

The Analysis of the Inclusion and Discrimination of Confusing Action Verbs in Commonly used for Chinese Learning Dictionaries

Liu Shutong¹ and Xiao Yu²

¹Huachiew Chalermprakiet University ², Guangxi Normal University
Corresponding Author, E-mail: shutongliu@gmail.com

Abstract

Verbs, as the core component of sentences, occupy a pivotal position in Chinese language teaching. The abundance of action verbs and the prominent issue of errors associated with them present significant challenges. Furthermore, several existing Chinese learning dictionaries face limitations, such as inaccurate selection of discriminatory word pairs and complex interpretations. Consequently, this paper aims to examine five commonly used for Chinese learning dictionaries, focusing on the inclusion of easily confused action verbs and analyzing their lexical discrimination. Through this investigation, we hope to provide targeted suggestions for improving the compilation of Chinese learning dictionaries.

Keywords: The Inclusion and Discrimination; Confusing Action Verbs; Commonly Used; Chinese Learning Dictionaries

Introduction

For foreigners who are learning Chinese, a dictionary serves as an essential aid in their daily studies, helping them resolve difficult points and deepen their understanding of the language. However, our investigation has revealed certain issues in the compilation and usage of these dictionaries, particularly in the aspect of distinguishing confusing verb pairs.

First, the vocabulary of the learning dictionary is not consistent with the actual confusion in the corpus. This means that high-frequency vocabulary that are often confused in practical usage and require focused discrimination, do not receive adequate attention in the dictionaries. Instead, the dictionaries tend to include words with fewer confusion, which, to some extent, undermines the practicality and pertinence of the dictionaries.

Additionally, there are problems in the discrimination section of the dictionaries. For instance, the use of difficult vocabulary in explanations can pose new challenges for students when they refer to the dictionary. The lack of a clear hierarchy among entries and the cross-discrimination of multiple words can make it difficult for students to understand the distinctions between words clearly. Furthermore, the excessive presentation of similar usages may lead to students' inadequate understanding of the subtle differences between words.

Therefore, this paper intends to conduct a detailed examination of five commonly used Chinese dictionaries, comparing their compilation with actual confusion data from the corpus. Through a thorough analysis of their vocabulary discrimination, we aim to identify shortcomings and propose relevant suggestions to provide targeted improvement recommendations for the compilation of Chinese learner's dictionaries.

* Received: October 13 2024; Revised: November 2 2024; Accepted: November 3 2024

The five synonym dictionaries selected for our study are: 1) *Dictionary of Chinese Synonyms* by Ma Yanhua and Zhuang Ying (2002), which includes 381 sets of synonyms comprising 1,046 words; 2) *1700 Groups of Frequently Used Chinese Synonyms* by Yang Jizhou and Jia Yongfen (2005), containing 1,718 pairs of synonyms with over 3,500 words; 3) *The Commercial Press Guide to Chinese Synonyms* edited by Zhao Xin and Li Ying (2009), featuring 1,030 synonym groups with more than 2,400 words; 4) *Similar Chinese Words and Expressions Distinctions and Exercises* by Fang Xujun (2012), which includes 948 word groups with 2,044 words; and 5) *A Dictionary of Chinese Synonyms* edited by Wang Huan (2015), containing 420 word groups with 850 commonly used words. These dictionaries hold high authority in synonym and near-synonym research, with a wide range of vocabulary that is highly relevant to the confusing action verbs we are studying.

Inclusion in Learner's Dictionaries

We screened and counted action verbs from the *Global Chinese Interlanguage Corpus* that had been used 100 times or more and confused with each other 8 times or more. Global Chinese Interlanguage Corpus, featuring rich resources, large scale and comprehensive types. The total number of words in the original corpus has reached 22.75 million words, and the annotated reached 115 million words, including written, spoken, and video corpus of Chinese language learners from 111 countries and regions. For instance, there are sentences with verb confusion in the corpus:

① *孩子们很高兴，因为他们还没【做(zuò)】[Cb 坐(zuò)]过飞机。*The children are very happy because they haven't 【done】[Cb: sit] in an airplane before.

