

Shuquan Li and Manfei Xu*

Corpus literacy empowerment: taking stock of research to look forward for practice

<https://doi.org/10.1515/jccall-2022-0002>

Received October 26, 2021; accepted December 24, 2021; published online April 7, 2022

Abstract: Over the past decades, corpus linguistics, as an approach to the study of language, has attracted much attention. Applications of its method and various corpus resources have been made in the field of language education as well. However, research in corpus linguistics has not yet been sufficiently converted into a language teacher training perspective especially in the primary and secondary education settings. We therefore call for the development of the corpus literacy of EFL (English as a Foreign Language) teachers as part of their continuous professional development. In support of this call, we examine, from the perspective of a teacher educator, a tentative teacher training model based on the published literature, for student teacher trainees and in-service English practitioners. This paper also offers detailed instructions on how to facilitate the teaching of vocabulary, grammar and discourse with free and user-friendly corpus tools (i.e., SKELL, VersaText and AntConc) in pre-tertiary language classrooms. Some pedagogical implications of the cultivation of EFL teachers' corpus literacy and major concerns when implementing corpus-based language teaching and learning are also discussed.

Keywords: corpus-based language pedagogy (CBLP); corpus linguistics; corpus literacy; language teacher education; teacher training model

1 Introduction

The increasing number of corpus resources available, including the growing body of related published literature, attests to the growing popularity of corpus-based language teaching and learning. According to Reppen (2010), a corpus is a large, structured collection of naturally occurring language stored electronically. It is therefore a machine-readable collection of authentic spoken and written texts that

***Corresponding author: Manfei Xu**, South China Normal University, Guangzhou, China,
E-mail: xumanfei@m.scnu.edu.cn

Shuquan Li, South China Normal University, Guangzhou, China,
E-mail: harris_lee@m.scnu.edu.cn

are systematically compiled and sampled to be representative of a particular language or language variety. The accessibility of these naturally occurring texts appears to significantly augment the authenticity of the language environment by showcasing multiple examples of attested language use and providing access to specialised registers. When we attempt to explore the vast amount of language data available, we might generally abide by four principles postulated by Biber, Conrad, and Reppen (1998): (1) analyse the actual patterns of language use in natural texts; (2) utilise a large and structured collection of natural texts as the basis for the analysis; (3) use computers extensively for the analysis and (4) employ both quantitative and qualitative analytical techniques.

Corpus tools and techniques came to be used in language teaching and learning shortly after they took on their modern form in the 1960s (Vyatkina & Boulton, 2017), first in indirect applications (i.e., use of corpus-derived information for improved language descriptions leading to new dictionaries and other resources) and later, in direct applications (i.e., search for and use of corpora by language teachers and learners) (Boulton & Vyatkina, 2021). For the direct applications of different types of corpus tools—ranging from *generating a word list or keyword list*, *producing clusters and N-grams* and *calculating frequencies of search terms* to *offering statistical information on the strength of word associations*—it seems that corpus investigators tend to favour the use of *concordancing*, which displays all the occurrences of word(s) or phrase(s) from the corpora. Concordancing may be preferred because by observing concordance lines with the search term (i.e., the node) highlighted and centred in each line, certain lexicogrammatical patterns of specific word(s) or phrase(s) can be visualised, obtained, retained and summarised (Ma & Mei, 2021; Pérez-Paredes, 2022).

Technically speaking, the use of corpora has two pedagogic benefits. First, as asserted by Vyatkina (2020), discovery learning is made possible, as corpus consultation by different learners encourages them to take different learning paths through access to open-ended language data. This type of inductive, interactive, experiential and analytical exploration facilitates language learners' investigation of socially established and idiomatic expressions in authentic language data. Second, the prevalence of varied corpus resources and corpus-based activities in language teaching and learning promotes language awareness of learners and facilitates their language acquisition if the corpora are used in language classes in a way that encourages data-driven learning (DDL) (Çalışkan & Gönen, 2018; Leńko-Szymańska & Boulton, 2015; Poole, 2020). In fact, the potential pedagogical value of corpora is now unleashed, and a case in point is the newly devised corpus-based language pedagogy (CBLP) which argues that corpus queries allow corpus

investigators to formulate questions, devise searching strategies, interpret raw data based on keen observations and draw solid conclusions about patterns and meanings of language phenomena (Ma, Tang, & Lin, 2021).

Despite the foregoing breakthrough in corpus-based language teaching and learning, this field has “by no means reached full maturity” (Zaki, 2020) and still has “a long way to go before [its] mainstream acceptance in pre-tertiary language education” (Crosthwaite, Luciana, & Schweinberger, 2021). Several reasons for this delayed adoption of corpus-based language teaching and learning have been repeatedly observed. The main reason is the poor availability and functionality of step-by-step guidelines and directions that match learners’ needs across different cultures, language proficiencies and learning contexts (Boulton, 2017; Leńko-Szymańska, 2017), since “corpora may be challenging for young learners to use due to their lack of basic computer skills or knowledge of how to use corpus tools” (Ma & Mei, 2021). Another reason for the delayed adoption of corpora in language teaching and learning is the rational fear of technical issues in operating corpus analysis software, as reported in Farr (2008), Breyer (2011), and Zareva (2017), among other studies. Many of these studies also claimed that interfaces of some corpora are not tailor-made for non-specialist users like language teachers and learners, which may lead to misunderstandings or confusion because some concordance lines may display in segments without concrete contexts. Notably, moreover, a lack of support for training of language teachers on corpus-based language teaching and learning (McCarthy, 2008; Römer, 2011) makes EFL (English as a Foreign Language) teachers much more sceptical of corpora and even resistant to them because corpora might demand from teachers much time and effort (e.g., in buying and installing new software, booking computer labs, building corpora and even teaching learners how to make their own corpora), but they do not guarantee good student learning outcomes.

A more viable way to “bridge the research-practice gap” (Chambers, 2019) in corpus-based language teaching and learning is through the concerted efforts of different stakeholders such as policy-makers, teacher educators and the EFL teachers themselves. Designers of syllabi and curricula should incorporate corpus literacy training into EFL teacher education programs and their continuous professional development programs. To fast-track the successful implementation of corpus-based language teaching and learning in the classroom, it is advisable for teacher educators to provide their trainees step-by-step instructions for preparing corpus-based learning tasks and activities for their students that match their students’ proficiency level. If language teachers themselves remain hesitant about the effectiveness of corpus-based language pedagogy (CBLP), it would be better for them to first participate in workshops that promote the application of corpora in language classroom settings and then embark on corpus-based activities that are

divided into smaller steps with concrete teaching and learning objectives. Once language teachers experience the pedagogic benefits of using corpora, they would be more likely to use corpus-related resources to address their language teaching difficulties.

This review study aims to evaluate studies conducted in the last 20 years on the empowerment of EFL teachers' corpus literacy so as to identify gaps in such research and determine areas of further study. Thus, this study is intended to be of interest and help to EFL teachers, particularly to those of them who seek to integrate corpora in their teaching practice. It is also hoped that this paper would have certain implications for educational policy makers, instructional designers, curriculum developers, and teacher educators for the effective implementation of CBLP. This study is probably the first attempt to call for corpus literacy empowerment for pre-tertiary EFL teachers in a telecollaborative mode after the outbreak of the COVID-19 pandemic. In the ensuing sections, we will focus on state-of-the-art research on corpus literacy empowerment for pre-service teachers and service teachers respectively. We will also consider practical applications of corpus tools for learning vocabulary, grammar and discourse. We will conclude this paper by discussing the pedagogical implications and future directions for the wider use of corpora for language teaching and learning.

2 Corpus literacy: where we are and where we are heading?

In this section, we present a synthesis of studies in literature that had implemented corpus literacy training of EFL teachers. Based on our discussion of these studies, we explore a tentative telecollaborative corpus literacy training model that highlights the connectivity between pre-service teachers and in-service teachers. Then, we will present three step-by-step corpus-based activities centred on teaching vocabulary, grammar and discourse, with reference to the core principles subsumed under CBLP.

