

A corpus-based study of Korean vocabulary acquisition and teaching strategies

Ming Zhang

Shandong Vocational College of Science and Technology, Weifang, 261053, Shandong, China

2024190583@qq.com

Keywords: corpus, Korean vocabulary, vocabulary acquisition, teaching strategies, language teaching

Abstract: This study aims to explore corpus-based Korean vocabulary acquisition and teaching strategies. Through in-depth analysis of Korean corpora, it reveals the characteristics of Korean vocabulary, such as frequency of use, collocation patterns, semantic changes and difficulty of acquisition. As one of the three elements of language, vocabulary occupies an important position in teaching Chinese as a foreign language. The teaching of Chinese and Korean vocabulary has its own particularities compared to other countries. Although more than half of modern Korean vocabulary is made up of Chinese characters, Chinese and Korean belong to different language families and have inherent differences in language structure and other aspects. Moreover, due to the passage of time and social changes, many Chinese characters in Korean that originally had the same meaning or form as Chinese words have undergone significant changes. These Chinese characters that have undergone changes are prone to negative transfer in the process of Korean students learning Chinese vocabulary. The study first provides an overview of the application of corpus in language learning and vocabulary research, and then analyzes the vocabulary acquisition process of Korean learners, and finds that vocabulary frequency, context, collocation and word meaning discrimination play a key role in vocabulary acquisition. On this basis, this study proposes strategies for optimizing Korean vocabulary teaching using corpus, emphasizing the combination of real language data from corpus and teaching practice to improve the effectiveness and pertinence of vocabulary teaching. The research results provide a new theoretical basis and practical guidance for Korean vocabulary teaching, which is of great significance for promoting Korean language teaching reform and improving learning outcomes.

1. Introduction

With the acceleration of the globalization process, the demand for learning Korean is growing, and Korean language education has received widespread attention worldwide[1]. However, the acquisition of Korean vocabulary has always been one of the main difficulties for learners, and the degree of vocabulary mastery directly affects the efficiency and level of Korean learning[2]. The comparison of Chinese and Korean languages and the study of Korean students' errors in acquiring Chinese vocabulary are still relatively weak[3]. Therefore, comparing Korean Chinese character words with modern Chinese words, identifying similarities and differences, analyzing Korean students' errors in Chinese words due to the influence of Chinese character words, and identifying teaching methods to address these errors are of great significance for teaching Chinese vocabulary to Korean students[4]. In recent years, corpus technology, as an emerging tool for language research, has been widely used in language acquisition and teaching research. With its large data scale and authentic language materials, it provides rich resources and data support for language teaching[5].

This study aims to combine corpus technology with Korean vocabulary teaching, analyze the usage characteristics of Korean vocabulary in authentic contexts, and explore how to use corpus data to optimize vocabulary teaching strategies. Through research on the frequency, collocation, and semantic variation of words in the Korean corpus, the laws and difficulties of Korean vocabulary acquisition are revealed, providing effective guidance for Korean language teaching. The research in this paper not only helps to improve the efficiency and effectiveness of Korean

vocabulary teaching, but also provides a scientific learning path for Korean learners, helping them to master Korean vocabulary more quickly and accurately.

This paper introduces the application of corpora in language research and vocabulary acquisition; provides an in-depth analysis of the characteristics of Korean vocabulary acquisition based on corpora; and proposes corpus-based teaching strategies for Korean vocabulary to provide reference and guidance for teaching practice. This paper also analyzes the Korean students' errors in Chinese vocabulary related to Chinese characters and the related teaching methods. It is hoped that this research will provide a new perspective on Korean vocabulary teaching and practical help for Korean learners' vocabulary acquisition.

2. The application of corpus in language research and vocabulary acquisition

As a language research tool, corpus first emerged in the mid-20th century. With the rapid development of computer technology, the construction and application of corpus has been greatly promoted[6]. Through the collection and analysis of a large amount of authentic language data, corpus can reveal the real situation of language use and provide detailed information on vocabulary, syntax, semantics, etc. Korean words can be divided into inherent words and borrowed words according to their sources[7]. Korean native words refer to words that have existed in the Korean language since ancient times. Korean loanwords refer to words that have been “borrowed” from other languages into the Korean language[8]. These words are usually divided into two categories: first, Chinese characters, which refer to Chinese-derived words that use Chinese characters as the basis for their form and meaning but are pronounced in Korean[9]. These include Chinese words borrowed by the Korean language, Japanese kanji words, and words created by Koreans themselves using Chinese characters[10]. Second, there are words borrowed from other languages, such as English loanwords and French loanwords. It covers a variety of written and spoken forms and has become an important resource for linguistic research, providing scientific and objective data support for research in a number of fields such as lexicology, linguistics and semantics. Vocabulary Frequency Distribution :

