Application of COCA in EFL Writing Instruction at the Tertiary Level in China

Gengsheng Xiao(✉), Xin Chen
University of South China, Hengyang, China
`xgstudy@126.com`

Abstract—English writing plays an increasingly important role in the context of globalization. However, many Chinese college students' writing skills of English as a Foreign Language (EFL) are still quite deficient in spite of substantial time and effort devoted to college English writing instruction by teachers and researchers. In the present study, we suggest the application of Corpus of Contemporary American English (COCA) in college English writing instruction to resolve current issues such as learners' insufficient exposure to authentic linguistic materials, inactive thinking, lack of motivation to write, and teachers' heavy workload of making assessment. The application of COCA runs through the whole process of writing. At pre-writing stage, COCA is applied for idea inspiration and vocabulary preparation; at while-writing stage, COCA is used for eliminating erroneous usages and improving language quality, especially lexical accuracy and complexity; at post-writing stage, the application of COCA aims at giving effective feedback by improving the quality of different forms of assessment, i.e. teacher, self- and peer assessment.

Keywords—COCA, EFL writing, English writing instruction

1 Introduction

In Mainland China, great importance has long been attached to developing learners' language skills of English as a Foreign Language (EFL), among which EFL writing attracts considerable attention from educators and students. On the one hand, in the context of globalization, academic writing in English plays an increasingly crucial role in international academic communication, especially in the publication of original articles in international academic journals. Therefore, English writing has been highly emphasized as a significant productive skill in higher education in China. On the other hand, English writing never comes as an easy skill for Chinese EFL learners, which is instead proved to be the most challenging for many students at the tertiary level in China, according to the results of a wide variety of national English examinations designed to assess learners' comprehensive ability to use English. It is found that Chinese college students have great difficulties in using accurate language to express their thoughts with substantial contents and appropriate discourse structure [1].

With regard to EFL writing instruction at the tertiary level in China, there are also some serious problems which are in urgent need of solution. Firstly, language input provided for students is not authentic and sufficient enough to effectively trigger productive output and develop academic writing. Classroom activities of English writing course are mainly centered on the textbook, and thus students do not have access to adequate native-like language materials. As a result, students tend to overuse the language points obtained from the textbook and produce output in an improper way. Secondly, College students tend to have low motivation to write and their autonomous learning ability on the whole stays at a low level. Most of them never conduct writing activities as self-access learning except those compulsory writing assignments. They also rely heavily on the essay outline and the model essay to stimulate their thinking, and they consider as teacher's exclusive obligation revising their compositions and correcting all the mistakes for them. Students are not equipped with the autonomy to make use of the available language resources to develop their EFL academic writing.

Under such difficult circumstances of college English writing instruction in China, educators and scholars conducted pedagogical experiments or implemented teaching model reforms in English writing classroom. Pedagogical notions such as self-monitoring in writing [2], collective lesson planning in writing instruction [3], write-to-learn approach [4], and portfolio-based writing assessment [5] are put forward theoretically and put into application practically. It is undeniable that previous studies on English writing instruction is helpful for the improvement of learners' writing skills to a great extent. Nevertheless, it should be noted that those serious problems existing in English writing instruction have not been dealt with pertinently. With the rapid development of computer and Internet technology, corpora have been widely employed in second or foreign language teaching, which has dramatically changed language education and English writing instruction [6]. The present study aims to integrate Corpus of Contemporary American English (COCA), a large-scale online corpus, into English academic writing instruction, for it can provide the learners with abundant, natural, and authentic linguistic data, and by getting access to this online platform, they can develop their autonomy to discover, explore and generalize linguistic rules, and they can use the language knowledge obtained from self-discovery learning via COCA to produce high quality written output and develop their English writing skills.

2 Overview of COCA

COCA, which was developed by Professor Mark Davies from Brigham Young University and initially launched in the Internet on February 20, 2008, has become the largest freely-available corpus of English. This online corpus, with a size of more than 520 million words, is also currently the most comprehensive and equally-divided corpus of American English, in which each of the five genres of English, i.e. spoken, fiction, magazine, newspaper, and academic, takes up 20% of the whole corpus respectively. Moreover, from 2009 to 2015, the corpus has been updated once or twice

every year, with an expansion of 20 million words annually, which greatly guarantees the timeliness of language materials in the corpus. Table 1 presents basic information about COCA.