② *孩子们应该可以自己【坐(zuò)】[Cb 做(zuò)]自己的事情，父母在孩子的身后看着他们，给他们鼓励。*Children should be able to 【sit】[Cb: do] and handle their own tasks, while parents watch from behind and offer encouragement.

In Example ①, the misused word is "做(zuò)done", and the correct word should be "坐(zuò)sit". We represent this as "做(zuò)do→坐(zuò)sit", "→" indicates the direction of confusion, meaning "do" was mistakenly used instead of "sit". There are 29 instances of "做(zuò)do→坐(zuò)sit" misuse in the corpus, which meets the criteria.

In Example ②, the misused word is "坐(zuò)sit", and the appropriate word is "做(zuò)do". This is denoted as "坐(zuò)sit→做(zuò)do". There are 5 cases of "sit→do" misuse in the corpus, which does not meet the criteria.

Using this method, we ultimately obtained 59 pairs of high-frequency and easily confused word pairs.

We individually counted the inclusion status of these high-frequency easily confused vocabulary in learner's dictionaries. If the vocabulary was included, it was marked with √. If the dictionary included only partially corresponding vocabulary with the high-frequency easily confused words, it was marked with o, and the information about the vocabulary was recorded. If the dictionary did not include the relevant words, it was marked with ×.

After individual counting, the inclusion status of the high-frequency confusing words in the dictionaries is as follows:

1. Word Pairs Not Included in the Dictionaries

There are 7 pairs of high-frequency confusing words that are not included in any of the 5 dictionaries, accounting for 11.86% of the 59 pairs of confusing words. Notably, most of these unrecorded words belong to the category of path verbs: 学 (xué) → 上 (shàng), 去 (qù) → 回 (huí), 去 (qù) → 坐 (zuò), 去 (qù) ← 往 (wǎng), 去 (qù) ⇌ 来 (lái), 去 (qù) ← 出 (chū), 去 (qù) ← 出去 (chū qù)

2. Complete Inclusion of High-Frequency Confusing Word Pairs

After a thorough examination of the 5 dictionaries, we found that a total of 20 high-frequency confusing words pairs were completely included. These pairs accounted for 33.90% of the total 59 confusing word pairs. This proportion reflects that the inclusion of high-frequency confusing words in dictionaries is not universal. The specific inclusion status is as follows:

Table 1. Word Pairs fully included in the dictionaries

Confusing Word Pairs in Global Corpus	Ma Yanhua (2002)	Yang Xianzho u(2005)	Zhao Xin (2009)	Fang Xujun (2012)	Wang Huan (2015)
帮助bāngzhù⇌帮bāng	√	√	√	×	o
帮助bāngzhù⇌帮忙bāngmáng	√	√	√	√	o
变biàn→改变gǎi biàn	o	√	o	o	o
表示biǎo shì→表现biǎo xiàn	√	×	√	×	o
表示biǎo shì→表达biǎo dá	√	√	√	√	√
穿chuān→戴dài	×	×	×	√	√
带dài→戴dài	×	√	×	×	√
访问fǎng wèn→拜访bài fǎng	√	o	√	o	×
给gěi→送sòng	√	√	o	o	o
见jiàn⇌见面jiàn miàn	√	√	√	√	×
开kāi→打开dǎ kāi	o	×	×	√	×
看kàn→见jiàn	√	o	o	×	×
看kàn→参观cān guān	×	o	o	×	√
看kàn⇌看见kàn jiàn	√	√	√	×	×
聊天liáo tiān→聊liáo	o	×	√	√	×
旅游lǚ yóu→游览yóu lǎn	√	√	o	o	×
旅行lǚ xíng←游行yóu xíng	o	o	o	√	×
去qù⇌走zǒu	√	×	×	×	×
去qù←上shàng	√	×	×	×	×
学xué⇌学习xué xí	×	√	√	√	×