2.1 What is corpus literacy and why do we need it?

Mukherjee (2006) coined the term *corpus literacy* and purported that corpus-based data-driven learning requires high learner autonomy and can only be successful for learners with basic corpus literacy. He defined *basic corpus literacy* as: (1)

understanding what a corpus is; (2) knowing what can or cannot be done with a corpus; (3) knowing how to analyse corpus data and (4) knowing how to draw conclusions about language use based on corpus data. Recent studies have enriched his construct of corpus literacy. Heather and Helt (2012) modified the term and contended that corpus literacy comprises a multifaceted set of complex skills in using the technology of corpus linguistics to investigate language and to enhance the language development of students. Callies (2019), in an attempt to integrate corpus linguistics into a curriculum for pre-service English teachers, updated the four key components of corpus literacy according to Mukherjee (2006) and argued that “using corpus output to generate teaching materials and instructional activities” should be included as a significant component.

Over the last few decades, an interesting area in language teacher education programs has been educating teachers on computer assisted language learning (Hubbard & Levy, 2006). In a survey by Mukherjee (2004) of 250 English language teachers in Germany, over 95% of the respondents agreed that their teaching could profit from the introduction of corpora and corpus tools. Recent studies (Frankenberg-Garcia, 2012a; Leńko-Szymańska, 2014; Viana & Lu, 2021) also highlighted the efficacy of improving teachers’ corpus literacy through consciousness-raising tasks at all levels of teacher education. It seems that language teachers with the requisite corpus literacy will be able to use corpora to complement or evaluate other sources of information (e.g., textbooks, grammar references, and dictionaries). For example, aided by user-friendly corpus tools, teachers can develop an enquiring and critical outlook on the cultivation of coursebooks users’ creative thinking by examining how instructional language in teaching and learning activities is organised in textbooks (Li & Xu, 2021). In a bid to maximise the potential of corpora as tools for language teaching and learning, corpus literacy training sessions can be conducted to help EFL teachers learn new skills and to encourage them to use corpus tools in the classroom. Additionally, EFL teachers who have already used corpora for teaching and are convinced of their usefulness can expand their methodological repertoires. This affirms that “empowering teachers with the necessary tools, skills and knowledge in using corpora will lead to the day when corpus resources and their use are no longer the exclusive preserve of corpus linguists” (Huang, 2018).

2.2 Relevant studies on corpus literacy training

The number of corpus literacy training programs across the globe for teachers at all levels have increased significantly in recent years, and the cohorts have become

more diverse: university students (e.g., Ebrahimi & Faghih, 2016), tertiary educators (e.g., Chen, Flowerdew, & Anthony, 2019), primary and secondary teachers (e.g., Crosthwaite, Luciana, & Wijaya, 2021) and non-language-oriented professionals (e.g., Viana & Lu, 2021). In the following subsections, we roughly divide these studies into two cohorts (pre-service teacher trainees and in-service teacher trainees) and investigate if these teacher training sessions could contribute to the success of corpus applications in language education.

2.2.1 Pre-service teacher training

Research on corpus literacy training for pre-service teacher trainees has focused on raising awareness of the use of corpora for language teaching and unveiling their reactions and attitudes on the merits and drawbacks of employing corpus tools and materials in real classroom settings. Most studies have also examined the effectiveness of raising trainees' language awareness and exploring their idiosyncratic responses. O'Keeffe and Farr (2003) offered examples of corpus-based tasks for increasing students' understanding of word classes and of register-related and socioculturally conditioned grammatical choices. They recommended that corpus linguistics be made a component of the education of new language teachers to enhance their language awareness. Similarly, O'Sullivan and Chambers (2006) integrated corpora into language teacher education programs and discovered not only positive reactions, such as access to real and up-to-date language use, but also negative reactions mainly on limited access to appropriate corpora and insufficient technical support. Farr (2008), to promote corpus-based instruction in language teacher education programs in Ireland, launched a program to raise the language awareness of 28 MA students' over two semesters. In the first semester, the educator mainly presented the preliminaries of corpora and demonstrated the use of corpus resources through workshops and tutorials. In the second semester, the students, armed with certain knowledge and expertise on corpora, collaborated on the conduct of corpus-based investigations. A post-course survey uncovered the students' positive predisposition toward corpora, and the findings were generally in line with those of the two studies reviewed earlier.

Corpus literacy training also attracted attention in other European countries besides Ireland. Breyer (2009) introduced to 18 future secondary school teachers in Germany the use of a home-built corpus of EFL textbooks and different English L1 corpora. In this 11-week course, students were trained to directly apply corpus-based activities in the classroom by creating worksheets to supplement the activities in textbooks. The results of the classroom discussions and reflective writing tasks suggested that the students had become more aware of the effectiveness of corpus-based language pedagogy, although they expressed

reservations about the difficulties of corpus queries and analyses. Callies (2016) instructed future English teachers in German primary and secondary schools to use native-language corpora and learner corpora to compare different modes and registers of reporting verbs in English. The student teachers were divided into groups that examined different registers and searched for 10 reporting verbs from the list given in the corpora. They were also required to generate concordances, filter them manually and calculate the lemma frequencies. In Poland, Leńko-Szymańska (2014) offered corpus literacy training to graduate students majoring in applied linguistics for one-semester. The training had three components: (1) an overview of corpus resources; (2) the application of corpora in the teaching of vocabulary, phraseology, grammar, discourse organisation and reading and writing and (3) the compilation and analysis of a self-built small corpus and the production of corpus-based materials and activities. The trainees reported the workshop as useful but requested more training time and more guidance on how to create their own activities. In a follow-up study, Leńko-Szymańska (2017) analysed 53 end-of-semester corpus-based projects submitted by teacher trainees and examined if and to what extent the trainees mastered the sets of skills necessary for independent use of corpora in language instruction. She argued that the trainees were reported as lacking the skills required to integrate corpus use into classroom practice despite their demonstration of a certain mastery of the corpus software for their own use. The study indicated that a semester-long course is not sufficient for pre-service teachers.

In recent years, corpus literacy training has also expanded to North America. Zareva (2017) integrated corpus training into a grammar course for 21 TESOL (Teaching English to Speakers of Other Languages) trainee teachers at a US university. The participants gave the same positive feedback about the training as those in the other studies reviewed. Negative aspects of short-term corpus literacy training that had been reiterated in other studies were also highlighted. Nasmith (2017) conducted a two-month-long investigation in Western Canada of how the benefits of corpora could be introduced during the initial pre-service training. Sixteen trainees in this study reported such short yet intensive pre-service training courses as helpful, especially the course that introduced non-concordancing corpus tools used in the classroom. Poole (2020) revisited the attitudes of pre-service teachers in the Southern U.S. towards corpus-based language learning and teaching in an undergraduate writing course. The results of open-ended questionnaires suggested that the use of ready-made corpus activities afforded opportunities for engaging students and heightening their language awareness. The study not only reassessed the increasing presence of corpus literacy training in TESOL education programs but also shed much light on how an emerging generation of pre-service teacher trainees perceived CBLP.

It is worth mentioning that the inclusion of corpus literacy training in the education program for new language teachers is now on the agenda in Asia and has already generated an increasing number of literature. Ebrahimi and Faghih (2016) introduced free online corpus resources to 32 Iranian MA students of TEFL (Teaching English as a Foreign Language) in a seven-week online course. The evaluations by the pre-service teachers of the corpus-based instruction in the initial stages of the language teacher education degree program were positive, suggesting that online corpus literacy training is a feasible alternative to face-to-face training. Abdel Latif (2021) conducted a longitudinal study on the responses of Arab EFL student teachers to corpus literacy instruction. The study revealed that the learner-centred corpus literacy instruction was generally perceived as positive. The student trainees were actively engaged in the discussions, guided by questions raised by the educator. Their immediate and long-term responses implied that they maintained optimistic expectations about implementing pedagogical applications of corpora in their workplaces.

From the studies we have reviewed so far, it is evident that corpus literacy training is often conducted for a batch of about 15–55 pre-service teacher students. However, more studies have also adopted qualitative perspectives in corpus literacy training, as such in-depth descriptions “help clarify and detail the dynamic, moment-by-moment contextual factors that impact upon the success or failure of CALL (Computer Assisted Language Learning) implementations” (Levy & Moore, 2018). Heather and Helt (2012) evaluated a semester-long introductory grammar course for six pre-service language teachers. The research used a case study approach to examine the development of multiple components of corpus literacy training. The results showed that while this course was generally effective in enhancing the trainees’ corpus literacy, its effectiveness varied among the trainees, which implies that the application of corpora in language teaching could depend largely on the trainees’ individual learning capabilities. The newly published qualitative research of Crosthwaite, Luciana, and Wijaya (2021) reported the initial attempts of Indonesian pre-service primary and secondary school teachers to integrate corpus consultation activities into their EFL lesson plans after following an extensive online corpus literacy training regimen (i.e., online courses, a series of live workshops and individual guidance on lesson planning provided by teacher educators). The trainees’ perceptions of how the teacher educators developed TPACK (Technological Pedagogical and Content Knowledge) (Kohler & Mishra, 2009) in English language teaching practice within the Indonesian formal school context were examined. This may increase the future *buy-in* of Indonesian language teachers in CBLP as well as in a wider range of professional development contexts within and outside corpus linguistics.

