ReCALL (2020), 32: 3, 272–290
doi:10.1017/S0958344020000105

ARTICLE

Toward a pattern-based referencing tool: Learner interactions and perceptions

Shu-Li Lai
National Taipei University of Business, Taiwan (shulilai@ntub.edu.tw)

Jason S. Chang
National Tsing Hua University, Taiwan (jschang@cs.nthu.edu.tw)

Abstract

In this process-based study, we introduced a collocation tool with a new interface and advanced search features and examined how a class of EFL college students interacted with it. To elicit their tool consultation behaviors, a vocabulary test with collocation questions was designed. The students' use of the tool to answer the vocabulary questions was screen-recorded for further analysis, serving as the major data source. One-on-one interviews with selected students were then conducted to clarify issues related to the study and their experience in using the tool. The findings indicated that the pattern-based tool was efficient in helping students solve collocation problems. This paper concludes with some pedagogical implications and suggestions for further research.

Keywords: pattern-based referencing tool; collocation; corpus consultation; learner interaction

1. Introduction

Over the last 20 years, there have been considerable advances in vocabulary learning and writing tools, including learners' dictionaries and corpus-based referencing tools. Language learners now have more tools at their disposal, and yet finding information on collocations, word patterns, and usages for productive use remains a challenge due to the limitations of the content and layout of the information provided. In the following, we will discuss the strengths and weaknesses of current dictionaries and corpus tools and introduce a new type of corpus tool that solves collocation problems through pattern-based searches.

1.1 Dictionaries and vocabulary learning

Dictionaries are essential tools for foreign or second language (L2) learners (Hyland, 2003). There has been much discussion on which types of dictionaries benefit L2 learners more (Laufer & Levitzky-Aviad, 2006; Nation, 2003; Tono, 2000). In general, bilingual dictionaries are more accessible and convenient for L2 learners, as they provide information in bilingual form, which makes interpretation easier and faster for the first encounter and for receptive use of the vocabulary. However, due to the fact that a bilingual dictionary often provides less detail and fewer example sentences in comparison with a monolingual dictionary (Nesi & Haill, 2002), if learners are looking up words for productive use, a learners' dictionary is often recommended (Nation, 2003).

Even with the support of various dictionaries, finding information on collocations and lexical grammar in a bilingual or a learner's dictionary is not always easy. The presentation of

Cite this article: Lai, S.-L. & Chang, J.S. (2020). Toward a pattern-based referencing tool: Learner interactions and perceptions. *ReCALL* 32(3): 272–290. https://doi.org/10.1017/S0958344020000105

© European Association for Computer Assisted Language Learning 2020.

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press

information in dictionaries sometimes makes it difficult for L2 learners to find the usage of the words quickly. For example, if learners want to find an appropriate verb to go with "v. advantage of something" to mean "make use of something well" using the online Oxford Advanced Learner's Dictionary (OALD),[1] after typing the word *advantage*, they need to read through several lines before finally finding *take advantage of something* in the idiom section. Sometimes the dictionaries fail to provide the information needed due to the arrangement of the content. For example, if students want to find a verb to go with *pressure* to mean "to take appropriate action to face pressure," they may find idioms like *under pressure* or *to put pressure on somebody*, but may not find any instances of verb + *pressure* if they type the keyword *pressure* in either the Cambridge or Oxford Learner's Dictionary search box. In fact, *to deal/cope with pressure* are very common expressions and patterns; in the dictionaries, students need to type *deal with* or *cope with* to find the phrase *deal/cope with pressure*. As pointed out by Frankenberg-Garcia (2014), dictionaries in their current form might be sufficient for language comprehension, but for language production, they should also focus on collocation and colligation. Arguably, there is a need for supplementary referencing tools, such as corpus tools that provide information on collocations and vocabulary patterns.

1.2 Corpora and language learning

A corpus is a large and principled collection of natural texts in digital format (Biber, Conrad & Reppen, 1998). With a computer program, the data can be processed and rearranged to provide information on frequency, phraseology, and collocation, offering new perspectives and adding understanding to language.

Teaching and learning vocabulary using corpora is often associated with "data-driven learning" (DDL; Johns, 1991) and corpus-based language learning (Johns, Hsingchin & Lixun, 2008). It refers to the ability to use corpus data to figure out what words mean and how they are used in context (Boulton, 2011; Johns, 1991), with either direct consultation of corpora via concordancers (e.g. Dolgova & Mueller, 2019; Li, 2017; Liou, 2019; O'Sullivan & Chambers, 2006; Quinn, 2015; Sun, 2003; Tono, Satake & Miura, 2014; Yoon, 2008; Yoon & Hirvela, 2004) or indirect use of corpora using teacher-prepared printed materials (e.g. Boulton, 2008, 2009, 2010).

Engaging learners in DDL facilitates language learning (Boulton, 2017; Boulton & Cobb, 2017), especially vocabulary learning. Research findings have indicated at least three major benefits of DDL for vocabulary learning: (1) observing concordance outputs raises learners' lexical and contextual awareness (Frankenberg-Garcia, 2012; Tribble, 2002); (2) the process of corpus observation, hypothesis-making, and testing encourages autonomous learning and improves critical thinking skills (Kirk, 2002; O'Sullivan, 2007; Yoon, 2008); and (3) observing concordances helps L2 learners see the relationships and new connections between forms and meanings, which leads to knowledge reconstruction (Sinclair, 2004).

1.3 Limitations of the direct use of corpora

Despite the advantages of DDL and the affordances of concordance tools, there are challenges involved. Learners' direct and independent use of corpora and concordance lines often requires considerable language proficiency so they can formulate questions, decide on the keyword to search, and observe the examples to find patterns (Chang, 2014; Kennedy & Miceli, 2010).

When using a corpus tool as a writing reference, concordance lines could be effective feedback forms, providing authentic examples for error detection and revision (Dolgova & Mueller, 2019; Gaskell & Cobb, 2004; Liou, 2019; Sun, 2003). It may be useful at the final proofreading or revision stage to improve the naturalness of writing (Gilmore, 2009). However, if learners are at the stage of

[1]https://www.oxfordlearnersdictionaries.com/us/definition/english/advantage_1?q=advantage

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press

writing the first draft, observing a large number of authentic sample sentences to generate word patterns and usages can be time-consuming and frustrating (Yoon, 2016; Yoon & Hirvela, 2004). At the initial writing stage, any interruption may disrupt the flow of words and thoughts (Frankenberg-Garcia, 2020; Frankenberg-Garcia *et al.*, 2019; Yoon, 2008). An ideal referencing tool for word use should be able to provide a quick solution, with explicit patterns, so that writers can focus on the content of their work, quickly solve their word problems, and get back to writing. According to Tarp, Fisker and Sepstrup (2017: 496),

> Any consultation of an external information resource inevitably represents an interruption of the activity in question. It may be assumed that most users of these resources just want to go back, as quickly as possible, to what they were doing in order to maintain the focus.

