Construction of Adverbial-verb Collocation Database Based on Large-scale Corpus

Dan Xing[1], Endong Xun[1], Chengwen Wang[1], Gaoqi Rao[1], Luyao Ma[1]

[1] BeiJing Language and Culture Univercity, School of Information Science, Xueyuan Road 15, 100083 Beijing, China
xingdan1@126.com

Abstract. This paper constructs a high-quality adverbial-verb collocation database based on the large-scale corpus. First, we established a knowledge system of adverbial-verb collocations based on previous studies and linguistic rules. Then, we designed and implemented a knowledge acquisition model of adverbial-verb collocation based on a large-scale corpus. Finally, we evaluated and analyzed the extracted results. The main purposes are to obtain high-quality adverbial-verb collocations by formal means and to provide data support for natural language processing and basic linguistic and applied research.

Keywords: Large-scale corpus, Knowledge extraction, Adverbial-verb collocation

1 Introduction

Because of the lack of morphology in Chinese, phrases are combinations of multiple words with complex internal hierarchical relations, which makes it difficult to use word information effectively. However, collocation is a combination of two words, which lies between words and phrases, and builds a bridge between words and phrases. According to the syntactic structure between the components of collocation, collocation can also be divided into types such as subject-predicate collocation, verb-object collocation, modifier-head collocation, adverbial-verb collocation and verb-complement collocation [1]. In the chain of language units composition, word collocation is one of the important links [2], especially in dependency grammar analysis, since collocation describes the combination of words and does not involve phrase structure, it has become the most basic knowledge source of many automatic dependency grammar analysis systems [3]. The construction of the collocation knowledge base can improve the ability of the computer to process language, and also provide collocation cases for language ontology, teaching and application research. In this paper, we divide the resource construction of the adverbial-verb database into three parts: First, we construct a knowledge system of collocation based on previous studies; Second, we design and implement knowledge acquisition from a large-scale corpus; Finally, we analyzed the extracted collocation knowledge statistically.

2 Related Work

The word collocation was first introduced by Firth, who emphasized the concept of word pairing and pointed out that vocabulary in natural language is not used alone or in isolation, but in combination with other words [4]. Firth did not give a clear definition, however, Halliday defined collocation in the lexical framework [5]. Based on this definition, many scholars have inherited and developed the definition of collocation [6-9]. Benson et al. defined collocation as "an arbitrary and recurrent word combination", and classified collocations into a lexical group and a grammatical group [9]. Xu et al. gave a more specific definition: "a collocation is a recurrent and conventional expression containing two or more contents that hold syntactic and/or semantic relations [10]." This definition emphasizes the syntactic and semantic relations between words but excludes the function words [11]. The earliest study of Collocation in China mainly focuses on whether the essence of collocation is syntactic or semantic [12-16]. Besides, there are several practical achievements, such as Common Words Collocation, Concise Chinese Collocation Dictionary and Modern Chinese Substantive Collocation Dictionary [17-19]. However, researches at this stage have no consensus because of the lack of quantitative analysis [20].

The earliest collocation extraction was Choueka et al., who used the corpus of about 11 million words in the New York Times Weekly to extract word collocation by calculating the co-occurrence frequency of repeated adjacent word strings [21]. Church et al. used a news corpus of about 44 million words to calculate the mutual information between the words to get the ideal word collocation [22]. Smadja used the Xtract system to extract collocations from a stock market news corpus of about 10 million words and the results showed that the accuracy of collocation extraction can reach 80% [23]. The earliest research on collocation knowledge extraction in China was Sun et al., who referred to relevant foreign knowledge extraction and made a quantitative analysis of collocation in Chinese, introducing the definition and nature of collocation [24]. After that, Sun summarized 14 simple rules by using the method of statistics and rules to judge under what circumstances the "V + N" sequence in Chinese labeled corpus is a legitimate collocation combination [25]. However, these rules do not use any dictionary information; Qu et al. proposed a framework-based automatic word collocation extraction method [26]; Huang et al. extended Sketch Engine to Chinese and found POS-based rules were efficient in extracting grammatical information [27]. Sun et al. used the Maximum Entropy Markov Model to conduct basic block recognition experiments on CPTB and MSRA of 470,000 words scale[28]; Zhou developed a basic automatic analyzer for Chinese real text based on rule-based method [29]; According to Lin, rules-based and statistics-based methods have their strengths and weaknesses and can complement each other, but both rely on reliable linguistic knowledge. Previous studies on collocation knowledge extraction based on statistics, linguistic rules and the fusion of various statistical methods are usually based on a single lexical association method, which is based on the co-occurrence number of candidate collocations and constituent words in the corpus, often omitting the high-quality collocation with some co-occurrences, and the timeliness of data information leads to the extraction of "outdated" collocation combinations [30]. This paper is based on a BCC corpus which is relatively new, and the extracted data has relative timeliness compared with previous studies. Meanwhile,