More recent studies on corpus literacy training for pre-service language teachers have established frameworks or models that could serve as prototypes for future studies. Krajka (2019) conducted corpus-based investigations in a graduate TEFL module. The author attempted to establish an *apprenticeship model* in two-steps: (1) the students worked in groups to produce corpus-based activities based on the teacher's directions and guided questions and reported their research findings to the entire class and (2) they reflected on what they had done based on the *question-search-conclusion* strategy at the end of the training. Farr and O'Keeffe (2019) developed the following PENSER (meaning *think* in French) framework to assist novice student teachers in developing corpus literacy: (1) **P**roblem identification based on personal teaching practice; (2) **E**mbracing the accepted challenges; (3) **N**oticing the challenges that teachers experienced through observations; (4) **S**olving the problems and (5) **E**xploring and **R**esearching if the challenges have been appropriately overcome or if further engagement is needed. Ma, Tang, and Lin (2021) investigated how a group of TESOL teacher trainees developed their corpus literacy and CBLP in a two-step training framework. In step 1, corpus literacy was initially cultivated in the student teachers through physical classroom training via a combination of lectures on knowledge of corpus linguistics and workshops where they performed hands-on corpus searches. In this scenario, corpus data were used as learning tools. In step 2, TESOL teacher trainees participated in online lessons to consolidate their initial corpus literacy in the form of lesson design. In this context, corpus data were treated as teaching tools. The effectiveness of this four-week corpus-based teacher training scheme and the initial competence of the trainees in using CBLP were measured through (1) analyses and rating of their lesson plans for vocabulary teaching in primary and/or secondary classrooms; (2) a self-designed survey and (3) interviews. The findings may offer some theoretical guidance for providing effective corpus-based training for pre-service language teachers via an online collaborative learning mode, which corroborates with our exploration of a telecollaborative model for pre-service teachers and in-service teachers (to be specified in Section 2.3).

2.2.2 In-service teacher training

Studies on the interface between corpus literacy training and the continuous professional development of practising teachers examined mainly two cohorts (i.e., pre-tertiary school teachers and university academics or professionals). Earlier studies focused on the promotion of corpus literacy among primary and secondary school teachers. Mukherjee (2004) conducted workshops to familiarise 248 in-service English teachers in secondary schools in Germany with CBLP. He collected data through questionnaires that were distributed before and after the

training sessions to investigate the participants' awareness of corpora and the effectiveness of the training. The findings revealed that 80% of the teachers were not aware of corpus before the training; but after the workshops, most of them regarded corpus data as useful for teaching and proposed that CBLP be added to their teaching agenda. The study findings pointed out the need to sufficiently train secondary teachers in order to help them incorporate corpus data into their teaching experience. Drawing on a similar context, Römer (2009) surveyed 78 German secondary in-service English teachers to find out their needs for quality teaching materials and corpus-based reference resources. The study has built a bridge between corpus research and pedagogical practice by focussing on the situation of language teaching practitioners on the ground and their need for support in their workplace through corpus linguistics. Another area that has been explored in the interface between corpus literacy and in-service teacher education is teachers' use of corpus resources over a longer span of time. Tsui (2004) conducted a seven-year longitudinal study to support primary and secondary English language teachers in Chinese Hong Kong via *TeleNex*, where a number of discussion corners were set up to which teachers can send questions or comments. The study revealed that school teachers tend to use corpora to address language difficulties pertaining to English grammar rules and patterns such as synonymous lexical items, lexical collocations and prescriptive stylistic rules.

Recent studies have claimed the importance of continuous professional development to empower tertiary faculty members with a certain degree of corpus literacy. Two studies both examined a small group of in-service university teachers in Turkey. In the first study, Özbay and Kayaoğlu (2015) examined the awareness and perceptions, by six tertiary language teachers, of CBLP in Turkish language classroom teaching. An eight-week training program mainly concentrated on corpus applications in the teaching of vocabulary and grammar. The results of the semistructured interviews showed that they believed the training improved their language awareness. Those of them who had no prior experience in corpus and its use in classroom practice argued that corpus appeared to have expanded their teaching repertoires and could serve as a complementary teaching aid. In the second study by Çalışkan and Gönen (2018), three in-service EFL instructors at a Turkish state university were offered corpus literacy training for four-weeks on the design and implementation of corpus-based materials to enhance vocabulary instruction. Similar to the conditions in the study of Özbay and Kayaoğlu's (2015), the EFL university instructors did not know before the training how to use corpus-based materials in vocabulary instruction. After the training, they believed that their incorporation of CBLP into their vocabulary instruction could raise their awareness of specific vocabulary items.

The ensuing three studies had more participants and were conducted in other countries across the globe. In Al-Fadl's (2018) research, 19 randomly selected faculty members from the English department of a university in the Kingdom of Saudi Arabia were given a ten-week training course in corpus linguistics and its relation to language teaching. The author confirmed that his attempt to conduct this training course was generally effective even though further professional development of in-service teachers was required before corpus exploration entered mainstream education in language departments and teacher training institutions on a large scale. More recently, Chen, Flowerdew, and Anthony (2019) introduced corpus-based academic writing pedagogy to over 60 in-service English language educators from different major institutions in Chinese Hong Kong. Most of the workshop attendees showed great interest in sustaining their professional development and indicated that the training session was very useful as a starting point for further integrating direct use of corpora in English language teaching. The participants appreciated the benefits of the hands-on activities with step-by-step guidance and teacher demonstrations on the screen, yet they also pointed out concerns about the time and the potential difficulties involved in adopting corpora in the academic writing classroom. Viana and Lu (2021) examined the perception of 28 participants in a UK-based non-credit-bearing continuous professional development project targeted for academics and professionals from a range of disciplines. A questionnaire was administered to investigate (1) their corpus literacy background; (2) their motivations to participate in the project; (3) the profits and disadvantages they encountered as they adopted corpora in their teaching practice and (4) their evaluations of how corpus resources played a part in research practice. The results suggested that a majority of the participants showed strong inclinations to continue applying corpora to both their in-class and out-of-class activities. The study also pointed out the significance of embedding corpus approaches into teaching and research and potentially raising the interest of language teaching practitioners in fields other than language-oriented teaching and learning.

2.3 A telecollaborative model for corpus literacy training: a tentative exploration

In Section 2.2, we systematically reviewed 25 relevant studies of corpus literacy training (17 for pre-service teacher trainees either at the BA or MA level, 5 for in-service academics or professionals and only 3 for pre-tertiary in-service teachers) from 2003 to 2021. It appears that the research foci were lopsided and

less attention was given to training primary or secondary in-service EFL teachers on how to deploy CBLP in classroom teaching. Likewise, collaborations between pre-service teacher trainees whose future job orientation was primary or secondary education and in-service pre-tertiary school teachers somewhat escaped the researchers’ attention. To our best knowledge, it seems that existing corpus literacy training programs do not attempt to connect pre-service student trainees and in-service pre-tertiary school teachers. Although most corpus literacy training programs are implemented in universities, much can be done to promote them in non-university contexts. With the advent of emerging educational technologies, corpus applications in language teaching and learning may now be implemented in other educational environments. Online telecommunications and telecollaborations for corpus literacy training have been empirically supported in previous studies (e.g., Crosthwaite et al., 2021; Ebrahimi & Faghih, 2016; Ma et al., 2021; Tsui, 2004). In line with the claim that “online collaborative learning can energise and empower corpus-based teacher training” (Ma et al., 2021), we crafted a tentative theoretical framework for our proposed telecollaborative model for corpus literacy training between pre-service language teachers and in-service language teachers (see Figure 1 below). This model includes two major components: (1) the teacher educator, the participants and their correlation and (2) the events that will take place in the training.

