

[Studies in Linguistics]

The Effect of English Nominalization on Text Difficulty: Taking *Intensive Reading Textbooks of College English* (Third Edition) by SFLEP for Example

LIU Yaxuan

University of Leeds, UK

Received: October 24, 2023

Accepted: November 6, 2023

Published: March 31, 2024

To cite this article: LIU Yaxuan. (2024). The Effect of English Nominalization on Text Difficulty: Taking *Intensive Reading Textbooks of College English* (Third Edition) by SFLEP for Example. *Asia-Pacific Journal of Humanities and Social Sciences*, 4 (1), 078–096, DOI: 10.53789/j.1653–0465.2024.0401.009

To link to this article: <https://doi.org/10.53789/j.1653–0465.2024.0401.009>

Abstract: English nominalization has been a topic of academic interest as an important linguistic phenomenon in English, and its types and frequency of use have different effects on vocabulary density and semantic understanding, resulting in changes in the difficulty of reading discourse. In this study, we propose to use AntConc statistical software to analyze the difficulty of English reading texts by studying the use of nominalization of English verbs, adjectives, and adverbs in the textbook, in order to reveal the influence of the use of English nominalization on the difficulty of the reading texts, and to provide references for the compilation and selection of English textbooks.

Keywords: English nominalization; textual studies; text difficulty; effect studies

Notes on the contributor: LIU Yaxuan, a postgraduate majoring in Professional Language and Intercultural Studies at the University of Leeds. She is interested in Modern English Corpus and Intercultural Communication. Her email address is 1561808284@qq.com.

1. Introduction

One of the most prevalent and significant linguistic phenomena in English is nominalization. Since academics and experts have conducted numerous studies on the topic, including how to handle it when translating English into other languages (see Somers et al., 1988; Bednárová-Gibová, 2016; Bisiada, 2018), its pragmatic

function (see Malchukov, 2006; Sušinskienė, 2009; Juznic, 2012; Mithun, 2016; Faure, 2021), and its formation (see Heyvaert, 2003; Hartmann, 2014; Bruening and Bruening, 2018; Shibatani, 2019; Kotowski, Schneider and Kawaletz, 2023), this study will adopt a more pragmatic approach by examining how English nominalization—including verb to noun, adjective to noun, and adverb to noun—affects reading difficulty. It will accomplish this by utilizing AntConc statistical software, which is more exact, accurate, and objective than earlier investigations, and by combining computer expertise with linguistic theory. The author aims to investigate the impact of English nominalization on the difficulty of reading texts in English, with a more quantitative and statistical method with the assistance of AntConc. Specifically, the author would first define nominalization and list its many manifestations. She would then discuss domestic and international studies on relevant studies of English nominalization. The author would later investigate the relationship between English nominalization and reading difficulty by using the intensive reading texts excerpted from *College English* (third edition) published by SFLEP^[1] as research materials. The AntConc corpus statistical software is used to count and verify the frequency and lexical density of nominalization in the texts of Books 1 and 5, respectively, of the *College English* intensive reading textbooks. Besides, the author also looks into the classification, function, and frequency of various types of English nominalization.

AntConc can provide the reader with nominal forms of verbs and adjectives by employing its indexing line retrieval, word list, and subject word statistics services, all of which are based on the noun-identification phenomena in AntConc. Statistical analysis and formative evaluation can also be conducted based on AntConc of word frequency distribution, word density, word collocation features, word combination collocation intensity, and speech mistakes.

Nevertheless, it is uncommon to investigate the relationship between nominalization frequency and students' difficulties in reading and comprehending English texts by applying AntConc as the tool. The attempted study would generate unmatched benefits over conventional approaches for the analysis and improvement of textbook compilation materials given the quantitative approach with the help of AntConc and qualitative discussion based on language typology. As it involves both qualitative and quantitative methods, the study would serve well in the academic context of English reading assessments, especially for L2 learners.

2. Concepts Explained and Literature Review

To further engage with the scholarly community, the following section will introduce some seminal concepts regarding the themed study and provide a critical overview of the current studies.

The study of English nominalization with the help of corpus has advanced quickly in recent years. A way of

[1] SFLEP stands for Shanghai Foreign Language Education Press, and the series of *College English* (third edition) was published in 2011. To Note that some studies have been conducted on the book; see for example Xin Yu and John Lowe (2007) initiated a study to evaluate the SFLEP college English oral test in China.

integrating textbook and corpus has been proposed by some researchers to assess how nominalization affects the level of difficulty of teaching materials, therefore helping teachers in the way of presenting necessary data while benefiting students by structured and gradual text reading instructions. Before diving into the literature, it is fundamental to go through some key concepts that need to be briefed about.

2.1 Definition of Key Terms

2.1.1 Nominalization

The act of changing a verb, adjective, adverb, or other lexical word into a noun—that is, a term that may be employed as a noun—is known as English nominalization. This can be done by either utilizing the word's form or by appending a certain suffix (such as -ment, -tion, -ity, -ness, etc.) to the end of the word. The original verb, adjective, or adverb may become more precise, emotive, and suitable for a certain situation via nominalization. Scholars' definitions of nominalization vary slightly from one another.

According to Haspelmath, M. (2002), nominalization is a morphological phenomenon that can be accomplished by employing the word's general form or by appending a particular suffix to the end of a verb, adjective, adverb, or other lexical term. When it comes to language use and structure, nominalization is crucial. O'Grady, W., Archibald, J., Aronoff, M., and Rees-Miller, J. (2010) took a step further to believe that nominalization is a linguistic phenomenon that converts lexical forms into nouns, this conversion can be accomplished by adding specific suffixes or using specific lexical forms. In a given context, nominalization can make the original verb, adjective, or adverb more specific, expressive, and appropriate for a particular context. This article, given the focus on the transfer nature in nominalization, would firmly follow the definition by scholars like O'Grady among others, taking the process as a conversion of lexical forms into nouns.

2.1.2 Classification

According to what is mentioned above, nominalization, as an important term in linguistics, is the process of using verbs, adjectives, and adverbs as nouns or nominalized phrases while modifying their word forms to some degree. Thus, given the method of lexical conversion, English nominalizations are presently primarily divided into three categories in academia: verbal, adjectival, and adverbial nominalizations (Blanco, 2023), which are addressed more specifically here.

(1) Verbal nominalization refers to a linguistic structure that converts a verb into a noun. By adding -ing or other frequent nominalizing suffixes like -tion, -ment, and -ity. Verbal nouns can be employed as subjects, objects, epithets, determiners, complements, and prepositional objects in sentences to represent the name of an action, state, thing, concept, or attribute. Take "happiness," "intelligence," "curiosity," and "strength," for example.

