



A Preliminary Study of Confusable Words in L2 Chinese Textbooks

Chun-Ting Chou and Li-ping Chang^(✉)

Graduate Program of Teaching Chinese as a Second Language, National Taiwan University,
Taipei 106319, Taiwan
lchang@ntu.edu.tw

Abstract. The present study adopts a corpus-based method to investigate the effectiveness and reliability of the confusable word sets listed in the L2 Chinese textbooks. Through the quantitative analysis based on the interlanguage corpus, we have found that, among 134 sets of words listed in the textbooks, only 50 sets have met the criterion we had set for “confusable words”. We hence provide a suggested word list based on the results of the analysis, a list of the sets that do not meet the criteria but are found in the corpus with more obvious mistaken usage, as well as a list of the sets that meet the criteria but are not the most typical ones. Through the preliminary research, it is hoped to provide teachers and textbooks editors with a more scientific method of selecting confusable words to improve the efficiency of teaching and learning.

Keywords: Confusable words · Chinese teaching · Corpus · Misuse rate

1 Introduction

Along with the expansion of L2 learners’ lexicons while learning a second language, more and more words sharing semantic similarities to some extent (i.e., the so-called “near synonyms (*jìnyìcí*, 近义词)”) may cause confusion to the learners, and subsequently can be misused by learners. Therefore, the ability of discriminating near synonyms is important for advanced learners. In order to raise the learners’ awareness of near synonyms, language teachers or editors of teaching materials have designed materials to remind the learners of the subtle differences among near synonyms, which should not be neglected.

A corpus-based approach can offer data and results for analysis based on massive and real language use. Some previous studies on differentiating synonyms have adopted the corpus-based approach to provide evidence for their analyses. For instance, Tsai et al. conducted a series of research since 1999 to analyze the lexical semantics and syntactic features of synonyms (e.g. discriminating *gāoxìng* 高兴 and *kuàilè* 快乐, both of which are often translated into “happy” in English). Xiao and McEnery (2006) analyzed collocations and semantic prosody of synonyms and they also set up criteria for discriminating synonyms, both in Chinese and English. They cited multiple sets of examples of synonyms to illustrate that both Chinese and English have semantic prosody, which

can be identified by native speakers' intuitions. L2 learners, however, lack such kind of intuitions. Thus, it is necessary to teach the learners to understand the differences among synonyms, in order to assist them to learn the language more effectively.

Near synonyms can be misused due to the semantic similarities among each other. Nevertheless, not every set of words misused by learners were caused by semantic similarities. For examples, (1)–(2) are the error sentences made by learners of Chinese as Second Language (henceforth CSL), which were found in the interlanguage (L2) corpus¹:

- (1) 那个 时候 世界 上 刚 受到 战争 的
Nàge shíhòu shìjiè shàng gāng shòudào zhànzhēng-de
 That time world up just receive war-PARTICLE
 *印象, 经济 起居 是 很 困难 的。
 **yìnxiàng, jīngjì qǐjū shì hěn kùnnán-de.*
 *impression economic living be very difficult-PARTICLE
 “At that time, the world had just undergone the *impression of wars,
 and the financial situations were very poor.”
- (2) 我 来到 中国 以后, 对 北京 的
Wǒ láidào Zhōngguó yǐhòu, duì Běijīng-de
 I arrive China after towards Beijing-PARTICLE
 *影响 非常 好。
 **yǐngxiǎng fēicháng hǎo.*
 impact very good
 “After I came to China, my *impact toward Beijing is very good.”

In the examples above, the misused words, *yìnxiàng* (印象, impression) and *yǐngxiǎng* (影响, impact) were not synonyms or near synonyms. They had no similarities in meaning. They were often similar in Chinese L2 learners' auditory sense instead. These examples reflect that language teachers and editors of teaching materials should not only put their attention to teaching how to discriminate near synonyms, but also the words that had the similar pronunciation, forms, etc.

Laufer noticed in 1988 that not only semantic features would cause confusion. In her studies, the concept of “synforms” was firstly mentioned. “Synforms” was defined as sets of words which have similar forms. The similarity on sound, script or morphology are what cause confusion. With similar concept of synforms, Carpenter (1993) collected 400 sets of “confusable words” from the perspectives of EFL learners. He indicated that the confusable words may “look or sound similar”. Taking the sentences (3)–(4) below for examples, which are found in the L2 corpus, *lǚxíng* (旅行, to travel) and *yóuxíng* (

¹ The examples in this paper are taken from HSK Dynamic Composition Corpus 2.0, <http://hsk.blcu.edu.cn/>. In the corpus, we have found 25 tokens of 印象 (“impression”) that should be replaced with 影响 (“impact”), which account for 80.6% of the 31 misuse instances of 印象 (“impression”) in total. Meanwhile, 13 tokens of 影响 (“impact”) have been found to be misused and should be replaced with 印象 (“impression”), accounting for 36.1% of the 36 misuse instances of 影响 (“impact”) in total.

