Multi-category Words of Verb-Noun in Chinese Learner's Mobile App Dictionary and Their Applications for Dictionary Compilation

Siqi Xie

College of Chinese Language and Culture, Beijing Normal University, Xinjiekouwai St, Beijing, China

201922090058@mail.bnu.edu.cn

Abstract: One of the most important and heated topics in modern Chinese studies is the classification of word types. Among all word types, the word of multi-categories is a special and difficult case, especially multi-category words of verb-noun (VNs) for students learning Chinese as the second language. To help them learn this kind of word, several teaching methods and textbook compilation methods have been proposed, with only a few studies focusing on dictionary compilation. However, as a significant tool for second language learning, the dictionary plays an important role in helping learners improve their productive ability in using multi-category words. This paper examined the definitions of the multi-category word of verb-noun by different scholars and chose Yu's criteria from the need of learning Chinese as the second language. Common colligations and collocations of five VNs in native Chinese were extracted from two corpora and *Online Chinese Collocation Assistant* tool. Then, taking 说明 as an example, this paper analyzed the differences in its use by native Chinese and Chinese learners with a particular focus on its colligations and collocations. The results reveal that it was overused by Chinese learners as a verb, while underused as a noun. There was an imbalance between the use of the part of speech as verb and noun. Finally, this study proposed a better way to represent 说明 in a learning dictionary to help Chinese learners use this word.

Keywords: multi-category words, Chinese learner's dictionary, dictionary compilation

1. Introduction

Multi-category word is a special type of word in Modern Chinese, which has received a great deal of scholars' attention. Although the definitions of multi-category words differ, there is a consensus that multi-category words have two or more parts of speech simultaneously. Words of verb-noun, verb-adjective and adjective-noun are three common types of the multi-category words, with the verb-noun words being the most frequently used two-category words [1]. According to Hu, 12.91%-19.33% of multi-category words are verb-noun words (VNs), which account for a large proportion of all multi-category words [2]. However, no unified criteria were reached in terms of VNs in modern Chinese, so it's necessary to confirm an operatable definition which is useful for learning Chinese as a second language. Due to its high frequency and specialty, the Chinese learners are confronted with many difficulties in using VNs in written and spoken Chinese. They cannot fully master the use of two

category of words and they tend to use only one part of speech. Furthermore, they cannot use the common colligations or collocations accurately.

To help the Chinese learners improve the capability of using VNs, researchers have proposed a series of methods in teaching those words [3]. In addition to these teaching suggestions, the dictionary is also an important tool for language learning. However, even though there are some classical Chinese learning dictionaries, they still have many problems in their compilation, such as the lack of frequency-based representation, easily understandable definition, and learner-supportive explanation. Therefore, the article is intended to find a suitable definition of VNs for Chinese language learning, to investigate the use of VNs by Chinese learners and to propose a solution to their problems of using VNs by providing the mobile app dictionary method. This method features a detailed corpus-based investigation of VNs usage of colligations and collocations between native Chinese and Chinese learners.

2. Literature review

2.1. Multi-category words of verb-noun

Multi-category Part of Speeches is one of the most urgent and complex problem in modern Chinese [4]. Multi-category words means that they often undertake the grammatical functions of different parts of speech of two or more and are related in meaning. Therefore, Multi-category words of verb-noun (VNs) means that they can be used as a verb, also as a noun. Take the word 说明 shuo1ming2 as an example, when we say that "说明情况 shuo1ming2 qing2kaung4" (explain the situation), it is used as a verb; when we say that "药物说明 yao4wu4 shuo1ming2" (the drug instructions), it is used as a noun.

Zhu proposed two criteria for identifying VNs: 1) they can directly modify nouns without the word "的"de or be modified by other nouns; 2) can be the object of the words "有 you3, 进行 jin4xing2, 受到 shou4dao4, 加以 jia1yi3, 予以 yu3yi3" and so on [4]. Similarly, Hu [5] concluded three standards for identification: 1) can be directly modified by nominal measure words; 2) can directly be the object of "有 you3"; 3) can be directly modified by a noun.

Criteria argued by Zhu and Hu are both based on grammar function. But the identification of words can't be separated from the meaning of words, and definitions may vary from case to case. Another scholar Lu gave three different definitions of VNs in terms of different perspectives：1) from the needs of ontological research - words that are homophonic and homonymous but have different parts of speech; 2) from the needs of Chinese teaching - words with the same word form, the same sound, and a very close relationship in meaning, but with different parts of speech; 3) from the needs of Chinese information processing - words with the same word form, the same sound but different meanings or different parts of speech [6].

However, the standard of "very close relationship" is vaguely defined, so Yu deemed that the verb and noun of the dual category need to be semantic roles of each other [7]. For example, in the sentence "用锁 2 锁 1 门 yong4 suo3 suo3 men2"(use the lock 2 to lock 1 the door), the noun "Lock 2"is the tool of the verb "Lock 1"; the verb "Lock 1"is the function and role of the noun "Lock 2".

