Qing Ma* and Fang Mei

Review of corpus tools for vocabulary teaching and learning

<https://doi.org/10.1515/jccall-2021-2008>

Received April 20, 2021; accepted June 15, 2021; published online September 6, 2021

Abstract: This review aims to introduce corpora as useful tools for facilitating vocabulary teaching and learning. Corpora have long been applied to improve learner language learning, but their direct implication in classroom teaching is rare. This review begins with providing basic concepts related to corpora and then illustrates how corpora can benefit language learning and teaching. To make better use of corpora, careful consideration needs to be given to how to choose an appropriate corpus and what specific corpus search functions should be used. To this end, a new corpus-based language pedagogy (CBLP) is introduced as a new pedagogy to integrate corpora into classroom to facilitate teachers' teaching. CBLP blends corpus linguistics with classroom pedagogy. In addition, four design principles are illustrated to help teachers design effective corpus-based lessons. Finally, a number of important issues are raised to help teachers improve their design of corpus-based lessons.

Keywords: corpora; corpus-based language pedagogy; lexical tools

1 Introduction

In recent decades, how to help students learn and discover language rules from authentic language data inductively has posed a significant challenge for English language teachers. Traditional tools (e.g., textbooks, dictionaries, reference books, etc.) and teacher intuition are unable to fully address this issue. This is an area where corpus linguistics can fill the gap. With the use of corpora, teachers and students can effectively study naturally occurring English language and summarise grammatical patterns and word usage. In addition, teachers can develop corpus-based hands-on activities that can cater to students at different proficiency levels. Students can also use corpora to explore authentic language data and

***Corresponding author: Qing Ma**, The Education University of Hong Kong, Hong Kong, China, E-mail: maqing@eduhk.hk

Fang Mei, The Education University of Hong Kong, Hong Kong, China, E-mail: ffmei@eduhk.hk

Open Access. © 2021 Qing Ma and Fang Mei, published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.

answer their own queries about English language usage and become independent and autonomous language learners.

For teachers, a corpus is a high-speed teaching tool and provides high-quality language samples that can be adapted to create various activities such as gap filling or cloze test for target learners. Moreover, corpora can help teachers encourage active and student-centred learning.

For learners, searching words in dictionaries can be laborious and time consuming. With the help of corpus technology, learners engage in a quicker and more efficient language learning experience. In addition, learners can enjoy a hands-on experience in which they take the initiative to search for and learn from authentic language examples, which can be a welcome break from traditional teacher-dominated classroom teaching where “sit and listen” is the main activity. Furthermore, corpus-based lessons with an appropriate amount of learner interactions and language use opportunities can stimulate learners’ interest and improve learner autonomy.

2 Introduction to corpus

2.1 Definition of corpus

According to Reppen (2010), a corpus “is a large, principled collection of naturally occurring language stored electronically”. It is a large collection of authentic spoken and written texts that are systematically compiled according to specific principles and presented in electronic form. Any genre of texts can be compiled into a corpus including newspapers, Facebook posts, recipes, novels, speeches, scripts, friends chatting, letters, books, magazines, lectures, compositions, memos, etc.

2.2 Definition of concordance line

A concordance line is a line of text taken from a corpus. Each concordance line in a set includes the keyword, that is, the word being studied, and the sentence where the keyword is situated. By reading concordance lines presented in key-word-in-context (KWIC) format, students can obtain and retain certain lexico-grammatical patterns (e.g., collocations) of the target word. Learners can also be guided to observe concordance lines and summarise the language use patterns, which help them to identify and correct their own lexico-grammatical errors.