(2) The process of changing an adjective into a noun is known as adjectival nominalization. Usually, to change an adjective into a new noun phrase, the definite article "the" or other qualifiers should be added before the adjective. As an illustration, "the blind" stands for those who are blind, "the elderly" for those who are

aged, and so on. We can more succinctly and precisely characterize a class of people or objects by using adjectives. It is also possible to communicate abstract ideas like “the beautiful” and “the absurd” by using adjectives instead of nouns.

(3) The term “adverbial nominalization” describes the grammatical structure that changes an adverb into a noun, much as how an adjective changes into a noun. The definite article “the” or other qualifiers are typically added in front of the adverb to accomplish this as well. For instance, “the fast” denotes those that have “speed”, whereas “the late” refers to those who have passed away. Additionally, adverbial nominalization can also describe a certain circumstance, such as “the impossible” and “the unexpected”.

2.1.3 Function

Stated differently, nominalization, also known as normalization, is the process of creating a noun or noun phrase outside of the context in which they are typically created. Nominalization is defined as “using a word which is not a noun (that is, a verb, an adjective, or an adverb) as a noun or as the head of a noun phrase with or without any morphological transformation” in specific definitions. Furthermore, nominalization is also understood by Transformational Generative grammarians to mean the derivation of a noun phrase from an underlying clause. An essential component of academic writing is nominalization. It gives texts more diversity, impartiality, and an impersonal tone. Because nominalization enables authors to fit a lot of information into sentences, it also improves writing’s reader-friendliness and conciseness. In the English language, verbal nominalization is the most prevalent. The nominalized statements are more formal and succinct than the original sentences, even if the two sentences convey comparable meanings. The formal level is also important in academic English writing. (Li, 2021)

With all of this information in mind, it is critical to define the role of nominalization as a general word in linguistic studies. This means that a shared concept of nominalization in the world can be justified with reference to the fact that other Indo-European languages, besides the English language, take on specific inflectional patterns for designating their diverse parts of speech. It makes sense that a claim like nominalization is a natural part of a language is difficult to refute, even when certain nominalizing occurrences are more evident than others depending on the linguistic context in which they occur.

2.2 Previous Studies on Nominalization in China and Abroad

The current studies that are related to the article’s theme, i. e. , nominalization, reading studies, and corpus-based L2 learning and teaching, should be categorized into three sections, respectively.

2.2.1 Types and Phenomena of Nominalizations

Domestically, Feng Shujian (1992) first systemically introduced nominal phrases’ characteristics in English and ways to translate them into Chinese. For defining nominalization, Xu (2009) believed that there are various types of nouns. In addition to the noun of verbs (i. e. , process), the noun of small sentences and the noun of environmental components belong to different types of nouns, and their internal grammar institutions have their

own characteristics.

Some Chinese scholars have provided answers to “Is it true that English quantitative adjectives can’t be nominalized?” (Zhong Shuneng & Fu Shuya, 2016). Other Chinese scholars are interested in the exploration of the nominalization classification of English verbs and their Patterns and in the phenomenon of phrasal verb Nominalization in Modern English (Liu Luying, 2009; Shi Penghai & Li Tong, 2002).

A special scholarly phenomenon in China is to deal with the Chinese translation of English nominal patterns in texts for special purposes. Fu Jingming (2007), Zhao Junfeng (2006), and Lv Yulin (2019) have approached the translation of nominal patterns in legal texts; Li Bingwu and Yan Jingmin (2002) discussed the translation of nominalization in science-related texts; Fan Xiaohui (2005) talked about the translation of the nominal patterns in medical thesis.

Regarding nominalization, foreign academics have differing opinions. In his introduction to the lexical structure and etymology of English, Bauer (1993) delves deeper into the construction of English words and compound words, analyzing a variety of word formation techniques such as prefixes, suffixes, transformations, and more. Plag (2003) examines a variety of English noun constructions, including compound terms, derivatives, and transformations. It also compares several forms of dynamic linguistic transformations and examines the prevalence and usage guidelines of these constructs. In addition to providing a thorough analysis of the development and changes made to terms in modern English, Stockwell & Minkova (2001) also look at the origins of variety and change in the English lexicon.

Given the basic studies with regard to what is nominalization, how to classify it, how to translate it, and how to understand the special linguistic case, the current studies are still in deficiency applying the understanding of nominalization and transfer within such a process from semantic, syntactic, functional, and even cognitive perspectives into studies like L2 language learning, text difficulties, reading assessment, etc. The current foreign language teaching and learning studies alike are, too, less systemic in introducing the framework and theories that concern nominalization (Zhai Ping, 2009; Cai Jigang, 2018; Li Wen & Guo Jianhui 2020).

2.2.2 The Association Between Nominalization and Reading Difficulty

According to Miao (2010: 10), the noun shift possesses the qualities of density, abstraction, and numerous meanings. To be particular, the words altered by the noun meanings are generally abstract nouns, which drastically increases the information density of the conversation, and the semantic relations frequently look unclear, which makes people difficult to comprehend and even misinterpret. Simultaneously, the conversation will become abstract and difficult due to the overuse of abstract terms. “The nominalized form’s semantic expression is vague and not nearly as vivid as the corresponding verbs or adjectives, despite its accuracy in grammar and practical matters.”

In the same paper, Miao (2010) also noted that some people find it challenging to read articles because of the particular nouns in English in four areas: vocabulary density, vocabulary structure, semantic expression, and contextual variables. According to Wang Yunsu (1999), “the reader’s familiarity with the language form of the

reading material and the information it contains directly correlates with the difficulty of reading.” The information substance and the linguistic form of the reading materials have a complementary connection. Effective reading relies on the reader’s ability to identify the least obvious but effective language cues and make accurate predictions right away, rather than having a thorough comprehension of the language component as a whole.

The reader’s degree of familiarity with the language of the reading material (known as Familiarity with the Language) and the information contained therein (known as Familiarity with the Information) are inversely correlated with the reading difficulty (known as Degree of Difficulty). Reading becomes easier the more accustomed the reader is to the material and language. Xie & Liu (2020) used a corpus to analyze the nominalization utilizing conditions in Chinese master’s theses and discovered that there is little nominalization in thesis theses and that greater nominalization results in more linguistic complexity in the essay. According to Ma, Zhu, and Chu (2021), nominalization is more frequently used in English academic discourse than it is in Chinese literary terminology writings.