In conclusion, taking grammar function, the real need in teaching Chinese, and meaning into consideration I will take Yu's criteria as a definition of VNs.

2.2. The difficulties of VNs use for Chinese learners

As a very special kind of word, the multi-category word has complex types and numerous numbers, which is the reason why it has always been the key and difficult points in Teaching Chinese as the second language (TCSL) [8]. For foreign students, especially those whose native language is Indo-

European, it is very difficult to learn multi-category words, because Indo-European languages have explicit morphological markers to indicate part of speech of words, whereas Chinese does not and tends to have words with multiple categories. Among multi-category words, VNs are the most frequent and should be paid enough attention in teaching, and so it has attracted the attention of many researchers. Hence a great number of researchers studied the errors of Chinese learners in learning VNs.

Long found two types of errors in VNs' use; one is the non-part-of-speech errors, and another is the VNs' part of speech errors [9]. He points out that the former can be divided into three types, of which the most common is the substitution of VNs by other words. Through the analysis, he drew to the conclusion that one of the causes of such errors is the improper collocation[1], but there aren't enough data to support his idea, with only several examples. Based on his study, Wang made corpus-based comparative analysis on the VNs' collocations between native Chinese and Chinese learners through the CCL corpus, Modern Chinese Corpus of state Language Commission, and HSK Dynamic Composition Corpus [8]. She found that most of the typical colligations[2] frequently used by native Chinese are underused by Chinese learners, account for a low proportion or even do not appear under different parts of speech in interlanguage corpus, and a few typical colligations are used too frequently, so that they're underused in terms of their frequency and variety. Besides that, there are many typical collocations that are used with low frequency under different parts of speech of each multi-category word, or the collocations are not typical enough.

Although there are a number of problems when Chinese learners use VNs, there is limited research to help them reduce those mistakes. Considering the collocation is an important factor in learning VNs, and dictionaries usually present collocations as illustrative examples, I will take some advice from the point of view of dictionary compilation.

2.3. Dictionary compilation for Chinese learners

2.3.1. The Learners' dictionary compilation system

Learners' dictionaries should be different considerably from those for natives in compiling style, word entry, lexical definition, and illustrative examples in order to help language learners use sentences accurately. Having this aim in mind, Laufer suggested a special dictionary for productive purposes. He considered an entry should including three parts: (1) L1-L2 translations, followed by information about the L2 word (definitions, examples, etc.); (2) semantically related words; (3) additional L1 meanings of the L2 translations [10].

These requirements can be fulfilled by an electronic dictionary since such a dictionary can combine the features of an L2-L1bilingual dictionary, an L1-L2 bilingual dictionary and an L2 monolingual dictionary [11]. Then Hurskainen proposed a new corpus-based approach to dictionary compilation, whose core features include (a) single-word headwords; (b) multiword headwords; (c) various types of cross-references; and (d) a user-defined selection of use examples in context, including controlled random selection, and selection based on frequent contexts [12].

But the electronic dictionary is easy to trap into the "corpus only theory" [13], which is divorced from the real needs of ordinary learners. As learners' dictionaries serve users of different levels of proficiency, the examples they provide cannot always be cited exactly in the form in which they appear in the corpus, but need to be carefully edited before being placed in the dictionary [14].

---

[1] Collocation in this article refers to "any holistic lexical, lexico-grammatical or semantic unit which exhibits minimal recurrence within a particular discourse community" (Siepmann, 2006, p. 2), e.g. 说明情况 (explain the situation).

[2] Colligation is used in a phraseological rather than in a frequency-based sense, which means that it is used to denote a type of word combination (Nesselhauf, 2003). In this article it is a combination of a word and a part of speech, like 说明+v

2.3.2. The compilation system of Chinese learner's dictionary

Cheng proposed a "QEI" principle: Q (Quick), E (Easy), I (Immediate), and the method must consider three elements: (1) The time pressure of learning Chinese for non-native Chinese; (2) the entry arrangement is based on a syntactic-semantic unit that is easy to remember and use as a whole; (3) the setting and corresponding of the syntactic distribution pattern of terms clue 3 in bilingual illustrative example translation [15].

Furthermore, since the dictionary for English learners has received great success in the world, it deserves us to learn their compilation mothed. For example, Zhang has noticed that mainstream contemporary English Learner's dictionaries pay great attention to the application of linguistics, and show the semantic structure of the paraphrased words from various aspects of the language, including the information of morphology, syntactic collocation and semantic discrimination [16]. Therefore, there are more information items in its microstructure, which are generally around 25-30 items, while the information items in TCSL dictionaries are much less, generally not more than 10 items. Although the Chinese learner's dictionary has noticed the communicative mode of words and tries its best to express the rules of the use of words in terms of style and interpretation, the way of expression is vague and the syntactic pattern is separated from the concept, so that ordinary users can hardly understand the words, not to mention imitate and use. So, Zhang suggested incorporating information that reflects the distribution characteristics of words, such as syntactic structure, collocation, or colligation into the dictionary definition [16].