It is also important to understand that although the features of keywords in context facilitate learning, not all learners are able to induce information from concordance lines (Gabel, 2001; Lai & Chen, 2015); more advanced students may be successful, but students at lower proficiency levels may find this challenging. Also, some concordancing tools are quite complicated; for example, presenting the concordance outputs in formats that learners may find difficult to interpret or generalize (Yoon & Hirvela, 2004). As pointed out by Kennedy and Miceli (2001: 88), during corpus investigations by language learners, there is indeed "considerable room for error due to lack of knowledge of the target language." So far, little attention has been paid to the artifact and the interface of the tools that learners interact with (Park, 2012).

1.4 Corpora and corpus tools

Efforts have been made by researchers, program developers, and teachers to provide language learners with various corpus-based vocabulary tools. How concordance outputs are displayed and the sophistication of concordance functions vary depending on how the tool is programmed and the types of corpora being processed. The tools are therefore assigned different names reflective of these specific functions. In this section, we will briefly introduce the major types of corpus tools to show the development of related tools so far.

A typical concordancer can be used to search, access, and analyze language from a selected corpus. For example, Web Concordancer,[2] created by Greaves and Cobb in the early 2000s, allows users to enter a word or phrase and search for multiple examples. For L2 learners, however, observing a large number of authentic example sentences can be challenging. Bilingual tools with the concordance lines in the students' first language (L1) *and* L2 can ease the corpus observation process. An early attempt was made by Liou *et al.* (2006) who developed two user-friendly parallel corpus tools: TotalRecall[3] and Tango.[4] In addition to developing bilingual tools to help learners with more limited language proficiency, another research team developed a "non-scary" version of Sketch Engine called SkELL (Kilgarriff, Marcowitz, Smith & Thomas, 2015: 66), with a user-friendly interface and example sentences that are filtered to show "good examples" only, the so-called GDEX (Kilgarriff, Husák, McAdam, Rundell & Rychlý, 2008).

Another line of tool developers focused on tools for English for academic purposes. Word use in general contexts and specialized contexts may be different. In some circumstances, language is "context-sensitive" (Flowerdew, 2008) or discipline-sensitive. The Prime Machine[5] allows users to compare words, collocations, and even the same word in two different corpora (Jeaco, 2017).

[2]https://www.lextutor.ca/conc/eng
[3]http://candle.cs.nthu.edu.tw/totalrecall/totalrecall/totalrecall.aspx
[4]http://candle.cs.nthu.edu.tw/collocation/webform2.aspx?funcID=9
[5]https://www.theprimemachine.net

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press

AWSuM,[6] an academic word suggestion machine, combines rhetorical moves and lexical bundles, and then auto-suggests common lexical bundles for each move in a selected section of a text (Mizumoto, 2017). While the tools mentioned in this section require users to type the keyword into the search box, the recent development of the automated proofreading system Grammarly[7] is worthy of attention. This system can identify errors at both word and sentence levels and gives real-time feedback.

Advances in natural language processing and machine learning have enabled researchers in computational linguistics and language learning to develop advanced vocabulary learning and writing tools. However, most tools still require learners to observe concordance lines and generate patterns and usages of words by themselves – a bottom-up process. Some tools suggest collocation combinations (e.g. SkELL), but they usually show all kinds of combinations for different patterns. Take *pressure* as an example. The results of a word sketch showed different categories, such as verbs with *pressure* as a subject or as an object, adjectives with *pressure*, and nouns modified by *pressure*. This interface might be good for vocabulary learning, but if the user's purpose is to locate a verb to go with *pressure*, they still need to go through a long list of results first, find the correct category, and then find the correct word to use. To address this issue, the current study introduced a pattern-based tool that allows users to type the search keyword and assign the parts of speech of its collocates to narrow down the search. For ease of observation, this tool displays patterns and their frequency counts in a decontextualized way; if students are interested in the original example sentences, they can click on the "example" button to see more co-texts – a top-down process.

1.5 Toward a pattern-based tool: Linggle

The tool adopted in our study is Linggle,[8] a free web-based service that automatically generates and displays information of recurring word patterns. It allows writers to retrieve phrases that match a submitted query with parts-of-speech wildcards, accompanied with frequency counts to indicate how common a retrieved phrase is. Linggle facilitates fast and convenient access to a wealth of linguistic information embodied in a web-scale data set, Google Web 1T 5-grams[9] (Chang, 2013). It also provides example sentences from the New York Times Corpus. The first version was launched in 2013 (Boisson, Kao, Wu, Yen & Chang, 2013; Chang, 2013). Since then, there have been several revisions to improve its usefulness as a pedagogical tool.

Linggle supports intuitive and powerful queries with keywords, phrases, wildcards, and even parts of speech (see the supplementary materials for a detailed description of Linggle queries). It retrieves phrases that match the query and displays the results (i.e. patterns) in order of decreasing frequency counts. For example, if users want to know the kinds of verbs that collocate with *knowledge*, they simply type "to v. knowledge" or "v. knowledge" in the search box, and Linggle will display the common verb collocates of "knowledge" (see Figure 1). Linggle also allows users to use wildcards such as "*" (zero or more words) and the underscore "_" (one word) to explore words occurring with other words in a query. If users have two words in mind and are not sure which one is correct or better, they can use a slash (/) to separate them, and Linggle will show the percentage of each matching phrase to help make a better decision quickly. For example, if students are not sure whether to use *to reach agreement* or *to achieve agreement*, they can try the search combination "to reach/achieve agreement." As shown in Figure 2, there are fewer usages of *to achieve agreement* (4.3%) compared with the correct usage of *to reach agreement*

[6] http://langtest.jp/awsum

[7] https://app.grammarly.com

[8] http://www.linggle.com

[9] The Google Web 1T 5-gram database (Brants & Franz, 2006) consists of frequency counts for bigrams, trigrams, 4-grams, and 5-grams extracted from 1 trillion words of English web text. Details can be found at https://catalog.ldc.upenn.edu/LDC2006T13

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press



Figure 1. The search for "to v. knowledge"



Figure 2. The search for "to reach/achieve agreement"

(95.7%). This pattern-based tool is different from traditional concordancers, as it displays patterns directly and has the potential to make finding information on collocations and common phrases easier. In this study, we addressed the following research questions:

1. How did the students use the new pattern-based tool to solve the collocation questions, and how did they interact with this new tool?
2. How did the students combine the pattern-based tool with the dictionaries to solve the collocation questions?