A teacher educator with expertise in corpus linguistics is viewed as important in the successful application of corpora to language pedagogy in the pre-tertiary

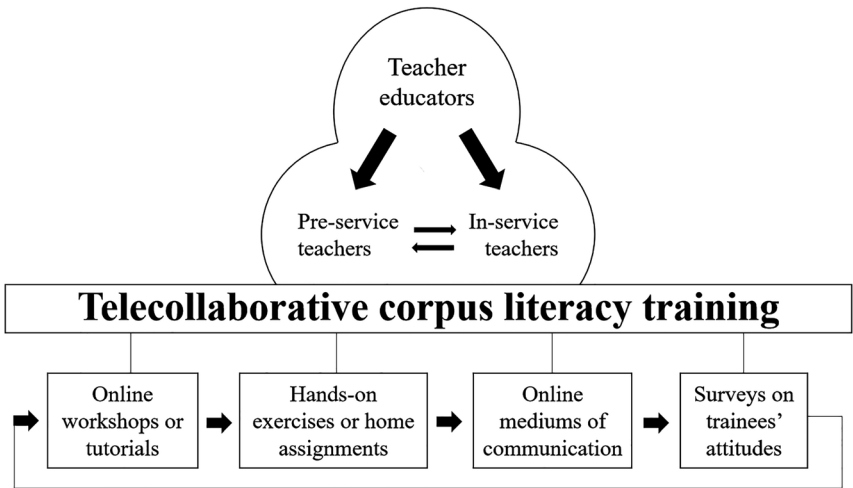


Figure 1: A telecollaborative framework for corpus literacy training.

classroom context because the teacher educator can guide the trainees in the training sessions. As for the participants, combining the pre-service teachers with the in-service teachers may be thought of as posing a challenge, as pre-service student trainees may have little or no teaching practice, whereas experienced in-service teachers are more likely to have different dispositions regarding the use of corpora given that they tend to judge how likely digital technology will work with students based on their classroom teaching experience. However, according to Bolton (2021), “collaborative rather than expert-to-novice training would be beneficial,” and the connection between pre-service student teachers and in-service language teachers could be established and strengthened via online communities. Nambiar and Thang (2015) argued that teachers can use blogs as an avenue to think, reflect and respond to views and comments regarding pedagogical practices and difficulties, thereby developing professionally. Inel Ekici (2017) used two different websites for science and math pre-service teachers to share their knowledge, experience, documents and views. Such online communities provide platforms for teacher educators and trainees to share experiences. Moreover, as the public health threat posed by the COVID-19 pandemic is still evolving globally, in agreement with Lomicka’s (2020) claim that “it is critical to establish a virtual language community and provides practical implications for creating virtual presence and engaging students in these virtual communities,” we argue that online communities can function as virtual hubs that disseminate essential and up-to-date information about CBLP.

To disseminate the use of corpora among the participants, workshops that present basic concepts on the classroom applications of corpora, along with other instructional procedures and techniques, are needed. Being provided with extensive knowledge, sufficient skills, user-friendly tools and free corpora for pre-tertiary language teaching would incline the teacher trainees to normalise corpus use in their language classrooms. To familiarise trainees with the interface and functions of corpus tools and online corpora, most training programs incorporate various hands-on exercises on corpus searches and analyses. In most studies reviewed above, take-away home assignments often require participants to design appropriate corpus-based classroom teaching activities either in the form of individual work or group projects. As suggested by Leńko-Szymańska (2014), corpus exploration cannot be left to a one-semester-long course within a teacher training program; sustainable and continuous follow-up must be extended outside the training sessions. When corpus users conduct searches and analyses individually at their own pace, they might come across some difficulties in their initial attempts to combine corpora with language teaching and learning. In this circumstance, online interaction through social media or websites (cf. Tsui, 2004)

could give a new impetus to the telecollaborations, as online interaction allows both pre-service teachers and in-service practitioners to gather in a virtual space to share knowledge, experience and practices. Avenues such as discussion boards, blogs, forums make it possible for the participants to think, reflect and respond to views and comments regarding pedagogical practices and difficulties on the design of corpus-based tasks or activities. Collecting participants' viewpoints on the implementation of corpus-based language pedagogy via online questionnaires, live video conferences or offline interviews helps examine the development of their corpus literacy after attending the training program. With their post-training assessment and feedback, online workshops or tutorials could be modified. Such a tentative and exploratory theoretical model for corpus literacy training is more likely to ensure that EFL teachers will not only be equipped with good technical and corpus linguistic skills, but will also develop the sound pedagogical skills needed for methodologically appropriate application of corpus-based tasks and materials in the classroom.

2.4 Practical applications of corpus-based language pedagogy (CBLP)

Implementing our proposed model and justifying its effectiveness involve systematic planning and convincing data from a series of empirical studies, which are outside the scope of this review study. Instead, in the following sections, we first summarise some key considerations or principles for designing corpus-based teaching activities in primary or secondary education, followed by detailed illustrations of three cases that we codesigned based on needs analyses for corpus-based language teaching from over 600 in-service teachers (Xu, Liu, & Zhou, 2018). We assume that these ready-made activities could be used to promote corpus-based language pedagogy (CBLP) in the classroom in that they focus on how English teachers could create corpus-based lessons and exploit corpus resources to effectively teach vocabulary, grammar and discourse in an attempt to address the needs of Chinese pre-tertiary English learners.

2.4.1 Issues to consider when designing corpus-based lessons

Corpus-based activities can be useful additions to teachers' arsenal as they can supplement and enhance existing syllabi with digital, flipped and modular content (Vyatkina, 2020). The rich corpus resources can be integrated into regular lessons and everyday teaching without disrupting the normal classroom routine,

and they can complement rather than replace the existing pedagogic materials and practices (Frankenberg-Garcia, 2012b). Ma and Mei (2021) elaborated four design principles of effective corpus-based lessons: (1) testing of student knowledge; (2) hands-on corpus searches by students to observe and analyse the language; (3) inductive discovery by summarising the language use pattern of students and (4) output exercise to practise using the language. With reference to these guiding principles, the following case studies provide step-by-step guidance regarding how to conduct corpus-based graded activities that require only basic internet skills to more complex applications. We also create diverse language use opportunities in the form of concordance lines, word clouds and concordance plots to motivate learners. To promote student engagement and to help reduce the monotony associated with learning with corpora, teachers may use different patterns of classroom interaction such as pair work or group work. when implementing the corpus-based lessons.

2.4.2 Case 1: Using SKELL to differentiate between lexical pairs *say* and *speak*

2.4.2.1 Lesson background

Chinese EFL students always feel puzzled at what follows *say* and *speak* because in Chinese, they share the meaning, “说 (shuō).” Given this concern, we investigate the use of these lexical pairs with SKELL (Sketch Engine for Language Learning), a tool with a simple interface that can easily check how a particular word is used by real speakers of English.

2.4.2.2 Teaching procedures

Step 1: Look up the meanings of *say* and *speak* in an online dictionary.

Before using SKELL, the teacher can test the student’s lexical knowledge of the two verbs by showing their definitions as retrieved from an online dictionary (e.g., <https://www.macmillandictionary.com/us>) (Figure 2).

Step 2: Search *say* and *speak* in SKELL.

Students open two web pages of SKELL (<https://skell.sketchengine.eu/>) and key in *say* and *speak* in the search bar on each page. Then, if they click *Word sketch*, they will see subjects and objects of *say* and *speak* (Figure 3).

say

DEFINITIONS AND SYNONYMS ★★★

VERB **US** /seɪ/

WORD FORMS

DEFINITIONS 11

1. express something using words

2. have opinion

3. mean something

4. give information/orders

5. show what someone/something is like

6. imagine something happening

7. use something as example

8. tell someone to do something

9. repeat words

→ phrases

speak

DEFINITIONS AND SYNONYMS ★★★

VERB **US** /spi:k/

WORD FORMS

DEFINITIONS 7

1. talk about something or to someone

2. use voice to talk

3. talk particular language

4. give formal speech

5. express ideas/thoughts

6. talk on telephone

→ phrases

→ phrasal verbs

Figure 2: Definitions of *say* and *speak*.

say

verb ▾ ✓ Show context

subject of <i>say</i>	object of <i>say</i>
1. official	officials said
2. report	report said
3. source	sources said
4. people	people say
5. expert	experts say
6. Jesus	Jesus said
7. spokesman	spokesman said
8. police	, police said .
9. company	the company said
10. statement	a statement saying
11. Police	, Police said
12. researcher	researchers say
13. Obama	Obama said
14. someone	someone says
15. other	others say

speak

verb ▾ ✓ Show context

subject of <i>speak</i>	object of <i>speak</i>
1. Jesus	Jesus spoke
2. God	God spoke
3. Lord	the Lord spoke
4. Paul	Paul speaks of
5. Obama	Obama spoke
6. voice	voice spoke
7. population	population speaks
8. prophet	prophets spoke
9. people	people speak
10. character	characters speak
11. Bible	the Bible speaks of
12. man	man spoke
13. resident	residents speak
14. Christ	Christ spoke
15. someone	someone speaks

subject of <i>say</i>	object of <i>say</i>	subject of <i>speak</i>	object of <i>speak</i>
1. official	officials said	1. Jesus	Jesus spoke
2. report	report said	2. God	God spoke
3. source	sources said	3. Lord	the Lord spoke
4. people	people say	4. Paul	Paul speaks of
5. expert	experts say	5. Obama	Obama spoke
6. Jesus	Jesus said	6. voice	voice spoke
7. spokesman	spokesman said	7. population	population speaks
8. police	, police said .	8. prophet	prophets spoke
9. company	the company said	9. people	people speak
10. statement	a statement saying	10. character	characters speak
11. Police	, Police said	11. Bible	the Bible speaks of
12. researcher	researchers say	12. man	man spoke
13. Obama	Obama said	13. resident	residents speak
14. someone	someone says	14. Christ	Christ spoke
15. other	others say	15. someone	someone speaks

Figure 3: Words that collocate with *say* and *speak* in SKELL.