2.3 Language information to be discovered through corpora

According to Reppen (2010), a corpus can provide insights into language use when intuition fails. Corpora show how people use the language and provide objective evidence of fresh and authentic language use. With authentic language data, a corpus shows which words or phrases are more frequently used than others and which words are typically seen together. In addition, the corpus reveals the register or style of a word, language changes over time as well as the meaning or form variants in different English varieties. For example, it is shown, through searching in a corpus, that the word “Facebook” mainly occurs after 2004, that is, after the online platform was established. In another example, in the American English corpora, the verb meaning “to rehearse” is spelled “practice”, but in the British English corpora, it is spelled “practise”.

2.4 Choosing an appropriate corpus

There are many different kinds of online corpora for users to choose from depending on their needs and level of proficiency. Language teachers may prefer *pedagogic corpora* consisting of “all the language a learner has been exposed to” (Hunston, 2002, p. 16) or *learner corpora* which mark the errors of learners to serve the teaching purpose. In particular, teachers who teach reading or writing may prefer *written corpora*, while teachers who specialise in listening or speaking may choose *spoken corpora* or *mixed corpora*.

Additionally, some corpora are free of charge and widely available, while others are commercialised. There are also corpora of different languages: monolingual, multilingual, parallel and comparable corpora as well as corpora of English varieties: British National Corpus (BNC) and the Corpus of Contemporary American English (COCA).

Furthermore, we may consider how to choose a suitable corpus to cater to the different skill levels of students. For example, the 2 k Graded Corpus in Lextutor is suitable for upper primary students and junior secondary students, while COCA is suitable for senior secondary or university students.

In this review, we will focus mainly on three free online corpora: Lextutor and COCA since they are widely available and can benefit vocabulary learning and teaching considerably.

2.5 Facilitating language learning and teaching with corpora

While the specific search functions of different corpus websites may vary, their major functions are to help users explore form, meaning and use of words and provide large amount of authentic evidence to generalize language use pattern. In this section, we will introduce some key functions of two online corpora to show how corpus tools may facilitate language learning and teaching.

2.5.1 The basic search functions of Lextutor

We can obtain ample examples of target words by searching the keyword (e.g., “comfort”) with the *equals* option or get a wider range of results by searching the keyword with the “family” option. In addition, we may also set *starts* or *ends* option to search for a word with different suffixes or words with the same suffix (e.g., *ness*) based on our needs, as shown in Figure 1 below.

Search Result: comfort

Keyword(s) : equals | comfort | In corpus: Brown (1 million wds)

Keyword(s) : family | comfort | In corpus: Brown (1 million wds)

Search Result: comfort; comfortable; comforting; comfortably; uncomforted; uncomfortable; comforts

Search Result: comfort; comfortable; comforting; comfortably; comforted; comforts

Keyword(s) : starts | comfort | In corpus: Brown (1 million wds)

Keyword(s) : ends | ness | In corpus: Brown (1 million wds)

Possible Result: absoluteness; aggressiveness; agreeableness; awareness; bitterness; blindness; brightness; business; consciousness; darkness; effectiveness; fairness; forgiveness; goodness; readiness; weakness

Figure 1: Options for searching a keyword in Lextutor.

In Lextutor, we can define how to sort the concordance lines using the *controls* options. We can either sort “xx word(s) to the left” or “xx word(s) to the right” to highlight the word before or after the target word for easy observation. For example, if we want to find the use pattern for the verb “insist”, we may search by “1 word to the right” in order to examine the use pattern closely (as shown in Figure 2). Teachers can also use the “gapped” function to create gap-filling exercises of the keyword for students.

Keyword(s): Max In corpus:

OPTION : With associated fam(s) within words to side
and NONE of these words anywhere in the line

CONTROLS : **NEW SORT** ↓ Sort Lines by Mean VP

Sorted By word(s) to of keyword Number of Lines Gapped?

