

A Discussion on Chinese Non-English Majors Use Corpus to Improve Academic Words Use in Academic Writing

Chen Wang

School of Education, Huaibei Institute of Technology, Huaibei 235000, China

Abstract

As international academic exchanges become more frequent, the importance of academic English writing skills for Chinese university and postgraduate students engaging in effective academic communication is becoming increasingly significant. However, poor paper quality due to insufficient language expression abilities and academic vocabulary proficiency is currently widespread. In response to this situation, this study explores the application of corpus linguistics in enhancing the academic vocabulary levels of Chinese students who are not majoring in English. By applying corpus linguistics directly (data-driven learning (DDL)) and indirectly (such as academic word lists (AWL)), practical corpus-based teaching strategies are proposed for both teachers and students, with the aim of providing effective practical guidance to teachers and students and improving overall academic English writing skills.

Keywords

Corpus Linguistics; Academic Vocabulary; Academic Writing; DDL; AWL.

1. Introduction

Academic English writing skills are playing an increasingly crucial role in university and postgraduate English education in China. The possession of high-level academic writing skills is an indispensable key competence for researchers seeking to gain recognition from international peers or to promote the publication, dissemination, and sharing of academic achievements. As demonstrated in the survey conducted by Yuan Xiaoming, 52.86% of non-English major students have already published academic articles in fields such as chemistry, physics, and engineering[1]. This finding suggests that some non-English major students have already acquired fundamental international academic communication capabilities. However, the proportion of Chinese scholars who can publish high-quality papers in English in international journals is relatively low, and the influence and citation rates of these articles are also insufficient. Cargill et al. have indicated that this phenomenon is related to the authors' language expression and writing skills, particularly in the use of academic vocabulary[2].

Academic vocabulary is of pivotal importance in academic communication, not only in constructing textual structure but also in expressing viewpoints and describing research methods[3]. Nevertheless, students who are not majoring in English typically encounter difficulties in acquiring proficiency in the academic vocabulary of English. Furthermore, conventional English instruction for non-English majors has disproportionately concentrated on fundamental language skills such as listening, speaking, reading, writing and translation, while disregarding the substantial disparities between academic writing and general writing with regard to structure, citation and register. Consequently, there is an urgent necessity for further research and optimisation of the teaching model for academic English writing.

The present study aims to explore the potential of corpus linguistics methods to assist Chinese non-English major students in enhancing their academic vocabulary proficiency. The present paper will first define the concept of academic vocabulary and its importance in academic

writing. It will then explore the basic concepts and teaching applications of corpus linguistics, review relevant research findings, and propose corresponding solutions to the actual problems existing in current teaching.

2. Research Background

Following the official release of China's 'Guidelines for University English Teaching' in 2016, academic English has become a significant component of the English curriculum system in universities nationwide. Institutions have proactively investigated the establishment of academic English teaching systems. Nevertheless, there are several significant issues that persist in current academic English writing instruction practices.

Firstly, the majority of universities still adopt the conventional teacher-centred classroom lecture model for their academic English writing courses. Whilst these courses encompass writing techniques, vocabulary, grammar, and textbook content, students' learning methods primarily involve passively receiving knowledge, with a paucity of practical training in writing innovation. It is often the case that students will mechanically complete writing tasks without actively participating in or flexibly applying the knowledge they have learned during the learning process.

Secondly, academic English writing is a rigorous and formal scientific genre that requires teachers to have a thorough understanding of the cutting-edge literature in students' fields of study. It is recommended that teachers employ supplementary materials to facilitate students in independent analysis and practice, thereby cultivating their innovative abilities, critical thinking skills, and academic writing capabilities. However, the majority of contemporary teaching materials are found to be inadequate in their alignment with students' interests, abilities, and professional needs. It is often observed that students have a somewhat indistinct understanding of the application of academic vocabulary in professional academic writing, which consequently renders the materials ineffective in providing practical guidance.

It is therefore proposed that the implementation of corpus-based teaching strategies will facilitate enhanced comprehension and proficiency in academic vocabulary among students, thereby ensuring effective enhancement of their academic writing skills.

3. Literature Review

The purpose of this section is threefold: firstly, to provide a systematic review of the relevant research findings of corpus linguistics in the field of academic English writing instruction; secondly, to clarify the theoretical foundation and application pathways; and thirdly, to propose the theoretical framework of this study and highlight its research positioning and innovations.

3.1. Academic Vocabulary

3.1.1. Definition of Academic Vocabulary

Farrell defines academic vocabulary as high-frequency words frequently used in academic articles that do not depend on specific contexts[4]. Paquot further defines academic vocabulary as a set of terms that often appear in academic texts but are not considered to be core vocabulary[5]. Despite the differences in their respective definitions, both of these concepts underscore the significance of standardisation and the extensive utilisation of academic vocabulary in academic writing.