游行, to parade) contain not only the same character “*xíng* (行)”, but also the characters *lǚ* (旅) and *yóu* (游) which share similar shape².

- (3) 每个人 来 中国 的 目的 不同，
Měi-ge-rén lái Zhōngguó-de mùdì bùtóng,
 Everyone come China-PARTICLE purpose different
 有的 来 *游行， 有的 来 谈 生意，
*yǒude lái *yóuxíng, yǒude lái tán shēngyì,*
 some come parade some come talk business
 还有 不少 外国人 来 学习。
háiyǒu bù-shǎo wàiguórén lái xuéxí.
 also many foreigner come study
 “Everyone came to China for different purposes. Some came for *parade and some for business. And yet, many foreigners came to study.”
- (4) 如果 自己的 国家的 人 去
Rúguǒ zìjǐ-de guójiā-de rén qù
 If self-PARTICLE country-PARTICLE person go
 外国 *游行 时 随地 扔掉 烟，
*wàiguó *yóuxíng shí suídi rēngdiào yān,*
 foreign-country parade time random throw cigarette
 这 不是 丢人 的 事 吗？
zhè bù shì diūrén-de shì ma?
 this not be shame-PARTICLE matter Q-PARTICLE
 “If people from one’s own country litter cigarettes during their *parade in foreign countries, it is a shame, isn’t it?”

Both *yóuxíng* (游行, to parade) in the sentences above should be replaced with *lǚxíng* (旅行, to travel). *Lǚxíng* and *yóuxíng* do not share semantic features, but they can still be misused. The sentences (1)–(4) illustrate that the errors made by the learners are not only caused by the near synonyms but also by the “synforms”.

To solve the problem in learning, the concept of “confusable words” (*yìhǔnyáocí*, 易混淆词) has been formed and adopted into Chinese teaching materials and lessons in recent years. Nevertheless, what are the criterion for identifying confusable words so that a more accurate list of confusable words can be included in teaching materials? Are the confusable words in the teaching materials worth being taught? These issues have not been discussed before.

In order to address these issues, the present study investigates the effectiveness and reliability of confusable words listed in L2 Chinese textbooks. By adopting a corpus-based approach, this paper finds out a set the criterion for the selection of confusable words for teaching materials.

² 5 tokens of 游行 (“to parade”) have been found in the corpus which should be replaced with 旅行 (“to travel”), which account for 62.5% of the 8 instances of misuse of 游行 (“to parade”) in total. No case has been observed that 旅行 (“to travel”) is misused to express 游行 (“to parade”).

2 Confusable Words and the Previous Studies (*yìhǔnyáocí*, 易混淆词)

The term “*yìhǔnyáocí* (confusable words)” was proposed by Zhang (2007). In her schema, confusable words should not be analyzed within the frameworks of synonyms or near synonyms. Zhang believed the research on near synonyms in Chinese had contributed to the theoretical aspect but had little help in Chinese teaching. For learners, not only near synonyms cause errors, some words which have no semantic relations also cause confusion. Therefore, the studies of words discrimination for learners should not be limited to near synonyms. They should turn the perspectives to learners’ interlanguage instead. Furthermore, the selection of confusable words should be based on the frequency of misusing, and the words that are frequently misused should be discriminated in priority. She also classified seven different confusable word categories according to the factors of learners’ misuse and misunderstanding, which are listed below:

Category A. Influenced by the forms, sounds and meanings of the words:

1. Words with basically the same meaning. E.g. *jiěshì* (解释, to explain), *shuōmíng* (说明, to explain).
2. Words with the same morpheme. E.g. *yǒu diǎnr* (有点儿, a little bit), *yī diǎnr* (一点儿, little bit); *kàn* (看, to see), *kànjiàn* (看见, have seen).
3. Words with the same or similar pronunciation. E.g. *wénmíng* (文明, civilization), *wénmíng* (闻名, famous); *yǒu* (有, there be), *yòu* (又, again).
4. Words with the similar script. E.g. *dà* (大, big), *tài* (太, too); *tíshì* (提示, hint), *jiēshì* (揭示, unfold).

Category B. Influenced by learners’ first languages (L1):

5. The word in learners’ first language while translated to Chinese is polysemy. E.g. “by” in English corresponds to *zuò* (坐, to sit) or *qí* (骑, to ride) in Chinese.
6. Characters in the learner’s first language consist of Chinese characters, e.g., “经验” in Japanese corresponds to the meaning of *jīngyàn* (经验, experience) or *jīnglì* (经历, background) in Chinese.
7. Two or more dialect words become one word while translated to Mandarin. E.g. Indonesian students with Hakka dialect background fuse *chī* (吃, to eat) and *shí* (食, to have food).