On STARTS or FAMILY searches, "sort by keyword" gives distribution of forms before concordance lines (+ again at bottom)

+ Scan for any recurring word (potential collocate) within words presenting <= times

DEMOS>>

+ **WILDCARDS** + **COLLO-DISAMBIGUATE**

☐ Link Extractor : As discussed [here](#) with examples [here](#)
Extractor URL moved to copiable hyperlink at top of concordance output

013. ☐ t you red handed again, Napper. Why do you always **INSIST** on burgling only second floor flats? Burgla [UK_bncw]
014. ☐ ery- that was the thing. In the bedroom she would **INSIST** on darkness. With his club foot he might we [US_brown]
015. ☐ fit by more faculty attention. Although faculties **INSIST** on governing themselves, they grant little [US_brown]
016. ☐ species in the United States alone. If you really **INSIST** on knowing their names, an excellent book o [US_brown]
017. ☐ off the empty nursery, although the girl tried to **INSIST** on moving back to the quarters to spare Kat [US_brown]
018. ☐ pionship so well could be broken up if Chelmsford **INSIST** on retaining all their players for friendly [UK_bncw]
019. ☐ r the impression he wishes to convey. He does not **INSIST** on telling all he knows about any given sub [US_brown]
020. ☐ t, and the Free Democrats, who picked up 23, will **INSIST** on that before they enter the government. M [US_brown]
021. ☐ f the reactions of previous married students, did **INSIST** on there being included some "hind sight" m [US_brown]
022. ☐ should accompany the shipments of plutonium. They **INSIST** that because the warships would be protecti [UK_bncw]
023. ☐ t all in a literary context, it seems perverse to **INSIST** that counting should stop at the number one [UK_bncw]
024. ☐ e of the processes. In this paper (as in [R]) we **INSIST** that each parallel process declares which g [UK_bncw]
025. ☐ bility will seem unacceptable only to critics who **INSIST** that election must be by votes cast directl [UK_bncw]
026. ☐ of the first program. In our normal form we will **INSIST** that if two expressions are both available [UK_bncw]
027. ☐ ol is no place for a shut trap. You should firmly **INSIST** that no bobby pins or hair pins be worn in [US_brown]
028. ☐ oviet Union has now rallied to her view. She will **INSIST** that no immediate decision is called for on [UK_bncw]
029. ☐ f we were creating a wholly new society, we could **INSIST** that our social, political, economic and ph [US_brown]
030. ☐ rd minimal acceptances, or in other words, why we **INSIST** that the formulas are incomparable. On the [UK_bncw]
031. ☐ waver in his belief in virile law enforcement to **INSIST** that there are other things in American lif [US_brown]
032. ☐ ade in the previous administration. Its spokesmen **INSIST** that there has not been time enough to inst [US_brown]
033. ☐ minimize the number of employee fill-ins? Do you **INSIST** that unneeded salary employees take their v [US_brown]
034. ☐ ility to perform miracles. Exponents of Zen often **INSIST** that very early Zen doctrine opposed the ra [US_brown]
035. ☐ ate income tax would have to go up. Deadman would **INSIST** there would be no need for that. Dandy woul [UK_bncw]
036. ☐ to stop Negroes from claiming equal rights. They **INSIST** they are ashamed of such violence and intim [US_brown]
037. ☐ hich I speak is, I know, not incorruptible. But I **INSIST** upon believing that even when it is lost, i [US_brown]
038. ☐ fingers firmly upon the pulse of their subjects, **INSIST** upon drawing out this legend, prolonging it [US_brown]
039. ☐ g in poetry is that of poets", he felt obliged to **INSIST** upon his duty to be hard to please when it [US_brown]
040. ☐ s have character. They stick to their principles, **INSIST** upon their due, but grow and bloom with dep [US_brown]

Figure 2: Defining the controls options in Lextutor.

2.5.2 Basic search functions of COCA

Apart from searching for words, collocates, synonyms and registers, COCA can search for phrases that cannot be accomplished in Lextutor. In addition, it not only shows the overall frequency of a word in different genres, such as spoken, fiction, magazine, newspaper or academic, but also its frequency at different times with the *chart* function (e.g. Twitter, as shown in Figure 3).