The word pairs included in 5 Dictionaries are:

表示 (biǎo shì) - 表达 (biǎo dá)

The word pairs included in 4 Dictionaries are:

帮助 (bāng zhù) - 帮忙 (bāng máng), 见 (jiàn) - 见面 (jiàn miàn)

The word pairs included in 3 Dictionaries are:

帮 (bāng) - 帮助 (bāng zhù), 看 (kàn) - 看见 (kàn jiàn), 学 (xué) - 学习 (xué xí)

The word pairs included in 2 Dictionaries are:

表示 (biǎo shì) - 表现 (biǎo xiàn), 戴 (dài) - 带 (dài), 戴 (dài) - 穿 (chuān), 访问 (fǎng wèn) - 拜访 (bài fǎng), 给 (gěi) - 送 (sòng), 旅游 (lǚ yóu) - 游览 (yóu lǎn)

The word pairs included in 1 Dictionary are:

改 (gǎi) - 改变 (gǎi biàn), 开 (kāi) - 打开 (dǎ kāi), 看 (kàn) - 参观 (cān guān), 看 (kàn) - 见 (jiàn), 聊天 (liáo tiān) - 聊 (liáo), 旅行 (lǚ xíng) - 游行 (yóu xíng), 去 (qù) - 走 (zǒu), 去 (qù) - 上 (shàng)

3. High-frequency confusing word pairs are partially included

The high-frequency confusing words correspond to some words of the dictionary. For example, the high-frequency word pair in the corpus is “保持 (bǎo chí) → 保护 (bǎo hù)”, but the dictionary included “保持 (bǎo chí) - 维持 (wéi chí)” and “保持 (bǎo chí) - 坚持 (jiān chí)”, indicating an inconsistency with the target word. We found that 26 words were partially included in the dictionaries rather than fully corresponding. We have made a detailed record of the confusing situation of some included words in the corpus. Statistical details are provided in the table below:

Table 2: Confusion Cases of Partially Included Word Pairs

Confusing Words in <i>Global Corpus</i>	Pre-word confusion	post- word confusi on	Partially Included Word Pairs	Pre- word confus ion	post- word confu sion
保持 bǎochí → 保护 bǎohù	10	0	保持 bǎochí - 坚持 jiānchí	1	6
			保持 bǎochí - 维持 wéichí	0	2
打 dǎ → 踢 tī	9	0	打 dǎ - 玩 - wán	1	2
打 dǎ → 做 zuò	12	8			
发表 fābiǎo → 发 fāyán	14	0	发表 fābiǎo - 发布 fābù	1	0
			发表 fābiǎo - 刊登 kāndēng	0	0
			发表 fābiǎo - 出版 chūbǎn	0	0
喝 hē ← 吃 chī	0	42	喝 hē - 饮 yǐn	0	2
喝 hē ← 和 hé	0	8			
叫 jiào → 交 jiāo	9	0	叫 jiào - 喊 hǎn	0	1
叫 jiào → 教 jiāo	8	0	叫 jiào - 吩咐 fēn fù	0	0
进入 jìn rù ← 入 rù	0	8	进入 jìn rù - 进去 jìn qù	0	3
进入 jìn rù ← 入学 rù xué	0	13			
开 kāi → 骑 qí	12	0	开 kāi - 驾驶 jià shǐ	1	0