this paper carries out the knowledge extraction project of adverbial-verb collocation based on the features of part of speech, word length, pause, rhythm and language rules, which is innovative in methods and provides a large-scale real data for further research based on statistics and machine learning and the examples of collocation for language ontology, teaching and applied research.

3 Construction of the Knowledge System of Adverbial-verb Collocations

There have been some previous studies on adverbial-verb structures. For example, Zhu classified adverbials into two categories: adverbial modifier and adjectival modifier. Adverbial modifier includes adverbs that are transformed by the verb and nouns with adverbial suffixes 的(*de, auxiliary*); adjectival adverbials include state adjectives, some compound words with state adjective suffixes can be used as adverbials. Zhu also thinks some substantives have the nature of the predicate so they are also modified by adverbials, such as numerals, quantifiers [31]. Xing thinks it usually defines adverbial as the modifier of verbs and adjectives. The adverbial verb can also be classified by noun or noun phrase. From the semantic point of view to classify the adverbial, it mainly includes the state adverbial, potential adverbial, degree adverbial, negative adverbial, condition adverbial, and object adverbial [32]. This paper establishes an adverbial-verb collocation system based on previous studies. The extracted adverbial collocations only refer to the simple non-recursive situations in which modifiers modify predicate headwords, such as 紧紧地抓住(*jin3jin3 de zhua1zhu4, hold tightly*), 不断地提高(*bu2duan4 de ti2gao1, improve constantly*), 很喜欢(*hen3xi3huan1, like very much*) and so on. Because of the particularity of the data extraction by computer, we do not consider the complex adverbial-verb structure for the time being, and the nouns which modify the predicate components such as 群众的支持(*qun2zhong4 de zhi1chi2, mass support*) and modifier which modifies the nominals, such as 很淑女(*hen3shu1nv3, very lady)* and 才周二(*cai2zhou1er4, only Tuesday*).

The adverbial-verb collocations in this paper are mainly centered on the predicative headword. It can be divided into three types in the form: adverbial + verb, adverbial+ adjective, adverbial + predicate pronoun. The adverbial-verb structure of complex adverbials and complex headwords is excluded for the time being. As shown in the table below:

Table 1. Classification table of adverbial-verb collocations

Adverbial collocation classification	Adverbial classification
Adverbial + Verb	Adverbial modifiers Adjective modifiers Verb modifiers Nominal modifiers Adverbial modifiers of predicative pronouns

	Quantitative phrase modifiers Modifiers of morpheme phrases Adverbial modifiers of preposition structures
Adverbial + Adjective	Adverbial modifier of adverbs Adverbial modifiers of adjectives Adverbial modifiers of predicative pronouns Quantitative phrase modifiers
Adverbial + Predicative Pronoun	Adverbial modifier