As the technology developed, how to make use of the advantages of the electronic dictionary is also an important issue. Ni summarized that the colorful icon, the multi-media assisted paraphrase, and the multi-level display will benefit learners a lot in the Chinese learner's mobile app dictionary [17].

Above all, the consensus reached in the academic circles is that TCFL dictionaries must reflect the characteristics of "extroversion", adhere to the principle of "practicality", ensure to meet the needs of non-native Chinese learners, and improve the efficiency and quality of Chinese learning.

3. Chinese learners' difficulties in their use of multi-category words of verb-noun

3.1. Methods

3.1.1. Target words

Target words were chosen according to the following four criteria: 1) medium level of difficulty so that they are easy for learners to acquire; 2) frequently used in both native Chinese and non-native Chinese; 3) frequently used in classical textbooks; 4) included in the classical Chinese learners' dictionaries[4].

Since this paper mainly focuses on VNs that of medium difficulty level, 27 VNs that are required of intermediate Chinese learners in the HSK4[5] were extracted from Long's [9] research.

According to the first criteria and previous research [9], VNs that are required for students in HSK4 were identified ($N = 27$) from the HSK Test Syllabus (followed by Yu's definition of VNs). Then I counted each word's frequency in the following four corpora: CCL Corpus, BCC corpus, HSK Dynamic Composition Corpus, and Corpus of Teaching Chinese as Second Language[6]. At last,

---

[3] In fact, it refers to the correspondence of chunk units.

[4] *Chinese dictionary of 8000 words, Learner's Dictionaries of Contemporary Chinese, The Commercial Press Learner's Dictionaries of Contemporary Chinese.*

[5] Hanyu Shuiping Kaoshi 4, Chinese Proficiency Test/Band 4, which is targeted at intermediate to advanced Chinese learners.

[6] The corpus gathered data from the mainstream textbooks for Chinese learners.

according to the frequency and the criteria above, 5 VNs were chosen as the target words: 保证 bao3zheng4、感觉 gan3jue2、经历 jing1li4、说明 shuo1ming2、规定 gui1ding4.

3.1.2. Corpus-based analysis of the use between native Chinese and Chinese learners

As mentioned before, many errors in the use of VNs result from improper use or misuse of colligations and collocations. Besides, collocations[7] are indispensable for language learners and second language teaching [18]. Therefore, the main focus of this research is the colligations and collocations' use of VNs.

To analyze the differences in VNs' use between native Chinese and Chinese learners, typical colligations and collocations of the five target words were extracted from BCC corpus, CCL corpus, Corpus of Teaching Chinese as Second Language (native Chinese corpus) and the HSK Dynamic Composition Corpus (interlanguage corpus).

High-frequency colligations and collocations of native Chinese were extracted with the help of the *Online Chinese Collocation Assistant*[8] tool. However, this tool only identifies colligations and collocation from two corpora (the Corpus of Teaching Chinese as Second Language and Chinese Wikipedia Corpus), so it can't fully represent typical usage by native Chinese. Therefore, as supplement, the use of colligations and collocations by native Chinese were further examined in the BCC and CCL Corpus. As for VNs' use by Chinese learners, the function of word collocation retrieval in the HSK Dynamic Corpus was used to extract the colligations and collocations by Chinese learners.

Taking 保证 bao3zheng4 as an example, firstly, typical colligations and collocations were extracted by *Online Chinese Collocation Assistant* (See Table 1).

Table 1: Typical Colligations and Collocations of 保证 in *Online Chinese Collocation Assistant*.

Colligation	Collocation	Frequency	Mutual Information	Typical Variants
Verb-Object	保证 质量	13	6.28	/
Adverb-Predicate	必须 保证	5	3.95	/
	就 保证	3	-0.73	/
Prepostion-Verb	向 X 保证	7	2.33	向 X 保证 过

Then I checked the BCC corpus and CCL corpus to see if there was any colligation or collocation that are not identified by the Online Chinese Collocation Assistant. Finally, the typical colligations and collocations of 保证 were summarized in Table 1. Following this procedure, the typical colligations and collocations of the remaining four target words were also extracted and presented in Table 2.

Table 2: Typical Colligations and Collocations of 保证.

Word	Colligation as a Verb	Typical collocation	Colligation as a Noun	Typical collocation
保证	~+Noun	保证质量	~+Noun	保证体系
	Adverb+~	必须保证		保证书
	Preposition+~	向 x 保证	V+~	有保证

---

[7] In that paper, the term 'collocation' is used in a phraseological rather than in a frequency-based sense, which means that it is used to denote a type of word combination. So, at here the definition of 'collocation' is the same as 'colligation' in my paper.