卖mài⇌买mǎi	55	39	买mǎi-购gòu	0	2
			买mǎi-购买gòumǎi	0	1
			买mǎi-采购cǎigòu	0	1
			卖mài-售shòu	0	0
			卖mài-出售chūshòu	0	0
拍pāi←拍照pāizhào	0	8	拍pāi-打dǎ	0	1
求qiú→救jiù	8	0	求qiú-请qǐng	0	0
			求qiú-请求qǐngqiú	0	0
			求qiú-要求yāoqiú	0	0
			求qiú-需求xūqiú	0	0
说shuō⇌说话shuōhuà	15	23	说shuō-告诉gàosù	22	18
说shuō←说明shuōmíng	0	11	说shuō-谈tán	1	7
说shuō←话huà	0	14	说shuō-讲jiǎng	3	2
说shuō→问wèn	12	0			
增加zēngjiā→提高tígāo	8	0	增加 zēngjiā-增长 zēngzhǎng	2	1
			增加 zēngjiā-增进 zēngjìn	0	0
			增加 zēngjiā-增添 zēngtiān	1	1
			增加zēngjiā-加jiā	0	6
做zuò→交jiāo	21	0	做zuò-作zuò	3	115
做zuò→坐zuò	29	0	做zuò-当dāng	2	16
做zuò→玩wán	8	0	做zuò-干gàn	0	1
做zuò→过guò	12	0	做zuò-搞gǎo	1	1
坐zuò⇌骑qí	19	10	坐zuò-乘chéng	0	1
坐zuò←用yòng	0	26			
坐zuò←上shàng	0	16			

It should be noted that words in red, such as “告诉” (gàosù), “当” (dāng), and “作” (zuò), are non-action verbs included in the dictionary. Although there are over 8 usage errors associated with these words, the confusion arises from the distinction between action verbs and non-action verbs, which is not within the scope of this study.

After analyzing, we identified a significant issue: high-frequency and easily confused word pairs in actual usage are not given adequate attention or included in the learning dictionary. On the contrary, the dictionary primarily includes words with relatively fewer cases of confusion.

For example, in our corpus, the pair “保持” (bǎo chí) and “保护” (bǎo hù) was confused and used interchangeably 10 times; however, the dictionary included the pairs “保持” (bǎo chí) - “坚持” (jiān chí) and “保持” (bǎo chí) - “维持” (wéi chí), which have lower rates of confusion. Specifically, the data show that there was only 1 misuse between “保持” (bǎo

chí) and "坚持" (jiān chí), and 6 misuses between "保持" (bǎo chí) and "维持" (wéichí) resulted in 0 and 2 misuses, respectively. A similar situation was found in the study by He Shan in 2018. We have compiled the statistics on the confusions and misuses of all partially included word pairs as follows:

Table 3: Number of Confusion Errors for Partially Collected Word Pairs

Confusion times	Number of phrases	Proportion
0	33	50.0%
1	16	24.2%
2	7	10.6%
3	3	4.5%
6	2	3.0%
7	1	1.5%
16	1	1.5%
18	1	1.5%
22	1	1.5%
115	1	1.5%

We found that cases with no misuse at all account for 50%, while non-high-frequency confusions with fewer than 8 occurrences account for 94%. Our research reveals a significant mismatch between the word pairs included in the dictionary and the actual confusion situations in the corpus. This finding highlights the obvious shortcomings of the learning dictionary in terms of word selection. There are no confusions at all in the corpus, yet the dictionary includes 10 pairs of differentiating words. Additionally, there are fewer than 3 instances of confusion in the corpus, while the dictionary includes 26 pairs of differentiating words.

Analysis of interpretation and Discrimination in the dictionary

After conducting an in-depth analysis of definitions and differentiations in learning dictionaries, we identified several significant phenomena:

1. Complex Definitions

The vocabulary used in the explanations is often more complex than the words being differentiated..For example, in Yang Jizhou (2005), the differentiation of the word pair "帮 (bāng)" and "帮助 (bāng zhù)" is noteworthy. "帮 (bāng)" and "帮助 (bāng zhù)" are classified as Level 1 and Level 2 vocabulary, respectively, but their explanations contain many advanced terms. For instance: 替 (tì) is a Level 4 word, 出力 (chū lì) is an out-of-scope term, 出主意 (chū zhú yì) is classified between Levels 7-9, 给以 (gěi yǐ) and 物质 (wù zhì)" are out-of-scope terms, while 支援 (zhī yuán) is a Level 6 word.