The classification demonstrates that the scope of confusable words is wider than near synonyms. Zhang pointed out that those words that have no similar semantic features, but cannot be classified as synonyms or near synonyms, were easily misused by learners. Thus, it can be helpful if such kinds of words can also be classified as confusable words.

In recent years, some textbooks have been designed to contain confusable words as teaching items to raise learners’ consciousness. However, confusable words can be caused by many factors, as mentioned, and the scope of confusable words is relatively wider. Learners can often spend only a few hours per week or per month in second language learning. How to maximize the effectiveness of the teaching materials is the key to success. If no scientific method is dedicated to determining which words should be

included in teaching, the selection might be arbitrarily decided by the teachers or textbook editors. The bias of asymmetry selection might happen. It may be counterproductive for L2 learning and the words that are not confusing in actual use might become problematic as a result.

As for the selection of confusable words, some studies have adopted the quantitative method. For instance, Zhang (2013) reckoned that one of the purposes of discriminating confusable words was to correct the learners' habitual misusing. For this reason, the selection of confusable words should be "misuse-driven". That is, the words with high error rate should be the confusable words to be discriminated. So, she proposed that the degree of confusion of the words should be determined by two indices. One was the frequency that the words were confused with others in the L2 corpus; and the other one was the ratio of the times that the words were correctly used to the times the words were used. Confusable words were ranked and given with scores based on these two indices respectively. The sum of these two scores represented the degree of confusion. She deemed that the words with a higher score should be included in the dictionary of confusable words.

Moreover, Xiao (2008) collected corpus data of Indonesian learners and conducted sample analysis according to their language background (e.g., whether they had ever studied in Chinese schools). The words which appeared at least 3 times and were misused for more than 2 times were defined as "quasi-confusable words". She then examined the "quasi-confusable words" in the whole corpus. If the words were misused for more than 15% in frequency, the words were considered as confusable words. By the quantitative method, 346 sets of confusable words for Indonesian learners were selected.

Conducting a contrastive analysis, Soontornthamnit (2018) made a prediction on confusable words based on the vocabulary in *New HSK Syllabus Level 6*. Through the results of the language tests and questionnaires for the Thai-native learners that surveyed whether they found the certain set of confusable words difficult or not, 96 sets of words with high degree of confusion were selected for "the table of Chinese confusable words for Thai-native speakers".

The previous studies mentioned above selected confusable words by quantitative methods. They analyzed the learners' corpus data, conducting tests, questionnaires, etc. The results suggest the word lists or dictionaries of confusable words. However, the time spent in learning is limited, the teaching materials may not be able to include an exhaustive list of all confusable words. Therefore, the effectiveness of the confusable-word list is more important than the comprehensiveness. The confusable words included in textbooks should be the words that the learners would misuse in the real situations. The frequency of misusing should also be taken into consideration.

3 The Research Method and Analysis

The present study examined the confusable words listed in *A Course in Contemporary Chinese* (Teng 2018), which is a series of textbooks being often used in Chinese language training centers in Taiwan. In addition, the fifth and sixth volumes of this series are currently the only two teaching materials that include the section for confusable words (*yìhǔnyáocí*, 易混淆词) in Taiwan. In such sections, the differences between each set of

words are interpreted with illustrative sentences and supplemented by word meanings, usages, collocations, etc. Besides, exercises are provided to help learners to be familiar with the main usages of the words. In total, there are 134 sets of confusable words collected in the two volumes³. Our study has examined the actual use of those words according to the L2 corpus, HSK Dynamic Composition Corpus 2.0. Based on the corpus, we aim to find out whether the word selection is appropriate.

For example, “*táotuō* (逃脱, to outrun), and *táolí* (逃离, to escape)” are listed as a set of confusable words in the textbook. In the L2 corpus, there are 66 tokens of *táolí* (逃离, to escape), among which 6 tokens are the cases of misuse. Yet, none of the six tokens include *táotuō* (逃脱, to outrun). Instead, *jiějué* (解决, to solve), *líkāi* (离开, to leave), *táobì* (逃避, to avoid), etc., are used. In other words, even though “*táotuō* (逃脱, to outrun), and *táolí* (逃离, to escape)” are listed as confusable words in the textbook, they are not confusing with each other in actual use. Similarly, “*chéngnuò* (承诺, commitment), *dāying* (答应, promise)” and “*fúchí* (扶持, to help to sustain), *zhīchí* (支持, to support)” were two sets of confusable words that are not really confusing with each other.