2. Methodology

2.1 Participants

The participants were a class of non-English major students ($N = 32$) at a public university in northern Taiwan. Each student was assigned a code, from S1 (the highest ranking) to S32 (the lowest ranking), according to their reading scores on the Test of English for International Communication (TOEIC) (ranging from 485 to 25, with a mean of 184 and a standard deviation of 118.2). These codes were used throughout the study for ease of description and to provide information on the students' language proficiency (see the supplementary materials for detailed information about the participants).

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press

Table 1. Research design

Time	Task
Week 1	• Orientation • Pre-test (no tools)
Week 2	• Vocabulary Test 1 (dictionaries only)
Week 3	• Tool-training session 1 • Introduce the concepts of corpora, collocations, lexical grammar, and word patterns • Introduce Linggle's syntax commands with example sentences(*, _, adj., adv., prep.) • Hands-on exercises • Group discussion on the hands-on exercises
Week 4	• Tool-training session 2 • Review the syntax commands learned in Week 3 • Introduce more syntax commands (/, ~, ?) • Hands-on exercises • Group discussion on the hands-on exercises
Week 5	• Vocabulary Test 2 (Linggle & dictionaries) • Questionnaire with 5 open-ended questions
After Week 5	• One-on-one interviews, 30 minutes each

2.2 Design and data collection

The study was conducted in a freshmen English class. To elicit the students' tool consultation behaviors, a vocabulary test with collocation questions was designed. Two tool-training sessions were scheduled over two consecutive weeks to ensure that the students were well trained in the use of Linggle before data collection. Their use of the tools to find the answers was screen-recorded for further analysis, serving as the major data source, as real-time data are better suited to answering process-oriented questions compared with retrospective data collected from reflective reports and questionnaires (Lai & Chen, 2015; Park, 2012; Pérez-Paredes, Sánchez-Tornel, Alcaraz Calero & Aguado Jiménez, 2011). By analyzing the screen-recorded files, the students' interaction patterns and the difficulties they encountered were revealed. One-on-one interviews with selected students were conducted to help clarify issues related to the study. Table 1 shows the research design and the procedures for the data collection.

In Week 1, the students were given a pre-test (58 items, 15 minutes) to ascertain their prior knowledge of the vocabulary presented. Without the help of any tools, the overall accuracy rate was only 42% (baseline test). This result suggested that the test was difficult enough, and would elicit some tool consultation behaviors.

In Week 2, they were given the same test again, now labeled Vocabulary Test 1, with the same 58 items but a longer test time of 25 minutes. This time, the students were allowed to consult the dictionaries provided, either the Yahoo! Kimo Bilingual Dictionary or the OALD. Dictionaries for general purposes do not always provide collocation information explicitly; thus students often need to find the information in example sentences, which requires more time. In a test-taking context with a time limit, students can find this challenging. The purpose of Vocabulary Test 1 was to raise students' awareness of collocations and direct their attention to the limitations of using a general dictionary to solve collocation problems.

In Weeks 3 and 4, two tool-training sessions on Linggle were scheduled, two hours per week. The students were taught the concepts of a corpus, corpus tools, collocation, and pattern grammar. Then, Linggle's syntax commands were introduced one by one, and were each followed by demonstrations, hands-on exercises, and pair and group discussions. Each week, after the teacher's demonstration, a hands-on exercise worksheet was distributed. When necessary, they combined Linggle with a dictionary of their choice. The training followed the computer-assisted

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press

language learning (CALL) training process suggested by Hubbard (2013), which involves strategy training, theory introduction, spiral teaching, and collaborative debriefing.

In Week 5, after two weeks of tool-training sessions, the students were given the same amount of time to finish Vocabulary Test 2, the same test as before. They were encouraged to use the newly learned Linggle tool to solve the collocation problems and were allowed to use the dictionaries if needed. The screen-recording program recorded the students' online tool consultation behavior for further analysis, including the keywords typed, the syntax commands used, the tools chosen, and the results retrieved. On the same day, a questionnaire with five open-ended questions was distributed, mainly to ask the students about their perceptions of and experiences in using Linggle. One-on-one interviews were conducted with 21 selected students to help clarify issues related to the study. After analyzing students' performances on the vocabulary tests, those who had the following characteristics were selected: (1) were able to make effective use of Linggle; (2) had great difficulties in using Linggle; (3) provided feedback on the questionnaires that needed further clarification.

2.3 Research instruments

2.3.1 Tools: Linggle and two dictionaries

Although the focus of the study was Linggle, the students could also use the OALD and the Yahoo! Kimo Bilingual Dictionary to solve the given problems or to conduct cross-referencing. The purpose of allowing the students to consult a dictionary was twofold. First, Linggle is a tool that provides information on patterns only, so the students may have needed the help of a dictionary to look up the meanings of the words or phrases retrieved. Second, having a dictionary at their disposal allowed the researchers to observe how the students made use of the two kinds of tools together.

2.3.2 Vocabulary test

The vocabulary test comprised five parts, with a total of 58 items that were mostly collocation questions that were incongruent in the students' L1 (i.e. with no direct L1 translation such as *fly kites*). According to the research, most collocation errors made by learners are L1 based, and incongruent collocations are more challenging for them (Laufer & Waldman, 2011; Wolter & Gyllstad, 2011); thus the test questions with mostly incongruent collocations were designed to prompt more tool consultation behaviors so we could investigate how the students interacted with Linggle. We also ensured that all the questions could be answered by Linggle if it was used appropriately. Table 2 shows a summary of the sections, the types of questions, and their purposes (also see the supplementary materials for the detailed descriptions about the test).

3. Results

Although the focus of this study was to explore how students interacted with Linggle, it was still important to know whether they succeeded in finding the answers using the new tool. The first part of the Results section presents the accuracy rates the students achieved in the three vocabulary tests, and then presents the interaction patterns in detail.