Table 1. Basic Information about COCA

Size (M) across genres	Spoken	Fiction	Magazine	Newspaper	Academic
	109.4	104.9	110.1	106.0	103.4
Size (M) across periods	1990-1994	1995-1999	2000-2004	2005-2009	2010-2015
	104.0	103.4	102.9	102.0	121.6

Besides its incomparable advantages of large size, genre balance, and constant updating, COCA, an online platform with a combination of language materials and language retrieval software, provides the users with an easy and convenient access to observation, exploration and research of the language data in the corpus. The majority of contemporary corpora such as British National Corpus (BNC) and American National Corpus (ANC) are not accessible for free. Furthermore, when accessing such corpora, users have to be equipped with the knowledge concerning the operation of relevant computer software tools such as *WordSmith*, *Antconc* and *Range*. Otherwise, they cannot analyze language materials and retrieve data. By visiting the website address *http://corpus.byu.edu/coca/*, the users just need to register and then log in, with all the language resources in COCA accessible to them for free.

3   Application of COCA in EFL Writing Instruction at Three Stages

Owing to their abundant language resources and easy accession, electronic corpora including COCA have been put into extensive application in the area of language teaching and learning. Language educators and teachers in many parts of the world have integrated corpora to classroom teaching and students' autonomous learning, and corpora have been proved to be very effective tools for language education [7]. Previous studies concerning the application of corpora including COCA into English education are conducted mainly in a general sense, while the integration of them into academic writing instruction has been unjustly neglected. How to effectively apply COCA into writing instruction still remains unknown to English educators and instructors. Therefore, the present study aims to address this issue by exploring the pedagogical application of COCA in EFL writing instruction throughout the whole process, which, based on process approach to writing, consists of three stages, i.e. pre-writing, while-writing, and post-writing [8].

3.1 Pre-Writing Application of COCA for Idea Inspiration and Vocabulary Preparation

When assigned a topic for writing, EFL students usually find it difficult to identify what to write about the topic and have great difficulties in generating ideas. According to second language acquisition theory, sufficient comprehensible input is an essential prerequisite to language acquisition and output production [9]. Therefore, at the stage of pre-writing, it is necessary to provide students with some language materials related to the essay topic. Teachers have been fully aware of the importance of providing linguistic input for EFL learners before they are engaged in specific writing tasks, and they usually present a model essay to the students in order to stimulate their thinking and expand their ideas around the writing topic. Meanwhile, reading the model essay carefully also enables the students to familiarize themselves with the expressions and the structural pattern that can be employed in their writing. Undoubtedly, model essays can play a positive and effective role in providing input for ESL learners at the stage of pre-writing, but when exposed to the model essay, students' creative and critical thinking is likely to be restricted, and their writing products tend to more or less resemble the model essay in terms of content, structure and language. Hence, in addition to the presentation of model essays, the teachers can introduce COCA to students, and select some key words for the writing topic and then search lexical items closely related with the key words, which can greatly inspire students' ideas and provide adequate lexical resources for them.
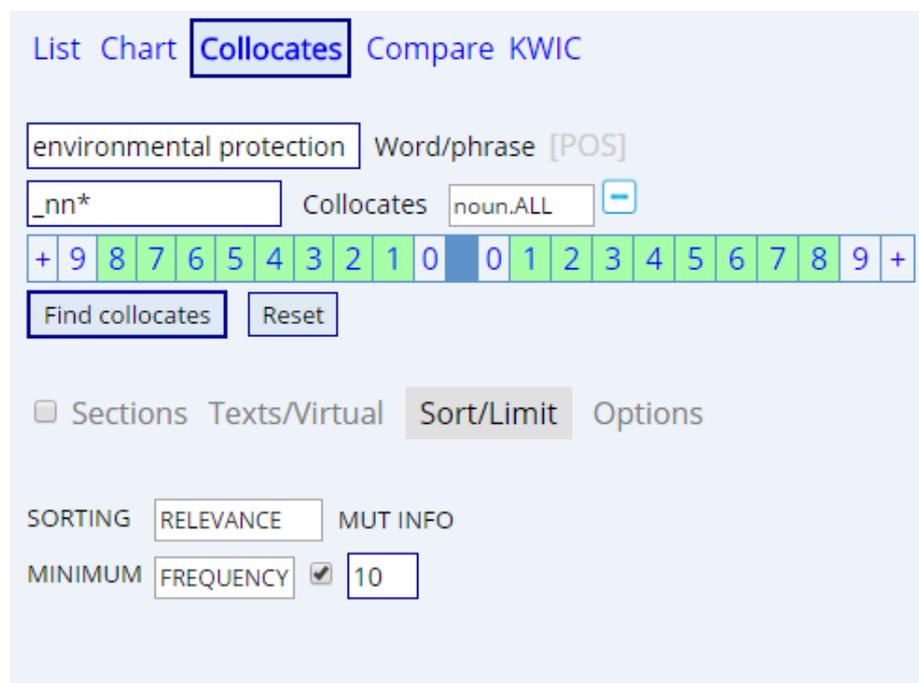