On one hand, based on these findings, we have found that the criterion for picking confusable words are mostly based on the judgement of the textbook editors. The asymmetric bias is obvious. On the other hand, we have found that some sets of confusable words listed in the textbooks indeed caused great confusion to the learners. Take “*dàodá* (到达, to arrive), *dádào* (达到, to reach)” for an example, of which the data of the usage is listed in Table 1 below. The total amount of tokens of *dàodá* (到达, to arrive) in the corpus is 52, of which 7 are the cases of misuse. The misuse rate is 13.5%. There are 6 other instances in which the learners used other incorrect words to substitute the *dàodá* (到达, to arrive).

As for *dádào* (达到, to reach), its frequency in the corpus is much higher than the frequency of *dàodá* (到达, to reach). 540 tokens of it are identified in the corpus. Among them, 38 tokens are found to be the cases of misuse, with a misuse rate of 7%; and 39 tokens are found to the cases of substitution with incorrect words by the learners.

In Table 1, the phenomenon of the learner often misusing the original word to express meanings of certain sets of words can be found. Furthermore, most of the time, the misuse situations are diverse in the sense that the substituting words are rather various. For instance, we found some cases in the corpus misusing *dádào* (达到) as *jiějué* (解决, to solve), *dádào* (达到) as *duōdá* (多达, much as), *dádào* (达到) as *huòdé* (获得, to gain) etc., each appeared only once. Since such unsystematic errors are not the focus of our study, we examine the systematic ones.

Back to the issue we have discussed: whether “*dàodá* (到达), *dádào* (达到)” listed in textbook are really confusable. We try to set a standard to solve this problem. As long as the misuse rate exceeds 10%, words are considered as confusable ones. For example, among the 7 error-usage in *dàodá* (到达), 4 of them should be replaced by *dádào* (达到). So, the misuse rate is 57% (4/7). Then, the pair “*dàodá* (到达), *dádào* (达到)” meets the criterion and is a set of confusable words. Besides, among 38 misuse cases of *dádào*

³ There are 141 sets in total, with 7 sets are repeated which are not counted as a result.

Table 1. The usage of *dàodá* (到达, to arrive) and *dádào* (达到, to reach) in the L2 corpus

<i>dàodá</i> 到达, to arrive (52 tokens)	
Situation A (incorrect uses)	4 times of misusing <i>dàodá</i> (到达) as <i>dádào</i> (达到) 3 times of misusing <i>dàodá</i> (到达) as the other words, once each
Situation B (using incorrect words as substitutes)	2 times of misusing <i>dádào</i> (达到) as <i>dàodá</i> (到达) 2 times of misusing <i>dào</i> (到) as <i>dàodá</i> (到达) 2 times of misusing the other words as <i>dàodá</i> (到达), once each
<i>dádào</i> 达到, to reach (540 tokens)	
Situation A (incorrect uses)	5 times of misusing <i>dádào</i> (达到) as <i>qǔdé</i> (取得, to get) 4 times of misusing <i>dádào</i> (达到) as <i>shíxiàn</i> (实现, to achieve) 3 times of misusing <i>dádào</i> (达到) as <i>dédào</i> (得到, to obtain) 5 times of misusing <i>dádào</i> (达到) as <i>dàodá</i> (到达) 5 times of misusing <i>dádào</i> (达到) as <i>dào</i> (到, to arrive) 22 times of misusing <i>dádào</i> (达到) as other words, once each
Situation B (using incorrect words as substitutes)	14 times of misusing <i>dédào</i> (得到) as <i>dádào</i> (达到) 4 times of misusing <i>dàodá</i> (到达) as <i>dádào</i> (达到) 3 times of misusing <i>dáchéng</i> (达成, to achieve) as <i>dádào</i> (达到) 2 times of misusing <i>chéngjiù</i> (成就, achievement) as <i>dádào</i> (达到) 16 times of misusing other words as <i>dádào</i> (达到), once each

(达到), only 2 should be replaced by *dàodá* (到达). The misuse rate does not exceed 10%, as a result, and this situation is not considered as a notable misused one⁴.

So far, we have examined 134 sets of confusable words in the textbooks in the method described above. Only 50 sets have met the criterion while the learners do not “obviously” misuse the words in other 84 cases. Appendix 1 displays more details of it.

4 Results and Discussion

Our research findings are summarized in the following, categorized into four aspects:

⁴ A set of confusable words could be divided into one-way misusing and two-way misusing, according to the directionality of misusing. The one-way misusing refers to the situations that only A is often misused as B, but B is not often misused as A. The two-way misusing refers to the situations that both A and B are often misused as each other.

A. 50 sets of confusable words listed in the textbooks have met our criterion. However, not all of them are the most typical sets.