3.1 Overall accuracy rates

In this study, the same vocabulary test was given three times: pre-test (baseline test), Vocabulary Test 1, and Vocabulary Test 2. The pre-test (15 minutes) was given before the study began to check students' prior knowledge of the vocabulary presented. Without the help of any tools,

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press

Table 2. Summary of the tasks on the vocabulary test

	Tasks	Question types
Part 1	Read a short news article and identify the parts of speech and the meanings of the underlined words.	8 words underlined for parts of speech The same 8 words underlined for meaning
Part 2	Choose the correct collocation between two distractors.	10 multiple-choice questions
Part 3	Fill in the missing words using the Chinese clues.	10 fill-in-the-blank questions
Part 4	Fill in the blanks with synonymous words.	10 fill-in-the-blank questions
Part 5	Correct 12 collocation errors in a short article.	12 error corrections

Table 3. Accuracy rates of each (sub)section

(Sub)sections of the test	Vocabulary Test 1 (%)
Part 1.1: Part of speech	79.3
Part 1.2: Meaning	79.3
Part 2: Collocations for perceptive use	87.2
Part 3: Collocations for productive use	56.6
Part 4: Synonyms	49.7
Part 5: Error corrections	34.1
M	62.3

the overall accuracy rate was only 42%. This result suggested that the vocabulary test would elicit some tool consultation behaviors in the next stage of the study.

For Vocabulary Test 1, the same items were included, but a longer test time was given as the students were allowed to consult dictionaries only. As shown in Table 3, the mean score reached only 62.3% ($SD = 15.1$). Ten of the 32 students' overall accuracy rates were still lower than 50% (see supplementary materials for more details). In terms of each subsection, the accuracy rates for Parts 3, 4, and 5 remained around 50% or lower. After the test, the students were aware of the limitations of the dictionaries and thus were motivated to learn the new tool.

For Vocabulary Test 2, as shown in Table 4, overall accuracy rates reached 76.6% ($SD = 14.2$) when the students used Linggle as their major vocabulary problem-solving tool. In terms of each student's performance, the lowest score was 48.3% and the highest score reached 94.8% (see supplementary materials). Most of the students used the new tool; only one student (S29) relied mostly on a bilingual dictionary when a Chinese clue was given to find an equivalent L2 in Part 3 of the vocabulary test. S29 pointed out in her interview that it was difficult for her to learn the syntax commands in Linggle, and deciding on the parts of speech of the words searched for was also difficult.

The results of a paired *t*-test indicated that the scores improved significantly in Vocabulary Test 2 ($t = 9.92$, $p < .001$). Although there might have been a practice effect or maturation effect, the improvements some students achieved were so significant that it probably could not be explained

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press

Table 4. Results of the *t*-test by accuracy rates

Test	*N*	*M* (%)	*SD* (%)	*df*	*t*-value	*p*
Vocabulary Test 1: Dictionaries only	32	62.3	15.1	31	9.92***	< .001
Vocabulary Test 2: Linggle and dictionaries	32	76.6	14.2			

***$p < .001$.

Table 5. Results of the t-test for each subsection in Vocabulary Test 2

Vocabulary Test 2		
Subsections	Scores (%)	*SD* (%)
Part 1.1: Part of speech	78.9	14.4
Part 1.2: Meaning	73.8	15.0
Part 2: Collocations for perceptive use	96.9	6.9
Part 3: Collocations for productive use	76.9	21.0
Part 4: Synonyms	85.3	19.5
Part 5: Error corrections	52.3	37.2
M	76.6	19.0

by the practice effect only. For example, S5 improved from 48.3 to 82.8; S12 improved from 70.7 to 91.4; and S16 improved from 56.9 to 87.9.

As each subsection shows, in general, the students managed to solve the various vocabulary problems using Linggle, except for the error corrections section (see Table 5). Part 5 reached accuracy rates of only 52.3%; this result will be explained in a later section when analyzing the students' interaction patterns.

3.2 Consultation patterns

This section reports each part of the test to provide a detailed picture of how the students used and interacted with Linggle (i.e. tool consultation behaviors). The screen-recorded files of the students' use of Linggle were analyzed along with the students' interview data. The results showed a general pattern: when looking for information on collocations and synonyms, the students turned to Linggle and assigned different syntax commands to solve the collocation problems. They also conducted cross-referencing, combining Linggle and a dictionary when needed. When looking for the meanings and parts of speech of the keywords only, dictionaries were still the students' first choice. Part 1 required the students to find the meanings and parts of speech of each keyword, and all the students used dictionaries.

3.2.1 Collocations for perceptive use

For Part 2, the task was to choose one correct collocation between two options. The syntax command that the students used most was the slash (/). For Question 1, *The meeting took almost five hours. It was impossible to (**pay/keep**) attention all the time*, the students typed either "to pay/keep attention" or "pay/keep attention" to retrieve the answer. As indicated by the number of entries (500,000 vs. 3,600) and their percentages (99.3% vs. 0.7%), the students quickly found their

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press

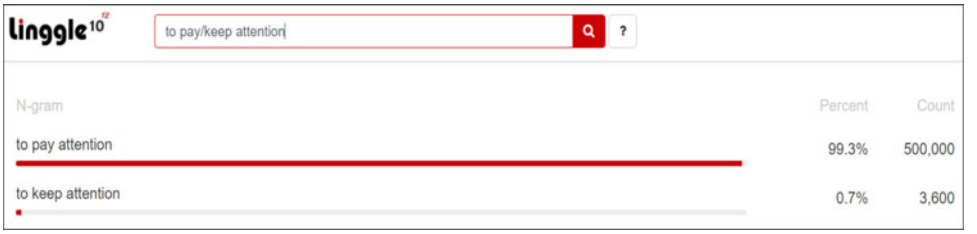


Figure 3. The search for "to pay/keep attention"



Figure 4. The search for "could you do/give me"

answer (Figure 3). A number of students[10] reported that they really liked this feature because this saved them a considerable amount of time.

In total, the students missed 10 items in Part 2, and the accuracy rates reached 98.6%. Of the 10 items the students got wrong, seven of them were caused by their incorrect understanding of the collocations; they answered these questions directly without consulting any tools. For the other three items, the students did consult a tool but still failed to answer the item correctly due to the incorrect chunking of the phrase. For example, for Question 5, *Could you (**do/give**) me a favor and post these letters on your way home*, instead of typing "do/give me a favor," S24 and S27 searched for "could you do/give me," as shown in Figure 4. To avoid this kind of mistake, more guidance is needed to raise students' awareness of patterns and chunking during the learner training session.

