A Corpus-Based Study on the Development of Second Language Learners' Vocabulary Knowledge and Use

Shan Li[1,a,*]

[1]*Southwest Petroleum University, Sichuan, Chengdu, 610500, China*
[a]*2682116624@qq.com*
[*]*Corresponding author*

***Abstract:** Pronunciation, vocabulary and grammar are the three basic elements of language learning. Vocabulary, as the building material of language, runs through the whole process of language learning. Understanding the meaning of Chinese vocabulary and mastering the usage of various parts of speech are of great significance to improving Chinese proficiency. This paper takes the words with both verb and adjective parts of speech as the research object, systematically studies the verb-form and class-like words at different learning levels in the inter-language corpus, and sorts out and classifies their correct usage and incorrect usage and analysis. Through the correct use cases of the corpus, the dynamic acquisition process and the characteristics of vertical development are summarized, that is, the correct rate of the use of most verbs and words will increase with the improvement of the learning level, but there are also some unbalanced phenomena. Furthermore, some suggestions for improving the learning and teaching of Chinese verbs and words are put forward.*

***Keywords:** Corpus; Second L language L earners; Acquisition Process; Chinese Vocabulary*

1. Introduction

Pronunciation, vocabulary and grammar are the three basic elements of language learning. Vocabulary, as the building material of language, runs through the whole process of language learning. Understanding the meaning of Chinese vocabulary and mastering the usage of various parts of speech are significance to improving Chinese proficiency. This article takes words with both verb and adjective parts of speech as the research object, and the research scope is the verb and form in the elementary vocabulary (levels 1 to 3) in the "International Chinese Education Level Standards for Chinese Proficiency" (herein after referred to as "level standards") word. "Grade Standards" is a standard issued by the Ministry of Education and the National Language Work Committee on July 1, 2021. It is aimed at overseas Chinese learners and can comprehensively inspect and evaluate learners' Chinese skills and Chinese proficiency standards. It is the standard for each Chinese education link mostly a according to [1]. It divides learners' Chinese proficiency into "three grades and nine grades", and contains a total of 11,092 words, which are arranged in the order of elementary (level 1 to level 3), intermediate (level 4 to level 6), and advanced (level 7 to level 9).

2. Corpus-based lexical correct use case analysis

This paper attempts to explore the mastery of Chinese second language learners' vocabulary knowledge by using concurrent words as the starting point. The inter - language research corpus of this paper mainly comes from the global Chinese inter - language corpus and the HSK dynamic composition, corpus and the verbs and words are classified and analyzed according to the learner's learning level from elementary, middle, and advanced.

According to table 1, the correct rate of use is positively correlated with the learning levelof the learners: "many""Hungry" "cold" "clear" "few" "Low" "move" "urgent" "cool" "effort" "Cheap" "careful" "success" "Mature" "Rich" "Responsible" "specification" "concentrated" "progress" "chaos" "behind" "trouble" "clear""Sure""protrude""unity""Complete" "Warm", a total of 28. The correct rate and the learning level of the learners are not positively correlated with the dynamic form and class words: "happy" "bad" "tired" "busy" "hot" "convenient" "Bright" "Full" "clear" "public" "break" "together" "deep" "Conscious", a total of 14. Because the correct rate of the mobile form and class word "fair" and "qi" is 100% in the three levels of elementary, middle and high levels, there is no

change in the correct rate, so it is classified into the phrases that are not positively related. According to the above statistical results, the correct rate of use of two-thirds of primary verbs and words is positively correlated with the learner's learning level. With the improvement of learners' learning level, the correct rate of using most of the words will increase accordingly. Table 2 is a one-way analysis of variance table based on the correct ratio of the use of verbs and class words, with the learner's learning level as a variable:

Table 1: Statistical table of the correct rate of verbs and class words for primary vocabulary in "Grade Standards"