As mentioned in session 3, we have found only 6 times that “*dàodá* (到达, to reach) and *dádào* (达到, to arrive)” were misused, whereas “*dádào* (达到, to arrive) and *dédào* (得到, to obtain)” were misused 14 times. The set of “*dádào* (达到, to arrive) and *dédào* (得到, to obtain)” is therefore more typical. Furthermore, the misuse rate of “*sīkǎo* (思考, thought), *kǎolù* (考虑, consideration)” met our criterion, while only 9 cases of misuse are found. However, as many as 22 cases of misuse of “*sīxiǎng* (思想, thought) and *sīkǎo* (思考, thought)” are found. For effectiveness, the set of “*sīxiǎng* (思想, thought) and *sīkǎo* (思考, thought)” is better than the set of “*sīkǎo* (思考, thought) and *kǎolù* (考虑, consideration)”. Therefore, in addition to investigating the practical usage of confusable words in the corpus, we list out the words that were more likely to confuse the learners based on the quantitative analysis as well. 15 new sets of words are recommended: “*biǎodá* (表达, to express), *biǎoxiàn* (表现, to display)”; “*dádào* (达到, to reach), *dédào* (得到, to get)”; “*guīzé* (规则, rule), *guīdìng* (规定, stipulation), *guīlǜ* (规律, regular pattern)”; “*yìnxiàng* (印象, impression), *yǐngxiǎng* (影响, impact)”; “*sīxiǎng* (思想, thought), *sīkǎo* (思考, thought)”; “*shíhòu* (时候, times), *shíjiān* (时间, time), *shí* (时, when)”; “*jiāowǎng* (交往, to associate), *jiāoliú* (交流, to contact)”; “*quèshí* (确实, indeed), *shízài* (实在, certainly)”; “*quēdiǎn* (缺点, drawback), *huàichù* (坏处, disadvantage)”; “*cānguān* (参观, sight-seeing), *fǎngwèn* (访问, visit)”; “*yǒu* (有, there be), *jùyǒu* (具有, have), *yǒngyǒu* (拥有, possess)”; “*tǎolùn* (讨论, to discuss), *yìlùn* (议论, to speak at)”; “*jiàoyù* (教育, to educate), *jiàoxùn* (教训, to teach someone a lesson)”; “*fāshēng* (发生, to happen), *fāxiàn* (发现, to discover)”; “*gòngtóng* (共同, in common), *hùxiāng* (互相, mutual)”.

B. 84 sets that do not meet our criterion. However, we found other candidates.

Take “*yǒuguān* (有关, have to do with) and *guānyú* (关于, about)” for examples. There are 714 tokens and 712 tokens respectively in the L2 corpus, without any case being misused as the other. On the contrary, the misusing of “*guānyú* (关于, about), *duìyú* (对于, for)” is serious. 712 and 861 tokens are found for “*guānyú* (关于, about) and *duìyú* (对于, for)” respectively in the corpus, among which 56 tokens and 64 tokens respectively are found misused. 13 examples of “*guānyú* (关于, about) being misused as *duìyú* (对于, for)” and the misuse rate was as high as 53.6%. 18 cases are found to be the misuse of “*duìyú* (对于, for) as *guānyú* (关于, about)”, and the misuse rate is 42.9%. The examples are listed as (5)-(6) below. It is obvious that “*guānyú* (关于, about), *duìyú* (对于, for)” have made a greater confusion to learners than “*yǒuguān* (有关, have to do with) and *guānyú* (关于, about)”. Clearer explanations on their subtle differences are more in need for L2 learners. In column B of Appendix 1, we list out the additional sets of words of which the misuse rate exceeds 10%, for reference.

- (5) 从 这 来 看， 父母 *关于
*Cóng zhè lái kàn, fùmǔ *guānyú*
 From this come see parent about
 孩子的 未来 发展的
háizi-de wèilái fāzhǎn-de
 children-PARTICLE future development-PARTICLE
 责任 是 很 明显的。
zérèn shì hěn míngxiǎn-de.
 responsibility be very obvious-PARTICLE
 “From this point of view, the parents’ responsibility *concerning their children’s future development is obvious.”
- (6) 我 以前 看过 *对于 “安乐死”的
*Wǒ yǐqián kàn-guò *duìyú “ānlèsǐ”-de*
 I before see-EXPERIENTIAL for euthanasia-PARTICLE
 电视剧。
diànshìjù.
 TV-show
 “I have seen TV shows *for euthanasia before.”

C. Instead of the misuse rate, the error rate of words is taken into consideration for the selection of sets of confusable words.