3.2.2 Collocations for productive use

For Part 3, the students needed to fill in the right word using the Chinese clues. All the students' primary tool of choice was Linggle, except S26 and S29. The students at different language proficiency levels managed to make some use of the syntax command for the parts of speech and assigned the correct keyword to solve their collocation problems. For example, for Question 3, *You should _____ a record of your progress. (做紀錄)*, most students typed "v. a record" in the query box and saw *keep a record* in the first line of the search results (see Figure 5). The students found this to be a useful and unique feature. As reported by S20, being able to narrow the search by directly assigning the part of speech of the searched word helped to retrieve all the possible combinations quickly. As commented by another student (S4), dictionaries did not always provide as many choices (combinations of words, patterns, etc.) nor frequency counts.

Frequency count also played a role. As Linggle displayed the patterns in order of frequency, the students tended to choose the patterns that had higher frequencies. For example, for Question 9, *The local government was accused (**of/at**) incompetence*, S13 typed "was accused *" and quickly decided to choose "was accused of" as it had the next highest frequency after "was accused" (Figure 6).

[10]S2, S5, S6, S19, S23, and S29.

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press

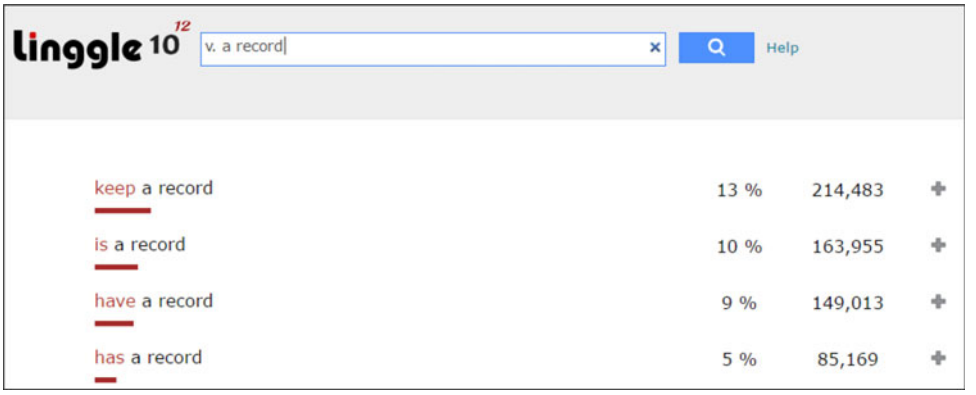


Figure 5. The search for "v. a record"

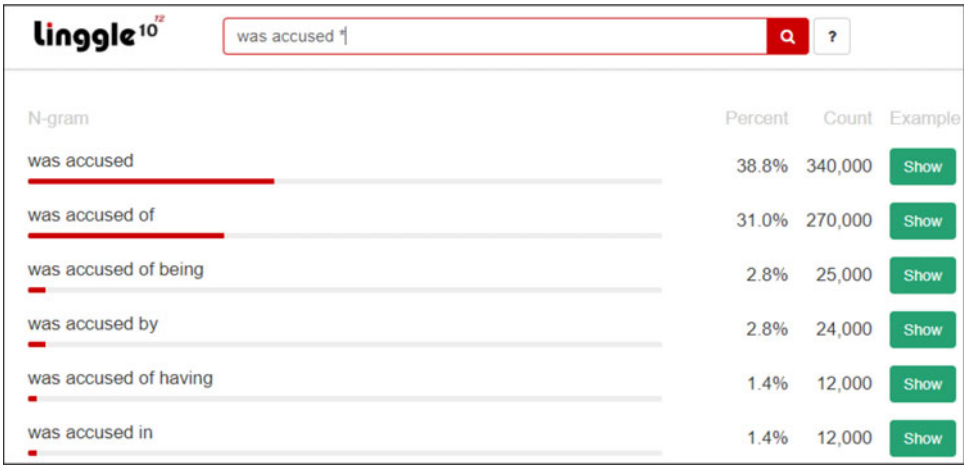


Figure 6. The search for "was accused *"

With regard to the students' consultation processes, 10 different combination patterns were revealed (see Table 6 for the top three patterns and the supplementary materials for the complete set of 10 patterns). Pattern 1, Linggle only, showed the highest frequency as the accuracy rate reached 88%. For Pattern 2 (Linggle + Yahoo!), after using Linggle, the students turned to Yahoo! to further confirm or to check the unknown words they retrieved from Linggle. For Pattern 3 (Yahoo! + Linggle), the students tried the Yahoo! bilingual dictionary first and then turned to Linggle if the dictionary did not provide the answer.

The syntax commands the students applied also showed some variations (see Table 7). The students assigned the part of speech to the searched word (98 times) and wildcards (97 times) to find most of the unknown collocation information, and the success rate reached 85%. The underscore (_) and slash (/) syntax commands were not used often, but the accuracy rates were good, ranging from 73% to 100%. This showed that the students preferred to either assign the part of speech to narrow down the search or use a wildcard to locate possible combinations.

It was interesting to find that there were four students (S2, S5, S14, and S25) who used only the wildcard function in Part 3, and their reasons for doing so were different. The two more proficient students (S2 and S5) favored wildcards because they themselves could quickly sort out the results

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press

Table 6. Patterns of cross-referencing in Part 3

Patterns	Combination of tools	Frequency	Accuracy rates (%)
1	Linggle	76	88
2	Linggle + Yahoo!	40	75
3	Yahoo! + Linggle	35	86

Table 7. Types of Linggle syntax commands the students used in Part 3

Type of syntax	Number of uses	Found the correct answer (frequency)	Accuracy rates (%)
Assigned part of speech (v. a baby)	98	83	85
Used wildcard (to * a record)	97	82	85
Used underscore (try to _ suicide)	11	8	73
Used slash (make/have a call)	1	1	100

from the long retrieved list; thus they did not bother to define the part of speech to narrow down the search. On the other hand, the two students with rather limited proficiency (S14 and S25) used more simple commands, such as a wildcard or an underscore, because they found that assigning the part of speech was too challenging.