4 Acquisition of Adverbial-verb Collocations

4.1 The Composition of Retrieval Formula

The acquisition of adverbial-verb collocation is based on the corpus of Beijing Language and Culture University, which contains 15 billion words [33]. To extract collocation from the large corpus, this paper used a series of retrieval methods. We added conditional or output statements based on basic retrieval. Statements are separated by ";" , such as Query {cond1; cond2;...; condi; print ($i)}."Query" denotes the basic search formula; the conditional statement in "{}" restricts the query content; the output statement restricts the output content, and only one output statement can be used in a high-level search formula. The limited part of the search formula needs to be enclosed in "()", and according to the order in which "()" appears, it can be obtained by using "$" and serial number. The first "()" is represented by "$1", for example, (a) (n) {len ($1) > 1; $2= [S_N_名词]; len ($2) > 1} denotes the definite collocation that the attributive adjectives precede the noun. Nouns and adjectives can be acquired by "$1" and "$2" respectively in the order in which "()" appears. The qualifications in "{}" indicate that the length of adjectives and nouns is greater than 1 and that nouns are limited to the list of "S_N_名词"(*ming2ci2, the noun list*), with default output of "$1" and "$2".The accuracy of extraction results can be improved by restricting part of speech, length, the word list, and pause.

Table 2 shows the partial extraction results based on large data.

Table 2. The partial extraction results based on large data

Retrieval Constitution	Results	Frequency
(d)地 (v){len($1)=2;len($2)=1}	默默地看(*mo4mo4 de kan4, look silently*)	132
	大声地说(*da4sheng1 de shuo1, speak loudly*)	111
	远远地看(*yuan3yuan3 de kan4, look distantly*)	99
	慢慢地走(*man4man4 de zou3, walk slowly*)	86

(a)的 (n){len($1)>1;$2=[S_N_名词];len($2)>1}	重要的作用(*zhong4yao4 de zuo4yong4, important role*)	3952
	重要的意义(*zhong4yao4 de yi4yi4, important significance*)	3101
	好玩的活动(*hao3wan2 de huo2dong4, fun activities*)	2347
	最好的朋友(*zui4hao3 de peng2 you3, best friend*)	1337
	美好的回忆(*mei3hao3 de hui2yi4, good memories*)	1094

4.2 The Construction of Adverbial-verb Retrieval Formula

We constructed the retrieval formula of the collocation by four means. First, we make it through the existence of 地(*de, auxiliary*), 着(*zhe, auxiliary*) and other forms of markers; Second, we establish the word collocation table and the exclusive table; Third, we use rhythm structure for length restriction; Fourth, through pause, punctuation W restrictions and other forms of means to construct retrieval.

A case study of adverbial-verb collocation retrieval in the form of verbs modified by the adverbial modifier: First, according to the existence and absence of "地"(*de, auxiliary*), the retrieval forms of adverbial modifiers modifying verbs are divided into two categories. For example, the retrieval formula *(d)地(v){print($1 $2)}* refers to querying the corpus of adverbs followed by "地"(*de, auxiliary*) modifying verbs from large data and outputting the adverbs and verbs; The retrieval formula *(d) (v) {print ($1 $2)}* refers to querying the collocation of adverbs followed by verbs from large data and outputting adverbs and verbs. Second, we construct the word collocation table or the exclusive table to construct retrieval formulas to make the retrieval results more accurate and comprehensive. For example, the retrieval formula *(~) 地 (v){$1=[S_D_重叠_DD2];print($1 地 $2)}*. Table *S_D_重叠_DD2* is compiled and supplemented according to the Peking University Modern Chinese Grammar Information Dictionary, which indicates the double-syllable adverb overlap such as 常常(*chang2chang2, frequently*), 渐渐(*jian4jian4, gradually*) and so on [34].

Then, we construct the retrieval formula by restricting the length of the prosodic structure. Lv mentioned that the three-syllable structures are divided into two categories except for a few cases: adverbial-verb structure and verb-object structure [35]. For adverbial-verb combinations, 2+1 is much more than 1+2 in three-syllable combinations [36]. Wang believed that adverbial-verb combinations, 2 + 1 is the majority, such as "慢慢走" (*man4man4zou3, walk slowly*) and so on. The length of the phonetic affects the combination of words. According to this rule, a series of retrieval formulas are summarized, such as *(d) (v) {len ($1) = 2; len ($2) = 1; print ($1 $2)}* indicates the collocation of two syllables adverbs and one-syllable verbs.