As described in Sect. 3, we have suggested that high-frequently misused words should be given priority in language teaching. However, we have also found that in some sets of confusable words, the frequency of each in the set may differ tremendously. Take “*gòngtóng* (共同, in common) and *gòngtōng* (共通, applicable to both/all)” as an example. 515 tokens of *gòngtóng* (共同, in common) are found in the corpus, but only 9 tokens of *gòngtōng* (共通, applicable to both/all) are found. The error rate of *gòngtōng* (共通, applicable to both/all) exceeds 10%, but the misuse rate of *gòngtóng* (共同, in common) is only 2%. It seems the learners have acquired the word *gòngtóng* (共同, in common). Nevertheless, of the 5 misuses of *gòngtōng* (共通, applicable to both/all), 4 should be replaced with *gòngtóng* (共同, in common). The misuse rate is actually very high. Although it seems to be a one-way misuse, the similarities on the form and sound still make them candidates of confusable words. While the error rate of words has not been set into the criterion for the selection, more further studies are needed.

D. According to our statistical data and results (refer to Appendix 1), the set of confusable words are mostly those with the same morpheme.

These confusable words are not influenced by L1 transferring. As most of the impacts of L1 transfer usually occurs in the early stages of learning and the writers of the compositions in the HSK (L2) corpus are advanced learners, the result is not too surprising. Our finding is actually consistent to the L2 language acquisition theory.

5 Conclusion

This study examines the confusable words used in textbooks through the learners’ authentic usages in an interlanguage corpus. We have found that some sets of confusable words are misused by learners, but only 50 of the 134 sets got misuse rate higher than 10%,

which is the criterion we set for confusable words. The confusable words selected by teachers or textbook editors might only be judged by their subjective experiences. The prediction is not really successful as the accuracy is less than 50%.

We are still working on the improvements of this study. As mentioned above, the directionality of misuse can definitely affect the teaching strategies. How confusable words should be treated in L2 Chinese teaching also requires further discussion. For example, should confusable words be taught or just listed as supplementary teaching materials? If confusable words should be included in teaching, how should they be taught? These questions still need further studying. After all, words that in similar forms without semantic relations in Mandarin Chinese are numerous. The misuse of words is inevitable in communication and writing.

The present study has provided a preliminary analysis and examination on the criterion for identifying confusable words used in textbooks, despite the unsolved problems. We have done the quantitative analysis. The qualitative analysis is still in progress.

Acknowledgments. The work was supported by the Ministry of Science and Technology, grant no. MOST 109-2410-H-002-189.

Appendix 1: Empirical Results of Confusable Words in Textbooks**A. Sets with misuse (50 sets)**

表达	‘to express’, 表示 ‘to show’
到达	‘to arrive’, 达到 ‘to reach’
具有	‘to have’, 具备 ‘to be equipped with’
权利	‘right’, 权力 ‘authority’
缺陷	‘defect’, 缺点 ‘drawback’
满足	‘to satisfy’, 满意 ‘satisfied’
现实	‘reality’, 实际 ‘practical’
发明	‘to invent’, 发现 ‘to discover’
相关	‘be related to’, 有关 ‘have to do with’
命运	‘fate’, 运气 ‘luck’
具有	‘to have’, 拥有 ‘to possess’
目标	‘goal’, 目的 ‘purpose’
真诚	‘sincerity’, 忠诚 ‘loyalty’
共通	‘applicable to both/ all’, 共同 ‘in common’
达成	‘to achieve’, 达到 ‘to reach’
面前	‘before the eyes of’, 前面 ‘in front of’
采访	‘interview’, 访问 ‘visit/ interview’
辽阔	‘vast’, 开阔 ‘broad’
交流	‘to contact’, 交换 ‘to exchange’
的确	‘certainly’, 确实 ‘indeed’
出现	‘to appear’, 发现 ‘to discover’
激发	‘to motivate’, 刺激 ‘to stimulate’
成立	‘to set up’, 建立 ‘to establish’
尽情	‘to one’s heart content’, 尽量 ‘as far as possible’
既然	‘since’, 竟然 ‘unexpectedly’
培养	‘to cultivate’, 培训, ‘to train’
兴趣	‘interest’ 有趣 ‘interesting’
谈论	‘to talk about’, 讨论 ‘to discuss’
证据	‘evidence’, 证明 ‘proof’
领悟	‘to comprehend/ to grasp’, 醒悟 ‘to come to realize’
温和	‘gentle’, 温柔 ‘tender’
思考	‘to think’, 考虑 ‘to consider’
意义	‘significance’, 意思 ‘meaning’
发展	‘to develop’, 发达 ‘developed’
平静	‘unruffled’, 安静 ‘quiet’
印象	‘impression’, 形象 ‘image’
接近	‘to approach’, 拉近 ‘to bring sth./ sb. closer’
回答	‘to answer/ respond’, 答案 ‘answer’