In the discrimination sentence, the meaning is the same, but the usage is different due to syllabic differences, for example, 帮 (bāng) is more colloquial, while 帮助(bāng zhù)

can be used in both written and spoken contexts, also involves advanced vocabulary like: 口语 (kǒu yǔ) (Level 4), 用法 (yòng fǎ) and 差异 (chā yì) (Level 6), 用于 (yòng yú) (Level 5), 书面 (shū miàn) (Level 7-9).

Clearly, the word pairs used for differentiation exceeds the complexity of the words themselves, making it very difficult for students newly introduced to Level 2 vocabulary to comprehend the dictionary contents.

Some scholars have proposed using purely situational example sentences to reduce the difficulty of differentiation, which means avoiding direct explanations of differences and instead allowing students to experience and summarize through real-life scenarios. However, we believe that although this mode can reduce the immediate cognitive load, it is not very efficient. Considering the richness of classroom content and limited time, the learning and differentiation of individual words should be precise and efficient, rather than relying on students' self-summarization and understanding.

2. English and Pinyin are less used

In the differentiation process of elementary vocabulary, we advocate for the primary goal to be clearly distinguishing differences between words and preventing usage errors. Proper use of English or Pinyin should not be neglected in this process.

Taking "帮 (bāng)" and "帮助 (bāng zhù)" as an example again: the English explanations provided in the dictionary both translate as help/assist/aid, but they do not offer further detailed differentiation. The Chinese explanations is relatively difficult, leading learners to look up dictionaries and understand the multiple meanings of words, which is clearly putting the cart before the horse.

Using English or the native language for differentiation can more effectively help learners comprehend precise differences between words, thereby enhancing learning efficiency. Therefore, in the differentiation of elementary vocabulary, we should not be confined to the principle of "using Chinese as much as possible." Of course, as learners' vocabulary increases and they progress to intermediate and advanced stages, the use of English or the native language can be gradually reduced.

It is worth noting that some dictionaries have already adopted bilingual annotation modes. For instance, Ma Yanhua (2002) included English explanations in the meaning descriptions and marked Pinyin next to Chinese sentences. This annotation approach allows learners to quickly obtain key information, significantly improves learning efficiency, and effectively reduces the fear of facing a large number of unfamiliar words.

3. Checklist-style Differentiation: Order of Items Requires Adjustment

After a thorough examination of existing dictionaries, we found that they generally adopt a method of listing checklists for comparative explanations, striving to elaborate on the various usages and differences between words comprehensively and in detail. However, when learners face with a large amount of information, this method fails to highlight key points effectively and causing confusion. Take Wang Hai's (2015) analysis of the differentiation between 帮 (bāng) and 帮忙 (bāng máng) as example:

1) 帮 has two meanings: ① To provide support, either mentally or materially, when others are engaged in certain tasks, or to actually participate in part of the work. ② Refers to engaging in physical labor for hire, such as “帮工” (bāng gōng) or “帮短工” (bāng duǎn gōng).

2) 帮忙 (bāng máng) is structurally a verb-object phrase, and semantically indicates specific assistance or support given when someone else is in need of help or labor. 帮倒忙

(bāng dào máng) means that while trying to help others but inadvertently adds trouble for them. 越帮越忙(yuè bāng yuè máng) has a similar meaning.

3)帮(bāng) and 帮忙(bāng máng) are generally not interchangeable. For example, “我要出去几天, 请你帮忙, 替我照顾一下家”(I will be out for a few days, please help me and take care of the house). If using 帮(bāng), it should be phrased as “我要出去几天, 请你帮我照顾一下家”(I will be out for a few days, please help me take care of the house).

