

Error Analysis and Teaching Suggestions for Corpus-Based Teaching of Disyllabic Homophones in Teaching Chinese to Speakers of Other Languages

Moshuhan Li¹ & Xiangdong Liang^{1*}

¹ College of Humanities and Education, Liaodong University, Dandong, Liaoning, China

*Correspondence: Xiangdong Liang, College of Humanities and Education, Liaodong University, Dandong, 118003, Liaoning, China. E-mail: 1773563426@qq.com

Received: August 22, 25 Accepted: September 26, 2025 Online Published: September 29, 2025

This research was supported by the 2024 Doctoral Research Start-up Fund Project of Liaodong University, titled "Research on the Enhancement Path of Chinese Language Learning Engagement of International Students in the Classroom Environment" (2024BS024).

This research also was funded by the Dandong Philosophy and Social Sciences Project in 2025, titled "Research on the Practical Path of Digital Empowerment for the Green Transformation of Chinese Language Teaching and International Cooperation in Liaodong Region" (2025DDSK142).

Abstract

Disyllabic homophones are the key and difficult points in the teaching of Chinese as a foreign language. Based on the theory of second language acquisition and interlanguage theory, this paper systematically combs the types of errors in the use of Chinese disyllabic homophones by foreign students based on the analysis of HSK dynamic composition corpus and foreign students' daily homework, oral recording and other corpora. The study found that the errors are mainly manifested in the confused use of homonyms, the wrong use of homonyms and the misuse of homonyms caused by the negative transfer of mother tongue. In-depth analysis shows that these errors are closely related to the complexity of Chinese homonyms, the mother tongue background of foreign students, improper learning strategies, and the lack of teaching materials and teaching methods. Based on this, this paper puts forward targeted teaching countermeasures, in order to provide reference for vocabulary teaching of Chinese as a foreign language.

Keywords: teaching Chinese as a foreign language, double syllable homonyms, error analysis, teaching suggestions

1. Introduction

Words are the carrier of the combination of sound and meaning. The number of language syllables in the world is limited, and the meaning that needs words to express is infinite. This contradiction of less sound and more meaning is common in language phenomena. The vocabulary of modern Chinese is vast, but all the words are only condensed in 414 syllables or 1333 tonal combinations.^[1] There are 45200 words in 'Chinese Pinyin Vocabulary', including 2342 groups of homonyms, including 5249 words, accounting for 11.6 % of the total number of words, and the proportion is still large.

As a language phenomenon, homophones in modern Chinese have its positive side in language use. It can express as many lexical meanings as possible with fewer syllables. This is an economic principle of vocabulary development^[2]. The existence of homophones can make the language expression implicit, euphemistic and vivid, and provide the possibility for rhetorical devices such as homophonic puns and poetry rhyme. The use of homophonic phenomena to form homophonic puns can express the speaker's wit or profound implicit meaning. On the other hand, the same phonetic form and different meanings are easy to cause confusion and interfere with verbal communication, which is the negative impact of homophony.

Yin Wengang's (2003) research shows that the homophony rate of Chinese characters (the ratio of homophone syllables in the total syllables) is 80.49 %, and the average homophony degree of Chinese characters (the

homophony degree of homophone syllables) is 7.85. The calculation method of the error rate is ‘error rate = frequency of errors in the corpus / frequency of use of the word’.^[3]

The study of homonyms has positive significance for the healthy development of language and the reduction of language barriers for Chinese learners. Corder (1967) put forward the error analysis program, that is, collecting data, identifying errors, describing errors, explaining errors, evaluating errors and so on.^[4] Through the analysis of the specific error corpus and the Chinese level of the foreign students who have errors, we can roughly understand the situation of students’ acquisition and teachers’ teaching of disyllabic homophones. Combined with the difference between learners’ mother tongue and Chinese, this paper analyzes the main causes of errors and the types and grammar points with the most errors, and provides corresponding teaching suggestions, which can help the teaching activities of Chinese as a foreign language classroom.

In-depth study of the errors in the use of disyllabic homonyms by foreign students will help to reveal the rules and characteristics of foreign students’ Chinese vocabulary acquisition and enrich the application of second language acquisition theory in Chinese teaching. At the same time, the teaching strategies based on error analysis can provide specific teaching guidance for teachers of Chinese as a foreign language, help foreign students better grasp and use homophones, and improve their Chinese level.

