

Exploring the Impacts of Learner Translator Corpus (LTC) on English Majors L2 Students

Jie Yang

Guangzhou Nanfang College, China

Email: 20210120047@fudan.edu.cn

Manuscript received December 12, 2024; revised January 13, 2024; accepted January 30, 2024; published May 14, 2024

Abstract—As a valuable resource for second language learners, Learner Translator Corpus (LTC) provides English majors with real-life contexts and offers authoritative, professional, and comprehensive translation guidance, thereby facilitating their acquisition of authentic English. This paper aims to provide a detailed introduction to Learner Translator Corpus (LTC), differentiate between LTC and Non-Learner Translator Corpus, draw a parallel between online LTC and local corpus, and elaborate on the usage and functions of Corpus of Contemporary American English (COCA) corpus and Chinese/English Political Interpreting Corpus (CEPIC) learner translation corpus. Results found that: 1) online corpora better cater to the needs of English majors for efficient knowledge acquisition, while LTC software is suitable for linguistic researchers to aid students in understanding and mastering specific language phenomena; 2) The grammatical errors caused by the influence of interlanguage can be avoided by utilizing LTC; 3) LTC significantly impacts English majors in language learning guidance, improving their language proficiency, and genre and professional language studies. By undertaking these analyses, this paper seeks to enhance English majors' understanding of corpus-related knowledge and enable them to effectively utilize LTC to address translation challenges.

Keywords—Corpus of Contemporary American English (COCA), Chinese/English Political Interpreting Corpus (CEPIC), English majors, Learner Translator Corpus (LTC)

I. INTRODUCTION

Learners of second languages, especially English majors, rely on a variety of tools to help them with the translation process in order to improve translation accuracy, with the Learner Translator Corpus (LTC) playing a key role. Scientific and technical texts are among the most prevalent materials encountered by English majors during their initial foray into translation, and they also represent the texts in highest demand within the market.

II. THREE PRIMARY CHALLENGES IN LEARNER TRANSLATOR CORPUS STUDY

A. Translators of Scientific and Technical Texts Often Encounter Inherent Subjective Elements

For excelling in scientific and technical translation, translators must possess interdisciplinary knowledge and advanced foreign language proficiency, enabling them to accurately convey the intended meaning [1]. They must utilize their linguistic and cognitive abilities to explain the source text, inevitably imparting their own subjectivity and individual style onto the target text. Since they are inexperienced translators, English majors need help from authorised and experienced Learner Translator Corpus (LTC) to ensure the proper usage of particular words and phrases.

Their translations will therefore always reflect their own personal viewpoints.

B. The Influence of Interlanguage Results in an Increased Occurrence of Grammatical Errors in English

Interlanguage refers to a linguistic knowledge system developed by second language learners during the process of acquiring a second language. It is constructed based on their limited knowledge of the target language and serves as an intermediary language for communication, distinct from both their native language and the target language [2]. In the context of Chinese-English translation assignments, junior English majors are susceptible to certain types of grammatical errors influenced by their native Chinese language. The auxiliary verb “Be” is frequently omitted, new words are creatively used based on reader comprehension, and definite and indefinite articles aren't used. The following concrete examples show these mistakes:

Example 1: The omission of the verb “Be”

1) Source language: 人们愈来愈富裕了。

Target language: people more and more affluent.

2) Source language: 有些风筝大到足以把人带到空中。

Target language: Some kite large enough to take people into the air.

Due to the influence of their native language and the tendency for “Be” verbs to be omitted in Chinese, English majors frequently encounter the issue of “Be” verbs omission in their translations. This phenomenon is evident in the aforementioned examples of student work, and it is a common problem encountered by English majors.

Example 2: The creation of new words and their subsequent usage based on individual comprehension

1) Source language: 中国人有节省的传统。

Paraphrase: Thrift is the traditional virtue of the Chinese nation.

Target language: Chinese have a conserve tradition.

In this specific instance, the student employs the noun form “conserver,” which deviates from established linguistic norms. Official dictionaries, such as the online “Longman Dictionary of Contemporary English,” do not recognize “conserver” as a valid noun form. However, an example sentence in Bing Dictionary defines it as “protector.” The etymology of the word “conserve” can be traced back to the Old French verb “conserver,” as indicated by the Online

Etymology Dictionary. Although the noun “conserver” is used wrongly by the student without being confirmed, the verb “conserve” is linguistically correct. On reputable websites like China Daily or NPR, the correct English phrase “saving traditions” is used to discuss the preservation of Chinese cultural practises.