Step 3: Guide students to observe and sum up the use patterns.

Based on Figure 3, the students are expected to discuss the subjects and objects of *say* and *speak* in pairs or groups. They may find that the subjects of both words can be *people* (e.g., *people*, *experts*, *police*, *Obama*, *man* or *God*) or *things* (e.g., *report*, *source*, *voice* or *Bible*). Then, they observe the objects of the two verbs. From the top 15 hits, they may discover that *say* is often followed by (1) infinite pronouns such as *something*, *thing*, *anything* or *nothing*; (2) time adverbials such as *yesterday*, *today*, *Thursday* or *week* and (3) concrete nouns such as *word*, *goodbye*, *government*,

company, people, man or official. With regard to the objects of *speak*, the students may discover that the majority of words belong to different types of language (e.g., *English, Arabic, French, Spanish or Mandarin*).

Step 4: Explore another lexical pair.

Upon clicking *Similar words*, word clouds of *say* and *speak* will be generated (Figure 4). The teacher can guide the students in examining other lexical pairs (e.g., *tell vs. talk*). This activity can enlarge students' vocabulary and draw their attention to the use patterns of lexical pairs that might share little semantic differences in Chinese.

2.4.3 Case 2: Adopting VersaText to teach *there be* grammatical structure

2.4.3.1 Lesson background

Chinese students sometimes get confused about the use of *there be*. Some of them may write an incorrect sentence like "There is two years since I last saw you."

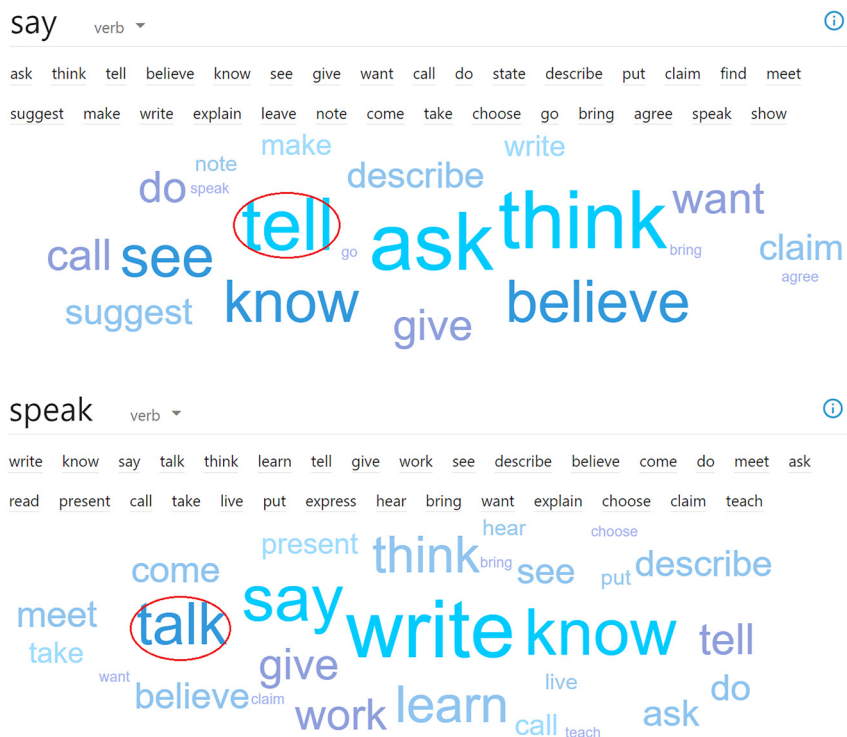


Figure 4: Word clouds of *say* and *speak*.

VersaText can better help them solve the problems. It is an online corpus tool that explores the language features of texts, consisting mainly of three functions: a word cloud, a concordancer and a lexical profiler. It is both part-of-speech tagged and lemmatised, which enable the exploration of many aspects of language, from word forms to genres. Users simply put texts into the *INPUT* box, and the website will analyse them automatically.

2.4.3.2 Teaching procedures

Step 1: Fill in the blanks of the following sentences with the appropriate form of *be*.

- 1) The Earth is a beautiful place. There _____ forests and rivers, mountains and fields.
- 2) A long time ago, there _____ a king in India. His favourite game was chess.
- 3) His eyes were fixed on Della and there _____ an expression in them that she could not read.

[Keys: 1) are, 2) was and 3) was].

The teacher creates some exercises on the use of *there be* to examine students' knowledge of this grammatical structure. Students may become aware of their misconceptions of the use of *there be* and thereby be motivated to explore its use.

Step 2: Extract the concordance lines of *there be* from the textbook corpus.

The teacher builds a small-scale textbook corpus and distributes it to the students. The students open VersaText (<https://versatext.versatile.pub/>) and paste the texts into the box (Figure 5). Students click the tab *WORDCLOUD*, then tick *LEMMA* and *OTHER* of *Include function words* (Figure 6). After clicking the word *there* in the word cloud, the website will produce the concordance lines. After hits in which *there* is used as an adverb are filtered out, 29 instances remain (Figure 7).

Step 3: Observe and sum up the use patterns of *there be*.

With the above concordance lines, the teacher can guide the students in observing the words and phrases used that are juxtaposed after the central node *there*. The students may notice that two types of nouns are adjacent to *there*: countable nouns and uncountable nouns. Countable nouns can be divided into singular nouns (e.g., *fishermen*, line 2; *animals*, line 18 and *many children*, line 23) and plural nouns (e.g., *an expression*, line 5; *one smile*, line 11 and *a flood*, line 29). The

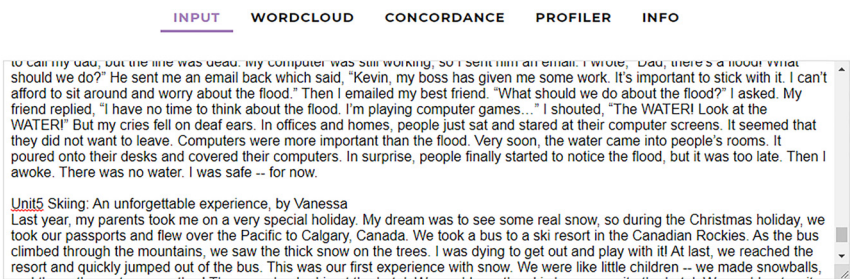


Figure 5: The interface of VersaText after loading the texts.