Clearly, the above arrangement has shortcomings in highlighting key points. The lexical differences in the first point are difficult for beginner students to understand. Seeing such an analysis might confuse them or even cause a fear of learning. The second point, “帮倒忙(bāng dào máng),” which refers to infrequent semantic differences, can be placed later in the order, to be explored in-depth once the students' Chinese proficiency improves. In contrast, the content of the third point is more straight forward and practical and should be given priority. Differences that are most important and easily understandable should be placed at the beginning so that students can quickly grasp basic usages and avoid major mistakes.

We propose the following improvements: First, lexical analysis should be handled in a tiered manner, clearly distinguishing between primary and secondary aspects. For those differences that frequently occur and which are crucial for learners, they should be prioritized for study; for those subtle and less frequently used differences, they can be addressed later. This arrangement aligns better with learners' cognitive patterns and helps improve learning efficiency. Second, during the revision of dictionaries, it is essential to make full use of corpus resources to identify real instances of confusion and error types. By doing so, we can more accurately grasp the actual needs of learners, thus producing dictionaries that better meet their requirements.

4. Differentiating multiple words simultaneously with complex relations of similarities and differences.

In existing dictionaries, considerations for economic use of space and the pursuit of portability, sometimes the analysis of three or more words are put together. While this approach saves space, it also increases the complexity of information and the difficulty of understanding, especially for beginners. For example, Zhao Xin (2009) demonstrates this in the analysis of the entry “帮-帮助-帮忙”(bāng - bāng zhù - bāng máng), where almost every item involves differences among multiple word groups, including “帮(bāng) - 帮助(bāng zhù)”, “帮(bāng) - 帮忙(bāng máng)”, and “帮忙(bāng máng) - 帮助(bāng zhù)”. Given such complex relations of similarities and differences, we suggest that dictionaries should adopt a more simplified and systematic method for differentiating words.

Specifically, word differentiation should be done in pairs, only one difference point is analyzed in depth and examples are given at a time. This not only reduces the steps for learners in summarizing and analyzing but also effectively improves learning efficiency.

Furthermore, with the prevalence of online dictionaries and mobile apps, the size of the physical page is no longer a limiting factor. Therefore, in future dictionary compilation, more flexibility can be applied to word differentiation without considering the space constraints.

5. Interpretation of individual words need to be improved

In Yang Jizhou's (2005) analysis of 帮-帮助(bāng - bāng zhù), the author uses eight sentences to illustrate the similarities and differences between the two. However, among these

examples, seven demonstrate scenarios in which the two can be interchanged, while only one shows a situation in which they cannot be interchanged.

The dictionary indicates that 帮 (bāng) is primarily used in spoken language, while 帮助 (bāng zhù) can be used in both written and spoken contexts. Coupled with numerous examples of interchangeability, this may lead students to feel that the two words are essentially interchangeable. As a result, students might say “谢谢你的帮” (Thank you for your help), as the example sentences do not showcase the nominal usage of 帮助 (bāng zhù). However, in the HSK Standard Course, an example of 帮助 is “看电视对学汉语有帮助” (Watching TV is helpful for learning Chinese), where 帮助 (bāng zhù) cannot be replaced with 帮 (bāng).

Even in certain sentences where 帮 (bāng) and 帮助 (bāng zhù) can be interchanged without causing semantic errors, native speakers' actual usage habits may render some expressions unnatural or overly formal. For instance, using 帮助 (bāng zhù) in everyday matters might seem excessive, while 帮 (bāng) is more appropriate. Therefore, when conducting synonym differentiation, we should place greater emphasis on demonstrating the differences between words and situations where they cannot be interchanged. What students genuinely need to learn and master are the unique usages of these words in different contexts and how to avoid confusion.