Fig. 1. COCA Searching Interface

For example, when assigning students a writing task on the topic of "environmental protection", the teacher can demonstrate the searching interface of COCA for the students, and enable them to use the searching of key words and collocates in COCA for stimulation of thinking and inspiration of ideas. The teaching procedures are portrayed in Figure 1.

As depicted in Figure 1, the key word or phrase on the writing topic needs to be keyed in in the box before "Word/phrase". In this instance, "environmental protection" is typed as the search string. Since the purpose of using COCA here is to obtain some useful language materials concerning environmental protection, and to acquire some ideas closely related to environmental protection, searching for the collocates of "environmental protection" can help the users to retrieve a list of words tightly related to the topic. Thus, the button of "Collocates" should be clicked, then the part of speech of the collocates is chosen as "noun. ALL", indicating that the searching results will be confined to nouns, because nouns are directly linked to ideas and messages concerning the topic. In order to get a more complete picture of "environmental protection", the span is set as from the left 8 words to the right 8 words, and as can be seen, those numbers are shadowed when the setting is done. In this way, the searching results will present the nouns co-occurring on both sides of "environmental protection" with a span of eight words. Then how the collocates are displayed can be set upon the hit of "Sort/Limit", and the "Sorting" of searching results is done according to "RELEVANCE" indicated by "MUT INFO", i.e. mutual information. Moreover, the minimum ""FREQUENCY'' can be automatically specified as "10". The searching results will be sorted in the order of relevance, i.e. mutual information score, which is an indicator to show how strongly two items are connected with each other. After the above procedures have been conducted, a list of collocates of "environmental protection" will be retrieved. Only the top 20 nouns with highest MI score are listed as presented in Table 2 due to the limited space.

As seen in Table 1, all the 20 listed nouns are strongly collocated with "environmental protection", because the MI scores are higher than 4. In other words, those nouns are closely related to the ideas or messages that are very likely to be taken into consideration as far as environmental protection is concerned. Taking the first word *agency* as an example, it is the collocate with highest MI, the MI score being 8.68, and the token number of *agency* in the whole corpus is 47149, among which 2798 are collocated with "environmental protection", accounting for 5.93% of all the tokens. From the statistics, it can be inferred that *agency* is significantly correlated with *environmental protection*. In order to make it clearer, we can hit the word *agency* to retrieve the specific context. Then it will be found that there is an administrative agency called the Environmental Protection Agency which aims to protect human health and the environment. Therefore, it is not surprising that *agency* is so strongly collocated with *environmental protection*.

The next step is to guide students to divide the scattered 20 nouns into several groups of certain conceptual domain, helping students form organized and systematic ideas. The first group is made up of the words that account for the reasons why environmental protection needs to be urgently implemented, e.g. *pollution, pesticide(s), emission(s), groundwater, pollutants, sewage, ozone, greenhouse*. It is apparent that

Table 2. Statistical Information about Collocates of *environment protection* Retrieved from COCA

