verbal model markers word formation grading scale is obtained. The entries in the new words electronic dictionary with the verbal model markers as search words can be compared with the examples of words in the word formation corpus. The results of this study can help to better identify the issues related to the selection and arrangement of new words in the electronic dictionary headings.

3.4 Study on the Semantic Development of Verbal Model Markers

It is required to illustrate the development and evolution of the semantics of Contemporary Chinese verbal model markers in six ways: new uses emerge from the derivation of model meanings, adjectives are labelled as verbs in the inflected use of adjectives, new meanings emerge from the influence of foreign models, new meanings emerge from the influence of abbreviated combinations, new uses emerge from lexical metaphors, and new uses emerge from lexical infection [10] [11]. The research results can help to better identify the semantic issues of some new words in the new word electronic dictionary, such as the possibility of adding or refining word meaning items accordingly.

4. CONCLUSION

Corpus can provide people with sufficient research materials. The establishment of new words corpus provides an extremely convenient tool for the compilation and revision of new words electronic dictionary, which is conducive to the tracking research of new words, the sharing of resources, the improvement of dictionary compilation means and the improvement of dictionary compilation quality. It has high application value. In this study, the application of the research results related to the Contemporary Chinese verbal model markers and their construction corpus in the compilation and revision of the electronic dictionary of new words can be divided into four aspects: First, the results of this study can help to better identify the morphemic origin of some new words in the electronic dictionary of new words, and make the word interpretation more accurate. Second, the results of this study can help to better determine and correct the part-of-speech problem of some new words in the new word electronic dictionary, and make the part-of-speech tagging more accurate. Third, the results of this study can help to better identify the issues related to the selection and arrangement of new words in the electronic dictionary headings, and make the entry arrangement more scientific. Fourth, the research results can help to better identify the semantic issues of some new words in the new word electronic dictionary, such as the possibility of adding or refining word meaning items accordingly, and make the word interpretation more accurate.

Due to the limited time and effort, the verbal model markers and their construction corpus in this study are not comprehensive enough, the categorisation of new word examples still needs to be improved, and some quantitative research work on Contemporary Chinese new words has not yet been fully carried out, all of which restrict the in-depth application of verbal model markers and their construction corpus in the electronic dictionary of new words. This will be the main direction of the next research.

AUTHORS' CONTRIBUTIONS

This paper is independently completed by Ming Zhao.

REFERENCES

- [1] Yang Wenquan, Yang Xuming. On the influence of the growth and decline of Neologisms on the vocabulary system of Contemporary Chinese. *Journal of Sichuan Normal University*, 2008 (1)
- [2] Li Yuming. Words model [a]. Edited by Xing Fuyi. *The features of Chinese grammar*. Beijing: Beijing Language and Culture University Press, March 1999
- [3] Su Xinchun. A bibliometric study of dictionaries and vocabulary. Shanghai: Shanghai Dictionary Publishing House, December 2013
- [4] Zhou Jian. Sixty years of neologism research and neologism dictionary compilation. *Dictionary research*, 2015 (2)
- [5] Kang Shiyong, Liu Hairun, editor in chief. *New words dictionary (1978-2018)*. Shanghai: Shanghai Dictionary Publishing House, December 2018
- [6] Le Jiamin, Kang Shiyong, editor in chief. *research on modernization of dictionary compilation*. Shanghai: Shanghai Dictionary Publishing House, September 2009
- [7] Kang Shiyong, et al. *Quantitative research and application of modern Chinese neologisms*. Beijing: China Social Science Press, August 2008
- [8] Peng Xiao. *Language contact and non-original factors in the new quality of Contemporary Chinese vocabulary*. Chengdu: Sichuan University Press, January 2015
- [9] Liu Danqing. Chinese is a verb language -- on the type differences between verb language and noun language [J]. *World Chinese teaching*, 2010 (1)
- [10] Li Xingjian, editor in chief. *Modern Chinese Standard Dictionary (3rd Edition)*. Beijing: foreign language teaching and Research Press, June 2014
- [11] *Modern Chinese Dictionary (7th Edition)*, edited by dictionary editing Office of Institute of language, Chinese Academy of social sciences. Beijing: Commercial Press, September 2016