This paper reviews the definition standards of Chinese homonyms in academic circles, such as the scope of the same pronunciation (sound, rhyme and tone are exactly the same), the meaning relationship (meaning is completely different or related), etc., and introduces common classification methods, such as homonyms, homonyms, etc. Using the real two-syllable homonym bias cases in the ‘HSK Dynamic Composition Corpus’ and the ‘Northern BBC Corpus’, the errors generated by Chinese learners in the learning process are sorted out and classified, and typical and regular errors are summarized. Types, and in-depth analysis to find out the causes of errors. This will help Chinese learners effectively avoid these errors and correctly grasp the usage and function of disyllabic homophones, so as to be able to use them properly in communication.

2. The Reasons for the Formation of Homonyms

2.1 Causes Of Pronunciation

2.1.1 The Coupling of Word Sounds

There is no logical relationship between pronunciation and vocabulary. In the process of language development, because people in different times and regions are creating new words through existing language materials, phonetic coupling cannot be avoided^[5]. There is almost no language in the world that does not contain homonyms. In order to bring convenience to memory and communication, human beings use limited phonetic forms to express infinite meanings. However, it is theoretically feasible to use only one phonetic form to express multiple meanings, but it is difficult to do in reality, which leads to the inevitable existence of many homonyms in our language. The number of syllables in modern Chinese is 419 without tone, and 1342 with tone. We can think that when people use these limited phonetic forms to represent a large number of vocabulary, it also leads to the existence of homonym in modern Chinese.

2.1.2 The Evolution of Word Sounds

Such as ‘青’ and ‘轻’. The original initials are different. According to ‘Guangyun’, ‘青’ is ‘仓经切’ in ancient times, and ‘轻’ is ‘去盈切’. But in Mandarin, the initials are the same, all of which are ‘q’. In this way, the number of homonyms in Chinese will be greatly increased, and some words with different sounds in ancient times will become homonyms in modern times^[6].

2.2 Lexical Reasons

2.2.1 The commonness of Chinese characters

The frequency of Chinese characters has a great influence on the memory of Chinese characters for foreign students. When learning Chinese characters on the basis of not having the background of Chinese characters, the characters that make them remember the most deeply must be those that are often seen and often used.^[7] If there are two homonyms, one of which appears much more frequently than the other, it is very likely that foreign students will only remember the homonym they often see and ignore the other.

2.2.2 The complexity of Chinese characters

A group of words with the same phonetic form, the word with a small number of strokes is easier to acquire, and the words with complex components and structures are prone to misuse^[8]

2.3 *Semantic Reason*

2.3.1 Semantic Change

The development of social life and the development of language itself and human thinking often make the words in the language produce new meanings. With the evolution of time, when the new meanings of these words slowly separate from the original meaning of the words and exist independently, it is difficult for people to perceive the relationship between each other. Therefore, it evolved into a word that is homonymous but independent from the original word, thus forming a homonym. There is a strong or weak connection between each meaning of a word with multiple meanings, but with the development of the meaning of the word, the connection between the meanings is lost. Under certain conditions, one or more of these meanings will be separated from the original word and become an independent new word. Then the new word will become a group of homophones together with the original word, which will increase the difficulty of foreign students' learning.

2.3.2 Similarity of Meaning

The similarity of word meaning refers to a group of homophones with one or more identical or similar meanings [10]. The same phonetic form of word has brought great difficulties for foreign students to learn Chinese. In addition, the meaning of the word also has a certain similarity, so the probability of misuse of this group of homophones will be very high. Even if we ourselves, it is easy to make mistakes. For foreign students, it is more difficult.

2.3.3 Transparency of Word Meaning

The transparency of word meaning refers to whether we can see the whole meaning of a word according to its partial meaning. Words with high semantic transparency are often widely used because of their word-building elements. Therefore, it is easy to see the meaning of the whole word through the combination of the meanings of its word-building elements. [11]. Words with high semantic transparency are easy to be mastered by learners, and it is also of great help for them to learn a language. On the contrary, words with low semantic transparency, because their word-formation elements are often relatively unfamiliar and difficult to understand, it is difficult to infer the overall meaning of words through the combination of their word-formation elements. When foreign students use some monosyllabic homonyms, they sometimes misuse them because of their low semantic transparency.