Example 3: The omission of articles

- 1) Source language: “春节”, “第一”
- Target language: “Spring Festival”, “first”

In the Chinese language, the omission of definite and indefinite articles, particularly in relation to traditional festivals like the “Spring Festival,” is a common grammatical error observed in bilingual conversions by students. This omission can be attributed to the fact that in Chinese, the ordinal word “first” is often represented by “第”, which primarily signifies word order. Consequently, English learners influenced by their native language tend to omit the corresponding definite article “the” in their bilingual translations, leading to grammatical inaccuracies.

C. There are Insufficient Conclusions in the Domain of Learner Translator Corpus (LTC) with a Practice-Oriented Approach

Scholars typically adopt a practice-oriented approach, focusing on the utilization of Learner Translator Corpus to identify and address translation problems encountered by English majors. However, there is a prevailing trend of hastily drawn conclusions and an inadequate summarization of relevant research challenges. Additionally, previous research has shown that English majors’ typical translation mistakes are only partially integrated, making it difficult to prescribe the right LTC to solve their particular problems.

To address the aforementioned gaps, this paper is structured into five main sections. The first section provides an introduction, outlining the research background, research

questions, and the paper’s framework. The second section focuses on the introduction of corpus analysis tools, delving into their definition, types, and specific tools for analyzing foreign corpora. The third section elaborates on LTC, specifically highlighting two prominent examples, Corpus of Contemporary American English (COCA) and Chinese/English Political Interpreting Corpus (CEPIC). The fourth section presents the conclusions, evaluating the achievements and limitations in the study of LTC.

III. TOOLS FOR CORPUS

Corpus tools are software applications designed for processing corpora [3], while corpus analysis tools encompass a range of software tools used to analyze corpus data. Examples of corpus analysis tools include Sketch Engine, BFSU Collocator, and others. These technologies aid academics in analysing the words, phrases, sentences, and their interactions that make up a corpus of language. They can make it easier to extract data about co-occurring words, lexical collocations, grammatical structures, and more, for example [4]. Maocheng Liang defines a corpus as an electronically gathered collection of texts that represent a language, language variant, or specific text category [5].

Several types of corpus analysis tools exist, including Concordancers and Query Tools (CQT), Annotation Tools (AT), Statistical Tools for Corpus Analysis (STCA), Specialized Corpus Tools (SCT), Data-Driven Learning Tools and Resources (DDLTR), and Corpus Discourse Tools (CDT) [3]. Additional corpus tools, such as Corpus Editing Tools (CET), Corpus Parallel Tools (CPT), and Multidimensional Analysis Tagger, can be found on the official website of Corpus Tools at Shanghai International Studies University [6]. Noteworthy foreign and domestic corpus analysis tools are as shown as Table 1:

Table 1. Noteworthy foreign and domestic corpus analysis tools

	Tools	Categor y	Website
Foreign Corpus Analysis Tools	Wmatrix	AT	https://ucrel.lancs.ac.uk/wmatrix/
	WordSmith Tools	CQT	https://lexically.net/wordsmith/
	Lexical Tutor website	CQT	https://www.lextutor.ca/
	AntConc	CQT	http://www.laurenceanthony.net/software/antconc/
	EmEditor	CET	https://www.emeditor.com/
	TreeTagger	AT	https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/
	Stanford Log-linear Part-Of-Speech Tagger	AT	https://nlp.stanford.edu/software/tagger.shtml
Domestic Corpus Analysis Tools	BFSU PowerConc	CQT	
	BFSU Stanford Parser	AT	
	Chi-square and Log Likelihood Calculator	STCA	http://corpus.bfsu.edu.cn/TOOLS.htm
	BFSU Collocator	SCT	
	BFSU Sentence Collector	DDLTR	
	SISU Aligner 2.0.0	CPT	Download: https://pan.baidu.com/s/1HsalJBsvqiR0L-I75sVzAQ Code: sisu
Wordless	CQT	https://github.com/blkserene/wordless#download	
Coh-Metrix	CDT	http://cohmetrix.memphis.edu/cohmetrixhome/	