Figure 3: The frequency of “Twitter” in different genres and from 1990 to 2019.

The word “Twitter” is frequently used in blogs and web and is almost non-existent in the 1990s. However, its frequency gradually increases from the 2010s. The *chart* function allows us to view frequency by section, so we may know how language develops, particularly the changes of lexical or grammatical patterns across time.

The *List* function allows users to search for grammatical forms and patterns. There are a variety of sub-search functions under the *list* function, which allows us to search for a word, phrase, part of the word, words that can be used to modify the word, and words that are synonyms with the target word. To begin with, the *wildcard search* function is highly recommended in searching for part of a word. For example, if we are not sure whether to use “incomplete” or “uncomplete”, we may input “*complete” in the search box and then obtain the answer easily by observing the search results, as shown in Figure 4.

Figure 4: Search result for “*complete” from COCA.

If we are not sure of the noun form of a verb, for example, “refurbish”, we may input “refurbish*” and get “refurbishment” but not words ending with -ness or -ity. In addition, we can also perform searches by using *part of speech* function. For example, if we want to know the prepositions that follow “prevent”, we may fill in

“prevent PREP” and results will show common prepositions that follow the word “prevent”, such as “prevent from”, “prevent in”, “prevent against” and “prevent by”. Besides, there is also the *synonym search* function in which we can search for synonyms by filling in the equal sign (=) before the word (e.g., “=caring”), then we may get a variety of choices (kind, concerned, sensitive, helpful) as substitutes. This function is very useful in helping students vary their lexis in writing.

The *Collocates* function is a core feature for COCA and can assist in finding collocations efficiently. For example, if we would like to find out what verbs normally come before “record” as a noun, we can define specifically the number of words before “record” ranging from one to four to know not only its immediate collocates but also collocates in a wider range as shown in Figure 5 below.

Figure 5: Defining the range of verb collocates for “record” as a noun.

Furthermore, the *Compare* function can be very useful. When we want to determine the difference in collocations between two words (e.g., “problem” and “question”), we can select “+1” to the right to see which prepositions are more suitable to be collocated with each word. Additionally, we can check the verb collocates preceding each word to look for meaning differences. As shown in Figure 6, the verbs collocated with “problem” have negative associations while those with “question” are neutral. In this way, we can infer that a “problem” may imply difficulty whereas a “question” may not.

WORD 1 (W1): PROBLEM (1.05)						WORD 2 (W2): QUESTION (0.95)					
	WORD	W1	W2	W1/W2	SCORE		WORD	W2	W1	W2/W1	SCORE
1	CORRECT	271	0	542.0	517.2	1	ASKED	2814	0	5,628.0	5,898.1
2	CAUSING	237	0	474.0	452.3	2	ASKING	1534	0	3,068.0	3,215.3
3	OVERCOME	201	0	402.0	383.6	3	BEGS	1022	0	2,044.0	2,142.1
4	FIXING	194	0	388.0	370.2	4	ANSWERING	1490	1	1,490.0	1,561.5
5	FIX	1530	4	382.5	365.0	5	ASK	4266	5	853.2	894.2
6	FIXED	158	0	316.0	301.5	6	RAISE	618	1	618.0	647.7
7	EXACERBATE	126	0	252.0	240.5	7	BEG	175	0	350.0	366.8
8	SOLVING	911	4	227.8	217.3	8	BEGGING	173	0	346.0	362.6

Figure 6: Verb collocates for “problem” and “question” respectively.