3. Analysis of the Causes of Errors in the Use of Chinese Disyllabic Homophones by Overseas Students

3.1 *The Complexity of Chinese Homophones*

The confused use of homonyms is the most common type of error [12]. For example, '公式 (gōngshì)' is mistaken for '攻势 (gōngshì)', and '权利 (quánli)' is mistaken for '权力(quánli)'. Foreign students often do not have a thorough understanding of the meaning and usage of words, and they are randomly replaced only according to the same pronunciation.

The number of Chinese homonyms is large, the meaning is different, and the usage of some words is more flexible, which brings great difficulties to the study of foreign students. For example, '工夫' and '功夫' not only have the same pronunciation, but also have a certain correlation in meaning (both related to time and ability), which is more likely to confuse foreign students.

This kind of bias is mainly due to the incomplete mastery of polysemous words by foreign students. For example, '花 (huā)' can mean both 'flower ' and 'cost'. International students may say the correct sentence 'I spent a lot of time looking at flowers in the park', but there may also be errors such as 'he spent a very beautiful flower', confusing the different meanings and usages of 'flower'.

3.2 *Negative Transfer of the Native Language of International Students*

Foreign students with different mother tongue backgrounds are affected by the negative transfer of mother tongue when learning Chinese homonyms [13]. For example, foreign students whose mother tongue is English may confuse some words with similar pronunciation but different meanings in Chinese, because there are relatively few homonyms in English, and they lack experience in dealing with homonyms. When a word in the mother tongue of foreign students is different from the homonym in Chinese in meaning or usage, it is easy to be interfered by the mother tongue and cause errors. For example, 'see' in English can mean both '看见' and '明白', while '看见 (kànjiàn)' and '看懂 (kàndòng)' in Chinese are two different words. Some foreign students may say 'I see what you mean', which is affected by the negative transfer of their mother tongue.

3.3 Improper Learning Strategies

When studying homophones, foreign students often adopt the method of rote memorization, lacking in-depth understanding of the meaning and usage of words. At the same time, when they encounter uncertain homonyms, they are easy to guess based on existing experience, resulting in errors [14].

3.4 Inadequate Teaching Materials and Teaching Methods

At present, some textbooks for teaching Chinese as a foreign language are not systematic and in-depth in explaining homonyms[15]. They often simply list the meanings and examples of words, and lack early warning and targeted exercises for errors. In terms of teaching methods, most teachers adopt the traditional teaching method, which lacks interactivity and interest, and it is difficult to stimulate the interest and enthusiasm of foreign students.

4. Corpus-Based Teaching Model of Disyllabic Homophones in Chinese as a Foreign Language

In view of the existing teaching situation, we propose a model that can use corpus to teach two-syllable homophones in Chinese as a foreign language. The two-way classroom interaction between teachers and students discusses the teaching content, explores the corpus needed for the teaching content in the corpus, selects and analyzes the corpus, creates the micro text needed for classroom teaching, and finally uses the corpus.

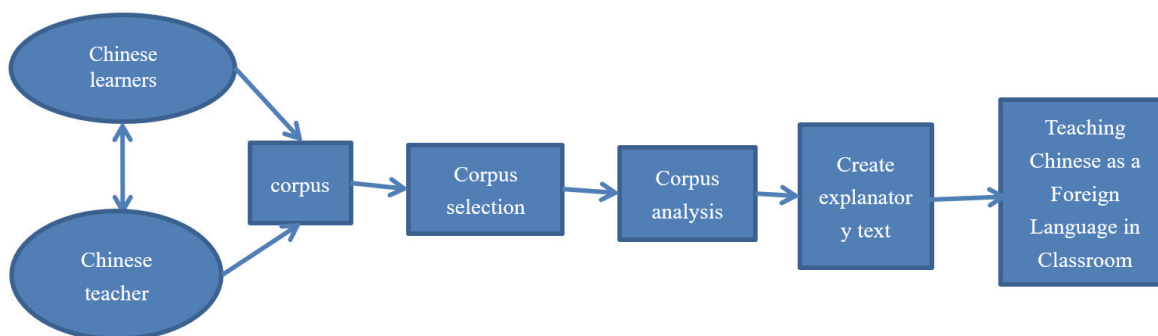


Figure 1. The process of teaching disyllabic homophones based on corpus

4.1 Corpus Selection

We have counted all kinds of corpus in the field of teaching Chinese as a foreign language (Table 1), and analyzed the current situation of the use of all kinds of corpus. In order to statistically analyze the errors of Chinese learners in the acquisition of two-syllable homophones, 71 groups of language materials related to two-syllable homophones in the 'Chinese Proficiency Vocabulary and Chinese Character Grade Outline' were extracted from the HSK Dynamic Composition Corpus and the Northern Chinese BBC Corpus for statistical analysis, and sorted according to the frequency of use [16]. Statistics and analysis of language materials that appear in the process of learners using two-syllable homophones (Table 2 and Table 3).