Figure 6: The query settings of *there be*.

students can also identify uncountable nouns such as *space* (line 6), *water* (line 9), *air* (line 27) and *gravity* (line 28). With these observations, the students can discuss in groups how to sum up the use patterns of *there be*. They may draw the conclusion that if the subjects are countable plural nouns, plural forms of the copula *be* (i.e., *are/were*) could be used; and if the subjects are either countable singular nouns or uncountable nouns, singular forms of the copula *be* (i.e., *is/was*) could be used instead.

people are interested in it. In 50 years, perhaps **there** will be no more cormorant fishermen in the world
 practiced in lots of places in South-East China, and **there** were many fishermen in the area. But today, few young
 difficult, Angela kept getting them right. Soon **there** was just a single question between Angela and victory
 You're wasting water." Dora looked around, but **there** was no one there. "Who are you?" "I'm a drop of water
 inside the door. His eyes were fixed on Della, and **there** was an expression in them that she could not read 5
 her into the living room and saw lots of toys there. **There** was hardly any space for the children to sit down
 worked hard to make their English Week a big success. **There** was an English book fair in the library and a treasure
 students took part in an English singing competition. **There** was also a speaking competition. The students had
 notice the flood, but it was too late. Then I awoke. **There** was no water. I was safe -- for now. Unit 5 Skiing:
 history) Unit 2 The King and the rice A long time ago, **there** was a king in India. The king's favourite game was 10
 very busy, with tired faces, And just one smile. Is **there** just one smile in this city? Or will you make it two
 example, to remember the word "smiles", we can imagine **there** is a "mile" between the first letter and the last
 Some computers are tiny. You may be unaware of them. **There** is probably one inside your TV or washing machine
 Houses: These houses are both "green" and "smart". **There** is glass outside the houses. Heat is held under the
 us with air, water and food. It is our home. Today, **there** is a lot of pollution. We burn things to make energy 15
 Protect The Earth The Earth is a beautiful place. **There** are forests and rivers, mountains and fields. Some
 Some places are very hot, and some are very cold. **There** are many different plants. Some are large. Some
 Some are small. All plants need light and water. **There** are different animals on Earth too. Some live on
 land. Some fly in the sky. Some live under the water. **There** are also many people like you and me on Earth. The
 history and natural beauty, you will love this garden. **There** are many beautiful buildings, bridges and ponds 20
 Helen's grandparents collect almost everything. **There** are eight doorbells on their front door. "This silver
 department stores. France is very famous for its wine. **There** are many vineyards in the centre of France and farmers
 some photos of it. Cindy used them for her painting. **There** are many children without parents. I met some of
 killing them for their fur. It is cruel of them to do so. **There** are not many pandas in the world now. It is important
 idea, says Emma. Keeping pet dogs is a good idea. **There** are lots of reasons for this. First, dogs are really 25
 . Here are their answers. Emily (aged 15): Well, **there** are just three of us: my mum, my dad and me. Have I got
 have to wear a spacesuit to help me breathe because **there** 's no air on the Moon. I'm going to take as many photos
 Earth, so it'll take us about four days to get there. **There** 's no gravity in space, so we'll all be able to float
 still working, so I sent him an email. I wrote, "Dad, **there** 's a flood! What should we do?" He sent me an email

Figure 7: Concordance lines of *there be*.

Step 4: Use *HIDE KWIC* to create exercises.

To consolidate the use of *there be*, students can return to the *WORDCLOUD* tab and tick the *BE + HAVE + DO* box only in the "Include content words" section (Figure 8). When students click *be*, concordance lines will show up and the students can sort them by clicking *Left context*. After *HIDE KWIC* is clicked, the central nodes are hidden (Figure 9). The students can immediately check if they have already mastered the use patterns of *there be*. If they want to generate more exercises for further consolidation, they can input other textbook corpora and follow the same procedures reported above.

protection was chosen and compiled from an English textbook written for senior high school students.

2.4.4.2 Teaching procedures

Step 1: Guess the meaning of keyword(s).

The selected reading passage mainly illustrates the protection of an endangered animal species: antelopes. The students may be unfamiliar with an antelope since it is uncommon to see in daily life. To help the students guess the meaning of *antelope* from the contextual clues, the teacher opens AntConc, clicks *File* to import the reading passage *case3.txt*, searches *antelope** (the asterisk * is a wild card and can represent more than one character) and obtains the concordance lines (Figure 10). The teacher can pose several guided questions such as (1) Where do antelopes live in? and (2) What happened to them between the 1980s and the 1990s? The students can either work in pairs or groups to observe the concordance lines and may conclude that antelopes are a kind of animal that mostly live in Tibet and whose number has dropped by over 50% at that time.

Step 2: Create a keyword list.

To create a keyword list, the students should import a larger reference corpus (e.g., BROWN). First, they click *Tool Preferences – Keyword List* and *Add Directory* to load the reference corpus. Finally, they just click *Apply* (Figure 11) and return to the interface. Next, they need to choose *Sort by Freq* and click *Start*. A keyword list is generated (Figure 12).

Step 3: Sum up the main ideas and discourse structure.

Concordance Hits 10			
Hit	KWIC		File
1	imals. This is why we're here -- to observe Tibetan antelopes. Tibetan antelopes live on the plains of Tibet, Xinjiang,		case3.txt
2	ve're here -- to observe Tibetan antelopes. Tibetan antelopes live on the plains of Tibet, Xinjiang, and Qinghai. Watch		case3.txt
3	he 1980s and 1990s were bad times for the Tibetan antelope. The population dropped by more than 50 percent. Hur		case3.txt
4	d by more than 50 percent. Hunters were shooting antelopes to make profits. Their habitats were becoming smaller		case3.txt
5	ction. Zhaxi and other volunteers watched over the antelopes day and night to keep them safe from attacks. Bridges		case3.txt
6	m attacks. Bridges and gates were added to let the antelopes move easily and keep them safe from cars and trains.		case3.txt
7	n cars and trains. The measures were effective. The antelope population has recovered and in June 2015, the Tibetan		case3.txt
8	lation has recovered and in June 2015, the Tibetan antelope was removed from the endangered species list. The gov		case3.txt
9	ction programmes, since the threats to the Tibetan antelope have not yet disappeared. In the evening, I drink a cup		case3.txt
10	< a cup of tea and watch the stars. I think about the antelopes and what Zhaxi told me. Much is being done to protec		case3.txt

Figure 10: The concordance lines of antelope(s).

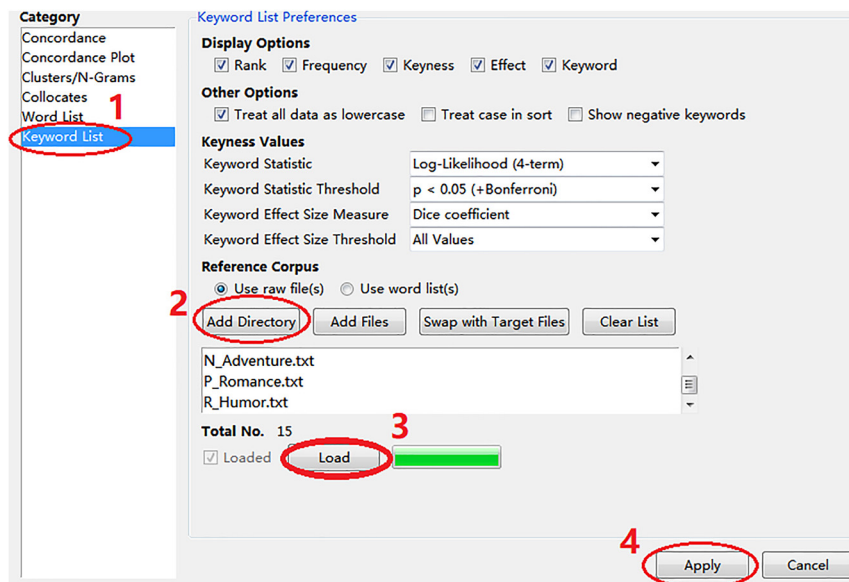


Figure 11: The process of loading a reference corpus.

The teacher can guide the students in grouping the keywords according to their themes. For instance, *antelope(s)* (lines 2 and 4), *animals* (line 7), *wildlife* (line 9), *protection* (line 12) and *species* (line 15) may denote the main subjects introduced in the passage. Places such as *changtang* (line 10), *reserve* (line 13), *tibet* (line 16), *habitats* (line 17), *qinghai* (line 18) and *xinjiang* (line 19) may indicate where the antelopes stay. Other keywords such as *we* (line 1), *save* (line 5) and *zhaxi* (line 6) may suggest the people who are protecting the endangered antelopes.

To verify these intuitive observations, the students can use *Concordance Plot* to search the above keywords in order to investigate their distributions in the reading passage. They can repeat the search process three times by keying in *antelope*|animal*|wildlife*|specie**, *tibet*|changtang|reserve*|habitat*|qinghai|xinjiang* and *we|zhaxi|sav*|protect**, respectively, and they will obtain three concordance plots (Figure 13). The bar represents the reading passage, and each line stands for the distributions of the keywords. The teacher can guide the students in observing the concordance plots and examining if their assumptions regarding the main ideas fit the statistical results. After an in-depth discussion, they may find that the structure of this reading passage generally has three parts: (1) the descriptive information of the antelopes across the whole text; (2) the introduction of places where the antelopes stay at the beginning of the text and (3) the measures taken to protect and save the endangered antelopes from the middle to the end of the text.