3.1.2. The Importance of Academic Vocabulary in Academic Writing

It is widely acknowledged that English academic writing is the predominant medium through which researchers engage in global academic discourse, thus serving as the universal language of international academic exchange. In this process, the employment of academic vocabulary is

a pivotal criterion for evaluating writing proficiency. Hou highlights that the majority of non-English major students receive inadequate and unsystematic training in academic vocabulary instruction[6]. As posited by Wu and Wang, although academic vocabulary may not directly convey core academic information, it is nevertheless indispensable in the expression of viewpoints, the description of research methods, and the presentation of research findings[7]. Consequently, targeted academic vocabulary instruction is imperative for enhancing the academic writing abilities of second language learners.

3.2. Corpus Linguistics and Its Educational Applications

3.2.1. Definition of Corpora and Corpus Linguistics

Biber and Conrad define a corpus as 'a large, representative electronic database of spoken and written texts'[8]. As Huston further observes, a corpus should comprise a wide range of sources unified by a shared structure, thereby demonstrating its ability to represent typical language usage patterns[9]. From these definitions, it is evident that several fundamental criteria must be met in constructing a corpus. Firstly, a corpus must consist of authentic texts that are stored electronically. Furthermore, it should contain a sufficient quantity of texts to adequately represent the diversity it aims to capture. Each text within the corpus should also be consistent in terms of size and format to reliably reflect the intended variety. Furthermore, texts should be drawn from multiple authors to prevent biases arising from an individual's distinctive style, thereby enhancing the representativeness and robustness of the data[10].

Corpus linguistics is the discipline of studying language through the systematic compilation of linguistic corpora of an appropriate size[11]. As corpora reflect language as it occurs naturally in authentic contexts, corpus linguistics provides an objective and comprehensive perspective on language use[9]. In other words, corpus linguistics can be understood as a methodology that utilises rich and informative linguistic data for systematic analysis, offering innovative approaches to investigating linguistic phenomena by dissecting and interpreting large-scale, authentic language data[12].

3.2.2. Application Pathways in Teaching

In the context of English as a Foreign Language (EFL), learners typically encounter limited exposure to authentic target-language input. Corpus linguistics addresses this limitation by enabling teachers to select vocabulary based on frequency of usage, thus focusing classroom instruction on the most critical language components[13]. Additionally, employing corpus data in language classrooms can provoke students' pragmatic awareness and engage them actively in genuine discourse[14]. Moreover, corpora offer relatively objective insights into actual language usage, shifting instructional reliance away from native speakers' subjective intuition and perceptions[15]. For instance, when teachers face difficulties explaining synonyms, their subtle differences, or their contextual usage, corpus data can provide empirical evidence drawn from authentic contexts. Consequently, EFL learners can more efficiently and effectively acquire language skills within a limited instructional timeframe.

There are primarily three direct applications of corpora in the classroom context: learner corpora, teaching-oriented corpora, and data-driven learning (DDL)[15]. Granger highlights the significant contribution of learner corpus research in bridging the previously isolated fields of corpus linguistics and foreign or second language acquisition[16]. Teaching-oriented corpora respond to substantial demand for tailored educational resources by assembling extensive language materials specifically suited for instructional purposes. DDL, meanwhile, involves learners actively exploring authentic corpus data to uncover and internalize linguistic features independently, significantly boosting their motivation and facilitating deeper language learning. It is especially effective for vocabulary instruction[15].

The most widely recognized indirect application of corpus linguistics in language teaching—particularly regarding academic vocabulary acquisition—is the development of vocabulary lists such as the Academic Word List (AWL). Below, a more detailed discussion of DDL and academic vocabulary lists is presented.

3.2.2.1 Direct Applications of Corpus Linguistics

DDL, a direct classroom application of corpus linguistics, refers specifically to "the direct use of corpora or corpus data in the classroom"[15]. Its core objective is to promote learners' active engagement with authentic language data, enabling them to observe, infer, and internalize linguistic patterns autonomously. For example, in designing vocabulary lessons, teachers can leverage corpus data to identify commonly occurring words and collocations, integrating these with realistic, context-rich examples. Moreover, corpus data can be selectively simplified or adapted to match the students' proficiency level, providing teachers with the flexibility to manage learning complexity effectively. Students themselves can also directly access corpus data, retrieving relevant information based on their personal learning goals and individual requirements, thereby enhancing both their motivation and overall language acquisition effectiveness.