4.2 Teaching Model

4.2.1 Teachers and Students Retrieve and Filter The Corpus

Based on the corpus, teachers and students analyze and summarize the actual errors and related factors of learners' Chinese interlanguage from the aspects of learners' nationality, mother tongue, age and Chinese level with the idea of data search and exploration.

4.2.2 Create Teaching Micro-Texts

In the process of creating a personal corpus, teachers can use the corpus to determine the content, focus and order of vocabulary teaching, and avoid or eliminate some outdated expression methods. In classroom teaching practice, the personal corpus created can be used to provide the real context of vocabulary use, carry out homonym discrimination, provide learners with collocation information of words, and so on.

Through two search tools, HSK dynamic composition corpus and BBC corpus, the text is selected, and the knowledge points of two-syllable homophones to be taught are concentrated and focused from its huge information

database. These corpuses need to meet the learning level and teaching needs of Chinese learners. Teachers also need to reprocess these extracted corpuses to ensure the effectiveness of corpus in teaching.

4.2.3 Comparative Analysis of Teacher-Student Corpora

Step 1: Retrieve all eligible collocates of a certain word from the corpus.

Step 2: the collocation word according to the word frequency statistics, summed up its type, and describe its collocation characteristics, and then according to the level of word frequency to give relevant examples.

Chinese teachers retrieve all the collocation words of a certain word through corpus index and teaching needs, then summarize the collocation features, establish the colligation between words, and find relevant reference books for verification. On the one hand, the purpose of this move is to see whether the reference book lists the collocation rules of the word in detail. On the other hand, we can also find the collocation rules of mother tongue. The generalization of the rules is mainly based on the collocation mode of notional words and notional words, especially the collocation mode of verbs, adjectives, and some nouns, adverbs and quantifiers.

Step 3: Let learners independently summarize the collocation rules of words according to their own Chinese language ability (this process can be completed before class).

Step 4: Explain the word, and then tell the students the summarized collocation features to help students establish the collocation mode of the words.

In such a teaching step, learners can have a deeper understanding of the two-syllable homophones they have learned, and teachers can explain them in a timely manner after students make mistakes, which is no longer as traditionally judged by experience or language.

4.2.4 Teachers and Students Independently Determine the Key and Difficult Points

'Writing' is an output link in the process of Chinese learning. It is the transition from language learning to language use. It is a process of reorganizing and selecting grammar and words that have been learned in the new context. It is also a comprehensive examination of the language knowledge that has been learned.

In the corpus-based two-syllable teaching classroom of Chinese as a foreign language, it is possible to achieve student-centered inquiry learning. Learners take the initiative to search for learning information and master the materials needed to acquire vocabulary. Chinese teachers are only assistants in answering questions.

4.2.5 Consolidation Exercises

After exhaustive information search, teachers and students show examples of two-syllable homophone phrases that need to be acquired in the acquisition of micro-text. The meaning of phrase is expressed by the meaning of the sentence. It is an effective way to acquire words by distinguishing two-syllable homophones through the context of sentence meaning. But at the same time, teachers should also let students know how to feel the differences of words in the face of real and vivid examples in the corpus.

There are some difficulties. Students can judge from the concept, color and scope of words, and need to let students master the necessary analysis methods. Teachers can print out the two-syllable homophones in the vocabulary outline to learners (Table 4), and the teaching process can also be displayed to Chinese learners in the form of courseware.

The important role of example sentences in the corpus is also reflected in the creation of context, which can stimulate students' learning motivation and autonomy^[17]. Students can independently create a sentence situation that discriminates two-syllable homophones. Learners can construct meaning according to what they are interested in, and then make a reasonable judgment on the relationship between homonyms.

4.2.6 Identify and Fill Gaps

According to Krashen's "Comprehensible Input Hypothesis"^[18] and Vygotsky's "Zone of Proximal Development" theory, the difficulty of corpus should be controlled within a reasonable range. Chinese teachers should be targeted in the selection of corpus, try to ensure that the content and difficulty of the corpus are compatible with the Chinese level, learning interest, learning needs and other aspects of the teaching object, and some special meanings embodied in the corpus can also be appropriately deleted.