While prior corpus tools have primarily focused on vocabulary and phrase studies, Professor Jiajin Xu emphasizes discourse research, introducing two detailed corpus tools, Coh-Metrix and TAACO, for articulation measurement and analysis. These tools enable the

investigation of the discourse characteristics in Chinese students’ English compositions. Moreover, four discourse research methods based on dedicated corpus tools are presented. The online text analysis tool, Coh-Metrix, automatically examines multidimensional textual features

such as words, sentences, and discourse units, facilitating the observation of linguistic features in translation, including articulation and other language-specific aspects [7]. Because translators have various degrees of expertise, English majors can use corpus discourse research methodologies to examine the lexical articulation traits in the translations of certain translators. By analysing frequent errors between English and Chinese, this analysis can help detect and address problems like excessive word repetition in Chinese to English translations or the overuse of personal pronouns and conjunctions in English to Chinese translations.

In conclusion, LTC software and online LTC websites each possess advantages and disadvantages. Considering the efficiency of knowledge acquisition and the fast pace of contemporary life, online corpora better cater to the needs of English majors for efficient knowledge acquisition. They are particularly suitable for English majors who require real-time resource updates and sharing, especially in the context of translation where timely updates are essential to meet contemporary requirements. On the other hand, the local version of LTC software is suitable for linguistic researchers who prioritize offline access, high performance, stability, and professionalism. Additionally, they can utilize the corpus samples and analysis tools within the software to aid students in understanding and mastering specific language phenomena, analyze language errors and problems, and provide personalized guidance and feedback.

IV. TYPES OF AVAILABLE LEARNER TRANSLATOR CORPUS AND RELATED STUDIES

A Learner Translator Corpus refers to a corpus generated by second language learners, which includes various types of learner-produced texts such as writing, speaking, and listening samples [8]. According to McEnery *et al.* [9], a Learner Translator Corpus consists of machine-readable texts in a non-native language, representing language samples written or spoken by learners in their second language. Notable domestic learner translation corpora include the Chinese Learner English Corpus (CLEC), the Chinese Student English Oral and Written Corpus, COLSEC, the translation sub-corpus of the Corpus for English Majors (CEM), the Standardized Terminology for Foreign Translation of Chinese Specialized Discourse, and CEPIC.

A corpus made up of texts written by non-learners, also known as native speaker corpora, is referred to as a non-learner translator corpus [8]. These corpora are collections of language samples produced by native speakers of a specific language, including news articles, books, academic papers, and more. Examples of Non-Learner Translator Corpus include BNCweb, the English Wikipedia corpus, the Gutenberg corpus, and the Reuters news corpus. Widely used foreign Non-Learner Translator Corpus include COCA and the British National Corpus.

Translation research commonly relies on different types of corpora, including translational corpora, comparable corpora, and parallel corpora. Translational corpora can be classified into large, medium, and small corpora based on their size. While literary texts are prevalent in the above-mentioned corpora, large and medium-sized translation research corpora often consist of encyclopedic texts or thematic corpora tailored to specific text categories or regional characteristics.

However, these specialized corpora are relatively few in number but hold significant potential for practical applications [10].

There are two types of translation corpora: historical corpora, which cover a period of time of at least 25 years, and synchronic corpora, which cover a period of time between 5 and 25 years. Additionally, corpora can be categorised as monolingual, bilingual, or multilingual based on the languages they contain, with bilingual and multilingual corpora also falling under the category of parallel corpora or comparable corpora. Additionally, corpora can be classified based on their usage as general corpora or specialized corpora [11]. General corpora cater to research needs across various disciplines, while thematic corpora refer to collections of texts centered around specific topics, including public language corpora.

Corpus research emerged in the 1960s [12], and in the 1990s, Mona Baker *et al.* utilized corpus-based research to investigate translation, resulting in the creation of the world's first translation research corpus, the Translational English Corpus (TEC) [13]. This development sparked a surge of interest in translation research and teaching abroad, leading to the emergence of corpus translation studies [11]. Domestic learner translation corpus studies primarily focus on areas such as interpreting pauses [14], teaching English for Science and Technology [15], translation of public language [10], and the assessment of translation quality among learners [16].

Wang *et al.* [14] utilizing the Parallel Corpus of Chinese EFL Learners, analyzed the occurrence of pauses in Chinese-English interpreting output. Zhu *et al.* [17] employed three main corpora to assist translators, including a large general monolingual reference corpus, a small specialized comparable corpus (LSP corpora), and a specialized corpus specifically designed to address interpreters' language expression learning needs. Jun *et al.* [18] built the Chinese/English Translation and Interpreting Learner Corpus (CETILC), which aims to investigate lexical articulation issues and explore the semi-automatic or automatic processing of learner texts by using Learner Translator Corpus (LTC).