Based on the aforementioned, it is theoretically viable to conduct nominalization-related studies on the reader’s comprehension and reading efficacy. Therefore, future studies could go to at least two pathways. First, researchers would identify, with enough corpus, the correlation between the frequency of nominalized words and the readers’ response, understanding, and feeling for the text, by which the study would benefit the verification of the linguistic role of nominalization. Second, the scholarly context of teaching studies would find out the relationship between the nominalization of the textbook and the efficacy of teaching, especially the L2 teaching and learning. This would be conducive to the course design and text selection, and the related data would be a reference to other studies related to writing and oral English tests and training. The point is the study of nominalization, regardless of linguistics or language teaching and learning, is in dire need of corpus as a proof since the related qualitative studies are sufficient enough for more empirical studies in order to verify and falsify. Therefore, the following is a review of the current corpus-based studies on the phenomenon of nominalization.

2.2.3 The Corpus-based Nominalization Studies

The type of study was mainly found abroad. Numerous scholars used big data as a language pool to figure out the evolvement and development of English nominalizations. Grefenstette and Teufel (1995) along with Cowie (1999) and Fonteyn (2019) have used corpus as a means to identify the features of derived nominalizations in the history of the English world, and in such a process, there was also devised related approaches for the automatic identification for nominalized words.

With the support of the above studies in terms of data and technology, there emerged further studies on forms and functions of nominalization in texts for special purposes. Shehzad (2019) probed into nominalization in PhD dissertations in English Studies in Pakistan with the help of a self-built corpus. Su and Qiu (2023) adopted a comparative perspective to look into the nominalization in English abstracts of Sci-Tech Papers by Chinese and Native English Authors. Liu, Fang, and Wei (2017) used corpus to find out the syntactic patterns of nominalization across Chinese and British media English. If translations can also be accounted for as a special

type of text, recently there also appeared corpus-based studies into nominalizations in translated texts. Hou (2013, 2015) researched, with the help of corpus, the nominalization in the English translations of Chinese prose, and used the frequency of nominalized words as an indicator of the translator's style. Liu, Liu, and Cheung (2023) took a more general step in comparing the nominalization in both translated and non-translated texts in order to find out the difference.

So far, it is necessary to make a conclusion that nominalization can be used as a pattern to investigate the textual features or as an indicator of the textual difference. Could we use the special linguistic phenomenon as an aspect of language teaching and learning? Given the relevant studies, we can make a claim that not only can the academia go for studies involving nominalization in teaching but can also take corpus as a tool to make rather objective observations. Wang (2012), Kim (2014), Zhang (2018), Lambert (2022), Bychkovska and Lee (2023) mainly focused on the L2 writing skills and how nominalization among other language features plays a role in defining competent writers in different levels such as pre-university and undergraduate, or for different purposes like academic or collaborative writing.

Although all the previously mentioned studies are credited for their systemic investigation into and application of the linguistic phenomenon, i. e. nominalization, they still lack combined ways of studies engaging both corpus-based searches for nominalized patterns and reading complexity. Therefore, the article would carry the related studies a step forward by using a statistical method to explore the correlation between nominalization and reading complexity, and the content is mainly of the texts for L2 teaching purposes.

3. Research Design

One approach to addressing the issue of teaching foreign languages is the computer-driven language research technique. Software for processing corpora, like AntConc, can offer context and frequency of word information, as well as validate grammatical rules, word usage, and linguistic phenomena like word collocation. These tools can be useful in resolving issues that arise during the teaching process. The article is based on the statistical tool as well as some language theories for further discussion. In a word, the study is a quantitative plus qualitative study for the L2 teaching reading materials.

3.1 Research Questions

Compiling and indexing word lists and obtaining high-frequency verb and adjective vocabulary are among AntConc's primary uses. This work attempts to answer the following three questions using AntConc for language research and experiment to gather and collect high-frequency words and keywords (such as verb nouns, adjective nouns, etc.):

(1) What is the distribution condition of the nominalization in the first five units of A and B texts in the third edition of the *College English Intense Reading Textbook* by the Shanghai Foreign Language Press?

(2) Is there a relationship between the lexical density and the nominalization phenomenon's frequency?

(3) How does the nominalization relate to the challenges associated with reading English texts comprehension?

3.2 Research Materials

The focus of the study is the nominalization phenomenon seen in texts A and B of the first five units of Book I and Book V of the third edition of SFLEP's intense reading *College English* textbook. This study focuses on the nominalization phenomena among the 5, 832 and 6, 429 words, respectively, in the self-compiled textbook corpus.

3.3 Research Tool

As the primary instrument of this investigation, we utilize three functions of the AntConc, i. e. contextual keyword (KWIC) distribution statistical analysis, word frequency distribution, and word density statistical analysis. AntConc, a corpus software, was selected as the primary instrument for language analysis because of its robust corpus analysis feature, which enables researchers to precisely and impartially count, examine, and validate the corpus from many aspects for the purpose of data retrieval and processing. Such an approach bears much significance in teaching and learning foreign languages. Specifically, it can make full use of some of the functions of AntConc analysis, such as index line retrieval, word list, subject statistics, etc., to show the nominal verbs, adjectives, and other noun phenomena. Through the word frequency distribution, vocabulary density, word collocation characteristics, word combination strength, and speech errors of statistical analysis and formative assessment, the software can be used to conduct a comprehensive and objective assessment of texts in case. Such an analysis method can provide more objective and accurate language analysis results, which can help students to better master the language knowledge and skills and help to solve the problems in language teaching. Besides, the software can also benefit the study in a way of visualizing the textual characteristics and figuring out the reading complexity in a quick and apparent manner.

3.4 Data Collection and Analysis

The following process is used to gather and analyze the study data:

First, using the texts of A and B from the first five units of Book 1 and Book 5 of *College English Intense Reading Textbook* by SFLEP, the author created a corpus.

Second, the author attempted to use the symbol <N> to identify the nominalized forms in the corpus; verbal phrases were labeled as <VN>, adjectival nominalization as <AdjN>, and adverbial nominalizations as <AdvN>.

Third, the author used the program AntConC version 4. 2. 0 to determine the total number of nominalizations, including VN, AdjN, and AdvN.

Ultimately, the author contrasted and examined the current nominalization study's data to provide a list of

conclusions.

3.4.1 Methods and Procedures of Retrieval

The noun list of associated verbs and adjectives is extracted using AntConc in units A and B of the first five units of the College English (third edition) (Dong Yafen, Zhai Xiangjun, 2011, College English 5). It is possible to obtain the reading material's maximum frequency as seen below (Figure 1).