4.3 Analysis of Teaching Cases

4.3.1 Example Word Selection

必须 (adv) /必需 (adj)

“必须”to indicate the necessity or necessity of a matter in terms of reason and reason, or to reinforce a

commanding tone, usually used as an adverb, often followed by a verb or verb phrase, such as: 学习必须刻苦钻研;

“必需” It means necessary and indispensable, equivalent to an adjective, usually used to modify nouns and act as attributive, such as: 煤和铁等是发展工业所必需的原料.

“需” Commonly used as a verb, it means need or demand. “须” It is an auxiliary verb, indicating ‘should’ and ‘must’.“必需”It is a verb that emphasizes the necessity and indispensability of something;“必须” It is an auxiliary verb that can only be used as an adverbial, indicating that it should meet certain conditions or requirements, such as:

A 牙膏、牙刷和毛巾是日常生活必需（一定要有）的东西.

B 我明天早上必须（应该要、一定要）八点钟起床.

4.3.2 Differentiate

These two are verbs and nouns respectively, both of which have the meaning of ‘response’. ‘Reflection’ can refer to telling the situation and opinions to the superior or the relevant part, and can also be compared to show the essence of objective things. ‘Response’ refers to the response caused by the event. The error rate of this group of two-syllable homonyms is 4.26 %. The main reason for the errors of foreign students is the confusion of word meaning.

5. Teaching Suggestions for Bisyllabic Homophones in Teaching Chinese as a Foreign Language

5.1 Optimize The Content of Teaching Materials

Textbook writers should strengthen the systematic arrangement and explanation of homonyms, and increase the content of error analysis and comparative exercises^[19]. For example, when introducing homonyms, we should not only list the meaning and usage of words, but also point out the common types of errors, and provide corresponding exercises to help foreign students deepen their understanding and memory.

5.2 Using Various Teaching Methods in Combination With Information Technology Tools for Teaching

In the teaching of homophones in Chinese as a foreign language, disyllabic homophones are difficult for international students to master due to their inherent characteristics and are prone to errors. The teaching of disyllabic homophones for international students should focus on pragmatics, semantics, and other aspects, and attach importance to the content of national research.

Traditional methods of teaching Chinese vocabulary to foreigners, such as translation method, collocation method, contextual teaching method, comparative analysis method, imagery method, direct teaching, chunk teaching method, mind mapping teaching method, task-based teaching method, etc., are well-known. However, in the new situation of the information age, vocabulary teaching still needs to be combined with information technology tools for teaching^[20].

When using the comparative teaching method, easily confused homophones are grouped together for comparative explanation. By analyzing their differences in meaning, usage, collocation, and other aspects, it helps international students distinguish and master them. For example, when explaining "公式" and "攻势", specific examples can be used to help international students understand that "公式" refer to "formulas that use mathematical symbols to represent the relationship between various quantities", while "攻势" refer to "the situation of attack".

When using situational teaching method, create real language situations for international students to learn and use homophones in specific contexts. For example, through activities such as role-playing and situational dialogue, international students can experience the different uses of homophones in practice and improve their language proficiency.

At the same time, utilizing multimedia resources such as images, audio, and video to enrich teaching content and forms. For example, by playing Chinese songs, movie clips, etc. that contain homophones, international students can learn homophones in a relaxed and enjoyable atmosphere, enhancing their learning effectiveness.

5.3 Developing Learning Strategies for International Students

Guide foreign students to adopt scientific learning methods, such as memorizing homophones through root affixes and semantic associations, and encourage them to think more, compare more and practice more in the learning process. At the same time, cultivate students’ autonomous learning ability, let them learn to use dictionaries, corpora and other tools to solve the problems encountered in learning^[21].

5.4 Strengthen Language Practice

Increase the language practice opportunities for foreign students, so that they can consolidate the homonyms they have learned in practical application. For example, organize Chinese corner, writing competition, speech competition and other activities to encourage students to actively participate in and improve their language expression ability and application ability.