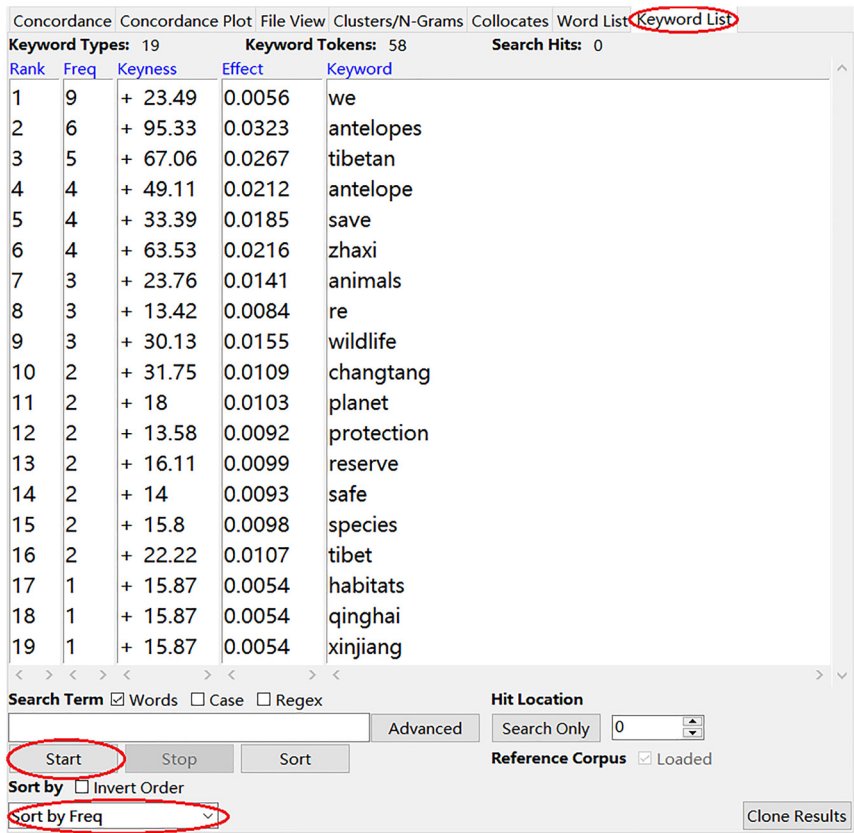


Figure 12: The process of creating a keyword list.

Step 4: Create a mind map to visualise the discourse structure.

The constant use of concordance lines may be monotonous and even boring for most young learners (Sealey & Thompson, 2007). Therefore, non-corpus-based output activities are also recommended in a corpus-based language teaching classroom. The teacher can list several components that should be covered in a mind map and manually draw a simple mind map in the classroom as a demonstration. Students can work in groups to create a mind map of the reading passage based on their discovery in the lesson by adding colourful elements (e.g., pictures of antelopes) to consolidate their corpus findings in a visibly vivid form.

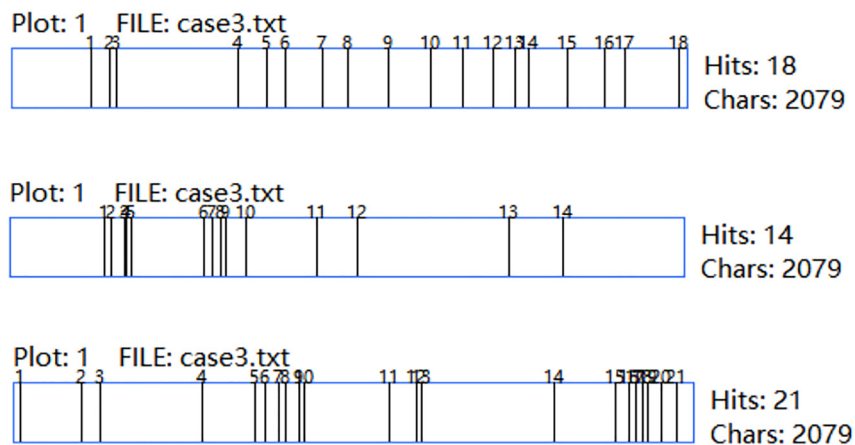


Figure 13: The distributions of keywords across the text.

3 Concluding remarks

This review paper has examined corpus literacy training in detail for EFL teachers and proposed a tentative telecollaborative model for corpus literacy training between pre-service language teachers and in-service language teachers. Three case studies with different angles of the applications of corpus-based language pedagogy were designed for the pre-tertiary language teaching and learning context, allowing EFL teachers to employ corpus resources in other similar contexts.

The ever-growing richness of the field of corpus-based applied linguistics has constantly called for the integration of corpus resources, consultation and analysis in the everyday teaching environment. Our tentative telecollaborative corpus literacy training model might stimulate other researchers' interest in conducting empirical studies to verify its validity and reliability, both of which were left unanswered in this review paper. The scope of the three corpus-based case studies only includes vocabulary, grammar and discourse. Our ready-made activities can be taken as a starting point for further integrating a direct use of corpora in pre-tertiary English language teaching and learning. Ultimately, trainees in the corpus literacy training programs need to be enabled and empowered to pursue their own corpus investigations. Indeed, it is not a task for one researcher or one paper alone. Instead, it requires concerted efforts of different stakeholders. Thus, it is hoped that practising teachers, corpus linguists and even learners themselves will collaborate to create corpus-based tasks or activities in other language areas (e.g., listening, speaking and writing).

Our paper offers a robust picture of corpus literacy training for EFL teachers across the globe over the last two decades and provides an overview of what has been achieved in integrating CBLP into language teaching and learning in our increasingly networked, technologised and mobile world. Corpora should be applied across a range of educational settings, and taking stock of previous research advances our knowledge of the perceived advantages and barriers of embedding corpus literacy training in the initial stages of language teacher education degree programs and continuous professional development projects for practising teachers. The practice-related significance of this review is also evident in that it points out the need to increase the number of corpus literacy training to be delivered telecollaboratively for Chinese EFL teachers.

It is hoped that this review paper will contribute positively to this promising and enticing field not only by putting corpus literacy studies in the limelight but also by empowering teachers with the knowledge, skills and tools they need for more successful and broader application of corpora in classroom teaching. More open-access, hands-on classroom activities (e.g., Le Foll, 2021) as well as paper-based materials (e.g., Crosthwaite, 2019; Friginal, 2018; He, 2017; He, Xu, & Zhang, 2020) may well bring us closer to the day when corpora are no longer viewed as resources only for researchers and corpus linguists. We call for more collaborations not only among teacher educators in CALL but also among practising teachers at all levels of language education so that they can dedicate themselves to promoting CBLP in a wider educational context.

Acknowledgments: This paper is based on a talk delivered by the authors at the 2021 International Conference on CALL on 17 October 2021. We wish to thank all conference participants, the anonymous reviewers, and the editors for their very insightful and constructive comments on an earlier draft of this paper.

References

- Abdel Latif, M. (2020). Corpus literacy instruction in language teacher education: Investigating Arab EFL student teachers' immediate beliefs and long-term practices. *ReCALL*, 33(1), 34–38.
- Al-Fadl, M. A. (2018). Corpus linguistics approach to in-service teacher development. *International Journal of Research in Teacher Education*, 9(3), 44–63.
- Anthony, L. (2020). AntConc (Version 3.5.9) [Computer Software]. Tokyo: Waseda University. Retrieved from <https://www.laurenceanthony.net/software>.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Boulton, A. (2017). Corpora in language teaching and learning. *Language Teaching*, 50(4), 483–506.