3.2.3 Consultation difficulties

With regard to the items the students got wrong, consultation difficulties were found. The first difficulty the students encountered was failing to assign the correct part of speech. V + N was easy, but ADV + V and ADV + ADJ were challenging for students. For example, for Question 8, *The two events are _____ related and the author wants to stress the relationship. (closely 密切相關)*, S17 believed that the missing word was an adjective and thus failed to find the answer, probably as a result of negative L1 transfer.

Section 1.1 of the vocabulary test measured whether the students could identify the parts of speech of eight chosen words in a short news article. The data indicated that the students who did not do well in this section encountered more problems when assigning syntax commands to make effective use of Linggle. For example, for Item 8 in Section 1.1 (Identify parts of speech: *Over the past week, they've blocked key highways and railway lines leading into Delhi*), S18, S24, S31, S29, S31, and S32 identified the highlighted word blocked as a noun instead of a verb; S30 even put "adjective" in the answer box. The lack of this ability influenced their Linggle use. As a result, the scores these students achieved in Vocabulary Test 2 were among the lowest as well.

With some questions, the students assigned the correct parts of speech and keyword in the Linggle search box but failed to "see" the answer in the search results; for example, for Question 7, *I heard that a woman tried to _____ suicide yesterday. (自殺)*, S24 typed "to v. suicide" in the search box (see Figure 7). The answer appeared on the first line, as shown in Figure 7, *to commit suicide*, but S24 was not able to "see" it because she did not recognize the word *commit*. For Question 3, *You should _____ a record of your progress. (做紀錄)*, S8 tried two searches, "* a record" and "v. a record"; the answers appeared on the screen, but she chose *take a record* as her answer because she reported that she believed it was correct based on her prior knowledge of the phrase.

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press



Figure 7. The search for "to v. suicide"

In the interviews, the students reported that they tended to choose words they recognized first when locating information from the search results. If their understanding of a word was not correct, or if they did not recognize the word, they failed to locate a correct word to use. Choosing the correct answer also depended on whether they used a dictionary to help them confirm the meaning of unknown words in Linggle. Some students conducted cross-referencing and thus increased their success rates. Two students (S26 and S29) continued to use a dictionary as their major tool for Part 3. S26 reported that his English was not good and had poor grammar skills; thus assigning the correct parts of speech in a Linggle search was difficult, which was why he mostly relied on a dictionary to answer the questions in Part 3. S29 had similar difficulties. She seemed to have no idea about the structure of a sentence, and it appeared to be difficult for her to learn the syntax commands of Linggle as well. Her language proficiency level was low, achieving a score of only 70 in the reading section of the TOEIC test. As reflected in her interview, it was difficult for her to read the bilingual sentences in the dictionary, not to mention in Linggle, a monolingual tool. These results indicated that the students' language proficiency level influenced how well they were able to use the tool, including assigning the correct keyword and syntax command and "finding" the answer in the list of retrieved patterns.

3.2.4 Finding synonyms

For Part 4, the students needed to find synonyms to fill in the blanks. Most students found this section to be much easier with the help of Linggle, including the students at a very limited proficiency level. Most of them assigned the part of speech or used a wildcard (*) to quickly retrieve possible synonyms. For example, for Question 1, *He did a* _____ *job.* (He did a good job.), 13 students typed "did a adj. job" and four students typed "do a adj. job" and successfully found adjectives such as *wonderful*, *fantastic*, *nice*, *terrific*, and *fabulous* to fill in the blanks (see Figure 8).

Table 8 shows the types of syntax commands and the number of uses for Question 1 in the synonyms section (see supplementary materials for all the questions). As the frequencies show, most students directly assigned the part of speech of the searched word along with the keywords and collocates, and the accuracy rate reached 85.3% ($SD = 19.5$).

Taking a closer look, we found two phenomena. First, the students with limited proficiency levels, such as S26, tended to use a wildcard (*) more often. As reported by one student (S12), wildcards were flexible in processing queries. The second phenomenon was that the students

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press

Table 8. Types of Linggle syntax commands the students used in Part 4 (Question 1 only)

Syntax	Syntax & frequency Question 1
Part of speech	did a adj. job (13) do a adj. job (4) a adj. job (3) adj. job (1)
Wildcard *	did a * job (6) He did a * job (3)
Underscore _	He did a _ job (1)
~	No use of the syntax

Note. The numbers in parentheses represent the frequency of uses of the syntax commands.

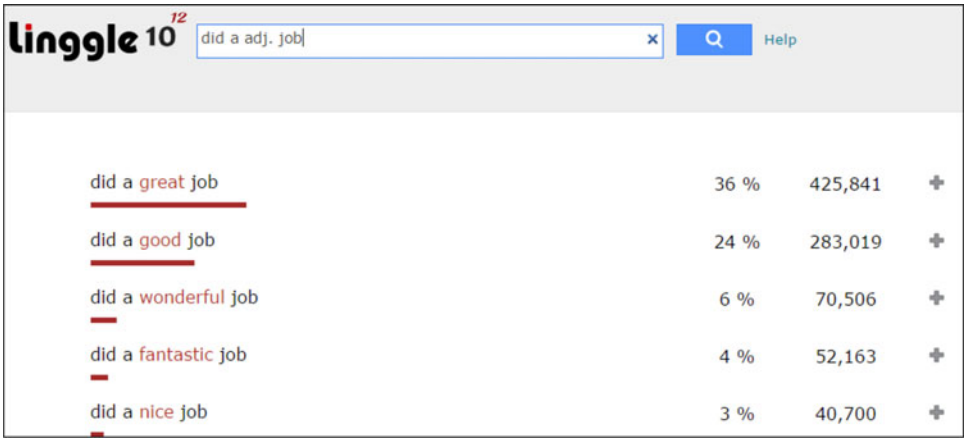


Figure 8. he search for "did a adj. job"

tended to choose words from the search results that they were familiar with. Although they took frequency count into consideration, to be on the safe side, they chose words that they were sure about. For example, S30 chose *nice*, *better*, and *superb* from the search results for "did a adj. job," which ranked fifth, seventh, and ninth in frequency, respectively. The students' prior knowledge influenced their decisions.