$$f(r) = \frac{c}{r^k} \quad (1)$$

The advantages of corpora are particularly evident in vocabulary acquisition research. Traditional vocabulary teaching often relies on the subjective experience of textbook writers, which makes it difficult to fully reflect the use of vocabulary in the real language environment. However, corpora provide a large number of real usage examples in natural language, which can help researchers and learners understand the frequency of vocabulary use in actual contexts, collocation patterns, and semantic changes. By analyzing the distribution of high-frequency and low-frequency words, learners can effectively master common vocabulary and improve learning efficiency. Corpus can also reveal the differences in the usage of the same word in different contexts, providing a more comprehensive reference for vocabulary acquisition.

With the deepening of language teaching research, corpus has gradually become an important tool for second language teaching, especially in vocabulary teaching. Through corpus, teachers can obtain real and rich information on vocabulary usage, providing data support for the design of teaching content. For example, in Korean language teaching, teachers can use corpus data to analyze the frequently occurring Korean words and explain the usage of words through collocations in real contexts, helping learners understand the polysemy and pragmatic functions of words. This corpus-based teaching method not only improves the pertinence of vocabulary teaching, but also effectively promotes learners' comprehensive language skills. Vocabulary Learning Curve:

$$V(t) = V_{max}(1 - e^{-kt}) \quad (2)$$

The application of corpus in Korean vocabulary acquisition has great potential value. As a language with significant differences from Chinese and English, Korean presents unique characteristics in its vocabulary system and grammatical structure, which pose a challenge to

learners. Through corpus technology, researchers and learners can gain an in-depth understanding of Korean vocabulary collocations, semantic changes, and common expressions, as well as obtain information on the practical application of vocabulary in a real language environment. This paper explores a new perspective on second language acquisition research using the acquisition of conjunctions by Korean students as an example. The purpose is to systematically summarize the regular acquisition characteristics of learners when using conjunctive adverbs, and while observing Chinese as a second language acquisition research, it also contributes to the study of Chinese as a native language. Secondly, it is hoped that through the generalization and refinement of the research object, and after a linguistic and morphological investigation of the overall use of correlative adverbs, a case-by-case analysis of their characteristics can be conducted. This will not only help learners master the correct usage of vocabulary, but also help improve their cognitive depth of vocabulary, thereby more efficiently achieving vocabulary acquisition.

3. Characteristics of Korean vocabulary acquisition based on corpus analysis

In the process of in-depth analysis of Korean vocabulary acquisition, corpora provide us with a large amount of real and natural language data, making the mastery of Korean vocabulary more scientific and comprehensive. Based on the statistics of the corpus of the intermediary language of international students, using corpus research methods and comparative research methods, the use of ten commonly used Korean collocations by international students was investigated. This part will explore three aspects: the frequency and usage patterns of vocabulary, the mastery of vocabulary semantics and context, and the difficulty of vocabulary acquisition and learning order. It will provide a detailed analysis of the characteristics of Korean vocabulary acquisition in a real language environment, with a view to providing more effective vocabulary learning strategies for Korean learners.

3.1. Frequency and usage patterns of vocabulary

There is a significant difference between the acquisition of high-frequency and low-frequency words in Korean vocabulary learning. Corpus-based analysis has found that high-frequency words appear more frequently in daily communication and reading, and are essential for learners to master at the beginning stage. These words usually cover basic daily expressions, common expressions and daily activities, and are therefore more likely to be repeatedly encountered and remembered in real contexts. However, low-frequency words often belong to specific fields or more formal expressions, and they appear less frequently. Learners need to gradually accumulate them in the learning process and master the usage of these words through extensive reading and practice.

Collocation is a key element in vocabulary acquisition. The corpus provides a wealth of information on vocabulary collocations, enabling learners to better understand the use of vocabulary in real contexts. Korean vocabulary collocations are significantly different from those of other languages such as Chinese and English. Many vocabulary collocations are fixed or have specific usage habits. The basic context of the investigation is “the frequency of use of each type of adverb”, showed in Figure1.