时刻 ‘moment’, 时候 ‘times’
 个人 ‘individual’, 各人 ‘each one’
 本性 ‘nature’, 本能 ‘instinct’
 教养 ‘upbringing’, 教育 ‘education’
 适合 ‘suitable for’, 适当 ‘appropriate’, 合适 ‘suitable’
 推广 ‘to popularize’, 推行 ‘to promote’, 执行 ‘to carry out’
 规则 ‘rule’, 规定 ‘stipulation’, 规范 ‘norm’
 产业 ‘industry’, 行业 ‘business’, 职业 ‘profession’
 通常 ‘usually’, 经常 ‘often’, 平常 ‘normally’
 念头 ‘idea’, 观念 ‘notion’, 概念 ‘concept’
 后果 ‘consequence’, 结果 ‘result’, 效果 ‘effect’
 渴望 ‘desire’, 希望 ‘hope’, 愿望 ‘wish’, 心愿 ‘aspiration’

B. Sets without significant misuse (84 sets)

(Words in brackets are additional sets suggested by this study)

缺乏 ‘be short of’, 缺少 ‘lack of’ (减少 ‘to reduce’, 缺少 ‘lack of’)
 明白 ‘to understand’, 清楚 ‘clear’ (知道 ‘to know’, 明白 ‘to understand’)
 理念 ‘concept’, 理想 ‘ideal’ (希望 ‘hope’, 理想 ‘ideal’)
 承受 ‘to bear with’, 忍受 ‘to endure’ (承受 ‘to bear with’, 受 ‘to receive’)
 此外 ‘in addition’, 另外 ‘besides’ (另外 ‘besides’, 还有 ‘also’)
 逃离 ‘to escape’, 脱离 ‘to get out of’ (脱离 ‘to get out of’, 离开 ‘to leave’)
 庇护 ‘to asylum’, 保护 ‘to protect’ (保护 ‘to protect’, 保持 ‘to maintain’)
 连结 ‘to connect’, 连络 ‘to contact’ (连络 ‘to contact’, 连系 ‘to communicate’)
 温馨 ‘warmth’, 温暖 ‘warm’ (温暖 ‘warm’, 温柔 ‘tender’)
 原本 ‘originally’, 原来 ‘it turns out that’ (本来 ‘at first’, 原来 ‘it turns out that’)
 信任 ‘to trust’, 相信 ‘to believe’ (信任 ‘to trust’, 信用 ‘credit’)
 有关 ‘have to do with’, 关于 ‘about’ (关于 ‘about’, 对于 ‘for’)
 引发 ‘to evoke’, 引起 ‘to lead to’ (引起 ‘to lead to’, 造成 ‘to bring about’)
 地点 ‘place’, 地方 ‘local’ (方面 ‘area’, 地方 ‘local’)
 热忱 ‘earnest’, 热心 ‘be enthusiastic about’
 (热情 ‘enthusiastic’, 热忱 ‘earnest’)
 阶段 ‘stage’, 阶层 ‘level’ (阶段 ‘stage’, 段落 ‘paragraph’)
 周遭 ‘around’, 周边 ‘surrounding’ (周围 ‘round’, 周边 ‘surrounding’)
 事业 ‘enterprise’, 职业 ‘profession’ (职业 ‘profession’, 行业 ‘business’)
 景象 ‘scene’, 现象 ‘phenomenon’ (现象 ‘phenomenon’, 影响 ‘impact’)
 远行 ‘go on long journey’, 旅行 ‘to travel’ (游行 ‘to parade’, 旅行 ‘to travel’)
 愕然 ‘stunned’, 已然 ‘be already so’, 偶然 ‘by accident’, 自然 ‘natural’
 (自然 ‘natural’, 当然 ‘certain’)
 风险 ‘risk’, 危险 ‘danger’, 冒险 ‘adventure’ (危害 ‘harm’, 危险 ‘danger’)
 身为 ‘in the capacity of’, 成为 ‘to become’
 (成为 ‘to become’, 当成 ‘to consider as’, 成 ‘have become’)
 应对 ‘to reply’, 应变 ‘to meet contingency’
 诠释 ‘to interpret’, 解释 ‘to explain’