3.2.2.2 Indirect Applications of Corpus Linguistics

Corpus linguistics also indirectly influences language instruction by providing standards and resources essential for academic English teaching and learning. Linguists have extensively analyzed and classified vocabulary within academic texts across diverse disciplines, leading to the development of specialized vocabulary lists intended to guide academic English instruction. Among these lists, Coxhead's AWL, initially developed in 1998 and revised in 2000, is particularly prominent and influential. The AWL, constructed according to rigorous corpus compilation criteria and comprising carefully selected word families derived from authentic academic English texts, significantly influences EAP curriculum design[5]. Not only does the AWL scientifically determine which vocabulary should be included in EAP materials, but it also provides essential guidance for sequencing vocabulary instruction and designing effective teaching materials. For instance, educators can use corpus-driven tools such as Vocabulary Tutor to efficiently analyze students' academic writing, promptly identifying non-academic lexical choices and enabling targeted feedback. This approach significantly enhances instructional quality and boosts the overall efficiency of both teaching and learning processes.

3.3. Empirical Research on Corpus-Assisted Academic Vocabulary Writing

As Schmitt noted, effective vocabulary learning methods include repetition and contextualization [18]. The DDL teaching method fulfils both of these conditions by employing corpus indexing to assist students in repeatedly encountering academic vocabulary and its real-world contexts, thereby facilitating a comprehensive understanding of the form, collocation, and semantic features of vocabulary. Hou Yuchen's analysis of a corpus of Chinese students' academic writing reveals that, while the coverage of AWL vocabulary met the required standards, there were issues with improper vocabulary usage [19]. This highlights the importance of contextual awareness and standardised expression in academic writing. Additionally, Zuo Xin identifies a noticeable colloquial tendency in Chinese students' English academic writing, further indicating the urgent need to strengthen students' understanding and proper use of academic style[19].

In summary, the application of corpus linguistics in academic writing instruction has achieved certain results, but there are still issues such as insufficient research on non-English major undergraduates, limited research on text structure and logical organisation, and room for improvement in aiding the widespread use of corpus tools in teaching. The present study aims to explore the effectiveness of corpus linguistics in academic writing instruction for non-

English major students through the DDL teaching method, using the AWL as a basic framework, and to provide practical teaching solutions.

4. Teaching Plan

The present study proposes to adopt the AWL framework and DDL teaching model proposed by Coxhead to explore how to enhance students' academic vocabulary application abilities through corpus linguistics. It is proposed that specific suggestions and measures will be put forward from both the teacher and student perspectives, to achieve better results in actual teaching.

4.1. To Teachers

Many words are provided in the textbook, but they are mostly in general English rather than academic vocabulary relevant to academic writing, with a limited number of examples of vocabulary-related sentences that are vague and contextually disconnected. To improve students' use of academic vocabulary in academic writing within the limited classroom time available, it is necessary to sift through the content. Firstly, teachers can use Lextutor, which is a corpus tool, to determine whether the vocabulary is academic and its frequency level before the lesson, to determine the objectives and the sequence of vocabulary teaching in the lesson. Teachers can choose the "CLASSIC" option, enter the textbook target vocabulary into the text box under "Input mode A", and based on the results, prioritise and focus on vocabulary with high-frequency levels and reduce or remove non-academic vocabulary depending on class time (See Appendix 1 for details of the steps taken in this activity and corpus analysis.). This activity is beneficial to classroom efficiency, but it can also increase teachers' workload and add to their teaching load. If done by a single teacher, it may be difficult to achieve in the short term.

There are differences in English language levels between students, and non-English majors have different areas of specialisation and therefore may have different requirements for academic vocabulary. So secondly, teachers can lead and demonstrate to students how to build their corpus after the class. Teachers can record videos to show students how to use Sketch Engine to build their corpora, and an online class group can be set up to discuss and solve technical problems in corpus-based learning through teacher-student and student-student interaction online after class (See Appendix 2 for details of the steps taken in this activity.).

4.2. To Students

Firstly, students can use English textbooks and classroom lecture notes, lectures published in English journals of their majors as material to build their corpus. This activity is conducive to increasing students' motivation and making them the subjects of learning. However, students' limited access to academic resources and their low ability to discern the quality of resources may be misleading to them. (See Appendix 2 for details of the steps taken in this activity.).