Compared to that of Korean students and native Chinese speakers; the correct use of correlative adverbs; errors in the use of correlative adverbs – an examination of errors from the perspective of subordinate clauses”. This process reveals the specific acquisition process and errors of each correlative adverbs and the errors, which is conducive to our understanding of the general acquisition characteristics of correlative adverbs from a microscopic case perspective, and to the process of describing, analyzing and explaining correlative adverbs “in general, one by one, and in the end”. The same verb can be used with different nouns or adjectives in different contexts. This collocation relationship determines the precise usage and expression effect of the vocabulary. Through frequency analysis of the corpus, learners can understand the collocation patterns of common words, thereby improving the accuracy and authenticity of Korean expression.

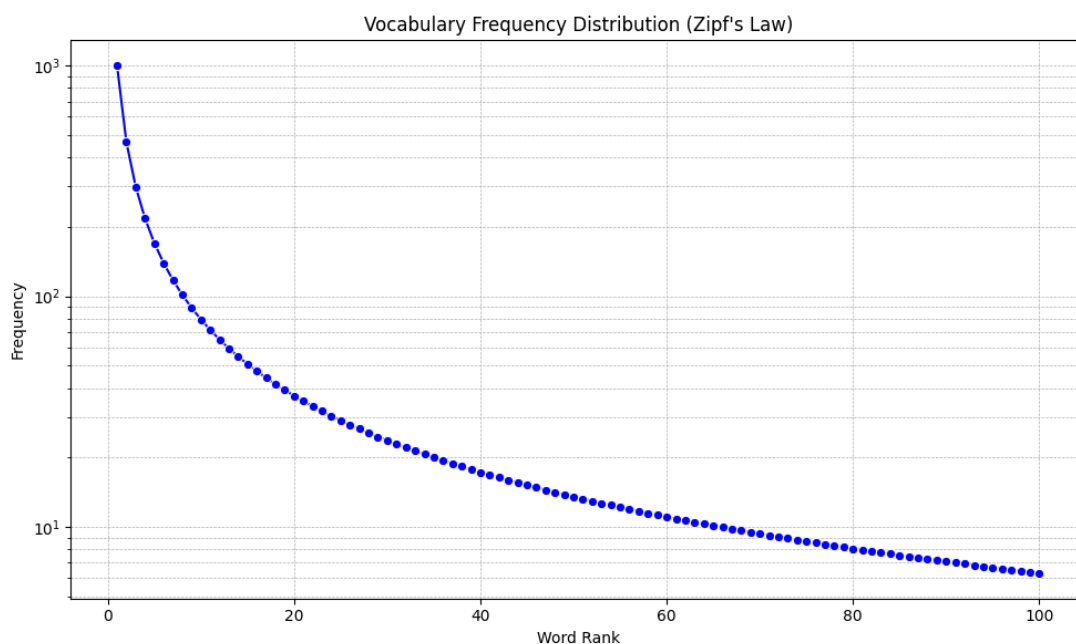


Figure 1 Vocabulary Frequency Distribution (Zipf's Law)

Korean vocabulary is highly polysemous, and the same word may have completely different meanings and usages in different contexts. With the help of the corpus, learners can observe the actual usage of these words in different contexts and understand the specific manifestations of their polysemy. This not only helps learners deepen their understanding of the meaning of words, but also helps them use words flexibly according to the context in actual communication, avoiding misunderstandings caused by unclear word meanings. For example, the meaning of a common word in everyday conversation and in a formal setting can be very different. The corpus provides learners with real-life examples of these usages to help them better grasp the polysemy of words, showed in Figure 2:

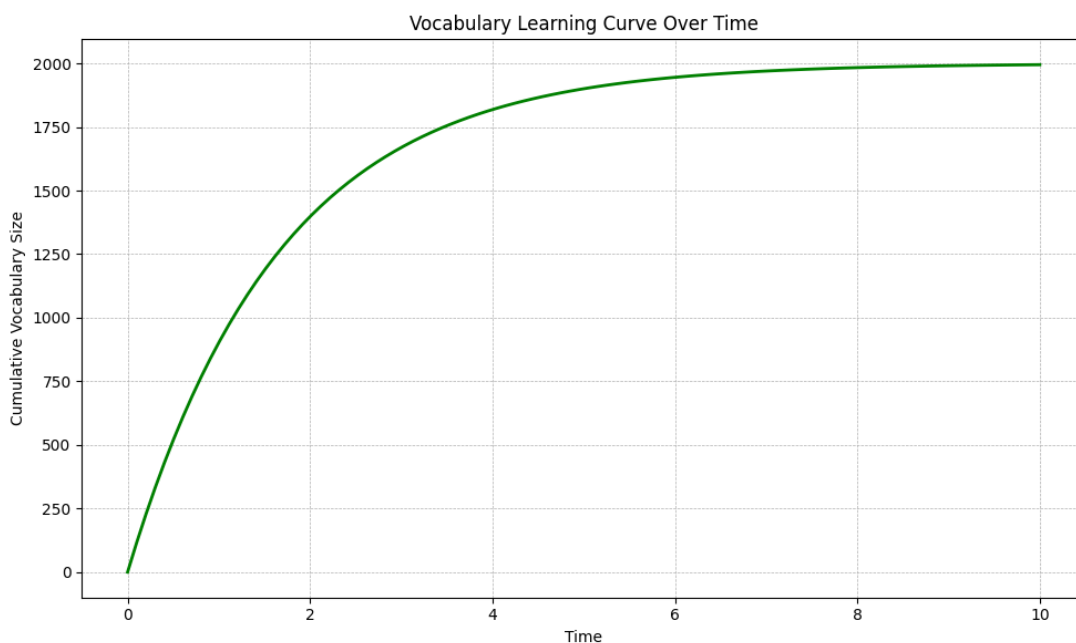


Figure 2 Vocabulary Learning Curve over Time

Korean has a large number of fixed expressions and idiomatic phrases, and mastering these expressions is essential for improving language fluency and expression. Corpus analysis shows that the frequency and collocation rules of these fixed expressions in different scenarios are very clear,

and they are the content that learners need to focus on mastering. Through the analysis of idiomatic phrases in the corpus, the intermediate language corpus data is examined using accuracy and relative frequency; the test data mainly uses the traditional accuracy method, supplemented by the latent scale method, and the order of the rankings is tested using the reuse coefficient or scale coefficient. The difficulty level of the acquisition of the ten collocations is divided into three levels. Learners can understand the usage occasions, collocations and expression effects of these phrases, so as to use Korean more freely in practical applications. Compared with simply memorizing vocabulary, learning fixed expressions and idiomatic phrases helps improve the efficiency of language learning and enables learners to reach the level of fluent expression more quickly.

3.2. Mastery of vocabulary meaning and context

In Korean vocabulary, distinguishing between synonyms and near-synonyms is a major difficulty in vocabulary acquisition. Many words show subtle differences in their use in different contexts. The corpus provides learners with a large number of practical examples of synonyms and near-synonyms, so that they can observe the collocation and use of these words in different contexts. Vocabulary teaching is an important part of Chinese teaching, but both beginners and intermediate and advanced learners of Chinese will experience a lot of vocabulary confusion during the process of acquiring Chinese. Previous academic research on this kind of lexical confusion and error has mostly been conducted from the perspective of synonyms/near-synonyms, but this perspective ignores inter-language contrastive analysis. By comparing the collocation partners, frequency of occurrence and pragmatic context of synonyms, learners can more accurately grasp their subtle differences and avoid misunderstandings or unnatural expressions caused by inappropriate choices in actual use.

Many Korean words are polysemous, meaning that they can have very different meanings in different contexts, which places higher demands on learners' vocabulary mastery. Corpora help learners understand the multiple meanings of words in different contexts by providing rich contextual information. A word may have different meanings and usages in everyday conversations, academic articles, or news reports. By analyzing these authentic corpora, learners can better understand the polysemy of words and accurately select and use words in specific contexts.

Context plays a decisive role in the use of Korean words. The meaning and collocation of a particular word may change in different contexts. With the help of corpus data, learners can observe the usage patterns of words in different contexts and understand which words are suitable for formal, informal, spoken or written contexts. The theoretical basis of the concept of confusing words comes from the interlanguage theory proposed by American scholar C. P. Selinker. This theory holds that when a second language learner learns a second language, they will produce a system of intermediary language that is different from both their first and second languages. This language system is in a state of flux, constantly approaching the target language. The same word may convey completely different emotional colors or levels of politeness when paired with different vocabulary. Through this in-depth analysis of the context, learners can develop a sensitivity to Korean vocabulary and improve the accuracy and appropriateness of using vocabulary in different situations.