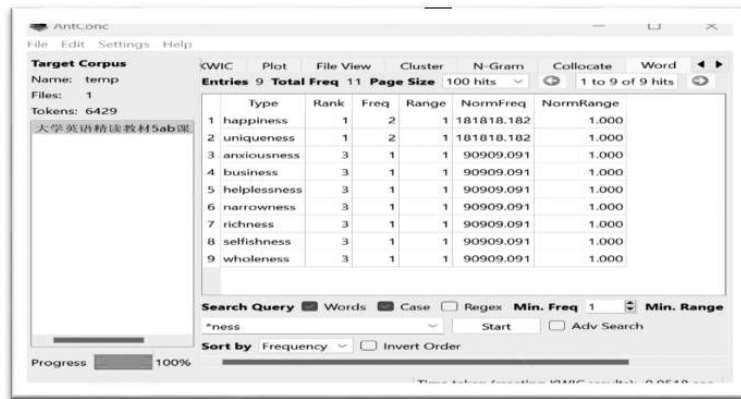


Figure 1. The High-Frequency Vocabulary Wordlist of AntConc

One special feature of the corpus software AntConc is the index marker map (concordance plot) (Figure 2), which uses visual stripes to display the distribution of a given word over the whole article or book.

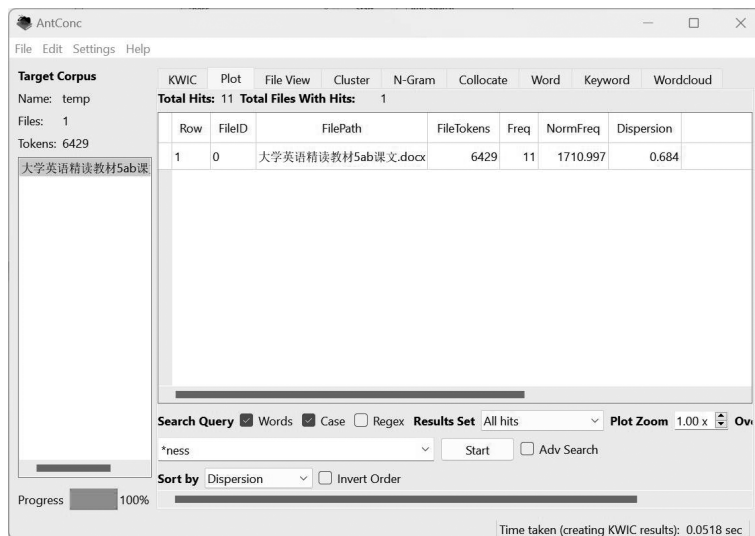


Figure 2. The Concordance of AntConc

One of the fundamental principles of writing materials is spaced “retrieval” which means that textbook editors in the design content need to intentionally make some target language vocabulary high frequency, let readers or learners constantly meet these expressions. This may have a wake effect on learners, and it is why the function of the index mark figure for textbook evaluation is important. It helps pupils acquire new words and explain what they have read in an article. It can increase the textbook reading difficulty for teachers as well as the systematic complexity of textbook reading for textbook readers.



3.4.2 Statistical Analysis of the Distribution of Keywords in Context (KWIC)

Using the various noun symbol — suffix to identify noun words, for example, we enter in the lower left box, the lower right corner to order according to the word frequency, and AntConc will automatically identify for us. Depending on the word frequency from high to low will contain term words, we choose the college English intensive reading textbook 1 in the first five units of A, B text as corpus A. It is important to note that AntConc cannot automatically identify word components; as a result, words that are not nouns must be manually excluded before statistics can be generated. Because many terms are inconsistent with our aim, like mental, they may surface via direct screening with noun letter combinations, this can have negative consequences. As the manual screening process is excessively labor-intensive and inefficient, we choose to use the input * —which appears at the conclusion of the terms—as the search object rather than the phrases themselves (Figure 3).

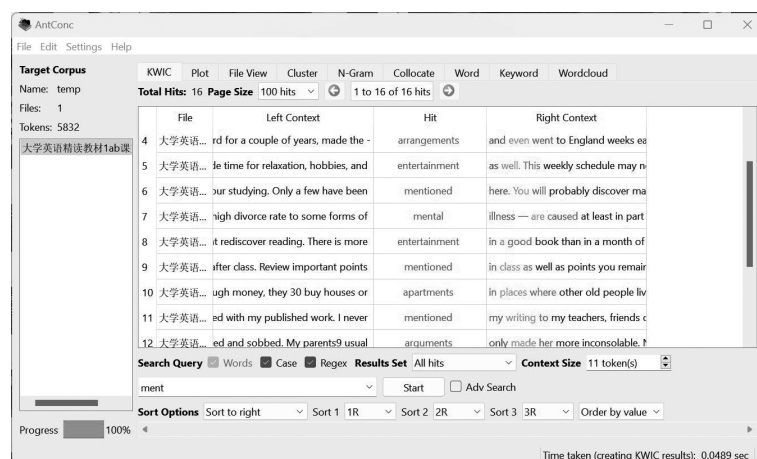


Figure 3. Statistics of -ment Words

The text box displays the suffix density, but some words are more formed like verbs than nouns. For example, some words ending with -ing are verbs rather than nouns, and some words ending in -er are prepositions rather than nouns. As a result, we must manually filter the words from the statistical range rather than using the word frequency displayed in the text box as the frequency of nouns, in case the experimental data is skewed. For instance, the article searches the whole U1A text for nouns with the age prefix, excluding six terms with similar forms, and only select verbs whose noun wording comes from it.

Double-clicking the blue portion allows you to view the index entry context in the original text as shown in the following figure (Figure 5.) via the import college English intensive reading textbook first U1A text txt file. The author clicks the keyword context (KWIC) button, checks words, enters * tion, and clicks on the start retrieval text with the end of the word number, the index would be shown as 6 (Figure 4).