Secondly, students can conduct academic writing exercises in English followed by searches through the corpus by themselves to improve and enhance the use of academic words. After students have completed the English academic writing exercise, they can use Lextutor to search the text. Students can choose the "CLASSIC" option, copy and paste the text they write into the text box under "Input mode A", and check which words are academic or if there is any non-academic vocabulary. If there are high-frequency academic words, students can incorporate the sentence in which the word is found into their corpus (See Appendix 2). If non-academic vocabulary exists, students can use the Sketch Engine to find close synonyms for substitutions. (See Appendix 1 for details of the steps taken in this activity and corpus analysis.)

5. Conclusion

In summary, this study uses corpus linguistics theory and methodology to thoroughly explore ways to improve the ability of Chinese college students majoring in non-English subjects to apply academic English vocabulary. The study demonstrates that mastering academic vocabulary is crucial for non-English majors to enhance their academic writing, yet the traditional teaching approach struggles to address students' individualised and specialised needs. For this reason, this study proposes that students' vocabulary use in authentic academic contexts can be effectively improved through the integrated application of the corpus-driven learning (DDL) model and academic vocabulary lists (AWL). In practice, teachers should use corpus tools such as Lextutor to identify and prioritise high-frequency academic vocabulary for teaching, and clarify the teaching objectives and sequence. Additionally, teachers should encourage students to compile personalised corpora related to their majors to enhance the autonomy and relevance of their learning. Students, on the other hand, should actively use corpus tools to analyse their own vocabulary use in writing and improve the accuracy and standardisation of their academic vocabulary through repeated practice and contextual understanding.

One limitation of this study is that it has not yet made a detailed distinction between non-English majors in different professional fields. Future research could expand the scope of the research subjects, increase the diversity of professions and disciplines, and incorporate more concrete case studies and evaluations of effectiveness in order to further improve and optimise the corpus-based teaching method for academic English writing.

References

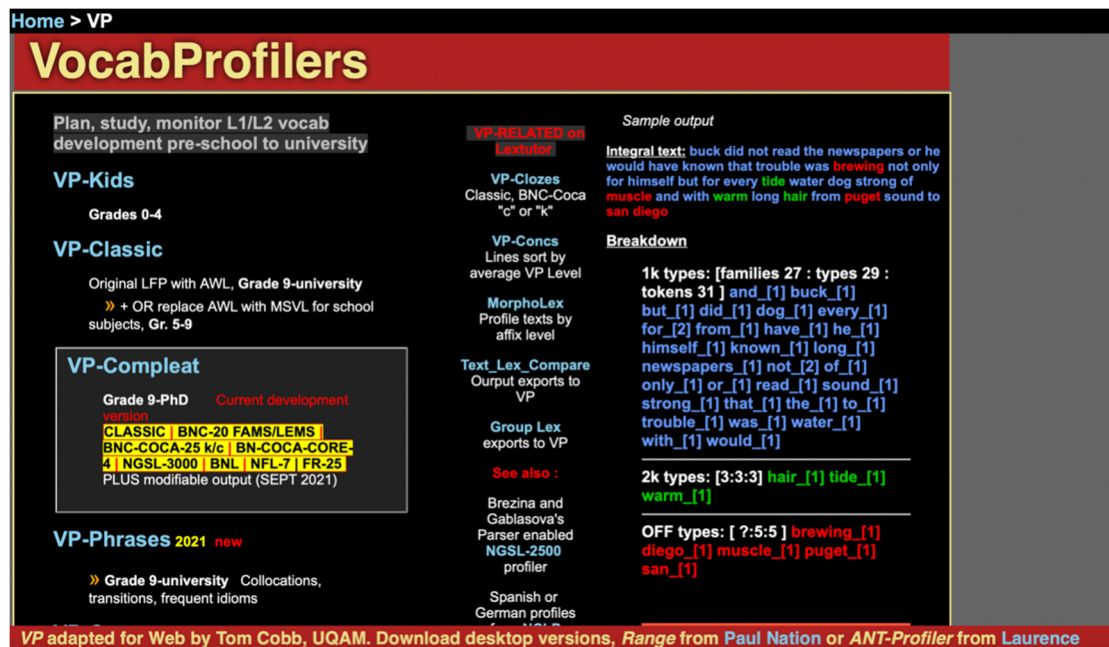
- [1] Yuan, B. Q., Wang, C. Y., Hu, X. and Feng, L. Q: How much do China's enrolled graduate students contribute to the publication of international high-level academic papers? --An Empirical Analysis Based on ES I Hot Papers (2011~2012), Degree and Postgraduate Education, (2014) No.2, p.57-61.
- [2] Cargill, M., O'Connor, P. and Li, Y. Y.: Educating Chinese Scientists to Write for International Journal: Addressing the Divide between Science and Technology Education and English Language Teaching, English for Specific Purposes, Vol. 31 (2012) No.1, p.60-69.
- [3] Han, L. N. and Liu, J.: An Analysis of Perspectives on Foreign Academic English Vocabulary Studies, Journal of the Party School of CPC Zhengzhou Municipal Committee, (2014) No.5, p.106-108.
- [4] Farrell, P.: *Vocabulary in ESP: A Lexical of the English of Electronics and a Study of Semi-technical Vocabulary* (Dublin: Trinity College 1990).
- [5] Paquot, M.: *Academic Vocabulary in Learner Writing: From Extraction to Analysis* (Bloomsbury Publishing 2014).
- [6] Hou, Y. C.: A corpus-based study of Chinese learners' use of AWL in academic English essay writing, Foreign Language Education, Vol.0 (2014) No.0, p.150-158.
- [7] Wu, J. and Wang, T. S.: A study of the applicability of Coxhead's "Academic Word List", Foreign language Education Abroad, (2007) No.2, p.28-33.
- [8] Biber, D., and Conrad, S.: Quantitative Corpus-Based Research: Much More Than Bean Counting, TESOL Quarterly, Vol.35 (2001) No.2, p.331-336.
- [9] Hunston, S.: *Corpora in Applied Linguistics* (Cambridge University Press 2002).
- [10] Coxhead, A.: A new academic word list, TESOL Quarterly, Vol.34 (2000) No.2, p.213-238.
- [11] Yang, H. Z. and Wei, N.: *Introduction to Corpus Linguistics* (Shanghai Foreign Language Education Press 2002).
- [12] McEnery, T. and Hardie, A.: *Corpus linguistics: Method, theory and practice* (Cambridge University Press 2011).