Korean contains a large number of idioms and set expressions, and these fixed expressions often have unique cultural connotations and context-dependence, which is a difficulty in Korean vocabulary acquisition. The corpus can show the use of these idioms and set expressions in different contexts, helping learners understand their specific meanings and situations of use. For example, some Korean idioms are commonly used in spoken language, while others are more commonly used in written or literary works. By studying the relevant usage examples in the corpus, learners can master the correct usage of these idioms and proverbs, and further improve their Korean expression skills and cultural understanding.

3.3. Difficulty of vocabulary acquisition and learning order

In Korean vocabulary learning, the priority of mastering basic vocabulary is an important step for learners to construct a vocabulary system. Corpus data shows that frequently occurring basic

vocabulary usually covers core concepts in daily communication, involving basic categories such as life, emotions, and behaviors. These words appear frequently in various contexts and are likely to be combined with other words. Therefore, learners should first master these basic words in order to quickly develop a certain level of language expression at the beginner level. Through corpus analysis, teachers can identify these high-frequency words to help learners establish a solid vocabulary foundation and lay a solid foundation for subsequent vocabulary expansion.

After mastering basic vocabulary, learners need to gradually expand to the learning of intermediate and advanced vocabulary, which usually involves more complex concepts and richer semantic features. Corpus analysis shows that intermediate and advanced vocabulary often appears in specific fields or more formal contexts, such as topics in science and technology, economics, and culture. At this stage, learners should expand their vocabulary in a targeted manner based on their own learning goals and interests, combined with real-world examples from the corpus. At the same time, analyzing the collocations and usage contexts of these words in the corpus can help learners understand the specific usage of these words, avoid memorizing in isolation, and improve learning efficiency.

There are some difficult words in Korean, which are usually closely related to Korean history, culture, customs and professional fields. There are differences between the vocabulary systems of different languages, and confused words are more likely to be affected by negative transfer from the mother tongue. Confused words produced by students with different native language backgrounds during the acquisition of Chinese have their own characteristics, so it is also necessary to carry out language-specific analysis for learners with specific native language backgrounds. Korean, as a language within the Chinese character culture circle, has a considerable number of Chinese character words, but there are still significant differences between its specific meanings and usage and Chinese vocabulary, which has a significant impact on native Korean speakers when acquiring Chinese verbs. Mastering these words not only requires learners to have a high level of language proficiency, but also a certain understanding of Korean culture. A corpus can demonstrate the application of these difficult words in real contexts and help learners understand their cultural background and inherent meaning. For example, some words related to traditional Korean festivals and customs appear more frequently in specific contexts. Through corpus analysis, learners can better understand the cultural connotations of these words and improve the depth and breadth of their language learning.

Based on corpus data analysis, the order of Korean vocabulary learning should follow the principle of going from easy to difficult and from commonly used to less commonly used, to ensure that learners can gradually improve their vocabulary and application skills. Learners should focus on mastering high-frequency basic vocabulary at the beginning, and then gradually expand to intermediate and advanced vocabulary, and finally delve into difficult vocabulary related to specific fields or cultural contexts. This scientific learning sequence helps to avoid the learning burden caused by a sudden expansion of vocabulary, and can continuously strengthen the understanding and memorization of vocabulary through the real language environment provided by the corpus, so as to achieve systematic and effective vocabulary learning.

4. Corpus-based Korean vocabulary teaching strategies

In teaching Korean vocabulary, teachers should make full use of the high-frequency vocabulary data in the corpus to design teaching content that meets the actual needs of learners. By analyzing the frequency distribution of Korean words in the corpus, teachers can determine the core vocabulary that should be taught first to ensure that students can quickly accumulate sufficient vocabulary at an early stage. The teaching of these high-frequency words should be combined with real-life contexts and common collocations, so that students can understand how to use the words in different contexts while grasping their basic meaning, thereby improving the efficiency and effectiveness of vocabulary learning.

Understanding vocabulary collocations and contexts is an important part of Korean vocabulary acquisition. The corpus provides teachers with a large number of real-life examples of vocabulary

collocations. In teaching, teachers can use the corpus to demonstrate the collocations and usage patterns of vocabulary in different contexts, helping students understand the diversity and practical application of vocabulary. For example, teachers can select some typical examples of vocabulary collocations for explanation, guide students to imitate and use them in practical exercises, gradually cultivate their sensitivity to vocabulary collocations, and improve the accuracy and naturalness of their Korean expression.