Then, the author checks the index bar number after entering -term, -age, -ing, -ance, -er / or, -ence, -y, -th, -ure, -ant, and -ness. For the upper-class teaching level, the author replaces the retrieval materials with those from the *College English Intense Reading Textbook 5* and tests the noun keyword (the verb is displayed in Figures 6–7, Figure 8) throughout the entire text in the same manner. The suffix for retrieval is mainly as

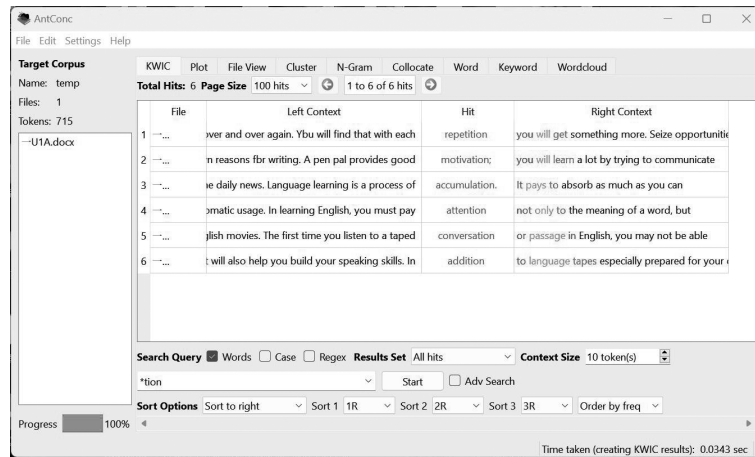


Figure 4. Keywords Statistics Ending with -tion

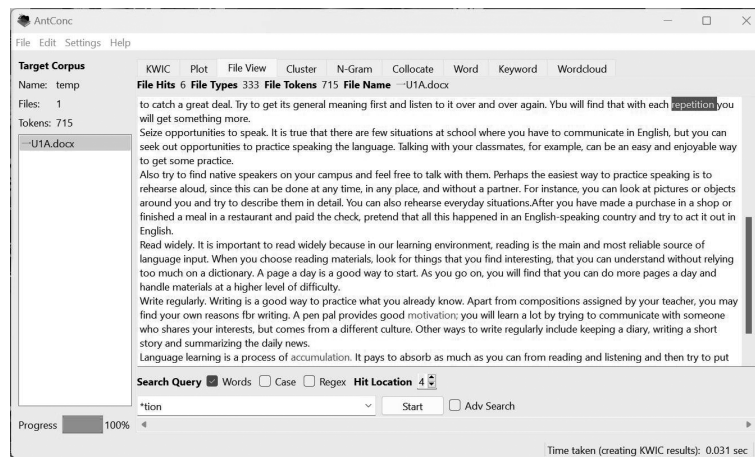


Figure 5. The Context of the Nominalized Verb Index in the Original Texts

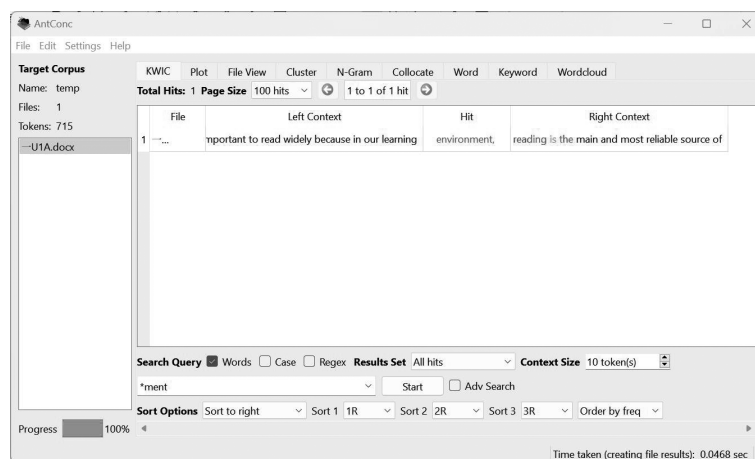


Figure 6. Nouns with -ment Suffix in the Entire Text of U1A

follows: -ence, -e, -y, -th, -ure, -ant, -ness, -term, -age, -ing, -sing, -on, -ance, -er.

3.4.3 Nominalized Word Frequency

AntConc English word frequency and theme word are listed in the target file. Two distinct corporawere

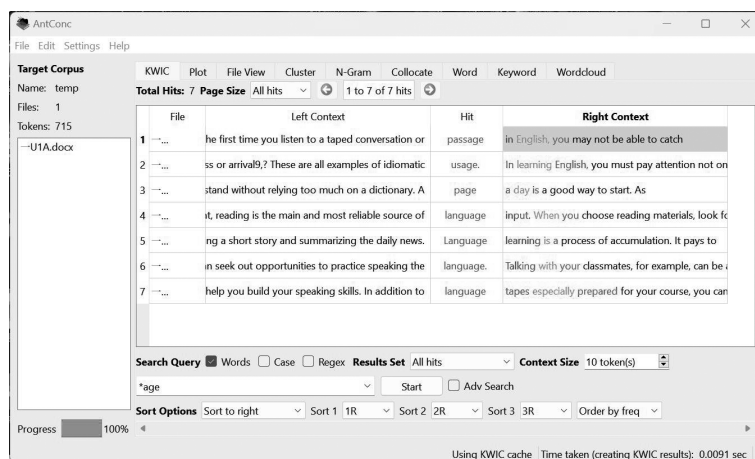


Figure 7. Nouns With -age Suffix in the U1A Text

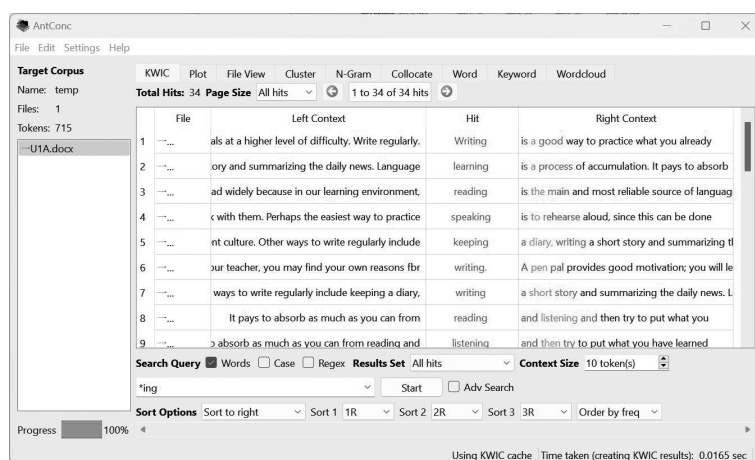


Figure 8. Words with -ing in the Full Text of U1A

gathered from the A and B texts of the first five units of *College English Intensive Reading Textbook 1* and the A and B texts from the first five units of *College English Intensive Reading Textbook 5*. To obtain the target word's frequency in the text, the author imports the target file, clicks the word to display the tab directly, and then clicks again. To see the target words' distribution in the text visually, plotting can be switched to visual bars. We choose key terms such as *ation, *ism, *ment, *ness, *ity, *ship, *hood, *ing, *ance/ence, etc. to search for all kinds of nominalization.