- [13] Sinclair, J.: *How to Use Corpora in Language Teaching* (Amsterdam: John Benjamins Publishing Company 2004).
- [14] Gavioli, L.: Enriching reality: Language corpora in language pedagogy, *ELT Journal*, Vol.55 (2001) No.3, p.238-246.
- [15] Timmis, I.: *Corpus linguistics for ELT: Research and practice* (Routledge 2015).
- [16] Granger, S.: A Bird's Eye View of Learner Corpus Research, In Granger, S., Hung, J. and Tyson, S. P. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, John Benjamins Publishing Company, (2002), p.3-33.
- [17] Gu, X. Q.: A Corpus Linguistics Investigation into English Synonyms "trip" "journey" "voyage", *Overseas English*, (2020) No.7, p.246-248.
- [18] Schmitt, N. and Schmitt, D.: *Vocabulary in Language Teaching* (Cambridge University Press 2020).
- [19] Zuo, X.: Colloquial Tendency of Chinese College Students' Academic Writing in English —— A Corpus-based Study on Lexical Bundles, *Journal of Educational Institute of Jilin Province*, Vol.37 (2021) No.6, p.55-59.

Appendix 1.

In the preparation stage, teachers are faced with many academic words and do not know how to order their explanations and weight them in the academic vocabulary class. Tools can be used to rank the academic nature of words. Detailed steps and explanations follow.

1. Visit Tom Cobb's website: <https://www.lexutor.ca/vp/>



(Figure 1. VocabProfile Home interface)

2. Choose the "CLASSIC" option (i.e. the first option highlighted in bright yellow in Figure 1) and it covers the content of corpora AWL.



(Figure 2.1. VocabProfile interface with the text for the analysis)

Input mode A Type or paste small to medium size text (max about 25 000 words/25 seconds) into window, choose options, and click yellow Submit_window button for Lexical Frequency Profile.

Title: Untitled | Eng+Enl Cognates (Lists) | Sentence Count | Bar Chart | Count Index | Basic

TEXT SET-UP
General: Include an empty space after every comma or full stop.
Research: Deal with spelling errors and proper nouns.

Demos: Lit (1) (2) | Graded (1) (2) | Science (1) (2) | News NYT WP Rex Speech Adults Kids | Legal | GSL+AWL 1k 2K AWL | French

Count+ List->Text

Words to recategorize => 1k (type, dbt-click)
(E.g. known cognates/compounds)

PROPER NOUN HANDLING ?
Off-list mid-sentence caps...
Class as offlist
Class as 1k/1c
Eliminate

COMPOUNDS HANDLING ?
When Off word is 2 wds...
Class as offlist
Break apart

SUBMIT_window

Input mode B For large files, to 1 million words (or top at ≈ 500k wds) >> LARGER TEXT = MORE BASIC INFO >>

CORPUS Brown Corpus (1 million) x LIST BNC/COCA Fams 25k

Submit_Corpus

UPLOAD LARGE TEXT FILE (with ~.txt file extension)
选取文件 未选择文件
Submit_File

(Figure 2.2. VocabProfile Text to be analysed submission page)

3. Delete the instructions in the text box under “Input mode A”, paste the target vocabulary that you want to select into the box (see Figure 2.1), and click on “Submit-window” (i.e. the bright yellow button in Figure 2.2). The program will analyze the vocabulary according to the CLASSIC lists.
4. Scroll down the page and look for the section type list. You will see that the words are categorized into different frequency levels.