Finally, we construct the retrieval formula by a pause. For example, the retrieval formula *(d) (v) W {len ($1) = 3; $2!= [是 有]; print ($1 $2)}* refers to adverbs with a length of 3, modifying the verbs except "是"(*shi4, be*) and "有"(*you3, have*), and the punctuation should be followed by verbs closely, such as "不由得哀号" (*bu4you2 de*

ai1hao2, couldn't help wailing), "好容易安置" (*hao3rong2yi4an1zhi4, It's hard to settle down*), "背地里安抚" (*bei4di4li3an1fu3, appease in the back*).

Table 3. The partial extraction and evaluation results based on large data

Modifier	Head	The retrieval of adverb-modified collocations	Evaluation of results
Adverbial modifiers of adverbs	Verb	(d)地(v){len($1)>1;print($1 地 $2)}	5
		(~~)地 v{$1=[DD_ABAB]}	4
		(~)地(v){$1=[S_D_重叠_DD2];print($1 地 $2)}	4
		(~)地(v){$1=[S_D_重叠_AABB3]}	4
		(~)(v){$1=[S_D_情态词_情];print($1 $2)}	4
		(~)(v){$1=[S_D_重叠_AABB3];print($1 $2)}	4
		(~)(v){$1=[S_D_重叠_DD2];len($2)=2;$print($1 $2)}	4
		(~)(v){$1=[S_D_重叠_DD2];print($1 $2)}	4
		(~)(v){$1=[S_N_兼类_d];print($1 $2)}	3
		(d)(v){len($1)=2;len($2)=2;$1!=$2;print($1 $2)}	4
		(d)(v)W{len($1)=3;$2!=[是 有];print($1 $2)}	4
		(d)(v){len($1)=3;$2!=[是 有];print($1 $2)}	4
		(d)(p)W{$2=[P_p 兼 v];print($1 $2)}	5
		(d)(v){len($1)=2;len($2)=1;print($1 $2)}	4

4.3 Evaluation of adverbial-verb retrieval formula

As shown in Table 3, the number refers to the accuracy rate, for example, 5 refers to 90%, 4 refers to over 70%, 3 refers to over 50%, 2 refers to 30%, and 1 refers to 10%. The results of this evaluation are combined forces of machine and human. Machine verification mainly excludes low-frequency invalid retrieval formulas. Human verification mainly includes two aspects: one is to observe the retrieval results, delete and improve the retrieval formulas through formal markers, such as the length of the word, and the other is to screen the retrieval results manually.

In constructing retrieval formula and extracting adverbial-verb collocations, we find that linguistic rules summarized by linguists can help us get higher quality adverbial-verb collocations. Take the retrieval of adverbial-head collocations "v+v" as an example, Sun thinks there is a great difference between the direct modification of verbs and the post-modification of verbs with auxiliary words, and direct modification

is limited [37]. Verbs are often modified by adverbials with auxiliary words such as “地” (*de, auxiliary*) and “着” (*zhe, auxiliary*), which highlight the modifiability of adverbials. According to this rule, we construct the retrieval formula (v) {$1=[P_feixinliv]; print ($1 $2)}, which improves the accuracy by 40%. Shen believes that verb modified as adverbials is very obvious, monosyllabic verbs as adverbials should be marked with “着” (*zhe, auxiliary*), disyllabic verbs and overlapping verbs should be marked with “着” (*zhe, auxiliary*) or “地”(*de, auxiliary*) [38]; Zhang believes that the number of verbs modified as adverbials is much more than the number of verbs modified adverbials directly [39]. Therefore, according to the rules summarized by linguists, we classify and refine the retrievable formulas, and classified the retrieval formulas of the verb modified by verb into three parts: V 着 V, V 地 V, and V-V. As shown in Table 4 below:

Table 4. The Retrieval Formulas of Adverbial-verb Collocation and Their Evaluation