6. Conclusion

Corpus-based teaching of two-syllable homophones in teaching Chinese as a foreign language is a useful supplement to traditional teaching Chinese as a foreign language. Based on the analysis of the errors in the use of Chinese two-syllable homonyms by foreign students, this paper finds that the errors are mainly manifested in the confused use of homonyms, the wrong use of homonyms and the misuse of homonyms caused by the negative transfer of mother tongue. These errors are the result of many factors, such as the complexity of Chinese homonyms, the mother tongue background of foreign students, the improper learning strategies and the lack of teaching materials and teaching methods. The source of corpus and the number of samples studied are limited, which may not fully cover all the errors in the use of homonyms by foreign students.

Future research can further expand the source of corpus and the scope of samples, and further explore the differences in homophone errors of foreign students with different levels and different mother tongue backgrounds. At the same time, we can conduct empirical research on the proposed teaching strategies, test their effectiveness, and provide stronger support for the teaching of homonyms in Chinese as a foreign language.

References

- [1] Yu, Y. (1992). *Analysis of homophones*. Beijing Language and Culture University Press.
- [2] Yin, W. (2003). The law of decreasing homophones, homophones, and the number of syllables in homophones as the degree of Chinese characters increases. *Language Science*, 2(4), 3.
- [3] Dong, F. (2010). *A comparative study of polysyllabic homophones and homographs in the fourth and fifth editions of the Modern Chinese Dictionary* [Doctoral dissertation, Shandong University].
- [4] Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching*, 5(4), 161–170. <https://doi.org/10.1515/iral.1967.5.1-4.161>
- [5] Hu, Y. (1981). *Modern Chinese* (3rd ed.). Shanghai Education Press.
- [6] Yao, Y. (2018). *Research on homophonic errors in "Developing Chinese"* [Doctoral dissertation, Central China Normal University].
- [7] Chen, X. (n.d.). *Research on disyllabic homophones in teaching Chinese as a foreign language* [Doctoral dissertation, Shandong Normal University].
- [8] Lin, X. (n.d.). *Research and teaching strategies on homophonic spelling errors based on HSK dynamic composition corpus* [Doctoral dissertation, Hunan Normal University].
- [9] Cao, W. (2002). *Manhua Hanzi*. Chengdu Tian Di Publishing House.
- [10] Sun, Y. (2020). The division and combination of modern Chinese disyllabic polysemous words and homophones. *Journal of Jiangsu Normal University: Philosophy and Social Sciences Edition*, 46(1), 17.
- [11] Wu, H. (2008). *Research on Chinese object words* [Doctoral dissertation, Soochow University].
- [12] Yu, Y. (n.d.). *Research on teaching strategies for bisyllabic homophonics in Boya Chinese* [Doctoral dissertation, Guangxi University for Nationalities].
- [13] Liu, R. (2021). *Research and design of homophonic words teaching in "Bridge: Practical Chinese Intermediate Course"* [Doctoral dissertation, Chongqing Normal University].
- [14] Peng, Y. (2015). *Research on homophones in the New HSK syllabus* [Doctoral dissertation, Anhui University].
- [15] Khuat, A. (n.d.). *Research on homophonic errors of international students* [Doctoral dissertation, Hebei University].
- [16] Office of the National Chinese Proficiency Test Committee Examination Center. (2001). *Revised Chinese vocabulary and character grading outline*. Economic Science Press.
- [17] Wang, Y. (2021). An analysis of the errors and strategies in acquiring homophones among Korean international students with intermediate Chinese proficiency. *Journal of Language Teaching and Research*, 12(3), 39–41.

- [18] Krashen, S. (1985). *The input hypothesis: Issues and implications*. Longman.
- [19] Fu, Y. (2012). *Corpus based analysis of discourse coherence errors among Korean international students and teaching suggestions for discourse writing* [Doctoral dissertation, Hebei Normal University].
- [20] Jiang, Y., & Liu, H. (2023). A comparative study of discourse organization patterns in Chinese second language writing based on corpus. *Chinese Language Learning*, (2), 83–92.
- [21] Liu, L. (2010). A preliminary exploration of corpus based teaching methods for Chinese synonyms. *Teaching and Management*, (15), 124–125.