- Boulton, A. (2021). Research in data-driven learning. In P. Pérez-Paredes, & G. Mark (Eds.), *Beyond the concordance: Corpora in language education* (pp. 9–34). Amsterdam: John Benjamins.
- Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology*, 25(3), 66–89.
- Breyer, Y. (2009). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning*, 22(2), 153–172.
- Breyer, Y. (2011). *Corpora in language teaching and learning: Potential, evaluation, challenges*. Frankfurt: Peter Lang.
- Çalışkan, G., & Gönen, S. I. K. (2018). Training teachers on corpus-based language pedagogy: Perceptions on vocabulary instructions. *Journal of Language and Linguistic Studies*, 14(4), 190–210.
- Callies, M. (2016). Towards corpus literacy in foreign language teacher education: Using corpora to examine the variability of reporting verbs in English. In R. Kreyer, S. Schaub, & B. Gildenring (Eds.), *Angewandte Linguistik in Schule und Hochschule* (pp. 391–415). Frankfurt am Main: Peter Lang.
- Callies, M. (2019). Integrating corpus literacy into language teacher education. In S. Götz, & J. Mukherjee (Eds.), *Learner corpora and language teaching* (pp. 245–263). Amsterdam: John Benjamins.
- Chambers, A. (2019). Towards the corpus revolution? Bridging the research-practice gap. *Language Teaching*, 52(4), 460–475.
- Chen, M., Flowerdew, J., & Anthony, L. (2019). Introducing in-service English language teachers to data-driven learning for academic writing. *System*, 87, 102148.
- Crosthwaite, P. (2019). *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*. New York: Routledge.
- Crosthwaite, P., Luciana, L., & Schweinberger, M. (2021). Voices from the periphery: Perceptions of Indonesian primary vs secondary pre-service teacher trainees about corpora and data-driven learning in the L2 English classroom. *Applied Corpus Linguistics*, 1(1), 100003.
- Crosthwaite, P., Luciana, L., & Wijaya, D. (2021). Exploring language teachers' lesson planning for corpus-based language teaching: A focus on developing TPACK for corpora and DDL. *Computer Assisted Language Learning*, 1–29. <https://doi.org/10.1080/09588221.2021.1995001> [Epub ahead of print].
- Ebrahimi, A., & Faghih, E. (2016). Integrating corpus linguistics into online language teacher education programs. *ReCALL*, 29(1), 120–135.
- Farr, F. (2008). Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language Awareness*, 17(1), 25–43.
- Farr, F., & O'Keefe, A. (2019). Using corpus approaches in English language teacher education. In S. Walsh, & S. Mann (Eds.), *The Routledge handbook of English language teacher education* (pp. 268–282). London: Routledge.
- Frankenberg-Garcia, A. (2012a). Raising teachers' awareness of corpora. *Language Teaching*, 45(4), 475–489.
- Frankenberg-Garcia, A. (2012b). Integrating corpora with everyday language teaching. In J. Thomas, & A. Boulton (Eds.), *Input, process and product: Developments in teaching and language corpora* (pp. 33–51). Brno: Masaryk University Press.
- Friginal, E. (2018). *Corpus linguistics for English teachers: Tools, online resources, and classroom activities*. New York: Routledge.

- He, A. (2017). *Introduction to corpus-assisted English language teaching*. Beijing: Foreign Language Teaching and Research Press.
- Heather, J., & Helt, M. (2012). Evaluating corpus literacy training for pre-service language teachers: Six case studies. *Journal of Technology and Teacher Education*, 20(4), 415–440.
- He, A., Xu, J., & Zhang, Q. (2020). *Selected cases of corpus-assisted secondary school English teaching*. Beijing: Foreign Language Teaching and Research Press.
- Huang, L. (2018). Taking stock of corpus-based instruction in teaching English as an international language. *RELC Journal*, 49(3), 381–401.
- Hubbard, P., & Levy, M. (2006). *Teacher education in CALL*. Amsterdam: John Benjamins.
- Inel Ekici, D. (2017). The effects of online communities of practice on pre-service teachers' critical thinking dispositions. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(7), 3801–3827.
- Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education*, 9(1), 60–70.
- Krajka, J. (2019). Teacher language awareness and world Englishes—where (corpus) linguistics, digital literacy and teacher training meet. *Crossroads*, 25, 28–51.
- Le Foll, E. (2021). Creating corpus-informed materials for the English as a foreign language classroom. *Open Educational Resource*. Retrieved from <https://elenlefol.pressbooks.com/>.
- Leńko-Szymańska, A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL*, 26(2), 260–278.
- Leńko-Szymańska, A. (2017). Training teachers in data driven learning: Tackling the challenge. *Language Learning & Technology*, 21(3), 217–241.
- Leńko-Szymańska, A., & Boulton, A. (2015). Introduction: Data-driven learning in language pedagogy. In A. Leńko-Szymańska, & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 1–14). Amsterdam: John Benjamins.
- Levy, M., & Moore, P. J. (2018). Qualitative research in CALL. *Language Learning & Technology*, 22(2), 1–7.
- Li, S., & Xu, M. (2021). Features of instructional language for cultivation of creative thinking: A data-driven analysis of Chinese senior high school English textbooks. *2021 international symposium on educational technology (ISET)* (pp. 48–52). IEEE.
- Lomicka, L. (2020). Creating and sustaining virtual language communities. *Foreign Language Annals*, 53(2), 306–313.
- Ma, Q., & Mei, F. (2021). Review of corpus tools for vocabulary teaching and learning. *Journal of China Computer Assisted Language Learning*, 1(1), 177–190.
- Ma, Q., Tang, J., & Lin, S. (2021). The development of corpus-based language pedagogy for TESOL teachers: A two-step training approach facilitated by online collaboration. *Computer Assisted Language Learning*, 1–30. <https://doi.org/10.1080/09588221.2021.1895225> [Epub ahead of print].
- McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching*, 41(4), 563–574.
- Mukherjee, J. (2004). Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In U. Connor, & T. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective* (pp. 239–250). Amsterdam: Rodopi.
- Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art and beyond. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods* (pp. 5–24). Frankfurt am Main: Peter Lang.

- Nambiar, R. M. K., & Thang, S. M. (2015). Examining Malaysian teachers' online blogs for reflective practices: Towards teacher professional development. *Language and Education*, 30(1), 43–57.
- Nasmisth, B. (2017). Integrating corpus tools on intensive CELTA courses. *ELT Journal*, 71(3), 273–283.
- O'Keeffe, A., & Farr, F. (2003). Using language corpora in initial teacher education: Pedagogic issues and practical applications. *TESOL Quarterly*, 37(3), 389–418.
- O'Sullivan, Á., & Chambers, A. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, 15(1), 49–68.
- Özbay, A. Ş., & Kayaoğlu, M. C. (2015). EFL teachers' reflections towards the use of computerized corpora as a teaching tool in their classrooms. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi*, 44(1), 85–104.
- Pérez-Paredes, P. (2022). A systematic review of the uses and spread of corpora and data-driven learning in CALL research during 2011-2015. *Computer Assisted Language Learning*, 35(1–2), 36–61.
- Poole, R. (2020). "Corpus can be tricky": Revisiting teacher attitudes towards corpus-aided language learning and teaching. *Computer Assisted Language Learning*, 1–22. <https://doi.org/10.1080/09588221.2020.1825095> [Epub ahead of print].
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge: Cambridge University Press.
- Römer, U. (2009). Corpus research and practice: What help do teachers need and what can we offer? In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 83–98). Amsterdam: John Benjamins.
- Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205–225.
- Sealey, A., & Thompson, P. (2007). Corpus, concordance, classification: Young learners in the L1 classroom. *Language Awareness*, 16(3), 208–223.
- Tsui, A. B. (2004). What teachers have always wanted to know and how corpora can help. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 39–61). Amsterdam: John Benjamins.
- Viana, V., & Lu, L. (2021). Corpus linguistics and continuous professional development: Participants' prior knowledge, motivations and appraisals. *Revista de Estudos da Linguagem*, 29(2), 1485–1527.
- Vyatkina, N. (2020). Corpora as open educational resources for language teaching. *Foreign Language Annals*, 53(2), 359–370.
- Vyatkina, N., & Boulton, A. (2017). Corpora in language learning and teaching. *Language Learning & Technology*, 21(3), 1–8.
- Xu, M., Liu, X., & Zhou, R. (2018). Internet-based exploration and practices in cultivating English pre-service teachers. *Journal of Guangdong University of Education*, 38(6), 48–54.
- Zaki, M. (2020). Corpus-based language teaching and learning: Applications and implications. *International Journal of Applied Linguistics*, 31, 169–172.
- Zareva, A. (2017). Incorporating corpus literacy skills into TESOL teacher training. *ELT Journal*, 71(1), 69–79.

Bionotes

Shuquan Li

South China Normal University, Guangzhou, China

harris_lee@m.scnu.edu.cn

Shuquan Li is currently a research assistant at South China Normal University. His main research interests are corpus linguistics, language policy and computer assisted language learning (CALL).

Manfei Xu

South China Normal University, Guangzhou, China

xumanfei@m.scnu.edu.cn

Manfei Xu is a professor at South China Normal University. Her main research interests include corpus linguistics, educational technology, corpus-based language teaching & learning and teacher training.