Overseas research on Learner Translator Corpus primarily focuses on cross-disciplinary corpus construction. Espunya [19] developed a learner translation corpus for translation teaching, allowing for quick retrieval of examples and illustrating students' mistranslations, thereby facilitating discourse pattern analysis. Reem [20] established an Undergraduate Learner Translator Corpus (ULTC) for parallel corpus analysis studies in Arabic, English, and French, utilizing computational linguistics, Computer-aided Translation (CAT), and Machine Translation (MT) tools. Koyama *et al.* employed a Learner Translator Corpus to assess grammatical errors made by second foreign language learners of Japanese [21].

A. Learner Translator Corpus COCA and Related Studies

Professor Mark Davies created the Corpus of Contemporary American English (COCA) electronic corpus, which is overseen by Brigham Young University's Centre for Language Studies. It is a priceless linguistic resource for academics, teachers, and students and is made up of samples of modern American English from many genres and

linguistic areas. Although COCA was not specifically designed for learners, it can serve as a reference corpus for LTC. Researchers and educators can utilize COCA to compare the language usage of learners and proficient speakers, investigating differences and developmental patterns between learner language and native-like language. Consequently, COCA plays a crucial role in Learner Translator Corpus research by offering substantial reference materials and facilitating comparative analyses.

The COCA corpus of contemporary American English encompasses three fundamental features: Size, Speed, and Annotation [22]. The COCA corpus of contemporary American English covers five major types of corpus which include spoken language, fiction, popular magazines, newspapers, and academic journals [23]. Its extensive size and rapid updates enable users to access a wide range of linguistic data in real-time. Additionally, COCA's annotated nature provides valuable metadata and linguistic annotations that further enhance the corpus's utility.

COCA offers English majors valuable references in three key areas:

1) *Vocabulary acquisition*

COCA aids students in encountering and understanding various terms encountered in their translations, even those not commonly found in standard dictionaries. By searching and analyzing term usage and contexts in COCA, students can grasp the precise meanings of these terms. Furthermore, COCA assists students in developing a rich vocabulary repertoire, learning word collocations and semantic nuances, and enhancing vocabulary accuracy and expressive capacity.

2) *Grammar and pragmatics learning*

COCA contains a wealth of English grammar and pragmatic samples, primarily derived from authentic official English foreign publications. Through the analysis of sentence structures, grammar rules, and grammatical variations in COCA, students can deepen their understanding of complex and challenging English grammar concepts. In cases where students encounter difficulties comprehending example sentences, they can refer to the original text provided in COCA, facilitating understanding and translation. COCA also enables students to identify the frequency and usage patterns of grammatical phenomena, enhancing their grasp of grammar rules, minimizing common grammatical errors, and improving the accuracy of grammatical expressions. Additionally, the corpus facilitates the learning of English discourse conventions and social terminology. By observing and analyzing communicative contexts, polite phrases, and rhetorical devices in COCA, students can enhance their sensitivity and adaptability in English language communication, improving their effectiveness in various language environments and consequently enhancing translation accuracy.

3) *Academic writing and research*

COCA serves as a valuable reference resource for English majors engaged in academic writing and research. Students can utilize COCA to locate examples and citations from relevant literature, gain insights into academic writing conventions and paradigms, and enhance the quality and scholarly rigor of their papers. Furthermore, COCA supports

students in conducting background research, data analysis, and corpus-driven research, such as example queries and translations in lexicography, thereby providing substantial support for translation scholarship.

B. *Translation Corpus CEPIC and Related Studies*

The Chinese/English Political Interpreting Corpus (CEPIC) is an accessible e-Learner Translator Corpus designed specifically for interpreters working with political texts [18]. The corpus serves three primary functions: keyword search, word collocation analysis, and contextual keyword expansion. Additionally, CEPIC enables parallel display of the same speech fragments in up to six versions, ensuring paragraph-level consistency and facilitating horizontal comparison for enhanced learning opportunities. Officially, the CEPIC corpus comprises 6.5-million-word tokens, sourced from two main channels: original speeches and their translations collected from government websites, and audio/video recordings obtained from government websites and TV show archives, with subsequent revision or re-transcription of the speeches and interpreted texts.