Appendix

Table 1. Names of Chinese Interlanguage Corpus Construction and Universities

Pen language corpus		
Number	Name	University
1	HSK 动态作文语料库 (HSK Dynamic Essay Corpus)	Beijing Language and Culture University
2	首都外国留学生汉语文本语料库 (Chinese Text Corpus for Foreign Students in Capital)	Beijing Language and Culture University
3	留学生中介语语料库 (Student Interlanguage Corpus)	Sun Yat-sen University
4	外国学生汉语中介语偏误信息语料库 (Informational Corpus of Chinese Interlanguage Errors in Foreign Students)	Nanjing Normal University
5	韩国留学生汉语中介语语料库 (Korean Interlanguage Corpus for International Students in China)	Ludong University
6	外国留学生汉语笔语语料库 (Chinese Writing Corpus for Foreign Students)	Beijing Chinese Institute
7	外国人汉语习得动态作文语料库 (Dynamic Corpus of Chinese Acquisition by Foreigners: A Case Study of Composition,)	Shanghai Jiao Tong University
8	TOCFL 学习者语料库 (Shanghai Jiao Tong University)	National Taiwan Normal University

Oral Corpus		
Number	Name	University
1	汉语学习者口语语料库 (Chinese Learners' Oral Corpus)	Beijing Language and Culture University
2	小型外国学生口语中介语语料库 (Small foreign students' oral interlanguage corpus)	Soochow University
3	语言习得汉语口语语料库 (Language Acquisition Chinese Oral Corpus)	The Chinese University of Hong Kong
4	根据电话口语考试建设的语料库 (Corpus built on telephone speaking test)	Peking University
5	汉语中介语口语语料库 (Chinese Interlanguage Oral Corpus)	Nanjing University

Chinese Character Error Corpus		
Number	Name	University
1	华语文学习者汉字偏误数据资料库 (Chinese	National Taiwan Normal University

	Character Bias Database for Chinese Language Learners)	
2	汉字偏误标注的汉语连续性中介语语料库 (Chinese Continuous Interlanguage Corpus with Biased Chinese Character Labeling)	Sun Yat-sen University
3	非汉字文化圈国家学生错别字数据库网络应用平台 (Online application platform for student typo database in non-Chinese cultural circle countries)	Beijing Language and Culture University

Other corpora		
Number	Name	University
1	留学生汉语中介语语料库 (Chinese Interlanguage Corpus for International Students)	School of Chinese Language and Literature, Jinan University
2	Guangwai Lancaster 汉语学习者语料库 (Guangwai Lancaster Chinese Learner Corpus)	Guangdong University of Foreign Studies —— Lancaster University
3	全球汉语中介语语料库 (Global Chinese Interlanguage Corpus)	Beijing Language and Culture University
4	汉语单音节语音语料库 (Chinese Monosyllabic Speech Corpus)	Beijing Language and Culture University
5	面向计算机辅助正音的汉语中介语语音语料库 (A Chinese interlanguage speech corpus for computer-aided orthography)	Beijing Language and Culture University

Table 2. Classification of disyllabic homophones by part of speech in the "Level Outline"

Bisyllabic homophones with the same part of speech		
part of speech	Disyllabic homophones	quantity
Conj-conj	以致 - 以至	1
Pron-pron (-pron)	其他-其它、他们-地们-它们	2
v-v	报道-报到、分辨-分辩、发言-发炎、融化-溶化、散步-散布、有力-有利、制定-制订	7
adj-adj	可以-可以、谈话-谈话、爆发-暴发、变换-变幻、承办-惩办、出事- 出世、个别-各别、激动-机动、旅行-规行、年轻-年青、谨蔑-污美、行驶-行使、预定-预订、遇见-预见、震动-报动、祝愿-住院	16
n-n (-n)	地方-地方、精神-精神、利害-利害、笔迹-笔记、保健-宝剑、电器- 电气、电薄-店员、工大-功大、公式-工事、公园-公元、焦点-交点、界线-界限、例子-栗子、权利-权力、启示-启事、人士-人事、人身-人參、世纪-事连、水利-水力、数日-树木、食物-实物、事物-事务、体质-体制、议员-译员、终点-钟点、作法-做法、植物-职务、杂志-杂质、势力-视力-事例、形式-形势-刑事	30
Bisyllabic homophones have different parts of speech		
part of speech	Disyllabic homophones	quantity
n-adj	董事-懂事、风力-锋利、文明- 闻名、心事-新式、蒸汽-争气	5