[8] Collocation Search - Online Chinese Collocation Assistant - Free online Chinese collocation dictionary, corpus, search (cca.xingtanlu.cn)

Table 3: Typical Colligations and Collocations of the Other Four VNs.

Word	Colligation as a Verb	Typical collocation	Colligation as a Noun	Typical collocation
感觉	**Adverb+~**	突然感觉	**Classifier+**	种感觉
		明显感觉	**Verb+~**	有 感觉
	~+Adjective	感觉好		能 感觉
		感觉不错		
	~+Verb	感觉到		
		感觉是		
		感觉不到		
规定	**Noun+~**	法律规定	**Classifier+~**	条规定
		国家规定	**Verb+~**	违反规定
	Adverb+~	明确规定		有规定
		严格规定		符合规定
		统一规定	**Adjective+~**	具体规定
				新规定
经历	**Verb+~**	要经历	**Classifier+~**	段经历
		会经历		次经历
	Adverb+~	亲身经历	**Verb+~**	有经历
		都经历		讲经历
		没有经历		说经历
	~+Noun	经历风雨	**Adjective+~**	类似经历
		经历人生		痛苦经历
				真实经历
				难忘经历
说明	**~+Noun**	说明问题	**Noun+~**	药品说明
		说明道理		情况说明
		说明情况	**Verb+~**	使用说明
	Noun+~	事实说明		看说明
		例子说明		
	Verb+~	举例说明		
		能说明		
	Adverb+~	就说明		
		也说明		
	Preposition+~	向 x 保证		
	~+Classifier	说明一下		

3.2. Results

Since 说明 is most frequently used among the five target words, it was analyzed in depth in terms of its colligations and collocations. There are 168,448 concordance lines of 说明 in BCC corpus and 401 concordance lines in HSK Dynamic Composition Corpus. The proportion was calculated respectively by dividing the total number of raw occurrences of the collocations by the number of concordance lines of 说明 in the two corpora.

Table 4: The Log-likelihood and Effect Size of Collocations of 说明.

Item	O1	%1	O2	%2	LL	LogRatio	Difference
说明问题 (Be illustrative)	1644	0.98	1	0.25	+3.09	1.97	-
说明情况 (Explain the situation)	804	0.48	4	1.00	-1.72	-1.06	+
事实说明 (The truth indicates)	1511	0.90	1	0.25	+2.63	1.85	-
情况说明 (The situation indicates)	619	0.37	2	0.50	-0.17	-0.44	/
举例说明 (Illustrate by examples)	518	0.31	4	1.00	-3.87*	-1.70	+
能说明 (Can explain)	1980	1.18	5	1.25	-0.02	-0.09	/
就说明 (Just show)	3332	1.98	17	4.24	-7.76**	-1.10	+
也说明 (Also explain)	3889	2.31	6	1.50	+1.31	0.63	-
向 X 说明 (Explain sth. to sb.)	281	0.17	5	1.25	-11.39***	-2.90	+
说明一下 (To explain)	817	0.49	21	5.24	-61.38***	-3.43	+
药品说明 (The medicine directions)	721	0.43	0	0.00	+3.43	1.78	-
看说明 (Read introduction)	183	0.11	0	0.00	+0.87	-0.20	-
按说明 (According to the introduction)	817	0.49	0	0.00	+3.89*	1.96	-
使用说明 (Operation instructions)	1214	0.72	0	0.00	+5.77*	2.53	-

Note. O1 means the BCC corpus and O2 means the HSK Dynamic Composition Corpus. "*" means $p < .05$; "**" means $p < .01$; "***" means $p < .001$. The Log Ratio statistic is an "effect-size" statistic, not a significance statistic. It represents the magnitude of difference between two corpora for a particular keyword. Every extra point of Log Ratio score represents a doubling in size of the difference between the two corpora for the keyword under consideration. "-" means underuse, "+" means overuse, "/" means equal use.

As shown in Table 4, in terms of the collocation of 说明, 按说明 and 使用说明 are underused by Chinese learners ($p < .05$) , 举例说明，就说明，向 X 说明 and 说明一下 are overused ($p < .01$).

Table 5: The Log-likelihood and Effect Size of Colligations of 说明.