The retrieval formulas of verb modified verb	Evaluation	Notes
(~)地(v){$1=[S_V_单作状语_地]}	5	The table *S_V_单作状语_地* is summarized according to the grammar information dictionary [34].
(v)地(v){$1=[P_vdv_v1];$1!=$2;print($1 $2)}	4	The table *P_vdv_v1* is summarized according to Sun and Zhang [37][39].
(v)地(v){len($1)=2;len($2)=1}	4	
(v)地(v){len($1)=2;len($2)=2}	4	
(v)地 v{len($1)=2;print($1 地 v)}	4	
(~)(v){$1=[S_V_单作状语_可];print($1 $2)}	3	The table *S_V_单作状语_可* is summarized according to the dictionary of grammatical information [34].
(~)(v){$1=[S_V_单作状语_可];$2!=[P_notv1v2x];$1!=$2;print($1 $2)}	4	The table *P_notv1v2x* is summarized according to retrieval results from big data.
(~)(v){$1=[P_V_双音节_状];len($2)=2;$1!=$2;$2!=[P_notv2v2];print($1 $2)}	4	The table *P_V_双音节_状* is summarized according to the dictionary of grammatical information [34]. The table *P_notv2v2* and the table *P_notv1v1* are summarized according to Sun [37].
(~)(v){$1=[P_V_单音节_状];len($2)=1;$1!=$2;$2!=[P_notv1v1];print($1 $2)}	3	
(~)(v){$1=[P_V_双音节_状];len($2)=1;$2!=[P_notv2v2];print($1 $2)}	4	

We can also use the retrieval results from large data to see whether the existing language rules are generally applicable. By observing the adverbial-verb retrievals of verbs modified by verbs, we can find that the retrievals restricted by part of speech or exclusive table and the retrievals with markers such as"地" (*de, auxiliary*) and “着” (*zhe, auxiliary*) are better than those without markers. In V-V retrievals, the retrieval effect of the query with first verb being two syllables is better than that with first verb being one syllable, which shows that there are a large number of double-syllable verbs used as adverbial modifiers to modify the verb head in V-V type and the regularity is obvious.

5 Statistical Analysis of Retrieval Results

Observing the statistical analysis Table5 as followed, we find that the proportion of adverbial modifiers modifying verbs is the highest, regardless of the proportion of the number of headwords or the collocation items, the proportion of “adverbial modifiers+verb” collocations in the total number of collocation items is as high as 44%, followed by the prepositional structure as the adverbial modifier which reflects the grammatical function of adverbs as adverbials.

Table 5. The Retrieval Results of Adberbial-verb collocations

Classification of Adverbial Collocations	Adverbial classification	Numbers of the headwords	The proportion of the headwords	Numbers of the items	The proportion of the item	examples
Adverbial + Verb	Adverbial modifiers	18922	0.1561	3851121	0.444028	偷偷地哭 (*tou5tou5 de ku1, cry secretly*)
	Adjective modifiers	17032	0.1405	1213775	0.139946	早知道 (*zao3zhi1dao4, know early*)
	Verb modifiers	17982	0.1484	811103	0.093519	同情地看 (*tou2qing2 de kan4, look at it with sympathy*)
	Nominal modifiers	14195	0.1171	233582	0.026932	现场表演 (*xian4chang3biao3yan3, performing live*)
	Adverbial modifiers of predicative	15592	0.1287	140417	0.01619	怎样做 (*zen3yang4zuo4, how to do*)

	Quantifier modifiers	5320	0.0439	10648	0.001228	一天天地过 (*yi4tian1tian1 de guo4, go through life day by day*)
	Morpheme phrases modifiers	132	0.0011	132	0.000015	小口地吃 (*xiao3kou3 de chi1, eat with a small mouth*)
	Preposition structures modifiers	16465	0.1359	1385811	0.159782	把苍蝇赶跑 (*ba3cang1ying1 gan3pao3, drive away the flies*)
Adverbial + Adjective Modifier	Adverbial modifier	4430	0.0366	823888	0.094993	尽情地放松 (*jin4qing2 de fang4 song1, relax as much as you like*)
	Adjectives modifiers	4201	0.0347	165403	0.019071	惊人地相似 (*jing1ren2 de xiang1si4 , similar astonishingly*)
	Predicative pronouns modifiers	4747	0.0392	26062	0.003005	那么美(*na4me mei3, so beautiful*)
	Quantitative phrase modifiers	2155	0.0178	6440	0.000743	一抹红 (*yi4mo3hong2, hints of red*)
Adverbial + Predicative Pronoun	Adverbial Modifier	9	0.0001	4766	0.00055	究竟怎样 (*jiu1jing4zen3yang4, how exactly*)
Total number		121182	1	8673148	1	

6 Conclusion and Prospects

This paper extracted 259 5108 adverbial-verb collocations based on BCC corpus, which mainly improves the retrieval method by observing the retrieval results and combines machine verification and manual screening to acquire knowledge.