Word and *Browse* are two new functions of COCA which are very useful and informative. The *Word* function can show full word sketch for the top 60,000 words in the corpus. Its “home page” demonstrates basic information of the searched word, including its distribution across genres, definitions, related topics, collocates, etc., with links to other pages such as “collocates, clusters, topics, dictionary, websites, and concordance lines” with more detailed information. We may study a list of “synonyms” of the keyword to understand the meaning nuances and use patterns for the keyword and its related words. In addition, the definitions of a word and common collocates of the word will be displayed for each searched word, which allows students to examine the accurate collocations that can help them correct their collocation errors (e.g., “exercise one’s body” instead of “practice one’s body”). Furthermore, teachers can also help students distinguish synforms (e.g., classic versus classical) since meanings are distinguished not only by definitions but also by collocations (as shown in Figure 7).

Figure 7: Home page for “classic” versus “classical” under the “word” function.

2.6 Implications for teaching

As shown above, corpus tools can assist teachers and students in checking whether they use words in widely accepted ways and help them identify which words are commonly used together to self-correct their own errors and consolidate their learning. In addition to increasing the accuracy of language use, teachers may also help widen students’ lexical diversity through corpus-based inductive

“data-driven learning” as advocated by Johns (1991) and McEnery and Wilson (1997). By searching for the synonyms of a word overused by students, teachers can present words in the same semantic field to help students diversify their vocabulary based on specific contexts. Moreover, by comparing the definition and collocations of the easily confused words, teachers may deepen students’ understanding of the meaning and use of words in multiple contexts so as to help them differentiate and understand lexical meanings. Furthermore, students can own the learning process themselves if teachers guide them to do hands-on corpus search to self-discover and summarise inductively the language use pattern. Afterwards, some output tasks as well as after-class learning homework can be provided to consolidate students’ learning with corpora and help them develop this new learning strategy. Therefore, it is necessary for teachers to recognise the great potential of the corpus-based linguistic approach to improve their teaching design and implement the corpus-based and student-centred approach in their English classrooms.

3 An innovative corpus-based language pedagogy (CBLP)

As pointed out by Mukherjee (2006), corpus-based data-driven learning requires high learner autonomy and can only be successful for learners with basic corpus literacy (CL), which involves understanding what a corpus is, knowing what can or cannot be done with a corpus, knowing how to analyse corpus data and knowing how to draw conclusions about language use based on corpus data. Ma, Tang, and Lin (2021) argue that CL is about using corpus as a learning tool but may not automatically lead to pedagogical skills to apply corpora to classroom teaching. They further point out that teachers need both corpus-based language pedagogy (CBLP) in order to integrate corpora into classroom teaching. CBLP is built upon CL and is defined as: “the ability to integrate corpus linguistics technology into classroom language pedagogy to facilitate language teaching” (Ma et al., 2021, p. 2).

3.1 Four design principles for corpus-based lessons

A corpus-based language pedagogy (CBLP) can be defined as the ability to use the technology of corpus linguistics to facilitate language teaching in a classroom context (Ma et al., 2021). As a new pedagogy, some guidelines are needed to help teachers implement this pedagogy into language classrooms. Based on Gass,

Behney, and Plonsky's (2013) classic L2 acquisition model, Ma et al. (2021) propose these four design principles for corpus-based learning and teaching materials:

- 1) Test student knowledge (detect lexical errors)
- 2) Hands-on corpus searches by students (observing and analysing the language)
- 3) Inductive discovery by students (summarising the language use pattern)
- 4) Output exercise (practise using the language)

The following is an example of a corpus-based lesson designed according to the aforementioned design principles.

The aim is to help upper primary students (P5 or P6) differentiate two verbs ("make" versus "do") which Chinese learners often mix up, leading to haphazard collocations such as "make homework" or "do a mistake". The first step in this lesson is a group matching activity where the teacher can check whether students know the common collocations for "make" or "do" respectively, which aims to identify students' lexical gaps. The next step is to provide students with the collocations for "make" or "do" in edited tables obtained from COCA. Only the collocations that are likely to be known by the targeted students are selected. By observing which words can be collocated with each word in question, students are encouraged to find out the meaning difference between the two. Then the teacher will guide the students to summarise the difference and language patterns. This is followed by a hands-on corpus search for more collocations each word can take from Word and Phrase. Once interest is aroused, students are motivated to do the hands-on search guided by the teacher.