Item		O1	%1	O2	%2	LL	LogRatio	Difference
Verb	**~+Noun**	5904	3.50	331	82.54	-1440.79***	-4.56	+
	Noun+~	5552	3.30	59	14.71	-84.59***	-2.16	+
	Verb+~	3321	1.97	61	15.21	-142.25***	-2.95	+
	Adverb+~	2433	1.44	3	0.75	+1.63	0.95	-
	Preposition+~	773	0.46	8	2.00	-11.14***	-2.12	+
	~+Classifier	817	0.49	21	5.24	-61.38***	-3.43	+
Noun	**Noun+~**	4124	2.45	0	0.00	+19.61***	4.30	-
	Verb+~	732	0.43	2	0.50	-0.04	-0.20	/
	Preposition+~	1176	0.70	1	0.25	+1.54	1.49	-
	VN+~	1551	0.92	0	0.00	+7.38**	2.88	-

In Table 5, it can be noted that the verb usage of 说明 by native Chinese is widely distributed in various colligations, such as ~+noun, noun+~, ~+verb，and ~+Classifier, the proportion of which is 35.05‰, 32.96‰, 19.72‰, and 4.85‰ respectively, while the verb usage of 说明 by Chinese learners in those colligations is much more frequent, the proportion of which is 825.44‰, 147.13‰, 152.12‰, and 52.37‰. Besides, the verb usage of 说明 by Chinese learners occurs much more frequently than noun usage. As for the noun usage of 说明, Chinese learners apparently underuse it at a low frequency, and there are even two colligations (i.e. Noun+说明 and VN+说明) which are not used.

In terms of the verb usage, the colligations of 说明+Noun, Verb+说明, Noun+说明, Preposition+说明, 说明+Classifier are greatly overused by Chinese learners ($p < .001$, the absolute value of LogRatio > 2). In terms of the noun usage, the colligations of Noun+说明 and VN+说明 are greatly underused by Chinese learners ($p < .01$, LogRatio > 2). Overall, there is a huge difference between the usage of 说明 by native Chinese and Chinese learners.

3.3. Discussions

The present study examined the differences of VNs use between native Chinese and Chinese learners. As shown by the above tables, the noun-part-of-speech is underused by Chinese learners, while the verb-part-of-speech is overused. The findings of the present study indicate that verb-part-of-speech of VNs were used more frequently than noun-part-of-speech. The findings are consistent with Wang's study [8] which investigated the VNs use. She found that Chinese learners tend to use a limited range of colligations and collocations, and she also observed that there is an imbalance between the verb and noun part of speech use, which corroborates current study [8]. To acquire the knowledge of word usage, language learners usually depend on classroom teaching and dictionaries. Therefore, if the textbooks or dictionaries don't represent the usage of VNs properly, Chinese learners may easily adopt negative learning strategies, which means that they will underuse some difficult or unfamiliar patterns to avoid making mistakes.

Moreover, this study shows that corpus-based analysis of colligations and collocations can be of great help for Chinese learners. Although there are only a few corpus-based studies on Chinese VNs colligations and collocations (e.g. [8]), studies on other languages demonstrate the importance of colligations and collocations on language acquisition [18-21]. The collocations can be a source of difficulty for non-native speakers of a language [22]. A common collocation typically used in the target language sometimes cannot be translated word by word, and its meaning depends on the specific context [21]. Besides, the colligations and collocations are essential for fluency and accuracy in spoken and written language, learning typical colligations and collocations rather than individual words helps learners to understand the meaning of lengthy utterance effectively [23]. Therefore, colligations and collocations indeed play a great role in second language acquisition and call for more attention in language teaching [18]. Identification and extraction of colligations and collocations have been facilitated by corpus linguistics which proposed a variety of methods to automatically extract collocations from the text corpora. Collocation coding based on corpus linguistics can effectively provide useful lexical information about language habits, which is useful for second language acquisition [21].

Collocations and colligations have been considered as an important component in dictionaries, since the publication of *Collins Cobuild English Language Dictionary* [24], which is the earliest learning dictionary complied by the corpus-driven approach, paying special attention to the typical colligations and collocations revealed by authentic collection of texts. From then on, learning dictionaries have entered the corpus-based compilation era. However, there still lack classical Chinese learning dictionaries that systematically represent typical colligations and collocations based

on real corpora. Considering the lack of specialized dictionaries in certain fields, especially for teaching Chinese as a second language, a new way of dictionary compiling through corpus-based analysis is proposed in the following section.

4. Suggestions for the Compilation of Chinese learner's Dictionary for VNs (Heading 3)

4.1. Introduction of the target dictionary APP *Pleco*

Pleco (Official website: http://www.pleco.com/) is one of the most famous and frequently used Chinese learning mobile app dictionary [25]. The dictionary was developed by a foreign student named Michael Love studying in China. Currently, the Pleco dictionary has more than one million users in 180 countries.

4.2. Analysis of the problem in dictionary compilation

In the era of Internet and smart phones, Chinese learner's dictionary APPs have come into being and are favored by many learners. However, there are still some demerits in those dictionary APPs, so this article will analyze some of the problems existing in dictionary APP compilation, taking Pleco as an example, with particular focus on the illustrative information.