grade	Vab.	correct use case						Correct rate (%)		
		primary	total	intermediate	total	advanced	total	primary	intermediate	advanced
Level 1	many	1173	1307	892	954	949	1007	89.8	93.5	94.2
	Hungry	146	174	60	65	71	76	83.9	92.3	93.4
	Happy	723	766	696	733	546	574	94.9	94.9	94.8
	bad	297	314	158	168	256	317	94.6	93.5	80.8
	tired	355	363	315	330	268	271	97.8	95.5	98.9
	cold	233	243	291	293	316	318	95.9	99.3	99.4
	busy	447	457	325	329	210	213	97.8	98.8	98.6
	clear	242	251	267	271	200	200	96.4	98.5	100.0
	hot	265	268	307	311	167	170	98.9	98.7	98.2
	few	224	261	184	195	225	227	85.8	94.4	99.1
Level 2	Low	107	129	100	109	83	90	82.9	91.7	92.2
	convenient	341	350	222	226	230	240	97.4	98.2	95.8
	move	121	155	182	207	122	127	78.1	87.9	96.1
	urgent	45	52	51	58	34	37	86.5	87.9	91.9
	cool	26	31	36	36	35	35	83.9	100.0	100.0
	Bright	14	14	66	72	22	25	100.0	91.7	88.0
	Full	72	80	80	87	74	84	90.0	91.9	88.1
	effort	1223	1362	1053	1117	676	704	89.8	94.2	96.2
	Cheap	167	173	143	143	129	129	96.5	100.0	100.0
	clear	121	132	103	114	80	86	91.7	90.4	93.0
	careful	59	70	58	67	66	72	84.3	86.6	91.7
Level 3	success	240	278	321	349	323	338	86.3	92.0	95.6
	Mature	80	83	83	83	63	63	96.4	100.0	100.0
	Rich	145	159	125	130	154	158	91.2	96.2	97.5
	Responsible	79	84	65	67	44	45	94.0	97.0	97.8
	public	13	13	14	14	3	3	100.0	100.0	100.0
	specification	5	7	7	7	7	7	60.0	100.0	100.0
	concentrated	40	55	28	30	19	20	72.7	93.3	95.0
	progress	212	225	196	202	138	142	94.2	97.0	97.2
	chaos	95	96	116	117	63	63	99.0	99.1	100.0
	behind	58	61	31	31	28	28	95.1	100.0	100.0
	trouble	217	223	154	157	103	104	97.3	98.1	99.0
	clear	37	48	25	28	15	15	77.1	89.3	100.0
	break	28	28	35	38	29	29	100.0	92.1	100.0
	together	28	28	27	27	22	22	100.0	100.0	100.0
	Sure	19	25	22	25	29	31	76.0	88.0	93.5
	in-depth	35	41	25	26	25	29	85.4	96.2	86.2
	protrude	26	29	12	12	4	4	89.7	100.0	100.0
	to unite	53	58	48	50	42	43	91.4	96.0	97.7
	Complete	27	31	25	26	12	12	87.1	96.2	100.0
	warmth	58	64	57	58	76	76	90.6	98.3	100.0
	conscious	28	28	17	22	5	5	100.0	77.3	100.0
total		7924	8616	7022	7384	5963	6239	92.0%	95.1%	95.6%

Note: The first-level vocabulary, second-level vocabulary and third-level vocabulary in the table are classified according to the level of vocabulary, and the primary, intermediate and advanced are classified according to the learning level of the students.

According to Table 2, the standard deviation, variance and coefficient of variation all show the same dispersion trend, that is, with the improvement of the learner's learning level, the difference in the correct rate of the same level of verbs and words is getting smaller and smaller, gradually tending to more stable. If the standard deviation, variance, and coefficient of variation reflect the stability of the data, then the maximum, minimum, median, and average reflect the upward trend of the data. The minimum value, median and average value in the table have the same trend of increasing with the learner's learning level, combined with the degree of dispersion, we can know: with the improvement of the learner's learning level, the acquisition accuracy rate of verbs and words gradually increases and the value gradually stabilizes.

Table 2: One-way analysis of variance table for the correct use cases of verbs and words

variable name	sample size	maximum value	minimum value	median	average value	standard deviation	variance	coefficient of variation
primary	42	1	0.6	0.916	0.905	0.087	0.008	0.096
intermediate	42	1	0.77	0.961	0.949	0.049	0.002	0.051
advanced	42	1	0.81	0.980	0.964	0.045	0.002	0.046

In general, these verbs and words are basically in line with the learners' gradual learning rules. Taking the second-level vocabulary "moved" as an example, the correct rate in the primary-level corpus is 78.1%, and the most frequent errors are the addition of the dynamic particle at the end of the sentence and the modal particle "le", followed by the misuse of the structural particle and the missubstitution of words with the same morpheme, such as " Thanks, excitement, gratitude" and so on; at the intermediate level In the level corpus, the correct rate increased to 87.9%, the frequency of misplacement of "le" at the end of the sentence and the misplacement of similar words were greatly reduced , and the frequency of misplacement of structural particles "de, di, de" was somewhat reduced. increase; in the advanced-level corpus, the correct rate is 96.1%, there is no misrepresentation of homomorphic words and structural particles, and the mistaken addition of dynamic particle and modal particle "le" at the end of sentences is also less . There are many reasons why the correct rate of using a small number of verbs and words does not increase with the improvement of learners' learning level. After the learners enter the advanced level, they may reappear the errors that occurred at the elementary level, which are mostly found in the first and second level vocabulary [2]; with the increase of learning content, learners may confuse certain usages, resulting in errors in advanced corpora that are not found in primary corpora. This kind of error is more common in second and third level vocabulary, such as second the first -level vocabulary "bright" has overlapped errors in the high-level corpus, which is not found in the primary corpus. Of course, these laws cannot be generalized, and the corpus of some vocabulary is not sufficient enough to be completely summarized by the above laws.