Previous research on this corpus has primarily focused on political discourse interpretation, gender differences in interpretation, and linguistic features. Regarding political discourse interpretation, Pan *et al.* [24] examined contrast markers in English and Cantonese and observed that interpreters tend to use fewer negation and contrast markers when translating Cantonese speeches by senior government officials, instead opting for additional contrast markers indicating concessions and topic shifts. Sha [25] analyzed value-laden keywords in Hong Kong political discourse. In terms of gender differences in interpreting, Xiong [26] explored gender-specific variations in interpreting vague words, finding that women tend to employ vague words when describing individuals and events, while men mainly utilize "seems" to describe events. Finally, Zhao and Yan [27] investigated the linguistic characteristics of Chinese-English visual interpretation by examining the CEPIC corpus, shedding light on the linguistic traits of Chinese-English interpreters.

The CEPIC Learner Translator Corpus offers several benefits for English majors in the following areas:

1) *Language imitation and model construction*

The CEPIC corpus provides genuine language samples from the realm of political interpreting, including speeches and press conferences. English majors can enhance their interpreting abilities by imitating and assimilating the language features present in these samples. Consequently, they can familiarize themselves with the terminology, language styles, and expressions employed within the political domain. By drawing on the exemplary sentences in the corpus, students can construct their own language models and apply them to their own political interpretation practices. Notably, the LTC's search functionality (Table 1.) allows users to specify speaker gender, interpreter gender, language, interpretation form, and speech time-frame, enabling targeted academic research on linguistic phenomena and developmental patterns in political interpretation. For instance, if you search for "patriotism" in the search box, as shown in Table II (from the official website: <https://digital.lib.hkbu.edu.hk/ceplic/search.php>), you can

select not only the city, gender, and name of the speaker of the term on the right, but also the gender of the interpreter. Therefore, students can investigate Chinese and Western speech styles, gender-based language style research, and the impact of different interpretation forms on the interpreted language for specific vocabulary, thereby aiding the distinction and emulation of language styles employed by speakers of diverse genders and nationalities. Moreover, teachers can leverage the Learner Translator Corpus to offer students tailored teaching materials and feedback, helping them overcome language-related challenges.

Table 2. Search bar

Items	Options (alphabetical order)
Part of Speech	All/Cantonese/English/Putonghua
Location	All/Beijing/Hague/Hong Kong/Lima/London/Wahingtong DC
Speaker Name	All/A...B...
Speaker From	All/21st Century Economic Report...
Speaker Role	All/Chair...
Speaker Gender	All/Female/Male
Speaker Language	All/Cantonese/English/Putonghua
Delivery Mode	All/Dialogue/Monologue
Interpreter Gender	All/Female/Male
Interpreter Language	All/Cantonese/English/Putonghua
Interpreting Mode	All/SI
Time	from 1997 to 2017 (optional)

2) *Accumulation of professional knowledge and vocabulary*

Political interpretation is a specialized field that demands interpreters possess extensive professional knowledge and vocabulary related to politics, international relations, and law. Through the analysis of political interpretation samples in the CEPIC corpus, English majors can gain insights into common terms, expressions, and relevant background knowledge within the political domain, thereby enhancing their professionalism in political interpretation.

As depicted in Table 3, conducting a search for “patriotism” yields 16 results, presenting sample sentences arranged chronologically. For a more visually accessible presentation, students can download an Excel table containing all the sample sentences by clicking on the download bar. This facilitates comprehensive study and organization. Additionally, the corpus provides collocations for each word, with related example sentences readily accessible by clicking on the corresponding word, highlighted in bold font. This feature streamlines the retrieval and mastery of professional interpretation examples associated with the target words, enabling students to learn correct expressions and enhance their interpretation proficiency.

1) *Acquisition of translation strategies and techniques*

Political interpreting necessitates the application of various translation skills, such as language shifting, background supplementation, and cultural adjustment. The samples in the LTC exemplify the utilization of different translation strategies, providing students with valuable examples and references. By analyzing political interpretation samples in the LTC, students can comprehend diverse translation decisions and techniques, thereby honing their own translation skills in political interpreting.