Examples of word collocation and example sentences are mixed, as shown in Figure 1:

Figure 1: Screenshot of the Entry for 说明 in Pleco

The first four entries are phrases or collocations, and the fifth entry is an example sentence. In addition, the colligation patterns are not included in the compilation system of *Pleco* dictionary APP.

Word collocations are not arranged hierarchically, and there is no clear logic in the entry's order, neither by difficulty nor by frequency. Taking Figure 1 as an example, there are four collocations 说明理由，说明真相，说明用法，举例说明，however, 说明用法 and 举例说明 are easier than the first two phrases according to the HSK Test Syllabus. The entries are also not ordered by frequency, since the concordance numbers of the four collocations in BCC corpus are 471, 122, 18, and 518 respectively.

Some of the entries are not accompanied by example sentences, and some example sentences in *Pleco* lack context. As shown in Figure 1, there are only collocations of 说明理由，说明真相 and 举例说明 without example sentences, so that learners can still face difficulties in knowing how to use the collocations. Besides, the usage of 说明 in sentence "代表团认为有必要说明自己的立场" is not very typical, and the context of this sentence is not clear enough, so maybe it is a little bit confused for learners to understand when and where they should use 说明 in such sentence.

The example sentences are not very practical, some of which presented in the dictionary are rarely used even by native Chinese. For example, the collocation 说明立场 occurs only once in BCC corpus, so this sentence is not very useful for Chinese learners.

4.3. The Suggestion of dictionary compilation

4.3.1. The principle of compilation

There are two kinds of dictionary for language learners. One is called the introverted dictionary, which aims to serve the native language learners, and another is the extroverted dictionary, which is targeted for second language learners. For example, *Pleco* is an extroverted dictionary targeted for Chinese learners. Introverted dictionaries focus on language understanding, mainly on interpretation, and are supplemented by collocation information. Extroverted learning dictionaries focus on language generation, so the examples should strive to reflect the more comprehensive and typical usage of words, rather than providing evidence and supplement for interpretation like the introverted dictionaries. Therefore, how to present collocation information should be considered in the first place when compiling collocation examples of export-oriented learning dictionaries.

1. Orderliness: The orderliness principle requires that the arrangement of examples should show the hierarchical order from frequently-used to infrequently-used, from simple to complex, and from concrete to abstract. The orderliness will make learners clearly and directly know the most common use, and it should be combined with the law of learners' cognitive development [26].
2. Representativeness: The colligations and collocations must be retrieved from each type of source corpus, and every collocation should be complained with a representative example sentence, so that the examples can show its most typical usage [26].
3. Generativity: If the dictionary is cleverly compiled, the example sentences should enable the users of the dictionary to make sentences generatively [27]. Learners should be able to form some correct sentences in their mind by combining examples with their existing language basis.
4. Practicality: Language learning cannot be divorced from real life, so the examples should be practical and useful by Chinese learners in daily life, and are supposed to be put into a context that learners feel natural, understandable and willing to apply [28].
5. Context: Each example should be easy to grasp the meaning, with a complete and clear context [29]. If the example sentences lack context, the learners can hardly understand how to use it in their writing or speaking.

4.3.2. A sample

Based on the above principles, I proposed a sample for the dictionary compilation. The colligations and collocations were extracted from BCC corpus, CCL corpus, and Online Chinese Collocation Assistant tool, and were ordered by frequency. The example sentences were adapted from HSK Dynamic Composition Corpus oriented to Chinese learners.

Table 6: The Sample of 说明 for Dictionary Compilation.

| Head word Part of speech Meaning of the word Common colligation Common collocation The example sentences The translation of example sentence | 说明
Verb
1. to show that something is true by providing facts, information etc.
➢ **~+n.**
shuōmíngwèntí
a) 说 明 问 题
数字是无聊的，但却往往最能**说明**问题。
Numbers are boring, but they are often most telling.
shuōmíngdàolǐ
b) 说 明 道 理
这个故事**说明**了一个道理，人一定要诚实。
This story indicates a truth that people must be honest.
➢ **n.+~**
shìshíshuōmíng
a) 事实 说 明
事实**说明**：付出多的人赢的机会大。
The truth shows that the people who pay more have more chance to win.
qíngkuàngshuōmíng
b) 情 况 说 明
这些情况说明如果没有自己的主见，最后可能什么也得不到。
These cases prove that if you don't have your own ideas, you may end up with nothing.
➢ **v.+~**
néngshuōmíng bùnéngshuōmíng
a) 能 说 明 /不 能 说 明
你举的例子能说明他很受人欢迎，但不能说明所有人都喜欢他。
Your example shows that he is popular, but it doesn't show that everyone likes him.
➢ **adv.+~**
jiùshuōmíng
a) 就 说 明
我认为个人买车是好事，买得起汽车的人多，就**说明**我们的生活水平提高了。
I think it is a good thing for individuals to buy cars, because the fact that more people can afford cars shows that our living standard has improved.
2. to make something clear or easy to understand by describing or giving specific examples about it
➢ **v.+~**
jǔlìshuōmíng
a) 举例 说 明
请你举例**说明**你的观点。
Please give an example to illustrate your point of view.
➢ **prep.+pron.+~**
xiàng shuōmíng
a) 向 X 说 明
他不好意思向父母**说明**他点的都是这个饭店最便宜的菜。
He was ashamed to explain to his parents that all he ordered were the cheapest dishes in this restaurant. | **3. to announce something clearly or officially**
➢ **~+classifier**
shuōmíngyíxià
a) 说 明 一下
我得先**说明**一下，现在我不是学生，是老师。
Let me make it clear that I'm not a student now. I'm a teacher.