When selecting examples of word collocations, we should focus more on their typicality and actual usage frequency to provide more accurate and practical learning guidance. Thus, in synonym differentiation, we should emphasize on listing the differences and should avoid using examples of interchangeable as much as possible. In other words, the similarities and interchangeable situations between vocabulary do not need to be memorized by students, what they need to learn are the distinct usages of the vocabulary and how to avoid confusion.

6. Lack of Corresponding Exercises

After differentiating words, corresponding exercises play a crucial role in reinforcing and enhancing the differentiation effect. However, traditional paper dictionaries, restricted by the limits of space and cannot include related exercises. Taking the lexicon with the largest vocabulary, compiled by Yang Jizhou (2005), as an example, it spans an enormous 1600 pages. If additional exercise content were to be included, it would necessarily entail a reduction in the number of words analyzed, presenting a difficult decision for dictionary compilers.

In response to this situation, scholars have already made beneficial attempts. Fang Xujun (2012) simplified the number of analysis based on frequently used vocabulary while incorporating rich exercises. He subdivided the analytical dictionary into elementary and intermediate versions, making corresponding adjustments in vocabulary selection and example design. This compilation strategy is not only innovative but also highly practical, worthy of in-depth study and reference.

Particularly, new tools such as online dictionaries or mobile applications provide extensive space for adding more diverse exercises. These platforms can utilize situational examples, images, and other multimedia forms to guide students more intuitively in differentiation of words, thereby enhancing learning effectiveness.

Suggestions

1. Expand the Scope of Word Selection

There are certain limitations in the selection of vocabulary in synonym dictionaries, which fail to reflect the confusion that arises in actual language usage. The selected vocabulary mostly consists of traditional synonyms. However, through our statistics and analysis, we have found that many semantically distant words also exhibit a significant amount of confusion, and these words are not included in the dictionaries.

Our statistics reveal that word pairs with relatively distant but still somehow related meanings account for 66%, indicating that although these word pairs are not strictly synonyms, they can still cause confusion due to some semantic connection.

To address this problems, we propose the following suggestions for improving the vocabulary selection strategy and expanding the scope of word inclusion: shifting the focus from synonyms to confusing words. Such a shift can more comprehensively reflect the actual situation of language usage and help learners more accurately master Chinese vocabulary.

2. Optimize Contrastive Word Pairs

Through our statistics analysis, we found that some of the words included in the dictionaries do not align with the confusion patterns in actual corpora. In other words, some synonym word pairs listed in the dictionaries are rarely mixed in real language applications. Conversely, the words that frequently cause confusion in the corpora are often not included in the dictionaries, leading to a lack of practicality.

In order to ensure the practicability of the dictionary, the words to be selected and determined according to the actual confusion in the corpus. Such an adjustment will make the dictionary closer to the actual needs of learners and improve their efficiency and accuracy in the process of Chinese learning.

3. Optimize Annotations and Examples

Some learner dictionaries face the challenge of providing difficult explanations in Chinese, the using vocabulary that is even more complex than the words being analyzed. It is important to use simple and commonly used vocabulary to explain the similarities and differences between words in a clear and accessible manner. Additionally, there are instances where English annotations consist of only one or two words, making it difficult to distinguish their differences. For example, the words 帮(bāng) and 帮助(bāng zhù) are both annotated in the dictionary with help/ assist/aid. In the analysis of elementary vocabulary, we should not be limited by the principle of “trying to use only Chinese.”. In the primary stage, we can use more detailed sentences to explain the differences in their meaning and usage. As learners' vocabulary expands and they reach the intermediate or advanced stages, we can gradually reduce the use of English or their native language. Some dictionaries also lack typical examples, failing to adequately demonstrate the actual usage differences between words; therefore, it is recommended to strengthen the typicality of example sentences.

4. Optimize Layout

In the above examination, we found that there are a variety of similarities and differences in the existing discrimination dictionary. Such discrimination, lack of focus and fails to highlight the key difference points.