3.2.5 Error corrections

In the last section, the students had to correct 12 highlighted collocation errors in a short article. As this was the last section, 12 of the 32 students were unable to finish the section in the allotted time, which was why the overall accuracy rate was rather low (52.3%; $SD = 37.2$). For the 20 students who did finish this section, the accuracy rate reached 77.9%. All of these students used Linggle as the major tool to locate collocation information. Two students successfully corrected all 12 errors and nine missed only one or two items. These students' language proficiency varied, ranging from S2, the most proficient one, to S22, a low-intermediate student (see Table 9). This suggests that students with various language proficiencies all managed to benefit from this pattern-based tool.

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press

Table 9. Students' performance in the error correction section

No. of correct items	Accuracy (%)	Students
12	100	S2, S4
11	92	S9, S12, S13, S22
10	83	S3, S6, S7, S8, S17
9	75	S5, S10, S20
8	67	S1, S19, S21
7	58	S16
6	50	S11
5	42	S23

4. Discussion

In this process-oriented study, we examined students' tool consultation behaviors and search patterns. The students applied both metacognitive and cognitive strategies when solving collocation problems. Metacognitive strategies refer to "higher-order strategies aimed at analyzing, monitoring, evaluating, planning, and organizing one's own learning process" (Dörnyei, 2005: 169). They decided which tool to begin with (i.e. a dictionary or Linggle), whether they should combine different tools to find the answers, and which syntax commands and keywords they should type into the query box. When reading the long list of retrieved patterns, cognitive strategies were applied. They went through the patterns and located the one that best matched their needs. As the evidence from the screen-recorded files showed, the process was relatively straightforward and required less cognitive load compared with traditional concordancers where students need to read several sentences to assess how words are used. They did not seem to encounter too many difficulties either: only a few students at a lower language proficiency level had difficulties in using the part-of-speech search function and assigning correct keywords.

The data also showed that the students applied several compensation strategies to increase their success rates and make their searches faster. According to Oxford (1990: 47), "Compensation strategies enable learners to use the new language for either comprehension or production despite limitations in knowledge. Compensation strategies are intended to make up for an inadequate repertoire of grammar and, especially, of vocabulary." In this study, the action students took to make up for the limitation of the tools and to make up for inadequate grammar knowledge when using the tools were considered as the compensation strategies. When they had no confidence in identifying the parts of speech of the search words, they used the underscore or wildcard syntax commands. When they did not find satisfactory results, they tried different keywords in combination, different parts of speech, and even different tools.

Using Linggle seemed to be "natural" for the participants, as it was very similar to the way they used the Google search engine to locate information. They typed the keyword in the search box and the tool showed the results immediately. Some corpus tools are more sophisticated, requiring users to select the corpus type, set the number of associate words on the left and right, and tell the system to sort the results according to the words on the left or right. With its relatively simple interface, Linggle is easier to use, although tool training is still needed. For most of the participants in this study, the syntax commands were not difficult to learn, although they favored different syntax commands. Examining the results retrieved was not difficult either because Linggle narrowed down the search results and showed the possible patterns "explicitly." Mistakes happened, though, when the students' prior understanding of the retrieved words were not correct; this in turn influenced their decision-making.

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press

More importantly, the tool lists patterns in decreasing order of frequency, thus directing the students to "see" the common usages and patterns first, although how much they transferred the input to intake and whether they actually "learned" the patterns is unknown. As a vocabulary problem-solving tool, Linggle was handy and effective as the students' consultation process and feedback showed good usability of the tool. When designing a tool or CALL program, it is crucial to ensure its usability. According to Hémard (2003: 23), a tool needs to be "easy to learn, effective in what it claims to do and sufficiently motivating for the users to work with it and accept its validity."

The research focus on corpus-assisted language learning has been on DDL in the past few years. Despite its significant contribution to language learning and teaching (Boulton & Cobb, 2017), teachers and researchers are aware of its challenges. For instance, it is common that learners experience varying degrees of frustration with respect to cognitive processes (Yoon, 2016), not to mention that autonomous corpus consultation requires long-term training (Chambers & O'Sullivan, 2004). It is time to turn our attention to the tools themselves.

In fact, learners need different tools and different methods for different purposes, and it is crucial to consider learners' needs (Frankenberg-Garcia, 2020). So far, corpus research has paid little attention to the artifact (Park, 2012) and the interface the learner is interacting with. The interface of a corpus tool and the features it provides influence how learners interact and how much they benefit from the tool. With a different construct, Linggle as a collocation referencing tool for productive use has shown promise.

4.1 Limitations and further research

Linggle has shown promise, although the study has certain limitations. First, for the purposes of this study, the test items were checked in advance to make sure that all the answers could be found in Linggle. It is very likely that when the students actually start searching for words and phrases on their own, their success rates might be a bit lower. Second, the study was conducted in a test setting. Further research could investigate how learners interact with the tool in a free-writing situation and whether it would help to increase the quality of wording and collocation use. Third, to "disguise the corpus as a dictionary" (Kilgarriff *et al.*, 2015: 61) and make the tool "non-scary," Linggle was designed for simplicity and did not, for example, allow users to specify disciplines or genres. This could be a limitation because some collocations and lexical patterns are in fact register- and genre-specific (Biber, 2012). Finally, typical concordancers present complete concordance lines, and users need to observe the corpus to formulate the rules. Keywords in context and the experience of observing the output encourage deep processing and increase students' awareness of context (Kirk, 2002; Liu & Jiang, 2009; Schmidt, 2001; Tribble, 2002). Linggle, on the contrary, is a decontextualized tool that displays patterns and frequency counts directly. It is not clear whether the ease of using Linggle reduced retention rates due to shallow processing. Comparing students' vocabulary learning in a contextualized and a decontextualized situation would be an interesting topic to explore further.

5. Conclusion

This process-oriented study introduced a new pattern-based corpus tool and documented how students interacted with the tool to solve collocation problems. With its new interface and features, the consultation process is smooth and the search for collocation information is easier. If pattern-based tools can complement dictionaries and traditional concordancers, tool developers might consider developing tools of this kind for other languages to benefit more learners.

As students need collocation tools to compensate and complement the limitations of dictionaries, teachers probably need to introduce different kinds of vocabulary and writing tools to offer learners more support. Moreover, although Linggle is easy to use, appropriate training is still

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press

needed to help students make effective choices. Finally, as Linggle is in fact a tool that reveals patterns and n-grams, it can help learners to visualize the patterns and raises their attention to collocations. Integrating the tool into vocabulary and collocation teaching is worth trying.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S0958344020000105

Acknowledgements. We would like to thank the editor and the three anonymous reviewers for their constructive and invaluable feedback on the earlier drafts of this paper. We are also thankful to the students who participated in the study. Our thanks are also extended to the Linggle developers for designing a helpful tool and generously making it available for all users. Further information about the tools developed by Linggle team is available at https://home.linggle.com.