	CONTEXT	FREQ	ALL	%	MI
1	AGENCY	2798	47149	5.93	8.68
2	ADMINISTRATOR	90	6573	1.37	6.57
3	PESTICIDE	14	2281	0.61	5.41
4	POLLUTION	82	13927	0.59	5.35
5	REGULATIONS	91	16545	0.55	5.25
6	PESTICIDES	20	3777	0.53	5.20
7	EMISSIONS	53	11895	0.45	4.95
8	GROUNDWATER	11	2490	0.44	4.94
9	CLEANUP	19	4428	0.43	4.89
10	SUSTAINABILITY	13	3361	0.39	4.74
11	DIVISION	121	32719	0.37	4.68
12	DEPARTMENT	365	99925	0.37	4.66
13	POLLUTANTS	10	2801	0.36	4.63
14	EXPIDITURES	14	4197	0.33	4.53
15	EMISSION	10	3017	0.33	4.52
16	SEWAGE	10	3461	0.29	4.32
17	OZONE	13	4664	0.28	4.27
18	GREENHOUSE	16	5979	0.27	4.21
19	CONSERVATION	40	16459	0.24	4.07
20	AGENCIES	65	26751	0.24	4.07

such words are very useful language resources to expound different forms of environmental pollution, and their respective source and harmful consequence. The second group comprises of the words explaining who should be engaged in environmental protection. Besides individuals, *agency(agencies), administrator, division, department* take responsibility for protecting the environment. The third group includes the words explaining what efforts we can make to protect the environment, and obviously, *regulations* can be enacted and enforced, the *cleanup* of waste can be properly handled and strictly supervised, the growth of *expenditures* for environmental protection needs to be ensured. The two words left, i.e. *sustainability* and *conservation,* naturally fall into the fourth category, which can be used to state the goal of environmental protection. The categorization of the collocates is a process of making vocabulary preparation for essay writing and meanwhile building up connections between concepts. This process is bound to help students stimulate critical and even innovative thinking, and more importantly, to help them organize ideas in a logical and clear way.

3.2 While-Writing Application of COCA for Improvement of Lexical Accuracy and Complexity

When students are equipped with sufficient, organized and logical ideas and they have also received enough linguistic input after they retrieve a list of key words related to the topic for their writing task, the next step for them is to write the essay. During this while-writing process, many Chinese college students find it very difficult to use words accurately and even more difficult to achieve lexical variety and complexity. COCA can be a much more effective tool than English-Chinese and Chinese-English Bilingual Dictionary to help students to use words properly and even reach the standard of accuracy and complexity.

Firstly, COCA searching can be done to eliminate erroneous usage of words in students' writing. As EFL learners, Chinese college students tend to firstly think in Chinese and literally translate the sentences into English. In such cases, most of the learners have the linguistic sensitivity for such language errors caused by the mother language transfer. They can turn to COCA to check out whether the expression is grammatically and pragmatically acceptable. COCA is a large-scale corpus with a capacity of more than 520 million words, and if the expression is very rarely used by native speakers of English, it then can be basically considered as a language error. For example, in Chinese college students' essays on the topic of "environmental protection", there are some typical errors as illustrated in the following sentences.

1. We must carry on active propaganda for environmental protection.
2. Greater efforts need to be made to solve the shortage of fresh water.
3. The government is responsible for advocating public awareness of environmental protection.

The problem of the first sentence lies in the violation of semantic prosody, because *environmental protection* is a positive issue, while *propaganda* is semantically negative, they are two lexical items unlikely to co-occur harmoniously in a context. However, Chinese students tend to learn a new English word by memorizing its corresponding Chinese meaning without paying attention to its semantic prosody. To check whether the two words or phrases are compatible with each other in terms of semantic prosody, COCA searching of collocates can be very effective. By searching for the collocates of "propaganda" in COCA, it can be found that *propaganda* is strongly collocated with some nouns with negative semantic prosody, such as *disinformation, anti-American, anti-Semitic, Nazi, misinformation, agitation*. Thus, the error in this sentence can be successfully removed by changing *propaganda* into *promotion*, because the two latter have a positive semantic prosody.

Although there is no semantic prosody violation within the next two sentences, there are grammatical errors in them because of the mismatching verb-noun collocations. By searching for collocates, it can be found *solve* and *shortage* seldom co-occur in a context, and the same is true for *advocate* and *awareness*. Instead, according to the statistical information of collocates retrieved from COCA, i.e. frequency and MI score, *solve* is generally collocated with such nouns as *problem*, *mystery*, *crisis*, *crime*, *case*, *puzzle*, *issue*, and *riddle*, while *shortage* tends to co-occur with notional

verbs such as *alleviate*, *suffer*, *experience*, *face*, *ease*, and *address*. As for the other sentence, *advocate* is inclined to co-occur with such nouns as *approach*, *use*, *policy*, *system*, *program*, and *strategy*, while *awareness* tends to keep company with such verbs as *raise*, *increase*, *develop*, *promote*, *heighten*, *build*, and *enhance*. Therefore, the above two sentences can be corrected through replacing the inappropriate collocations with those typically co-occurring together, and thus the possible revisions of students' original sentences can be "*Greater efforts need to be made to solve the problem of fresh water shortage.*" and "*The government is responsible for raising public awareness of environmental protection.*"