Noun. explanation; directions; instructions
➢ **v.+~**
kànshuōmíng
a) 看 说 明
我还是不太明白要怎么安装，能给我看看**说明**吗？
I still don't quite understand how to install it. Can you show me the instructions?
➢ **Prep.+~**
ànshuōmíng
a) 按 说 明
你按说明去做，一步一步来就可以学会了。
You can learn it step by step by following the instructions.
➢ **n.+~**
yàopǐnshuōmíng
a) 药品 说 明
吃药之前一定要看药品**说明**
Be sure to read the instructions before taking the medicine.
➢ **VN.+~**
shǐyòngshuōmíng
a) 使 用 说 明
这个机器的使用**说明**很详细，看了以后马上就会用了。
The user's instructions for this machine are very detailed. You'll be able to use it immediately after you read them. |

The above entries were compiled according to the five principles: orderliness, representativeness, generativity, practicality, and context. Each part of speech of the word was accompanied with the most frequently used colligations and collocations, and collocations were ordered by the frequency of use, which is rarely represented in the classical Chinese learning dictionaries. Each word of the example sentences was consistent with the HSK syllabus difficulty, so that learners can easily understand the meaning of the collocation and the example sentences without having to search dictionary again and again. Besides, compared with the traditional dictionaries, the definition of an entry is not a list of synonyms, but a detailed description of their semantics and usage.

5. Conclusion

In this article, a creative method for compilation of VNs in Chinese learner's mobile app dictionary learning dictionary app for Chinese learners through corpus-based work was presented. Firstly, I sorted out scholars' discussions on the definition of the multi-category word of verb-noun and decided to take Yu's [7] opinion as the criteria. Then I summarized the differences between native Chinese and Chinese learners in their use of VNs according to the native Chinese corpora and interlanguage corpora taking 说明 as a target words. Next, I analyzed some of the demerits of the widely-used mobile app dictionary Pleco, and summed up five principles of dictionary compilation. At last, I proposed a dictionary compilation sample of the VN 说明 according to the extracted common colligations and collocations.

However, there are still some limitations in my research: 1) the representativeness of the corpora used: the interlanguage corpus only contains the sentences in the HSK tests, so the themes and topics are limited in a certain area and cannot fully represent the usage of Chinese learners. For example, the collocation 药品说明 is used in special theme or topic, whereas the HSK test is a general test, rather than a test oriented at some specific area; 2) the results of the study should be disgusted with reference to Markedness Theory in second language acquisition [30]. For example, the underuse of noun usage may result from the learners' extensive use of the target language. It is possible for English speakers to overemphasize the characteristics of Chinese verb usage, and thus neglect the noun usage.

Despite the limitations, this article is of great importance for learning dictionary compilation, which offers a new way of dealing with the problems that traditional dictionary cannot solve, and suggestions to compile a dictionary that is production-oriented, linguistically accurate, and user-friendly, like simplifying the explanation by colligations and collocations. In the future work, expanding the size of corpora and collecting some elicited sentences through questionnaires or homework is probably a good way to improve the representativeness and comprehensiveness of corpora. Besides, due to the limited time and energy, the study only analyzed one word in details and can be extended to other words in the future.

References

[1] Liu, P., & Ding, J. (2015). A Research into the multi-category words of verb-noun in modern Chinese based on the quantitative analysis of the collocating classifiers. In Q. Lu & H. H. Gao (Eds.), Chinese Lexical Semantics (pp. 294–306). Springer International Publishing. https://doi.org/10.1007/978-3-319-27194-1_30

[2] Hu, M. (1995). Verb-Noun Words' Quantitative Research. Language Studies, 02, 91-99

[3] Yin, C. M.. (2007). The overall study on concurrent words of the HSK vocabulary outline [Master's thesis, XinJiang Normal University]. CNKI.

[4] Zhu, D. X.. (1985). Reasearch on modern Chinese grammar. Beijing: The Commercial Press.

[5] Hu, M. (1995). Study in Problems of Modern Chinese Parts of Speech. Studies of the Chinese Language, 05, 381–389.