3.2 Important issues to consider when designing corpus-based lessons

The four design principles serve as the initial guidance for teachers and student teachers to design many successful corpus-based lessons available on the website Corpus-Aided Platform for Language Teachers (CAP) developed at the Education University of Hong Kong by a group of corpus linguists. An analysis of all the corpus-based lesson materials found that the well-designed lessons shared three characteristics, which were also important issues to consider when designing corpus-based lessons for classroom teaching, namely how to provide guidance/training for students, balancing the use of corpus and non-corpus resources and creating language use opportunities.

3.2.1 Providing appropriate guidance/training when requesting students work with corpus data

Traditionally, printed concordance lines are suggested to use if no previous corpus trainings has been provided or if computers, tablets or internet connection are not available in the classroom. In this regard, Lin and Lee (2015) provide some guidance, for example, reducing the number of concordance lines, using complete concordance lines whenever possible and providing students with focused guiding questions. In the design of the majority of the corpus-based lessons available on the CAP website, students are provided with hands-on corpus search opportunities instead of printed concordance lines. In addition, they are also provided with clear and step-by-step guidance regarding how to work with corpus data and how to conduct corpus searches.

3.2.2 Balancing the use of corpus and other resources

Working with concordance lines on one's own may be “monotonous”, especially for young or less motivated learners. Therefore, teachers may use different patterns of classroom interaction, such as pair work or group work, to help reduce the monotony associated with learning with corpora. In addition, teachers can use other alternative activities such as stories, games and oral or written activities to enhance the level of student engagement. Furthermore, it is important to provide a concrete context for the lesson by using different resources, such as movies, videos or pictures, to maintain student interest. Last, teachers should consider how to add fun elements to the lessons since learning since that is one of the best strategies to motivate students, especially younger learners.

3.2.3 Creating sufficient opportunities for students to use the target language

According to Gass et al.'s (2013) L2 acquisition model, successful language teaching should provide learners with sufficient language use opportunities at the end of the learning session. Apart from multiple choice questions, gap-filling or sentence-making exercises, some communicative tasks should also be incorporated so that students can use the language with their classmates in semi-authentic situations. Another useful piece of advice is to set a context or use a consistent theme for the whole lesson to activate students' schemata and help them better understand and use the language in context. For example, in a corpus-based lesson designed for P5 or P6 students available on CAP, a popular Disney film, *Frozen*, provides the context where students learn to differentiate two adjectives “boring” from “bored”. Towards the end of the lesson, more scenes from another

Disney film (e.g., *Finding Nemo*) are used to help students extend the learning to other similar pairs such as “frightening versus frightened”, “scaring versus scared”, etc.

3.3 Additional resources for teachers: CAP

The CAP is a website dedicated to training language teachers how to integrate corpus tools and resources into effective classroom teaching to facilitate students’ language learning. The website provides rich information and ample corpus-based resources targeting all language skills—speaking, listening, reading and writing—to help language teachers at all levels develop corpus literacy and take full advantage of this new corpus-based language pedagogy. On this website, readers find systematic teacher training information based on a corpus-driven approach, guidelines for how to design corpus-based teaching materials and useful corpus-aided materials.

The CAP website serves as the ideal pathway to transfer the expertise in corpus linguistics to the broad teacher community in Chinese Hong Kong and other regions/countries. Using this website as a portal, the corpus team of the Education University of Hong Kong has developed and promoted an innovative corpus-based language pedagogy to more than 230 schools in China in the past three years; 25 workshops (physical and online) for more than 800 language teachers have been conducted. Both of these have achieved a significant social impact. After winning the Silver Medal at the 47th International Exhibition of Inventions in Geneva in April, 2019, the CAP website won the 2020 Esperanto “Access to Language Education” Award, organized by CALICO, the Esperantic Studies Foundation and Lernet.net in May, 2020.