Types List [1]
type_[number of tokens]

VP-CLASSIC (1k, 2k + AWL)-1k types: [fams 39 : types 41 : tokens 46]
Morpholex Extract List

a [1] according [1] after [1] an [1] and [2] as [1] below [1] by [1] contains [1] deal [1] determined [1]
enter [1] every [1] for [1] from [1] full [1] general [1] here [1] how [1] include [1] many [1] of [1] one [1]
or [3] proper [1] reality [1] set [1] space [1] stop [1] tell [1] the [3] this [1] to [1] type [1] up [1] will [1]
window [1] with [1] words [1] you [1] your [1]

VP-CLASSIC (1k, 2k + AWL)-2k types: [fams 9 : types 9 : tokens 9]
Morpholex Extract List

bands [1] button [1] empty [1] frequency [1] nouns [1] paste [1] sample [1] spelling [1] yellow [1]

VP-CLASSIC (1k, 2k + AWL)-3k types: [fams 7 : types 8 : tokens 12]
Morpholex Extract List

analysing [1] demonstration [1] errors [1] instructions [1] research [2] submit [1] text [4] texts [1]

(Figure 3. Lexical profile output)

OFFLIST : [?: types 4 : tokens 4]
Extract List

click [1] comma [1] corpora [1] vocabprofile [1]

(Figure 4. Example of words in the off-list section)

5. From blue, green and then yellow, the words are classified according to their decreasing frequency. This means that all three categories are academic words, but blue is the most

frequently used word, green is less frequent than blue but more frequent than yellow, and yellow is the least frequently used word (see Figure 3).

- The red words in the “Offlist” don’t belong to academic words. The teacher can remove or reduce the classroom weighting of this part of the vocabulary depending on the actual classroom situation. If students are using Lextutor to retrieve their own academic writing exercises, then words in the “Offlist” can be removed to improve the academic nature of the text.

Appendix 2.

Each student has a different level of academic vocabulary and has different demands on it, so it is necessary to encourage students to build their own corpus. Detailed steps and explanations follow.

- Firstly, all the materials used to create the corpus are categorised by topic and, to ensure the academic nature of the corpus, articles published in core academic journals can be used as far as possible. A search was conducted on the Knowledge Network based on the keyword "teach vocabulary", and the ten most cited articles were selected and ranked according to the number of citations. (see Figure 5).

The screenshot shows the search results for 'teach vocabulary' on the Knowledge Network. The search bar at the top contains 'teach vocabulary' and the search button is labeled '检索'. Below the search bar, there are tabs for 'Literature', 'Journal', 'Ph.D.', 'Meeting', 'Newspapers', 'Foreign literature', 'Yearbook', 'baike', and 'Dictionary'. The 'Literature' tab is selected. The search results are displayed in a table with columns: '题名' (Title), '作者' (Author), '来源' (Source), '发表时间' (Publication Time), '数据库' (Database), '被引' (Cited), '下载' (Download), and '分享' (Share). The table lists 9 articles, with the first article being '略论对外汉语词汇教学的两个原则' (A Brief Discussion on Two Principles of Teaching Chinese Vocabulary to Foreigners) by Li Rulong and Wu Meng, published in 2005-03-25, with 435 citations and 10847 downloads. The table also includes a 'Group Browse' section with filters for 'Source Database', 'Discipline', 'Publishing Annual', 'Research Level', 'Author', 'Free subscription', and 'Customized retrieval'. The 'Publishing Annual' filter is set to '2023(1)', '2022(8)', '2021(1086)', '2020(1495)', '2019(1454)', '2018(1199)', '2017(1201)', and '2016(1078)'. The 'Research Level' filter is set to '2015(777)', '2014(1119)', '2013(11 28)', '2012(1151)', '2011(719)', '2010(495)', and '2009(549)'. The 'Group Browse' section also includes a '排序' (Sort) dropdown menu with options: '主题排序' (Sort by Topic), '发表时间' (Sort by Publication Time), '被引' (Sort by Cited), and '下载' (Sort by Download). The '排序' dropdown is currently set to '被引'. The search results are displayed in a table with columns: '题名', '作者', '来源', '发表时间', '数据库', '被引', '下载', and '分享'. The table lists 9 articles, with the first article being '略论对外汉语词汇教学的两个原则' (A Brief Discussion on Two Principles of Teaching Chinese Vocabulary to Foreigners) by Li Rulong and Wu Meng, published in 2005-03-25, with 435 citations and 10847 downloads. The table also includes a 'Group Browse' section with filters for 'Source Database', 'Discipline', 'Publishing Annual', 'Research Level', 'Author', 'Free subscription', and 'Customized retrieval'. The 'Publishing Annual' filter is set to '2023(1)', '2022(8)', '2021(1086)', '2020(1495)', '2019(1454)', '2018(1199)', '2017(1201)', and '2016(1078)'. The 'Research Level' filter is set to '2015(777)', '2014(1119)', '2013(11 28)', '2012(1151)', '2011(719)', '2010(495)', and '2009(549)'. The 'Group Browse' section also includes a '排序' (Sort) dropdown menu with options: '主题排序' (Sort by Topic), '发表时间' (Sort by Publication Time), '被引' (Sort by Cited), and '下载' (Sort by Download). The '排序' dropdown is currently set to '被引'.