创业	‘to start up business’, 创新 ‘to innovate’
灾害	‘disaster’, 灾难 ‘misfortune’
承诺	‘commitment’, 答应 ‘promise’
扶持	‘to help to sustain’, 支持 ‘to support’
安置	‘to place’, 设置 ‘to set up’
争相	‘to vie with each other’, 竞争 ‘to compete’
点出	‘to point out’, 指出 ‘to indicate’
制造	‘to manufacture’, 建造 ‘to build’
训练	‘to train’, 练习 ‘to practice’
根本	‘at all’, 原本 ‘at first’
品牌	‘brand’, 招牌 ‘sign(broad)’
洞察	‘to have insight into’, 观察 ‘to observe’
场所	‘place’, 场合 ‘occasion’
舞动	‘to dance/ to brandish’, 舞蹈 ‘dance’
特地	‘specially’, 特别 ‘special’
传承	‘to pass down through generations’, 传人 ‘successor’
预期	‘to expect’, 期待 ‘to look forward to’
传授	‘to pass on’, 传承 ‘to pass down through generations’
技巧	‘technique’, 技术 ‘technology’
力道	‘physical strength’, 力量 ‘power’
抱持	‘to hold’, 保持 ‘to maintain’
借着	‘by means of’, 借由 ‘through’
回覆	‘to reply’, 回答 ‘to answer/ respond’
紊乱	‘confusion’, 错乱 ‘confused’
焦急	‘vexed’, 焦虑 ‘anxious’
天分	‘natural gift’, 天才 ‘genius’
热衷	‘fall over oneself for’, 热心 ‘be enthusiastic about’
宣布	‘to announce’, 宣告 ‘to declare’
收益	‘profit’, 收入 ‘income’
名义	‘(in the) name (of)’, 意义 ‘significance’
盲点	‘blind spot’, 盲目 ‘blindly’
破绽	‘weak point’, 破洞 ‘hole’
开明	‘enlightened’, 开放 ‘open’
纪录	‘record/ documentary’, 记载 ‘to put down in writing’
展现	‘to show’, 表现 ‘to display’
宣称	‘to assert’, 宣布 ‘to declare’
庞大	‘huge’, 巨大 ‘giant’
预估	‘to estimate’, 预算 ‘budget’
亏本	‘to lose one’s capital’, 回本 ‘to earn its cost’
决选	‘final election’, 竞选 ‘to run for’
建设	‘to construct’, 兴建 ‘to build’
随着	‘in pace with’, 接着 ‘after that’

支撑	‘to prop up’, 支持	‘to support’	
维护	‘to maintain’, 保护	‘to protect’	
梦想	‘dream’, 理想	‘ideal’	
多重	‘multiple’, 多元	‘multiplex’	
景致	‘view’, 景观	‘landscape’	
家当	‘family belongings’, 家用	‘family expense’	
脱逃	‘to run away’, 脱离	‘to get out of’	
学术	‘academics’, 学问	‘learning’	
调控	‘to regulate’, 调整	‘to adjust’	
忧郁	‘depressed’, 焦虑	‘anxious’	
区域	‘region’, 领域	‘field’	
促使	‘to spur’, 促发	‘to trigger’	
野生	‘wild’, 野放	‘to free captive animals’, 野外	‘open country’
居留	‘residency’, 居住	‘to reside’, 居所	‘residence’
华裔	‘ethnic Chinese’, 华侨	‘overseas Chinese’, 华人	‘Chinese people’
行销	‘to market’, 销售	‘to sell’, 畅销	‘to sell well’
专精	‘specialized’, 专业	‘professional’, 专家	‘specialist’

References

- Carpenter E.: *Confusable Words*. HarperCollins (1993)
- Laufer, B.: The concept of ‘Synforms’ (similar lexical forms) in vocabulary acquisition. *Lang. Educ.* 2(2), 113–132 (1988)
- Soonthornthamnit, N.: *Tàiguó Xuéxízhě Hànyǔ Yì Hǔnyíáo Cí Yánjiū* (泰国学习者汉语易混淆词研究), Minzu University of China (2018)
- Teng, S.H. (ed.): *A Course in Contemporary Chinese Volume 5, 6*, Taipei. Linking Publishing, Taiwan (2018)
- Tsai, M.C., Huang, C.R., Chen, K.J.: From near-synonyms to the interaction between syntax and semantics. In: Yin, Y.M., Yang, I.L., Chan, H.C. (eds.): *Proceedings of the Fifth International Symposium on Chinese Languages and Linguistics*. Academia Sinica, Taipei, pp. 439–459 (1999)
- Xiao, R., McEnery, T.: Collocation, semantic prosody, and near synonymy: a cross-linguistic perspective. *Appl. Linguist.* 27(1), 103–129 (2006)
- Xiao, P.: *The Research on Confusable Words of Indonesian Students in Chinese Interlanguage*, Beijing Language and Culture University (2008)
- Zhang, B.: Synonymy, Near-Synonymy and Confusable Word: A Perspective Transformation from Chinese to Interlanguage. *Chin. Teach. World* (3), 98–107 (2007)
- Zhang, B.: Pertinence: a fundamental principle for compiling a dictionary of confused words. *Chin. Teach. World* 27(2), 214–231 (2013)