Based on the construction of large-scale corpus, this paper mainly focuses on the knowledge system establishment, extraction, evaluation and data statistics of the adverbial-verb collocation knowledge system. We find that certain linguistic rules can guide knowledge acquisition, but the lack of linguistic formal markers, the ubiquitous grammatical and pragmatic ambiguities of words and sentences, and the habitual,

dynamic and multi-domain characteristics of language increase difficulties in knowledge acquisition. The results of linguistic research, such as the part of speech, word length, and prosodic information, can help us extract simple adverbial collocations more accurately. Collocation resources can help us find new ideas, supplement or discover new rules. However, collocation knowledge acquisition based on large-scale corpus also has drawbacks. First, some retrieved results will have errors, mostly in the low-frequency part, because of the imperfect classification of parts of speech, part of speech tagging system and inaccurate segmentation. Second, the existing linguistic rules are not thoroughly studied, or the existing linguistic laws do not meet the needs of natural language processing. When retrieving large data, there are many collocation noises, which cannot guarantee the correctness of acquired collocations. Finally, the lack of context, the limitations of collocation identification, overlap between different collocations could influence the accuracy.

For the existing problems, the future work prospects are as followed: First, given the imperfect part of the part-of-speech tagging system, we are trying to screen and tag the verb lists of BCC corpus manually. Second, because of the incomplete laws of linguistic research, we can establish collocation tables or exclusion tables according to the results and retrieve them again.

Acknowledgments. This work is supported by the Beijing Advanced Innovation Center for Language Resources (TYR17001), Graduate Innovation Fund (19YCX119) and National Natural Science Foundation of China (16AYY007).

References

1. Li, Y.D.: Modern Chinese Word Collocation. Commercial Press, Beijing (1998) (in Chinese)
2. Sun, H.L.: Inductive Grammar Rules from the Annotated Corpus: Experimental Analysis of the "V+N" Sequence. Papers Collection of the Fourth National Joint Conference on Computational Linguistics. Tsinghua University Press, 157-170 (1997) (in Chinese)
3. Lin, J.F.: Word Collocation Extraction and Its Application in Information Retrieval. Doctor, Harbin Industrial University (2010) (in Chinese)
4. Firth, J.R.: Modes of meaning in Papers in Linguistics, Oxford University Press (1957)
5. Halliday M A K.: Lexis as a Linguistic Level. In Memory of JR Firth, 148-162 (1966)
6. Sinclair, J.: Corpus, Concordance, Collocation. Oxford University Press, Oxford (1991)
7. Singleton, D.: Language and the Lexicon: An Introduction. Routledge (2016)
8. Kjellmer, G.: Some Thoughts on Collocational Distinctiveness in Recent Developments in the Use of Computer Corpora in English Language Research. Costerus.45, 163-171 (1984,)
9. Ilson, M., Benson, E., Benson, R.: The BBI Combinatory Dictionary of English: A Guide to Word Combinations. Amsterdam, John Benjamins (1986)
10. Xu, R.F., Lu, Q., Wong K.F., Li W.J.: Building a Chinese Collocation Bank. International Journal of Computer Processing of Languages 22(01), 21-47 (2009) (in Chinese)
11. Hu, R.F., Chen J.Y., Chen K.H.: The Construction of a Chinese Collocational Knowledge Resource and Its Application for Second Language Acquisition. Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics, Technical Papers (2016)
12. Xing, G.W.: Is collocation a grammatical problem?. Journal of Anhui Normal University(Humanities and Social Sciences Edition) (4), 77-84 (1978) (in Chinese)