Besides the use of COCA for correcting lexical errors, COCA also can be used to differentiate synonyms for the purpose of precise lexical usage in writing. For example, when undertaking the writing task on the topic of "environmental protection", it is inevitable to discuss how harmful the pollutants are to human beings' physical health. Under such a circumstance, the students may feel confused about two synonyms, i.e. *poisonous, toxic,* and doubt whether they are interchangeable when used to describe the nature of some pollutants causing illness or even death. With COCA, it is very easy and quick to compare the two synonyms. We simply click on the "Compare" button, enter the two words, i.e. *poisonous* and *toxic* respectively as Word1 and Word2. Because the two compared words are adjectives, and one of the main grammatical function of adjectives is to describe nouns, here the usage differences between *poisonous* and *toxic* are reflected in what nouns are strongly collocated with one instead of the other. So the part of speech of collocates are specified as noun, and the span is set as four words to the both sides of the node word. Then we can click on the button "Sort/Limit", and keep the default setting. The COCA interface for comparing synonyms are shown in Figure 2.

Then, two lists of collocates (nouns) sorted by ratio between the two synonyms can be easily generated. While Figure 3 presents the nouns that have significantly stronger collocational relationship with *poisonous* than with *toxic*, Figure 4 illustrates that the collocation strength of the 10 nouns with *toxic* is significantly greater than *poisonous*.

From the two word lists of collocates, it can be seen that *poisonous* tends to be used to describe living beings such as *snake*, *duck*, *spider*, *flower* and *mushroom* while *toxic* is more likely to be used to describe the *emission* of harmful gas and smoke, *release* of chemicals, *site* of *waste disposal* or *dump*, *contamination, exposure* to pollutants. It is quite evident that *toxic* is more frequently used by native speakers of English to describe those harmful substances that cause different forms of pollution and cause illness or even death in human beings. Thus, *toxic* is more typical than *poisonous* for the context of environmental protection though the two synonyms are quite similar to each other in terms of meaning and usage.

Additionally, COCA can also be used to help students figure out whether the words or expressions are appropriate for formal academic writing. According to several corpus-based studies on Chinese students' writing, Chinese students including advanced EFL learners clearly employ a spoken type of discourse in their English writing, and in other words, Chinese students tend to have remarkable features of oral style in formal academic writing [10][11]. One of distinctive functions of COCA is
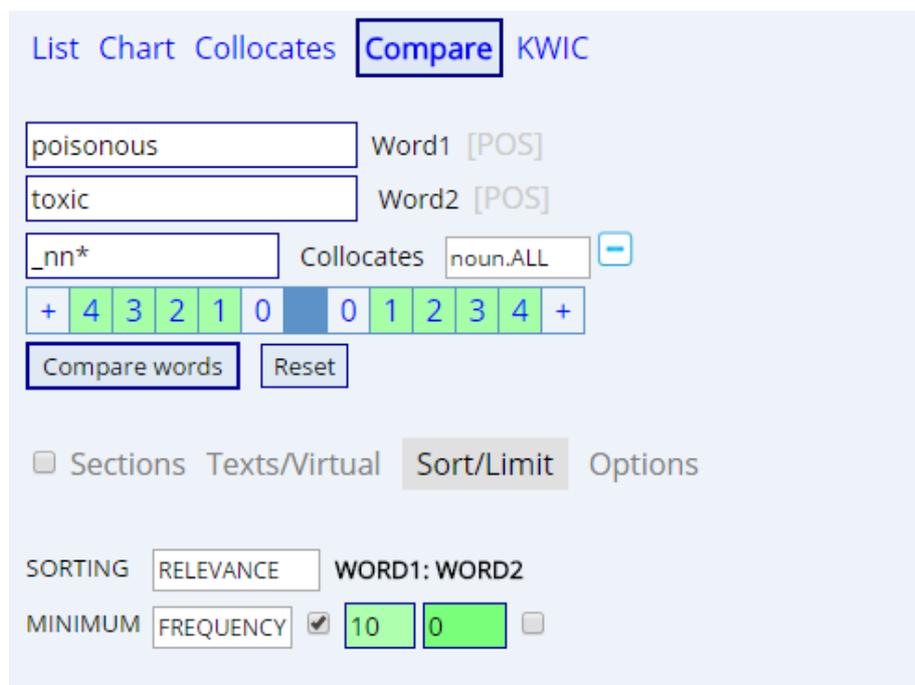
List Chart Collocates **Compare** KWIC