3. Analysis of the Causes of Problems in Vocabulary Acquisition

3.1 Native language negative transfer

Second language learners will be more or less disturbed by the mindset of their mother tongue, and they will substitute the language expression habits of their mother tongue into the Chinese they are learning, resulting in inter-lingual error transfer of mother tongue. The native language knowledge, skills, and even learning methods and attitudes that language learners have mastered may have an impact on learning new language knowledge and skills. If this impact is positive and positive, it is called positive transfer, and it can also be directly referred to as migrate [3]; On the contrary, it is called negative transfer, or interference. The negative transfer caused by the mother tongue is called the negative transfer of the mother tongue.

3.2 Generalization of target language knowledge

The generalization of target language knowledge, that is, the negative transfer of target language knowledge, refers to the fact that students apply the limited target language rules they have learned to new language phenomenon due to their lack of learning knowledge. After the learners have accumulated some Chinese knowledge and have the thinking ability to use Chinese knowledge to inductive reasoning, the old target language knowledge sometimes brings positive transfer after being mastered by students, which is helpful for foreign students to learn new knowledge; target language, foreign students often make mistakes due to over-generalization and errors due to the wrong application of their own knowledge of the target language. For example, the overlap error caused by the generalization of target language knowledge when using verbs and class words is more typical. Adjectives use the form of AABB to overlap and emphasize the degree, adding the meaning of "very". After learning this knowledge, foreign students often abuse the AABB form of adjectives.

3.3 Insufficient explanation of teaching materials

Textbooks are the link between teaching and learning, and are the direct basis for teachers to implement classroom teaching. The selection and use of a textbook can largely determine the effect of teachers' teaching and students' learning. The arrangement and annotation of textbook vocabulary is also the embodiment of its scientific and systematic nature. In the complete set of textbooks for the

comprehensive chapter of "Developing Chinese", there are 24 verbs and words that are studied in this article, including 20 for the elementary level, 4 for the intermediate level, and 0 for the advanced level. Vocabulary is mainly distributed in junior and intermediate textbooks, and it is mainly concentrated at the elementary level, which basically conforms to the distribution of the syllabus [4]. "Developing Chinese" explains and exemplifies the part of speech of adjectives in Lesson 10, and explains and gives examples of the part of speech of verbs in Lesson 27. This method is conductive to the acquisition of verbs and verbs. It is not necessary to expand too much beyond the required knowledge at the primary stage; the two commonly used parts of speech should be explained in detail and given examples. And with the improvement of the learner's learning level, for example, at the intermediate and advanced stage, the already learned verbs should be supplemented with another part of speech usage of the verbs, and the newly learned verbs should be comprehensively Part-of-speech tagging and word meaning explanation may be more conductive to increasing learners' understanding language and broadening their language proficiency.

4. Classroom teaching suggestions for second language learners of Chinese

4.1 Differentiated teaching for grammatical functions and parts of speech

In the process of second language learning, learners' understanding and mastery of vocabulary cannot be separated from classroom teaching, and part-of-speech teaching is the basis of concurrent word-class teaching. Through the analysis of the error corpus of verbs and class words, it can also be found that part of the errors are due to the lack of understanding or even confusion of the grammatical functions and collocation rules of verbs and adjectives by learners, resulting in improper use. Therefore, in the process of teaching Chinese, the grammatical functions and usages of various parts of speech should be clarified to help students distinguish the grammatical functions of verbs and adjectives. It is not only necessary to distinguish the usage of verbs and part-of-speech words when off they belong to different parts speech, but also learn to distinguish the part of speech of part-of-speech words in different contexts or semantic conditions. Only by being able to accurately distinguish and analyze the distribution of part of speech and sentence structure can the correct How to choose words and make sentences. The importance of learning the grammatical functions of parts of speech lies in the fact that proficiency in the difference between parts of speech and their usage is an important basis for making sentences [5]. Verbs and class words have two parts of speech usage at the same time, so the typical functions of verbs and adjectives can be taught respectively according to the conclusion in the second chapter of this article. In addition, targeted explanations and exercises should be carried out for the parts with a high error rate. According to the analysis of the reasons for the errors in Chapter 4, we can learn from the structural auxiliary words, degree adverbs, dynamic auxiliary words and modal particle "le", Targeted training on prepositions, willing verbs, and judging verbs.