13. Lin, X.G.: Research on Mr. Zhang Shoukang and the Collocation of Words. Journal of Capital Normal University (Social Science Edition) (1), 59-63 (1995) (in Chinese)
14. Lin, X.G.: The Nature and Research of Word Collocation. Chinese Language Learning (1), 7-13(1990) (in Chinese)
15. Lin, X.G.: On Semantic Classification and Word Collocation. Journal of Renmin University of China (5), 77-82 (1991) (in Chinese)
16. Zhu, Y.S.: The Semantic Basis of Collocation and the Practical Significance of Collocation Research. Foreign Language, 14-18 (1996) (in Chinese)
17. Zhang, S.K., Lin, X.G.: A Dictionary of Collocation of Common Words Used by Students. Hebei Children's Press, (1989) (in Chinese)
18. Zhang, S.K., Lin, X.G.: Concise Chinese Collocation Dictionary. Fujian People's Press, (1990) (in Chinese)
19. Zhang, S.K.: Modern Chinese vocabulary dictionary. Commercial Press, Beijing (1999)
20. Hu, Q.G., Gao, Q.Y.: Word Collocation and Teaching Chinese as a Foreign Language. Language and Translation (2017) (in Chinese)
21. Choueka, et al.: Automatic Retrieval of Frequent Idiomatic and Collocational Expressions in a Large Corpus. Literary and Linguistic Computing, 34-38 (1983)
22. Kenneth Ward Church, Patrick Hanks: Word Association Norms, Mutual Information, and Lexicography. Computational Linguistics, 22-29 (1990)
23. Smadja F.: Retrieving Collocations from Text: Xtract. Computational Linguistics 19(19), 143-177 (1993)
24. Sun, M.S., Huang, Ch.N., Fang, J.: A Preliminary Study on the Quantitative Analysis of Chinese. Chinese Language (1997) (in Chinese)
25. Sun, H.L.: Distribution Characteristics of Word Collocation in Text. Papers of the International Conference on Chinese Information Processing, 230-236 (1998) (in Chinese)
26. Qu, W.G. et al.: Frame-based Automatic Word Collocation Extraction Method. Computer Engineering 30(23), 22-24,195 (2004) (in Chinese)
27. Huang Ch.R. et al.: Chinese Sketch Engine and the Extraction of Grammatical Collocations. In Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing, 48-55 (2005) (in Chinese)
28. Sun, G.L., Huang, Ch.N.: Chinese Chunking Based on Maximum Entropy Markov Models. Computational Linguistics and Chinese Language Processing 11(2), (2006) (in Chinese)
29. Zhou, Q.: Automatic Learning and Extended Evolution of Chinese Basic Block Rules. Journal of Tsinghua University (Natural Science Edition) 48(1), 88-91 (2008) (in Chinese)
30. Lin, J.F.: Word Collocation Extraction and Its Application in Information Retrieval. Doctor. Harbin Industrial University (2010) (in Chinese)
31. Zhu, D.X.: Lectures on grammar. Commercial Press, Beijing (2016) (in Chinese)
32. Xing, F.Y.: Three Hundred Questions on Chinese Grammar. Commercial Press, Beijing (2014) (in Chinese)
33. Xun, E.D., Rao, G.Q., Xiao, X.Y., Zang, J.J.: Development of BCC Corpus under the Background of Big Data. Corpus Linguistics, Beijing (2016) (in Chinese)
34. Yu, S.W. et al.: Detailed Explanation of Modern Chinese Grammar Information Dictionary. Tsinghua University Press, Beijing (1998) (in Chinese)
35. Lv, Sh.X.: A Preliminary Study on Monosyllabic Problems in Modern Chinese. Chinese language, 10-22 (1963) (in Chinese)
36. Wang, L.J.: Rhythm Form of Chinese. Beijing Language University Press (2015) (in Chinese)
37. Sun, D.J.: An Investigation of Verbs as Adverbials in Modern Chinese. Language Teaching and Research, 116-129 (1997) (in Chinese)
38. Shen, J.X.: Asymmetry and Tagging Theory. Jiangxi Education Press, NanChang, 269-270 (1999). (in Chinese)

39. Zhang, J.: The Main Forms and Motivation Analysis of Modern Chinese Verbs as Adverbials. Journal of Central China Normal University (Humanities and Social Sciences Edition) 53(05), 100-108 (2014) (in Chinese)