poisonous	Word1 [POS]
toxic	Word2 [POS]
_nn*	Collocates	noun.ALL	⊟

+ 4 3 2 1 0 | 0 1 2 3 4 +

Compare words   Reset

☐ Sections Texts/Virtual Sort/Limit Options

SORTING RELEVANCE **WORD1: WORD2**

MINIMUM FREQUENCY ☑ 10 0 ☐

Fig. 2. COCA Interface for Comparing *poisonous* and *toxic*

WORD 1 (W1): **POISONOUS** (0.22)

	WORD	W1	W2	W1/W2	SCORE
1	SNAKES	139	2	69.5	318.9
2	DUCK	17	0	34.0	156.0
3	SNAKE	79	5	15.8	72.5
4	BITE	11	1	11.0	50.5
5	SPIDERS	20	2	10.0	45.9
6	FLOWERS	10	1	10.0	45.9
7	TOAD	10	1	10.0	45.9
8	REPTILES	10	1	10.0	45.9
9	MUSHROOM	16	2	8.0	36.7
10	MUSHROOMS	21	5	4.2	19.3
11	BERRIES	25	7	3.6	16.4

Fig. 3. A List of Nouns Significantly Collocated with poisonous in Contrast to toxic

WORD 2 (W2): **TOXIC** (4.59)

	WORD	W2	W1	W2/W1	SCORE
1	EMISSIONS	153	0	306.0	66.7
2	ASSETS	133	0	266.0	58.0
3	REGISTRY	124	0	248.0	54.0
4	SITE	88	0	176.0	38.4
5	RELEASE	69	0	138.0	30.1
6	SHOCK	67	0	134.0	29.2
7	DISPOSAL	64	0	128.0	27.9
8	CONTAMINATION	60	0	120.0	26.2
9	EXPOSURES	60	0	120.0	26.2
10	DUMPS	57	0	114.0	24.8

Fig. 4. . A List of Nouns Significantly Collocated with *toxic* in Contrast to *poisonous*

that it can statistically count and then show the frequency of a particular word, phrase, or construction across the following five different genres, i.e. spoken, fiction, magazine, newspaper, and academic [12]. In this case, if a word very frequently occurs in other genres especially in spoken section of COCA while there is a much lower frequency of the word in academic section, it is then too informal or even colloquial, and not appropriate for academic writing. For example, Chinese students tend to overuse *a lot of* to modify countable nouns in their essay, and by searching COCA, they can get Figure 5 indicating the frequency of *a lot of* across different genres.

According to Figure 5, it can be clearly seen that *a lot of* is very frequently used in spoken language while its frequency in academic section is significantly lower than that in other sections, especially in the spoken section. Therefore, *a lot of* is a colloquial expression and it should be avoided in formal, academic writing. On the contrary, *many*, though a very simple word, can be a better choice when used to modify countable nouns in formal writing. Figure 6 shows the frequency of *many* across different genres in COCA.

Figure 6 demonstrates that despite a relatively balanced distribution of *many* across genres in COCA, the normalized frequency of *many* is the highest in academic section. The frequency information of *many* across different genres in COCA reveals that *many* is not too informal for academic writing and it is compatible with academic discourse. In writing, to avoid oral style, *many* can be used to replace *a lot of*.

As mentioned above, during the process of academic writing, Chinese college students can make use of different functions of COCA to promote the accuracy, complexity and appropriateness of words and polish their writing.