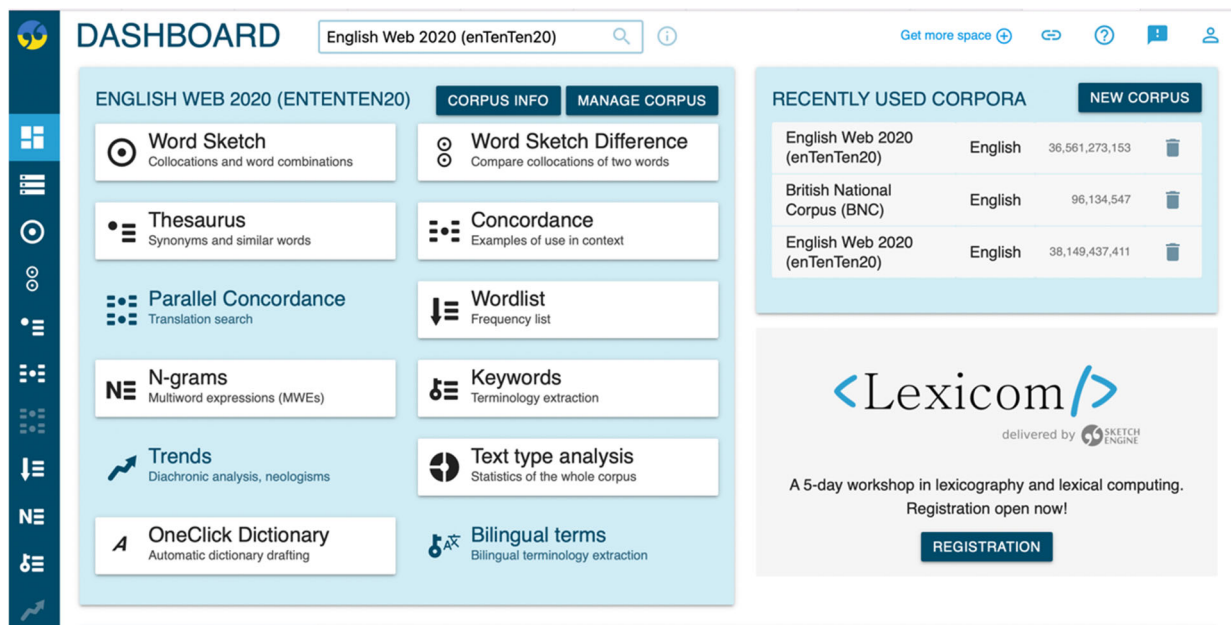
题名	作者	来源	发表时间	数据库	被引	下载	分享
1 略论对外汉语词汇教学的两个原则	李如龙; 吴茗	语言教学与研究	2005-03-25	期刊	435	10847	+
2 不同环境下的外国人汉语学习策略研究	吴勇毅	上海师范大学	2007-04-01	博士	415	22971	+
3 对外汉语教师素质研究	王晓音	陕西师范大学	2013-05-01	博士	218	20228	+
4 基于网络环境的大学英语自主学习监控理论与实践研究	何明霞	上海外国语大学	2012-05-01	博士	180	18069	+
5 大学英语词汇习得及对教学的启示	李莉文; 任长惠	外语与外语教学	2001-06-01	期刊	157	1667	+
6 外语教学中文化因素研究	马冬虹	上海外国语大学	2007-12-01	博士	135	16244	+
7 《(汉语水平)词汇等级大纲》双音合成词语素统计分析	邢红兵	世界汉语教学	2006-07-05	期刊	119	3204	+
8 语音意识和阅读能力相关性实验研究	邱建华	西北师范大学	2003-11-01	硕士	113	2027	+
9 思维导图在高中英语写作教学中的应用	崔贝贝	陕西师范大学	2012-05-01	硕士	111	3728	+