Table 3. Example sentences and collocations

Keyword in Context (16 records found)			
Year	Location	Speaker Name	Examples
1997	Beijing	Li Ping	... in/IN a/DT down-to-earth/JJ manner/NN,/PU education/NN on/IN patriotism/NN ,/PU collectivism/NN ,/PU and/CC socialism/NN,/PU ...
1998	Beijing	Li Ping	... /PU Vigorous/JJ efforts/NNS were/VBD made/VBN to/TO promote/VB patriotism/NN ,/PU collectivism/NN ,/PU socialism/NN ,/PU and/CC ...
2000	Beijing	Zhu Rongji	... and/CC ethical/JJ progress/NN ,/PU Efforts/NNS to/TO promote/VB patriotism/NN ,/PU collectivism/NN and/CC socialism/NN should/MD be/VB ...

Note: Top 20 collocates of “patriotism” based on this search.

(You can click on a collocate to narrow down the search)

A, And, Centered, Collectivism, Debates, Democracy, Education, Efforts, In, National, Not, Of, On

Table 4 illustrates the side-by-side comparison of the original text and its translation, with the Mandarin version of the original Chinese text and its annotated version displayed on the left, and the original English text and its annotated English version on the right. This layout facilitates a one-to-one correspondence between the source and target languages, while also highlighting the search terms. The overall arrangement streamlines student reference, reducing the likelihood of mistranslation caused by formatting issues. Furthermore, each word in the content is labeled, enabling students to swiftly grasp word usage. Through comparisons with exemplary samples in the LTC, students can analyze their own political interpretation samples, engaging in reflective self-assessment. This practice helps identify linguistic and translation issues, guiding them toward areas of improvement.

Table 4. Example sentences and collocations

Putonghua Raw	English Raw
推进文化改革发展。用中国梦和中国特色社会主义凝聚共识、汇聚力量，培育和践行社会主义核心价值观，加强爱国主义教育。.....	We will promote the reform and development of the cultural sector. We will draw on the Chinese Dream and socialism with Chinese characteristics to build consensus and pool strength, nurture and practice the core socialist values, and make greater efforts to foster patriotism.
Putonghua Annotated	English Annotated
[推进]推进文化改革发展。用中国梦和[中国特色社会主义价值观]中国特色社会主义凝聚共识、汇聚力量，培育和践行社会主义核心价值观，加强爱国主义教育。.....	We will promote the reform and development of the cultural sector. We will draw on the Chinese Dream and socialism with Chinese characteristics to build consensus and pool strength, nurture and practice the core [...]socialist values, [...]and make greater efforts to foster patriotism.

V. CONCLUSION

In summary, this paper aims to introduce and classify Learner Translator Corpus, with a specific focus on two online free corpora available for English learners: COCA and CEPIC. Besides, the LTC significantly impacts English majors in three main areas. Firstly, in terms of language learning guidance, the LTC provides English majors with a wealth of language samples from other learners,

encompassing writing texts, spoken dialogues, grammatical errors, and more. Secondly, regarding language knowledge and skill development, LTC assist students in improving their language proficiency. By observing the language usage patterns of other learners, students can acquire new vocabulary, syntactic structures, idiomatic expressions, and apply this knowledge to their own language expressions. Lastly, for genre and professional language studies, the LTC presents language with examples and English majors can utilize the LTC to explore common language features, vocabulary, and expressions within their respective professional domains. This knowledge holds significant importance for students preparing for future careers in professional writing, translation, and interpretation.