This is so because the suffix -ing denotes verb nominalization, -ation denotes the process of turning a verb or adjective into a noun, -ment denotes the process of turning a verb or adjective into a noun, -ness denotes the process of turning an adjective or adverb into a noun, and -ity denotes the process of turning an adjective into a noun. A noun can be transformed into a status or position by using the suffix -ship, a state by using the suffix -hood, a school or idea by using the suffix -ism, and so on. -ance/ence denotes how a word might become a state or characteristic. In the first text (the A text and the B text in the first five texts. The intuitive distribution figure is shown in the following figure (Figure 9.).

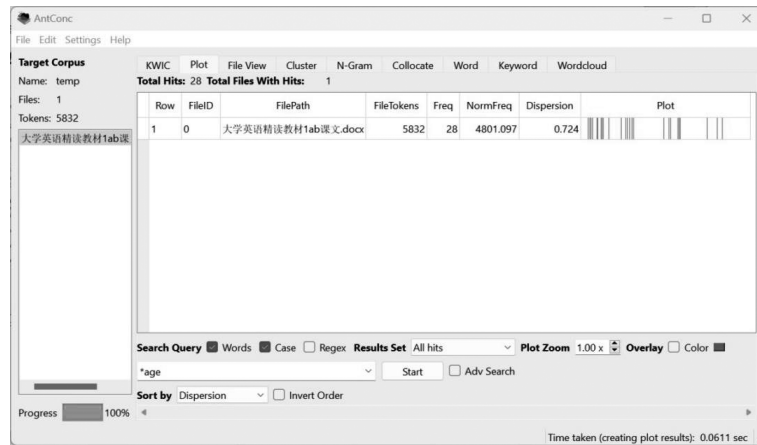


Figure 9. Frequency of Words Containing -age Suffix in the A and B Texts of U1–U5

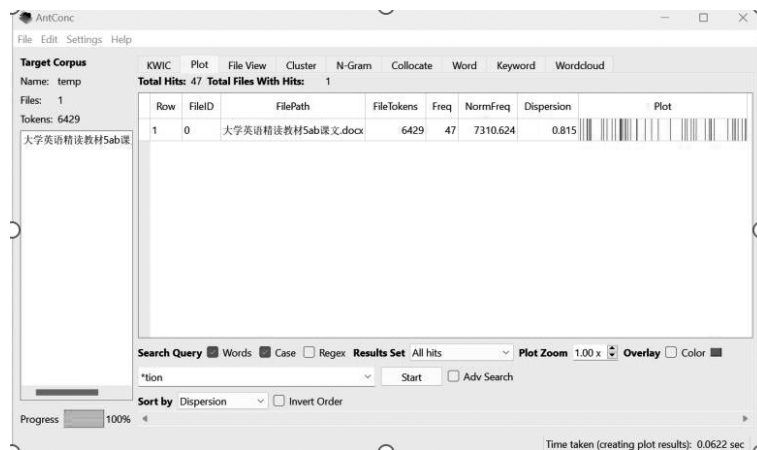


Figure 10. The Word Frequency of A and B texts in U1–U5

Similarly, the noun phenomenon of certain verbs is retrieved in Corpus 2 (the first five units of the *College English Intense Reading Textbook*). Figure 10 below displays the distribution.

Mathematical analysis and visual inspection of the image lead to the conclusion that corpus 2 has a higher overall frequency and is more densely packed with verb nouns than corpus 1.

3.4.4 Statistical Analysis of Vocabulary Density

The approach to applying the functions of AntConc corpus analysis tools given by Anthony (2014), includes importing corpus, using various analysis tools in the software, and using it for vocabulary analysis. To determine the word density in the first five units of the *College English* intensive reading textbook (third edition), the author chooses File> Open File (s) from the menu of AntConc. The text files for the first five units of the third edition of the *College English* intensive reading textbook are accessible from the local drive. Then the author chooses “Corpus> Word List” from the menu. Later, the author verifies the boxes for Words and Frequency under the Output Options, sorts the data in decreasing order. Then the author should check the Include All and Exclude Numbers boxes in the Words tab, and in the Word List tab, the author sets the Word List Output File to the file name and saves the result. Finally, the Start button should be clicked to start generating the vocabulary

list. Then the total number of words in the file should be counted. The author should select the “Corpus> File Information” option, set the “File Information Output File” to the name and path of saving the result, click the “Start” button to start generating the file information, and finally count the vocabulary density. The vocabulary list file should be opened, and the author should divide the number of words by the total number of words in the file to get the vocabulary density.

3.4.5 Analysis of the Nominalized Word Frequency

As per the previously mentioned procedures, within the *College English* textbook U1A text, for instance, the index bar number for -term, -age, -ance, -ing, -er, or, -ence, -y, -th, -ure, -ant, and -ness, are respectively 1, 7, 34, 0, 2, 21, 1, 57, 6, 1, 3, 0.

Here, 32 nouns that finish in -ing remain after non-noun words are manually eliminated. Simultaneously, the term nouns-tion, i. e., -term, -on, -ance, -er, -age, -ing, -sing, -ence, -e, -y, -th, -ure, -ant, -ness, which are used in the *College English Intense Reading Textbook 5* for the upper-class teaching stage have the index numbers 10, 1, 15, 21, 1, 1, 32, 35, 1, 84, 11, 8, 3, 2. Thus it can be seen that in the more academic, more objective, and rigorous *College English* intensive reading textbook like texts of U3A, there are more verb-noun phenomena, compared with the more daily and explanatory *College English* intensive reading textbook 1U1A. Using the first five units of texts A and B from the third edition of the *College English Intense Reading Textbook*, we can use this procedure to count the total number of instances of all kinds of nominalization. We can then repeat the process with book five.

4. Results and Discussion

4.1 Results

The results are largely the answers to the research questions proposed. The target of the data retrieval results is to answer the questions as to the distribution of the nominalized words in the texts A and B of *College English* (third edition), the relationship between the lexical density and the nominalization phenomenon based on the finding, and the relationship between the nominalization and reading comprehension.

4.1.1 Distribution the Verbal Nominalization

As previously mentioned, the data presented by AntConc indicates that in the first five units of texts A and B of the third edition of the college English intensive reading textbook, there are only 189 instances of all types of nominalizations overall, compared to 221 in book five. Inbook one, there is less verbal nominalization in distribution frequency than in book five.

4.1.2 Correlation between Nominalization Frequency and Lexical Density

The statistical vocabulary density in the five units of *College English Intense Reading Textbooks A and B* is as follows. Based on the aforementioned steps it could be calculated that the number of files total: 21523,

variations in vocabulary: 3586, and vocabulary density: 16.6% in the first five units (A and B) of the third edition, which means there are 16.6 distinct words for every 100 words.