[6] Lu, J. M. (2013). A Course on Modern Chinese Grammar Research (4th edition). Beijing: Peking University Press.

[7] Yu, S.W., Zhu, X.F. & Liu Y. (2020). The verb-noun cross-category issue revisited with reference to the generative lexicon theory. Lexicographical Studies, 04, 1-8+125.

[8] Wang P.. (2020). Investigation and analysis of collocation processing in the foreigners-oriented Chinese learning dictionaries: Take verb-noun, verb-adjective and verb-noun-adjective as examples [Master＇s thesis, Beijing Foreign Language University]. CNKI.

[9] Long H. (2016). Research on errors of multi-category words of verb-noun based on HSK dynamic composition corpus [Master＇s thesis, Minzu University of China]. CNKI.

[10] Laufer, B. (1995). A case for a semi-bilingual dictionary for productive purposes. Kernerman Dictionary News, 3, 2－4.

[11] Laufer, B., & Levitzky-Aviad, T. (2006). Examining the effectiveness of ＇bilingual dictionary plus＇ - A dictionary for production in a foreign language. International Journal of Lexicography, 19(2), 135－155. https://doi.org/10.1093/ijl/eck006

[12] Hurskainen, A. (2008). SALAMA dictionary compiler-A method for corpus-based dictionary compilation. Institute for Asian and African Studies, Box, 59.

[13] Kernerman, I., Gottlieb, H., & Mogensen, J. E. (2007). What＇s so good or bad about advanced EFL dictionaries. Gottlieb, Henrik－Jens Erik Mogensen (Eds.), 139－145.

[14] Adamska-Salaciak, A. (2013). Issues in compiling bilingual dictionaries. The Bloomsbury Companion to Lexicography, 213－231.

[15] Cheng, Ting-Au. (2010). On the basic problems of bilingual learning dictionary compilation. Lexicographical Studies, 04, 52-63+126.

[16] Zhang, Y. H. (2011). An investigation on the problems in the definitions in Chinese dictionaries for international learners: Based on a contrastive study of EFL and CFL lexicography. Chinese Teaching in the World, 25(1), 120－128.

[17] Ni, T. T. (2018). The analysis of paraphrase and illustration problems on Chinese learners dictionaries based on mobile terminal [Master＇s thesis, Nanjing University]. CNKI.

[18] Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. Applied Linguistics, 24(2), 223－242. https://doi.org/10.1093/applin/24.2.223

[19] Dziemianko, A. (2014). On the presentation and placement of collocations in monolingual English learners＇ dictionaries: Insights into encoding and retention. International Journal of Lexicography, 27(3), 259－279. https://doi.org/10.1093/ijl/ecu012

[20] Siepmann, D. (2006). Collocation, colligation and encoding dictionaries. Part II: Lexicographical aspects. International Journal of Lexicography, 19(1), 1－39. https://doi.org/10.1093/ijl/eci051

[21] Vaupot, S. (2021). Creating a bilingual dictionary of collocations: A learner-oriented approach. Indonesian Journal of Applied Linguistics, 10(3), Article 3. https://doi.org/10.17509/ijal.v10i3.31888

[22] McAlpine, J., & Myles, J. (2003). Capturing phraseology in an online dictionary for advanced users of English as a second language: A response to user needs. System, 31(1), 71－84. https://doi.org/10.1016/S0346-251X(02)00074-X

[23] Hunston, S., & Francis, G. (2000). Pattern grammar: A corpus-driven approach to the lexical grammar of English. Amsterdam/Philadelphia: John Benjamins Publishing.

[24] Sinclair, J. (ed.) (1987), Collins COBUILD English language dictionary. London/Glasgow: Collins Publishers.

[25] He Shan & Zhu, Rui Ping. (2018). A qualitative research of Chinese learners using dictionary app for autonomous learning: Based on Pleco. Ludong University Journal(Philosophy and Social Sciences Edition), 35(01), 59－64.

[26] Yu D.. (2019). On the principles of compiling examples in the foreign oriented Chinese-learning dictionary APP. Chinese As a second Language Doctoral Forum and the 12th Academic Forum on Teaching Chinese as a second Language For Graduate Students, 7.

[27] Farina, D. M. T. Cr. (1995). L. V. Shcherba: Towards a general theory of lexicography. International Journal of Lexicography, 8(4), 314－350. https://doi.org/10.1093/ijl/8.4.314

[28] Cheng, Ting-Au. (2004). Preliminary considerations on CFL pedagogical lexicography. Chinese Teaching in the World, 04, 85-94+4.

[29] Li Y. (2013). On the compilation of general-purpose Chinese dictionaries for foreign learners of Chinese. Lexicographical Studies, 05, 34-39+94.

[30] Callies, M. (Ed.). (2013). The Routledge encyclopedia of second language acquisition. In Markedness (pp. 406－409). Routledge.