First, with regard to the collocation of the structural particles "de、 di、 de", many grammar books have made strict specifications according to the syntactic components, that is, use "de" after an attribute, "di" after an adverbial, and "de" before a complement. ". But for Chinese learners, judging only based on syntactic components may be complicated and difficult. You can first teach students to use parts of speech to select structural auxiliary words, such as "X+de+noun, X+di+verb, X+de+adjective/degree adverb ", or combine the judgment of syntactic components with the judgment of part of speech, that is, "attributive +de+ noun, adverbial +di+ verb, X +de+ adjective (complement)", so as to improve the correct rate of students.

Second, regarding the application of degree adverbs, it can be analyzed from the perspective of sentence structure and part of speech. According to the sorting of the corpus related to degree adverb errors in Chapter 3, it can be seen that some fixed sentence patterns often have errors related to degree adverbs. "More" is used instead of other degree adverbs; "Too...too""More and more...." and other sentence patterns generally do not add degree adverbs. In terms of parts of speech, generally only adjectives can be modified by adverbs of degree, and adverbs of degree cannot be used before verbs.

Third, with regard to the dynamic particle and the modal particle "le", the main errors to be avoided are the wrong addition of the modal particle "le" at the end of the sentence and the omission of the dynamic particle "le". Such errors can be distinguished from the grammatical functions and sentence meanings of modal particles and dynamic particles. For example, "le" needs to be added when it needs to indicate that an action or trait has been realized, but "le" is not needed when simply describing a property or state. Teachers can use examples to compare and show the difference between the two syntaxes and sentence meanings to help students understand.

Fourth, with regard to the errors of verbs and prepositions, the most frequent occurrences are the collocations with the willing verb "will, can, want", the judgment verb "is" and the preposition "by". For the above types of errors, we can also start from the perspective of sentence patterns, and explain and practice the fixed sentence patterns "subject +will/can/want+ verb", "subject +is+ adjective" and "Passive voice".

4.2 Staged teaching according to the learning level

In view of the particularity of concurrent words, international Chinese teachers should pay attention to the dual teaching of lexical meaning and grammatical meaning when teaching concurrent words. It can neither simply explain the meaning of words nor only explain a single point, but at grammatical teaching that combines meaning and grammar, from easy to difficult, step by step. For learners at the primary stage, too much input of language knowledge is often likely to cause confusion, which makes it more difficult to understand and use concurrent words correctly. Therefore, when facing learners at the primary stage, you can follow the principle of gradual and orderly progress. For example, when you learn the verb usage of a certain verb and class word, you don't need to rush to explain its part of speech and expand its adjective usage. , but first consolidate the verbal knowledge needed at this stage through explanations and exercises [6]. For example, in what scene can the word be used, how should it be used, can the word be used as a subject, predicate or object, determine whether its position in the sentence can be accompanied by other components, what kind of components can be attached, and so on. In the intermediate and advanced stages, with the improvement of students' learning level and the accumulation of Chinese knowledge, they have a certain knowledge structure of Chinese pronunciation, vocabulary and grammatical structure, so that they can further carry out part-of-speech teaching. Part of speech teaching is the basis of part-of-speech teaching. Chinese learners first need to understand the grammatical features and usage of each part of speech before they can understand part-of -speech words on this basis, and then use verb part-like words correctly. At the intermediate and advanced stage, you can also systematically compare and sort out the verbs and similar words you have learned before, clarify the usage of verbs and adjectives from a macro perspective, and analyze easily confused words. Intensive training, correct and wrong analysis, and strengthen the impression of learners.

4.3 Use a variety of teaching methods to strengthen and consolidate exercises

The purpose of international Chinese education is not only to enable learners to master the basic knowledge of Chinese and the basic skills of listening, speaking, reading and writing, but also to cultivate learners to use Chinese accurately and proficiently, internalize the acquired knowledge and apply it to actual communication scenarios Ability. Therefore, Chinese teachers can use a variety of teaching methods when explaining concurrent words. First of all, you can use the most basic method of explanation to sort out the meaning and usage of the words you speak. Teaching is inseparable from the teacher's accurate explanation in the classroom, especially the teaching of verbs and words. Chinese teachers should clarify the semantics and usage of words during teaching, and expand and expand knowledge at an appropriate stage, use texts and correct demon examples to different part-of-speech usages, and help students have a more accurate and clear cognition and understanding g of the part-time words explained. For example, the verbs " moved, busy" that are easy to be mistakenly preceded by adverbs of degree, or words that are easily confused with words of different parts of speech, such as the noun " Responsibility" and the verb "responsibility", etc. Teachers should clarify both the usage of parts of speech and the parts of speech of words, so as to help students reduce similar errors in subsequent use.