SECTION (CLICK FOR SUB-SECTIONS) (SEE ALL SECTIONS AT ONCE)	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT (SEE ALL)
SPOKEN	102,484	109.4	936.85	
FICTION	13,932	104.9	132.81	
MAGAZINE	23,774	110.1	215.91	
NEWSPAPER	33,900	106.0	319.92	
ACADEMIC	4,715	103.4	45.59	

Fig. 5. Frequency in COCA by Genre: *a lot of*

SECTION (CLICK FOR SUB-SECTIONS) (SEE ALL SECTIONS AT ONCE)	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT (SEE ALL)
SPOKEN	99,448	109.4	909.10	
FICTION	47,457	104.9	452.40	
MAGAZINE	120,978	110.1	1,098.69	
NEWSPAPER	115,326	106.0	1,088.35	
ACADEMIC	114,308	103.4	1,105.26	

Fig. 6. Frequency in COCA by Genre: *many*

3.3 Post-Writing Application of COCA for Diversified Assessment and Effective Feedback

Without making persistent efforts to practice writing and accomplishing sufficient writing tasks EFL learners can never make constant progress in academic writing skills. Furthermore, teacher and peer feedback and self-assessment also play an indispensable role in writing development [13]. However, teacher assessment is usually very time-and energy-consuming. Giving feedback about all the students' essays is a heavy workload, especially for Chinese English teachers because the class size tends to be very large. Meanwhile, without timely assessment, students' enthusiasm for writing will be gradually weakened, and they cannot make effective revisions and accomplish essay improvement. COCA can be a very effective tool for alleviating teachers' workload of correcting students' essays and making assessment, and meanwhile the application of COCA can give full play to self-assessment and peer assessment, which in turn helps the students develop autonomous learning ability and cooperative learning ability.

First of all, the teacher can demonstrate how to make a COCA-assisted assessment, setting an example for self-assessment and peer assessment afterwards. The teacher can select two or three essays representing different writing proficiency levels, and then the feedback can be given mainly from three perspectives, i.e. content, structure, and language, pointing out the strengths and diagnosing the weaknesses.

Specifically, the quality of content can be evaluated by checking out to what extent ideas expressed in the essays are relevant, comprehensive and even systematic just as the collocates of key words in COCA show. Moreover, the categorization of the collocates in COCA can help assess the structure of the writing to check whether or to what extent the author organizes the ideas in a logical and clear way. Finally, the quality of language can be assessed mainly at syntactic and lexical levels. In above part, little has been discussed on the application of COCA in sentence writing. In fact, although we can obtain more information concerning the usage of language at the

lexical level, COCA can still provide us with useful information on syntactic features in academic essays. By reading sufficient texts in academic section of COCA, students can better understand the criteria for good sentences in formal writing, i.e. syntactic accuracy and complexity. By syntactic accuracy, it means that students are supposed to make grammatically correct sentences similar to what the native speakers of English use in COCA academic section. By syntactic complexity, it means that students should avoid single sentence structure and instead they need to make use of different sentence patterns, i.e. simple, compound, complex, and compound-complex sentences just as texts in academic section of COCA show. Therefore, the texts in COCA academic section can be taken as a benchmark for assessment of sentence writing. Finally, the assessment of lexical accuracy can be made through checking whether there are such errors as violation of semantic prosody, mismatching collocation, oral style, etc.*,* while assessment of lexical complexity can be made by examining whether synonyms are properly used to enrich language expression. From the above procedures, COCA can be used as a very effective tool to facilitate teacher assessment of students' essays and with the help of COCA, the teacher can give both comprehensive global feedback and specific local feedback on students' academic writing.

After teacher's demonstration of how to make assessment about an essay assisted by COCA, students will be able to make self-assessment. By following teacher's assessment procedures and searching for relevant information in COCA, students will learn how to assess their own essays at both macro and micro levels, evaluating the overall quality and identifying the merits and defects from various aspects, such as content, structure and language. However, self-assessment is not always so effective especially when the students are less proficient writers. When the teacher sets assessment criteria, peer assessment can be very practical and effective [14]. Thus, when making peer assessment, two students can follow COCA-assisted assessment steps demonstrated by the teacher, and give feedback on each other's essays, pointing out advantages and disadvantages. They can also put forward some suggestions for further correction and revision. Peer assessment and feedback can help students build up the patience and enthusiasm for writing of multiple drafts, leading to substantial essay improvement and writing proficiency development.