4 Limitations of corpora use and possible solutions

As a new language pedagogy, corpus applications in classroom use may increase teachers’ confidence in language teaching with ample authentic examples that help improve students’ language proficiency and learner autonomy. However, no pedagogy or teaching resource is perfect. Teachers should also understand the limitations of corpora use and know how to overcome the limitations with alternative solutions.

First, technical issues must be solved. Corpora may be challenging for young learners to operate since they may lack basic computer skills or have little knowledge regarding how to use corpus tools. Thus, teachers can extract the appropriate concordance lines and print them out for young students (e.g., primary school students).

Second, searching through corpus resources during a lesson may be time-consuming. As the corpus data can be immense, learners may encounter difficulty in reading through all the concordance lines and summarising the different usages of lexical items. To solve this problem, teachers can provide students with hands-on corpus training before implementing corpus-based teaching.

Third, some concordance lines selected by students may be in segments or without concrete contexts, which may lead to confusion or misunderstandings. Therefore, it is advisable for teachers to prepare step-by-step instructions as concrete guidance to help students work with concordance lines, that is, dividing the searching task into smaller steps with concrete objectives.

Last, but not least, corpora are not tailor-made according to students' skill levels. Some concordance lines may be too difficult for students to comprehend, let alone summarise the language use pattern. Therefore, teachers should be very careful in selecting the corpus and concordance lines appropriate to the proficiency level of their students. If necessary, some modification of words or parts of sentences can be made, such as replacing difficult words with easier ones or removing unnecessary parts of a sentence.

5 Conclusion

Corpus resources and tools can provide answers to questions of what, which, how and how often regarding word frequency, collocations, synonyms, register, etc. However, corpora cannot answer the questions of why, that is, why should language be used in a particular way and for what purposes; nor can corpora tell us directly how to distinguish between a new norm and a mistake in language. How to make use of corpora to answer the why questions and to create genuine language use opportunities are entirely at the hands of the teachers. Finally, teachers should be very selective in dealing with concordance lines and be strategic in integrating corpus resources into classroom teaching to facilitate student language learning.

Research funding: The article was funded by the CRAC project (Ref: 03AAB) by the Education University of Hong Kong and it was also supported by Project of Discipline Innovation and Advancement (PODIA)-Foreign Language Education Studies at Beijing Foreign Studies University (Ref: 2020SYLZDXM011).

References

- Gass, S., Behney, J., & Plonsky, L. (2013). *Second language acquisition*. New York and London: Routledge.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Johns, T. (1991). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, 10, 14–34.
- Lin, M., & Lee, J. (2015). Data-driven learning: Changing the teaching of grammar in EFL classes. *ELT Journal*, 69(3), 264–274.
- 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>.
- McEnery, T., & Wilson, A. (1997). Teaching and language corpora (TALC). *ReCALL*, 9(1), 5–14.
- 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, Germany: Peter Lang.
- Reppen, R. (2010). *Using corpora in the language classroom* (Cambridge language education). New York: Cambridge University Press.

Bionotes

Qing Ma

The Education University of Hong Kong, Hong Kong, China
maqing@eduhk.hk

Qing MA is an associate professor at the Department of Linguistics and Modern Language Studies, The Education University of Hong Kong. Her main research interests include second language vocabulary acquisition, corpus linguistics, computer assisted language learning (CALL) and mobile assisted language learning (MALL).

Fang Mei

The Education University of Hong Kong, Hong Kong, China
ffmei@eduhk.hk

Fang MEI is a Research Associate at the Education University of Hong Kong. She received an M.A. in TESOL from The Education University of Hong Kong. She has been a TOEFL teacher, a teacher trainer, and now a researcher. Her research interests include computer assisted language learning (CALL), especially corpus-based language teaching & learning, and teacher training.