Secondly, methods such as scenario method and intuition method can be used to further deepen students' understanding and mastery of the spoken words. Any language cannot be separated from the real language environment, and should be practiced in a specific communication environment to set up a real or simulated real language environment for students to practice the language knowledge they have learned. In the questioning or interaction with students, try to guide students to use different parts of speech to express, so that students can truly experience the difference and connection between the two parts of speech, and deepen students' understanding. For some verbs that can be visually presented by pictures, actions, etc., you can use the direct method to use intuitive visual stimulation to mobilize students' cognitive ability, which is more conductive to mobilizing students' learning enthusiasm and increasing students' perceptual knowledge. Vocabulary is directly linked to the students' first reaction, which saves the step of native language conversion, and is more conductive to the understanding, memory and use of the second language. For example, emotional words such as happy, sad, and moved

can be directly displayed with the head portraits of characters that can express the corresponding emotions, so that students can directly connect vocabulary with emotions. Proper use of this kind of heuristic teaching is conductive to speeding up the classroom rhythm and improving teaching efficiency [7]. In addition to the learning of new knowledge, practice and consolidation should also run through the language learning. After all, the teaching time in the classroom is limited. Through the drill and practice of the vocabulary learned, students can deepen their understanding of the meaning and part of speech, and better grasp the usage of verbs and words. However, consolidation and practice should also consider effectiveness, pertinence, difficulty and interest. The content of exercises should be based on what has been learned, closely followed the purpose of teaching, and focus on the practicality of exercises [8] .According to the learning level of the students, it should be neither too difficult nor too easy. In addition to practicality, we must also pay attention to the fun of practice. It should not be just a mechanical practice, but should be practiced from different angles and in different ways to induce students' ideal behaviors and provide positive feedback in a timely manner in a non-evaluative atmosphere. Feedback guides students to actively participate in the practice process, controls the development of students' practice in the right direction, ensures the success rate of the use of verbs and words.

5. Conclusions

This paper takes the verbal and similar words in the elementary vocabulary in the "International Chinese Education Level Standards for Chinese Proficiency" as the research object. By searching the global Chinese interlanguage corpus and the HSK dynamic composition corpus of Beijing Language and Culture University, the lexical corpus is classified and counted. On the one hand, it analyzes The correct use cases of the vocabulary are analyzed, and the common collocations in the correct use cases are dynamically analyzed; on the other hand, the real corpus is used to describe and analyze the problem, and try to draw some conclusions and suggestions. According to the statistical arrangement of these matching methods, we found that: first, from the perspective of the overall acquisition situation, with the improvement of learners' learning level, the correct use rate of learners is gradually increasing. Secondly, from the perspective of vertical development, it can be seen that the use of various usages by learners has gradually become more diverse, and the forms of expression have gradually become more complex. The shortcoming of this article lies in the lack of investigation and testing. It only examines the verbs and words in the primary vocabulary of the "Grade Standard" based on the corpus, and the scope of investigation is not wide enough. The suggestions are also relatively limited to elementary vocabulary. It is hoped that in the future, we can continue to investigate and study the verbs and words in the intermediate and advanced vocabulary of the "Grade Standards", and further test and confirm the research results of this paper through teaching practice.

References

[1] Liu Changqing. Research on light verb theory and Chinese related issues [M]. Wuhan: Wuhan Publishing House, 2008.

[2] Guo Rui. Course of Parts of Speech Research in Modern Chinese [M]. Beijing: Commercial Press, 2002.

[3] Qi Huyang. Grammar in Teaching Chinese as a Foreign Language [M]. Shanghai: Fudan University Press, 2005.

[4] Shen Jiaxuan. Nouns and Verbs [M]. Beijing: The Commercial Press, 2016.

[5] Zhang Guoxian. Research on Function and Cognition of Adjectives in Modern Chinese [M]. Beijing: Commercial Press, 2006.

[6] An Hualin. The treatment of names, verbs, and forms from the perspective of two lexicons[J]. Language Teaching and Research, 2005(04).

[7] Gao Hang. Investigation on the problem of concurrent classes in Chinese from the perspective of cognitive grammar[J]. Chinese Learning, 2009(02).

[8] Li Quan. On the Practicability of Teaching Chinese as a Foreign Language [J]. Language Teaching and Research, 2007(03): 28-35.