4   Conclusion

The present research analyzed the application of COCA in EFL writing instruction at the tertiary level in China. As a large-scale online corpus, COCA provides abundant and authentic English language materials, from which useful linguistic information can be retrieved for English learners. At present, college English writing instruction in China is confronted with enormous challenges, e.g. lack of sufficient comprehensible linguistic input to trigger productive output, heavy workload for the teacher to give timely feedback on students' writing, learners' lack of motivation for writing. In order to handle this unfavorable situation, we suggest the application of COCA in college English writing instruction. The integration of COCA into writing

instruction runs through the whole process of writing including three stages, i.e. pre-writing, while-writing, and post-writing.

First of all, at pre-writing stage, application of COCA can help college students inspire their ideas around the topic for writing and stimulate their active thinking. Besides, from COCA useful linguistic information such as words and phrases related to the writing topic can be retrieved to make necessary vocabulary preparation for subsequent writing task.

Secondly, at while-writing stage, different search functions of COCA can be used to eliminate erroneous usages of language, such as violation of semantic prosody, mismatching collocation, and genre inappropriateness. The primary purpose of COCA application at this stage is to improve language quality, especially lexical accuracy and complexity.

Finally, at post-writing stage, COCA can be used for improving assessment and providing effective feedback for students' writing. By demonstrating how assessment can be made according to the linguistic information obtained from COCA, teachers can alleviate heavy workload of correcting errors and giving feedback, and more importantly, guide students to make self-assessment and peer assessment, enabling them to promote writing skills and autonomous learning ability.

5 Acknowledgment

6 References

[1] Xiao, G. & Chen, X. (2015). English academic writing difficulties of engineering students at the tertiary level in China. World Transactions on Engineering and Technology Education, 13(3): 259-263.

[2] Xiang, W. (2004). Encouraging self-monitoring in writing by Chinese students. ELT Journal, 58(3): 238-246. https://doi.org/10.1093/elt/58.3.238

[3] Shi, L. & Yang, L. (2014). A community of practice of teaching English writing in a Chinese university. System, 42(1): 133-142. https://doi.org/10.1016/j.system.2013.11.009

[4] Wang, C. (2017). From write-to-learn to learn-by-CEC. Foreign Language Teaching and Research, 49(4): 547-556.

[5] Li, Q. (2016). Chinese EFL learners' perceptions of their experience with portfolio-based writing assessment: A case study. Chinese Journal of Applied Linguistics, 39(2): 215-230.

[6] Luo, Q. (2017). Data-driven learning in second language writing class: A survey of empirical studies. International Journal of Emerging Technologies in Learning, 12(3): 182-196. https://doi.org/10.3991/ijet.v12i03.6523

[7] Boulton, A. & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. Language Learning, 67(2): 348-393. https://doi.org/10.1111/lang.12224

[8] Xu, F. (2011). English writing pedagogy revisited: A multi-perspective review. Foreign Language World, (2): 57-64.

[9] Krashen, S. D. (1981). Second Language Acquisition and Second Language Learning. Oxford: Pergamon.

[10] Wen, Q., Ding, Y. & Wang, W. (2003). Features of oral style in English compositions of advanced Chinese EFL learners: An exploratory study by contrastive learner corpus analysis. Foreign Language Teaching and Research, 35(4): 268-274.

[11] Wang, L. & Zhang, Y. (2007). A corpus-based study of high frequency verbs in Chinese learners of EFL writing. Foreign Language Teaching and Research, 39(2): 110-116.

[12] Davies, M. (2013). Google Scholar and COCA-Academic: Two very different approaches to examining academic English. Journal of English for Academic Purposes, 12(3): 155-165. https://doi.org/10.1016/j.jeap.2013.01.003

[13] Miao,Y. Badger, R. & Zhen,Y. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. Journal of Second Language Writing, 15(3): 179-200. https://doi.org/10.1016/j.jslw.2006.09.004

[14] Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. Language Testing, 26(1): 75-100. https://doi.org/10.1177/0265532208097337

7   Authors

Gengsheng Xiao is currently an associate professor in School of Languages and Literature, University of South China, Hengyang, China. His research interests include second language acquisition and corpus linguistics. (xgstudy@126.com).

Xin Chen is currently an associate professor in School of Languages and Literature, University of South China, Hengyang, China. Her research interests include applied linguistics and interpretation. (englishow@126.com).