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

- [1] J. Xu and X. W. Yang, "A corpus study on the use of high frequency verbs in English translations of academic works by advanced Chinese English learners," *Shanghai Journal of Translators*, no. 1, pp. 77–80, Feb. 2012.
- [2] S. Gass and E. Varonis, "Input, interaction, and second language production," *Studies in Second Language Acquisition*, vol. 16, no. 3, pp. 283–302, 1994.
- [3] Tools. BFSU Corpus Research Group. [Online]. Available: <http://corpus.bfsu.edu.cn/TOOLS.htm>
- [4] What is Sketch Engine? Sketch Engine. [Online]. Available: <https://www.sketchengine.eu/>
- [5] M. C. Liang, W. Z. Li, and J. J. Xu, *Using Corpora: A Practical Coursebook*, Beijing: Foreign Language Teaching and Research Press, 2010, p. 14.
- [6] Corpus Tools. Corpus Application and Research. [Online]. Available: <http://cascorpus.com/tool-list>
- [7] J. J. Xu, *Corpora and Discourse Studies*, Beijing: Foreign Language Teaching and Research Press, 2019, p. 143.
- [8] Corpus. [Online]. Available: <https://zh.wikipedia.org/wiki/%E8%AF%AD%E6%96%99%E5%BA%93>
- [9] T. McEnery, R. Xiao, and Y. Tono, *Corpus-Based Language Studies: An Advanced Resource Book*, London: Routledge, 2006.
- [10] H. F. Lu, L. Jiang, J. B. Zhou *et al.*, "On the construction of a C-E/E-C translation corpus of signs," *Contemporary Foreign Language Studies*, no. 10, pp. 48–78, Oct. 2015.
- [11] K. B. Hu, *Introduction to Corpus Translation*, Shanghai: Shanghai Jiao Tong University Press, 2011, pp. 28–31.
- [12] T. W. Li, "Corpus and translation teaching," *Chinese Science & Technology Translators Journal*, no. 3, pp. 46–49, Aug. 2007.
- [13] M. Baker, G. Francis, and E. Tognini-Bonelli, *Corpus Linguistics and Translation Studies: Implications and Applications*, London: Routledge, 1993.
- [14] J. Y. Wang, D. F. Li, and L. Q. Li, "A study of pauses in EFL learner's interpreting based on PACCEL-S corpus," *Foreign Language Education*, vol. 40, no. 5, pp. 78–83, Sept. 2019.
- [15] L. Z. Zhao and A. G. Wang, "On the establishment and application of English for science and teaching corpora—A research based on an EST (aeronautics) corpus," *Journal of University of Shanghai for Science and Technology (Social Sciences Edition)*, vol. 36, no. 3, pp. 201–206, Sept. 2014.
- [16] G. R. Dai and S. J. Zuo, "The application and research of corpus in translation quality assessment," *Foreign Language Education*, vol. 42, no. 2, pp. 92–96, March 2021.
- [17] Y. F. Zhu, J. B. Wang, and X. H. Yang, "Corpora and translator education: Exploration and prospects," *Foreign Language Education*, vol. 37, no. 4, pp. 91–95, July 2016.
- [18] J. Pan, T. M. Wong, and H. Wang, "Navigating learner data in translator and interpreter training: Insights from the Chinese/English Translation and Interpreting Learner Corpus (CETILC)," *Babel*, vol. 68, no. 3, pp. 236–266, April 2022.
- [19] A. Espunya, "The UPF learner translation corpus as a resource for translator training," *Language Resources and Evaluation*, pp. 33–43, vol. 48, 2014.
- [20] R. F. Alfuraih, "The undergraduate learner translator corpus: A new resource for translation studies and computational linguistics," *Language Resources and Evaluation*, vol. 54, pp. 801–830, Sept. 2020.
- [21] A. Koyama, T. Kiyuna, K. Kobayashi, M. Arai, and M. Komachi, "Construction of an evaluation corpus for grammatical error correction for learners of Japanese as a second language," in *Proc. the Twelfth Language Resources and Evaluation Conference*, 2020, pp. 204–211.
- [22] M. Davies, "The advantage of using relational databases for large corpora: Speed, advanced queries, and unlimited annotation," *International Journal of Corpus Linguistics*, vol. 10, no. 3, pp. 307–334, Jan. 2005.
- [23] X. F. Wang, M. Davies, and G. H. Liu, "A good platform for English teachers and learners: The Corpus of Contemporary American English (COCA)," *Technology Enhanced Foreign Language Education*, no. 5, pp. 27–33, Sept. 2008.
- [24] J. Pan and B. T. M. Wong, "A corpus-driven study of contrastive markers in Cantonese–English political interpreting," *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, vol. 9, no. 2, pp. 168–176, May 2018.
- [25] L. Sha, "A corpus-based study of ideological manipulation from appraisal theory in Hong Kong political interpreting discourse," in *Proc. International Symposium on Second Language Acquisition-Based Language Pedagogy*, 2019, pp. 304–311.
- [26] B. Xiong, "Gender difference in vague words: A corpus-based study in Chinese-English political interpreting," in *Proc. the 2022 5th International Conference on Humanities Education and Social Sciences (ICHESS 2022)*, 2022, pp. 2470–2480.
- [27] H. Zhao and J. X. Yan, "Linguistic features of interpretese: A corpus-based analysis of Chinese-English interpreting products," in *Proc. the 1st UK-China Symposium on Translation Studies Conf.*, 2021, pp. 45–46.

Copyright © 2024 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).