n-v	抱负-报复、导弹-捣蛋、反应-反映、会关-攻关、功课-攻克、街道-接到、家具-加剧、手法-守法、算术-算数、声普-生育、团员-团回、行程-形成、香蕉-相交、修养-休养、意志-冲制、注释-注视	16
n-adv	童年- 月年	1
n-place words	心理- 心里	1
n-distinguishing words	县城- 现成	1
n-time words	钢材- 刚才	1
v-adv	必须-必雷、嫉妒-极度、进来-近来、忠干-络干	4
v-adj	步行-不行、兼任-坚韧、试用-适用、食用-实用	4
v-conj	不止 - 不只	1
v-distinguishing words	检查-检察	1
adv-time words	路年- 中年	1
adv-adj	有意-有益	1
Disyllabic homophones belong to different word classes		
part of speech	Disyllabic homophones	quantity
	人家-人家、过度-过渡、绘画-会话、经历-精力、蜜蜂-密封 预言-寓言、只是-指示、中心-衷心、结果-结果-结果	9

Table 3. Errors of disyllabic homophones in HSK dynamic composition corpus

Error Rate	Error Pause Frequency	Usage Frequency	Disyllabic Homophones
7.39%	68	920	必须-必需
3.56%	31	870	权利-权力
11.35%	32	282	反映-反应
1.44%	4	277	启示-启事
0.77%	2	260	事物-事务
0.99%	2	203	制定-制订
0.56%	1	178	中心-忠心
0.00%	0	167	人士-人事
6.34%	9	142	功课-攻克
0.00%	0	130	植物-职务
0.00%	0	115	世纪-事迹
4.81%	5	104	过度-过渡
2.94%	3	102	公园-公元
7.00%	7	100	以致-以至
2.22%	2	90	意志-抑制
4.49%	4	89	报道-报到
4.41%	3	68	形式-形势-刑事
16.17%	8	48	工夫-功夫
0.00%	0	42	散步-散布

0.00%	0	40	发言-发炎
2.78%	1	36	激动-机动
0.00%	0	36	生育-声誉
0.00%	0	31	势力-视力
0.00%	0	21	家具-加剧
8.33%	1	12	报复-抱负
0.00%	0	11	蜜蜂-密封
0.00%	0	10	注视-注释
0.00%	0	5	预言-寓言

Table 4. List of disyllabic homophones in the "Grade Outline"

1	地方-地方	2	精神-精神	3	可以-可以	4	利害-利害	5	人家-人家
6	谈话-谈话	7	报道-报到	8	报复-抱负	9	爆发-暴发	10	变换-变幻
11	保健-宝剑	12	笔迹-笔记	13	不行-步行	14	必须-必需	15	不止-不只
16	承办-惩办	17	出事-出世	18	导弹-捣蛋	19	电器-电气	20	懂事-董事
21	电源-店员	22	分辨-分辩	23	风力-锋利	24	反映-反应	25	发言-发炎
26	个别-各别	27	刚才-钢材	28	过度-过渡	29	工夫-功夫	30	攻关-公关
31	功课-攻克	32	公式-工事	33	公园-公元	34	绘画-会话	35	检察-检查
36	焦点-交点	37	嫉妒-极度	38	接到-街道	39	激动-机动	40	家具-加剧
41	进来-近来	42	经历-精力	43	坚韧-兼任	44	界线-界限	45	旅行-履行
46	例子-栗子	47	蜜蜂-密封	48	年轻-年青	49	启示-启事	50	其他-其它
51	融化-溶化	52	人身-人参	53	人士-人事	54	散步-散布	55	手法-守法
56	世纪-事迹	57	水利-水力	58	数目-树木	59	算数-算术	60	食物-实物
61	事物-事务	62	适用-试用	63	食用-实用	64	生育-声誉	65	童年-同年
66	团圆-团员	67	体质-体制	68	诬蔑-污蔑	69	文明-闻名	70	县城-现成
71	形成-行程	72	香蕉-相交	73	心理-心里	74	行驶-行使	75	心事-新式
76	休养-修养	77	预定-预订	78	遇见-预见	79	有力-有利	80	议员-译员
81	预言-寓言	82	有意-有益	83	以致-以至	84	意志-抑制	85	震动-振动
86	终点-钟点	87	制定-制订	88	作法-做法	89	中年-终年	90	蒸汽-争气
91	只是-指示	92	注视-注释	93	植物-职务	94	忠心-衷心	95	终于-忠于
96	祝愿-住院	97	杂志-杂质						
98	结果-结果-结果	99		权利-权力-全力	100	势力-视力-事例			
101	他们-她们-它们	102		形式-形势-刑事					

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).