Comparable statistical results may be found in Book 5, Units A and B text files of the *College English* intensive reading statistics: the overall word count: 19822; 3,354 is the distinct vocabulary count. The vocabulary density is 16.9% in the third edition's first five units (A and B), which means 16.9 unique words for every 100 words. Based on the statistical findings, the vocabulary density of the first five units A and B of the textbook's book (16.9%) is marginally higher than that of the first five units A and B of the first volume (16.6%), indicating a more complex and challenging-to-understand vocabulary use in the text of volume five, given the cognitive load theory (Sweller, 1988).

The reader will have a greater cognitive burden as a result of needing more time to comprehend each word's meaning and function in a book with a higher vocabulary density. Texts having a high word density will be more challenging, especially for English learners, because of their limited vocabulary knowledge. As a result, the corpus survey discovered that, in the *College English* intensive reading materials, there was a positive correlation between the frequency of nominalized verbs in the texts and the word density. This proves that the nominal phenomenon in English academic papers is a common linguistic fact and that readers would receive more intensive information from the nominalized materials (Zhang, 2017).

Furthermore, as per Guo Yun's questionnaire survey (2017, 106 - 107), the English science and technology literature likewise exhibits the same correlation pattern. However, it is clear that our process with the help of AntConc, which screens nominalized verbs and adjectives, differs from other approaches in that it is comparatively more precise, thorough, and objective.

4.1.3 Correlation between Statistics and Reading Difficulty

The statistical findings indicate that *College English* 1 has a lower frequency of nominalization and a lower vocabulary density than that of 5, because the texts in the latter are more objectively applied, more specialized than general, and more academic than those in Book 1.

According to cognitive load theory, more information density makes it easier for the reader to become burdened with comprehension, which raises the possibility of producing ambiguity or even misunderstanding. Therefore, from the standpoint of the students, the anticipation sample and the A and B texts in the first five units of the fifth volume of the *College English* intensive reading textbook have a higher difficulty factor.

4.2 Discussions

The discussions are further explanations based on the data findings with the help of AntConc, and the aims in the following section are to figure out why there was the differentiation and to give an account for the role of nouns in writing and reading. This part would further answer the research questions in 3.1.

4.2.1 Results Explanation Based on AntConc

From the specific content of the two research subjects, *College English Intensive Reading Textbook* book one

selected a more popular content; most of the texts are written from first-person perspectives; the U1A text is an instruction manual on English learning strategies, and the U1B text is a manual on optimizing study habits. Both texts must be easily understood to fulfill their intended purpose and be effective in guiding English language learners. Therefore, it is suggested to use fewer verb nouns in the writing to lessen the cognitive burden of the reader. This would effectively reduce vocabulary density and decrease the likelihood that readers who are not native English speakers would find the article difficult to grasp.

However, the materials in book five are chosen in a more scholarly manner. A biography, which is informative and narratively rigorous with a high frequency of verb to noun and higher information density—meaning that the article’s context is close and interlocking and its logic is strict and rigorous—is found in the majority of third-person descriptive or narrative articles, which are objective. This includes U2A texts. Many explanations may be used to prove the positive link between vocabulary density and the frequency of pronouncing English nouns from a linguistic standpoint. Among them, the hypothesis grounded on cognitive load theory maintains that a text with a high vocabulary density would have a higher cognitive burden due to the increased frequency of nouns.

As a result, reading these texts will become more challenging as the reader will have to expend more cognitive energy to comprehend the content. In addition, research based on corpus linguistics suggests that the frequency of utilizing nouns in high-density texts is connected to domain specialization. Nouns are typically employed more frequently in particular disciplines (such as science and technology, medicine, etc.) since these fields need the expression of more professional phrases and concepts nominally. When digesting texts in these disciplines, learners must be especially adept at these frequently used nominalized terms. In conclusion, a positive correlation was found between the frequency of English nominalization and the lexical density of texts, indicating that English nominalization is used more frequently in texts with a high lexical density. This should be given enough consideration in the learning and teaching of English as it is connected to the text’s complexity, cognitive load, and domain specialization.

4.2.2 *The Role of English Nouns*

Li Fengxian (2021) claims that even though the two phrases convey the same ideas, the noun-treated sentences are shorter and more formal than the original ones. The formal level is especially important in academic English writing. As is said, “written text with high vocabulary density is used more frequently because it is relatively less common in spoken language and more common in written language.”

In the first textbook

e. g. , 1: *Do not treat all new words in exactly the same way.* (Dong & Zhai, 2011, *College English* book one, p, 1)

In the fifth textbook

e. g. , 2: *It is because of this intimate connection between words and life itself that we have organized this small volume in a new way.* (Dong & Zhai, 2011, *College English* book five, p, 16)

The nominalized sentence, on the other hand, is more formal, scholarly, and authoritative than the original sentence. Thus, it is evident that there is a relationship between the sentence's nouns and formality. Because nominalization eliminates verb polysemy and adjective ambiguity, it can improve the precision and rigor of sentences. Furthermore, nominalization emphasizes the theme and point of view in academic publications by turning verbs or adjectives into themes. This helps to highlight some abstract and difficult notions.

Furthermore, the use of nouns in academic texts helps to improve the efficiency of information transmission, shorten complex sentence structures, make them easier to understand and remember, and keep readers from becoming bogged down by long, boring articles. All of these benefits help to reduce the barriers that readers face when reading academic texts. People are more likely to use simple verbs and adjectives to express ideas when speaking in oral English because these forms of expression are more important for achieving the goals of instant communication, communication, and emotion transmission. As a result, nouns are used less frequently.

4.2.3 Some Limitations of Using Nominalized Words

From the perspective of cognitive function, Miao (2010) made the following observation, "the use of nouns should be moderate; excessive noun use will make the language appear affective and even burdensome. If we put forward the limitation of English nouns, the problems and conditions that must be paid attention to in the use." Thus, some proposed guidelines must be followed while selecting and utilizing nominalized methods. Adjectives and non-noun verbs should never be unilaterally changed into nouns in language usage, let alone English nominalized meanings, which cannot be thought of as the ultimate key. Linguistic norms must be followed while using noun- words to avoid semantic or even actual ambiguity. To minimize students' comprehension hurdles, it is vital to avoid the exploitation of nouns.