(Figure 5. The search page of Knowledge Network)

- Make a folder on the computer namely 'Academic Writing'. In this folder, create a sub-folder namely "vocabulary teaching". Create a table in Excel to record the information of the articles as follows:

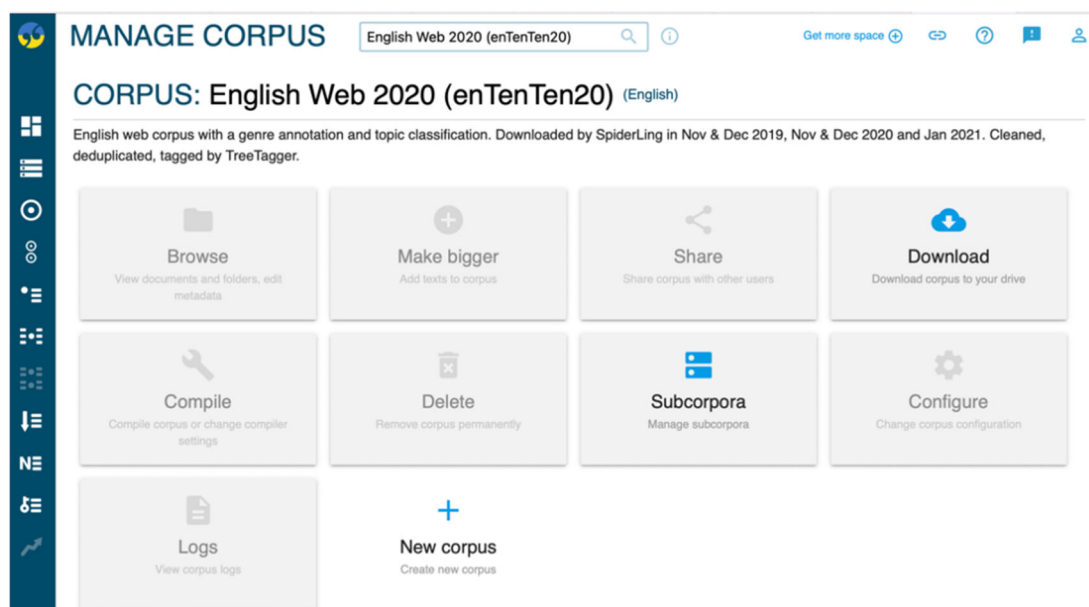
Name	Folder	Date
The Lexical Approach and English Teaching in China	vocabulary teaching	2007

3. Do similar steps with the remaining article and other topics.
4. When you finish the step3, you can upload the folder to Sketch Engine and analyze the language of your own corpus. First, visit Sketch Engine and log in.
(https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fententen20_tt31_1)



(Figure 6. Sketch Engine Home interface)

5. Choose 'Mange corpus' (see Figure 6).
6. Choose 'New corpus' (see Figure 7).



(Figure 7. Establish a new corpus)

7. Fill in the information about the name and press "Next" (see Figure 8).

Build your own private corpus from texts on the web or from your own documents.

The screenshot shows a form for creating a new corpus. At the top, it says "Build your own private corpus from texts on the web or from your own documents." The form has four main sections: "Name" with the value "Academic Writing", "Corpus type" with two radio buttons (selected: "Single language corpus", unselected: "Multilingual corpus"), "Language" with the value "English" and a search icon, and "Description" with the value "Vocabulary teaching". Below these is a storage usage indicator: "Storage used: 0 of 1,000,000 words (0%)". At the bottom, there is a dropdown menu labeled "Available features" and two buttons: "BACK" and "NEXT".

(Figure 8. Naming, information entry, and submission of new corpus)

8. Choose "I have my own texts" (see Figure 9)

The screenshot shows two buttons for selecting text sources. The left button is labeled "Find texts on the web" with a globe icon and the subtext "Automatically find and download relevant texts". The right button is labeled "I have my own texts" with a cloud upload icon and the subtext "Upload your own files (.txt, .pdf,...) or paste text".

(Figure 9. Selecting corpus text sources)

9. Then "Choose a file" upload your corpus and wait for Sketch Engine to process the files. And press "Next".
10. Then press "completed". Wait and select 'corpus dashboard'. Then you have your own corpus which is ready to use.