A corpus is not only a tool for teachers to teach with, but also an important resource for students to learn independently. In teaching Korean vocabulary, teachers should encourage students to use the corpus to look up and research words, and learn to obtain information about the use of words from a large amount of corpus data. By independently analysing the frequency, collocation and context of words in the corpus, students can gain a deeper understanding of word usage and develop independent learning skills. In addition, the authentic examples in the corpus can also help students remember words better and consolidate their grasp of vocabulary.

Korean learners have different vocabulary levels, and vocabulary teaching strategies should be developed accordingly for students at different levels. For beginners, teachers can help them lay a solid vocabulary foundation by selecting high-frequency and practical words from the corpus. For intermediate and advanced learners, the focus should be on guiding them to learn complex word collocations and the usage of polysemous words to deepen their understanding of vocabulary. The rich corpus data in the corpus can meet the learning needs of students at different levels, help teachers flexibly adjust the content and methods of teaching, and achieve personalized vocabulary teaching.

5. Conclusion

In the study of corpus-based Korean vocabulary acquisition and teaching strategies, it is clear that corpora not only provide rich and authentic language data for vocabulary acquisition, but also bring significant innovation and improvement to teaching practice. Confusing words are divided into confusions of synonyms/near-synonyms, confusions of similar meanings, and confusions across semantic categories, as well as “one-to-one” confusions, “one-to-many” confusions, and “many-to-one” confusions. In addition, as the concept of confusable words continues to develop in academia, learners can use Korean more confidently in a changing language environment by focusing on mastering high-frequency words, gaining a deeper understanding of word collocations and contexts, and cultivating independent learning skills. In addition, differentiated teaching strategies for learners at different levels can effectively improve their learning experience and vocabulary mastery efficiency.

With the continuous development of corpus technology, we should continue to explore its potential for Korean vocabulary teaching, especially in terms of the integration of intelligent teaching tools and online learning platforms. Strengthening teacher training in the use of corpora and improving their application in classroom teaching will help promote the continuous improvement of Korean vocabulary teaching. Through these measures, we hope to better meet learners' needs for Korean vocabulary and help them achieve greater success in their language learning journey.

References

- [1] Al-Seghayer K .The Effect of Multimedia Annotation Modes on L2 Vocabulary Acquisition: A Comparative Study[J].Language Learning & Technology, 2001, 5(1):202--232.
- [2] Pulido D .Modeling the Role of Second Language Proficiency and Topic Familiarity in Second Language Incidental Vocabulary Acquisition Through Reading[J].Language Learning, 2010, 53(2):233-284.DOI:10.1111/1467-9922.00217.
- [3] Wang M , Ko I Y , Choi J .The importance of morphological awareness in Korean - English biliteracy acquisition[J].Contemporary Educational Psychology, 2009, 34(2):132-142.

DOI:10.1016/j.cedpsych.2008.12.002.

[4] Chen C, Truscott J .The Effects of Repetition and L1 Lexicalization on Incidental Vocabulary Acquisition[J].Applied Linguistics, 2010, 31(5):693-713.DOI:10.1093/applin/amq031.

[5] Evans M A, Saint-Aubin J .Vocabulary acquisition without adult explanations in repeated shared book reading: An eye movement study. [J].Journal of Educational Psychology, 2013, 105(3):596-608.DOI:10.1037/a0032465.

[6] Xiao Feng. Applying Priming Methods to L2 Learning, Teaching and Research: Insights from Psycholinguistics[J]. Language Teaching Research, 2013, 17(2):269-271. DOI:10.1177/1362168813476175.

[7] Ji M G , Baek S .Native Korean-Speaking Children Learning to Read in English: A Structural Analysis of L2-English Literacy Acquisition[J].Journal of psycholinguistic research, 2019, 48(2):391-415. DOI:10.1007/s10936-018-9610-7.

[8] Taylor A .Learning Vocabulary in Another Language I[J].English for Specific Purposes, 2004, 23(1):87-90.DOI:10.1016/S0889-4906(02)00014-5.

[9] Laine M ,Tünde Polonyi,Kálmán Abari.More Than Words: Fast Acquisition and Generalization of Orthographic Regularities During Novel Word Learning in Adults[J].Journal of Psycholinguistic Research, 2014, 43(4):381-396.DOI:10.1007/s10936-013-9259-1.

[10] İsmail Hakkı Erten, Tekin M .Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets[J].System, 2008, 36(3):407-422. DOI:10.1016/j.system.2008.02.005.