Wang Yunsu (1980, 1999) noted that textbooks "unilaterally emphasize the importance of information, ignoring the difference and complementarity between the two in the relationship between language form and information content." The primary goal of designing a reading activity is information gathering. People are led to believe that mastering a language is equivalent to mastering its design and arrangement, and that mastering the information is equivalent to mastering the language. This not only obscures the distinction between the two but also ignores their complimentary connection. The result of the study shows that more nominalized words make it difficult for students to determine that they do not understand the grammatical structure of the sentence or the information content when they encounter reading disorder, which increases the difficulty of reading.

Additionally, verb nominalization can be decreased by speaking in clearer, shorter sentences, and by employing fewer passive assertions. For instance, "the team completed the project" might be shortened to "the team completed the project." However, Sanchez-Salcedo's (2019) study, which examined the impact of English noun phrases on Spanish students' reading comprehension, discovered that while unfamiliar noun phrases can impair reading comprehension, familiar noun phrases can enhance understanding. As a result, to aid students in understanding the noun phrases used in the text, a glossary or glossaries might be included in the textbook appendix.

Additionally, to lessen the strain on the reader, one should think about utilizing its representation or abbreviation. To put it briefly, eliminating superfluous noun phrases, giving definitions and explanations, and utilizing straightforward terminology can aid readers in understanding English texts especially difficult ones like literature.

5. Conclusion

Within the parameters of the text objects examined in this study, there is a positive correlation between the frequency of English nominalization and the text's lexical density. Also, the text becomes more challenging to read and comprehend at greater lexical densities.

The link of combining language study with information technology is corpus, which is now a very developed and often used language research tool. The extensive reading materials for *College English* (third edition) involve linguistics, teaching, and computer-related theory. Based on AntConc data analysis, noun phenomena that caused the reading difficulty of the text are discussed in quantitative evaluation, i. e., measurement, data statistics, and analysis functions. The technology is systematically discussed in the article, covering various noun phenomena and their relationship to specific application reading difficulties. The findings in the above article also allow textbook writers, users, and students to further develop and refer to.

References

- Anthony, L. (2014). *AntConc: A learner and researcher-friendly corpus analysis toolkit*. In Proceedings of the Fourth Workshop on NLP for Computer-Assisted Language Learning (pp. 1–8).
- Bauer, L. (1993). *English Word-Formation*. Cambridge: Cambridge University Press.
- Blanco, R. C. (2023). "Varying Degrees of Nominalization Found in Nominal-ing Forms and their Translation into Spanish." *Ideas*. Vol. 8. No. 8.
- Bouchard, D. (1995). *The Semantic of Syntax: A Nominal-ist Approach to Grammar*. Chicago: The University of Chicago Press.
- Fligelstone, S. (1993). Some Reflections on the Question of Teaching, from a corpus linguistics perspective. *ICAME Journal*. (17), 97–109.
- Haspelmath, M. (2002). *Understanding morphology*. Oxford: Oxford University Press.
- Ma, Y., Zhu, X., & Chu, M. (2021). A Corpus-Based Study of Nominalization in English and Chinese Academic Discourse. *Journal of English for Academic Purposes*, 51, 100992.
- Nation, P. (2013) Designing Reading Tasks to Maximise Vocabulary Learning. *Applied Research on English Language*, 3, 1–8.
- O'Grady, W., Archibald, J., Aronoff, M., & Rees-Miller, J. (2010). *Contemporary linguistics: An introduction*. Bedford/St. Martin's.
- Plag, I. (2003). *Word-Formation in English*. Cambridge University Press.
- Rupel, T. (2020). Nominalization in English for Academic Purposes: A Study of Cohesion and Readability. *Journal of English for Academic Purposes*, 43, 100842.

- Sánchez-Salcedo, M. P. (2019). The Effect of Noun Phrases on Comprehension of Authentic English Texts by Spanish Students of English. *Porta Linguarum*, 31, 139–152.
- Stockwell, R. P., & Minkova, D. (2001). *English Words: History and Structure*. Cambridge University Press.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257–285.
- Tavakoli, H., & Davoudi, M. (2018). The Effect of Written Corrective Feedback on the Use of Nominalization in EFL Writing: A Focus on the Development of Complexity, Accuracy, and Fluency. *Language Learning & Technology*, 22(1), 26–46.
- Xie, J., & Liu, Y. (2020). A Corpus-Based Study of Nominalization in Chinese Master's Theses. *English for Specific Purposes*, 60, 103–114.
- ZHANG Keren. (2021). A Corpus-based Study of Translator's Prosodic Styles: A Case Study of English Translation of Lisao by Yang Xianyi and Xu Yuancong. *Asia-Pacific Journal of Humanities and Social Sciences*, (1)1, 061–074.
- 董亚芬, 翟象俊:《大学英语精读 1》2011 年第三版。上海:上海外语教育出版社。
- 董亚芬, 翟象俊:《大学英语精读 5》2011 年第三版。上海:上海外语教育出版社。
- 高辛:《基于语料库的高中英语教材词汇教学研究》,2016 年教育硕士论文,杭州:浙江大学。
- 郭云:《名词化现象对英语科技文献阅读的影响》,《大众科技(学术版)》2017 年第 9 期,页 106–107。
- 李丰贤:《基于语料库工具 AntConc 对学生雅思写作词汇丰富性和搭配强度研究》,《国外英语考试教学与研究》2021 年第 3 期,页 71–91。
- 马俊:《基于语料库方法得〈新视野大学读写教程〉词汇评估》,2016 年教育硕士论文,沈阳:沈阳师范大学。
- 缪海涛:《英语名词化限制的认知功能研究》,《外语学刊》2010 年第 3 期,页 47–50。
- 王芸苏:《阅读难度与大学英语阅读教学》,《铁道师院学报》1999 年第 5 期,页 79–81。
- 徐玉臣:《名词化的生成机制、类型及功能的新视界》,《外语教学理论与实践》2009 年第 2 期,页 32–38。
- 杨卉卉, 陆军:《语料库技术在大学英语写作教学研究中的应用》,《扬州大学学报(高教研究版)》2016 年第 4 期,页 86–91。
- 张四红, 江莉, 程玲, 顾义俊:《基于 AntConc 的英语词汇教学研究——以〈大学英语(精读)〉课本文本为例》,《长江大学学报(社会科学版)》2011 年第 12 期,页 93–96。
- 张晓芳:《艺术史类英文学术论文中名词化现象的相关分析》,《中国高校美术学院学报》,2017 年第 1 期,页 100–103。
- 赵成浩:《英语词汇学习策略培训对大学生英语阅读能力的影响》,2012 年教育硕士论文,延吉:延边大学。

(Editors: JIANG Qing & KCJ)