Attention should be paid to the discrimination of vocabulary in the dictionary and should be graded and clear primary and secondary. The main discrimination of the confusion tendency should be ranked first. Instead of listing the comparison and explanation method of

checklist, we strive to comprehensively and elaborate the various usages and differences between words.

In the existing dictionaries, for the sake of layout and the pursuit of portability, they sometimes put three or more words together. Word discrimination should be carried out in pairs, with only one difference at a time and giving examples. This can not only reduce the steps of induction analysis in form, but also effectively improve the learning efficiency

5. Emphasize the Application of Mobile Apps and AI in Teaching

In recent years, the development of Chinese learning apps has progressed rapidly, greatly facilitating students' learning. Dictionary and translation apps allow for photo-based word searching and translation, as well as multilingual voice translation, making it rare for students to carry paper dictionaries anymore. Traditional paper dictionaries are limited by their size, unable to include too many words or examples, nor to incorporate exercises. However, these problems can be easily resolved in apps. We suggest that relevant dictionaries pay attention to the development of accompanying apps, ensuring that academic research can be put into practice and quickly transformed from theory to actual results. With the development of large-scale AIs like Chat GPT and Wenxin Yiyin, they are bound to play significant roles in Chinese teaching. We should leverage their advantages in answering questions and resolving doubts, while also being mindful not to rely excessively on AI, always maintaining the leading role of teachers.

Conclusion

In this study, we have conducted a detailed examination of five commonly used Chinese synonym dictionaries, comparing their vocabulary selection and discrimination with actual confusion data from the corpus. Our findings reveal issues in the compilation and usage of these dictionaries.

Firstly, there is a significant discrepancy between the word selection in learning dictionaries and the actual confusions occurring in the corpus. High-frequency, easily confused word pairs that require careful differentiation are not included in the dictionary; instead, the dictionary features words that exhibit less confusion. Among the word pairs included in the dictionary, 90% rarely show corresponding confusion in the corpus.

Secondly, the discrimination sections of the dictionaries exhibit several problems. Complex definitions using vocabulary more advanced than the words being differentiated pose challenges for learners. The lack of clear hierarchy and the cross-discrimination of multiple words further complicate the understanding of differences. The excessive presentation of similar usages also leads to inadequate understanding of subtle distinctions between words.

We propose several suggestions for improving the compilation of Chinese learner's dictionaries.

Firstly, the scope of word selection should be expanded to include not only synonyms but also words that frequently cause confusion due to semantic proximity.

Secondly, the contrastive word pairs should be optimized based on actual confusion data from the corpus to ensure the practicality of the dictionary.

Thirdly, annotations and examples should be simplified and optimized to use clear and commonly used vocabulary, and English annotations should be more detailed when necessary.

Fourthly, the layout of the dictionary should be adjusted to prioritize the main discrimination points and avoid overwhelming learners with excessive information.

Finally, we emphasize the potential of mobile apps and AI in improving dictionary usage and language learning efficiency.

In summary, through our analysis, we have identified some shortcomings in the current compilation of Chinese learner's dictionaries and proposed practical suggestions for improvement. These findings and recommendations aim to enhance the effectiveness and relevance of dictionaries for foreign learners of Chinese, helping them better distinguish confusing verb pairs and deepen their understanding of the language.

References

- Ma, Y. & Zhuang, Y. (2002). *Dictionary of Chinese Synonyms*. Beijing: Peking University Press.
- Yang, J. & Jia, Y. (2005). *1700 Groups of Frequently Used Chinese Synonyms*. Beijing: Beijing Language and Culture University Press.
- Zhao, X., & Li, Y. (2009). *The Commercial Press Guide to Chinese Synonyms*. Beijing: The Commercial Press.
- Fang, X. (2012). *Similar Chinese Words and Expressions Distinctions and Exercises*. Beijing: Beijing Language and Culture University Press.
- Wang, H. (2015). *A Dictionary of Chinese Synonyms*. Beijing: Beijing Language and Culture University Press.