Ethical statement. The study was original, and we followed the institutional requirements in conducting research. Before the study began, the first author explained the purpose and the procedure of the study to the participants. Effort was made to ensure the participants' anonymity. There is no conflict of interest.

References

Biber, D. (2012) Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory*, 8(1): 9–37. https://doi.org/10.1515/cllt-2012-0002

Biber, D., Conrad, S. & Reppen, R. (1998) *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511804489

Boisson, J., Kao, T.-H., Wu, J.-C., Yen, T.-H. & Chang, J. S. (2013) Linggle: A web-scale linguistic search engine for words in context. In Schuetze, H., Fung, P. & Poesio, M. (eds.), *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. Sofia: Bulgaria, 139–144.

Boulton, A. (2008) Looking for empirical evidence of data-driven learning at lower levels. In Lewandowska-Tomaszczyk, B. (ed.), *Corpus linguistics, computer tools, and applications: State of the art*. Frankfurt: Peter Lang, 581–598.

Boulton, A. (2009) Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, 21(1): 37–54. https://doi.org/10.1017/S0958344009000068

Boulton, A. (2010) Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60(3): 534–572. https://doi.org/10.1111/j.1467-9922.2010.00566.x

Boulton, A. (2011) Data-driven learning: The perpetual enigma. In Goźdź-Roszkowski, S. (ed.), *Explorations across languages and corpora*. Frankfurt: Peter Lang, 563–580.

Boulton, A. (2017) Corpora in language teaching and learning. *Language Teaching*, 50(4): 483–506. https://doi.org/10.1017/S0261444817000167

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

Brants, T. & Franz, A. (2006) *Web 1T 5-gram Version 1*. Philadelphia: Linguistic Data Consortium. https://catalog.ldc.upenn.edu/LDC2006T13

Chambers, A. & O'Sullivan, Í. (2004) Corpus consultation and advanced learners' writing skills in French. *ReCALL*, 16(1): 158–172. https://doi.org/10.1017/S0958344004001211

Chang, J. S. (2013) Introducing Linggle: From concordance to linguistic search engine. *Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation*. Taipei: National Chengchi University, 12–14.

Chang, J.-Y. (2014) The use of general and specialized corpora as reference sources for academic English writing: A case study. *ReCALL*, 26(2): 243–259. https://doi.org/10.1017/S0958344014000056

Dolgova, N. & Mueller, C. (2019) How useful are corpus tools for error correction? Insights from learner data. *Journal of English for Academic Purposes*, 39: 97–108. https://doi.org/10.1016/j.jeap.2019.03.007

Dörnyei, Z. (2005) Language learning strategies and student self-regulation. In Dörnyei, Z. (ed.), *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah: Lawrence Erlbaum Associates, 162–196.

Flowerdew, L. (2008) Corpora and context in professional writing. In Bhatia, V. K., Flowerdew, J. & Jones, R. H. (eds.), *Advances in discourse studies*. New York: Routledge, 115–127.

Frankenberg-Garcia, A. (2012) Learners' use of corpus examples. *International Journal of Lexicography*, 25(3): 273–296. https://doi.org/10.1093/ijl/ecs011

Frankenberg-Garcia, A. (2014) The use of corpus examples for language comprehension and production. *ReCALL*, 26(2): 128–146. https://doi.org/10.1017/S0958344014000093

Frankenberg-Garcia, A. (2020) Combining user needs, lexicographic data and digital writing environments. *Language Teaching*, 53(1): 29–43. https://doi.org/10.1017/S0261444818000277

Frankenberg-Garcia, A., Lew, R., Roberts, J. C., Rees, G. P. & Sharma, N. (2019) Developing a writing assistant to help EAP writers with collocations in real time. *ReCALL*, 31(1): 23–39. https://doi.org/10.1017/S0958344018000150

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press

Gabel, S. (2001) Over-indulgence and under-representation in interlanguage: Reflections on the utilization of concordancers in self-directed foreign language learning. *Computer Assisted Language Learning*, 14(3–4): 269–288. https://doi.org/10.1076/call.14.3.269.5792

Gaskell, D. & Cobb, T. (2004) Can learners use concordance feedback for writing errors? *System*, 32(3): 301–319. https://doi.org/10.1016/j.system.2004.04.001

Gilmore, A. (2009) Using online corpora to develop students' writing skills. *ELT Journal*, 63(4): 363–372. https://doi.org/10.1093/elt/ccn056

Hémard, D. (2003) Language learning online: Designing towards user acceptability. In Felix, U. (ed.), *Language learning online: Towards best practice*. Lisse: Swets & Zeitlinger, 21–46.

Hubbard, P. (2013) Making a case for learner training in technology enhanced language learning environment. *CALICO Journal*, 30(2): 163–178. https://doi.org/10.11139/cj.30.2.163-178

Hyland, K. (2003) *Second language writing*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511667251

Jeaco, S. (2017) Concordancing lexical primings: The rationale and design of a user-friendly corpus tool for English language teaching and self-tutoring based on the lexical priming theory of language. In Pace-Sigge, M. & Patterson, K. J. (eds.), *Lexical priming: Applications and advances*. Amsterdam: John Benjamins, 273–296. https://doi.org/10.1075/scl.79.11jea

Johns, T. (1991) From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *English Language Research Journal*, 4: 27–45.

Johns, T. F., Hsingchin, L. & Lixun, W. (2008) Integrating corpus-based CALL programs in teaching English through children's literature. *Computer Assisted Language Learning*, 21(5): 483–506. https://doi.org/10.1080/09588220802448006

Kennedy, C. & Miceli, T. (2001) An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology*, 5(3): 77–90. https://doi.org/10125/44567

Kennedy, C. & Miceli, T. (2010) Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning & Technology*, 14(1): 28–44. https://doi.org/10125/44201

Kilgarriff, A., Husák, M., McAdam, K., Rundell, M. & Rychlý, P. (2008) GDEX: Automatically finding good dictionary examples in a corpus. In Bernal, E. & DeCesaris, J. (eds.), *Proceedings of the 13th EURALEX International Congress*. Barcelona: Universitat Pompeu Fabra, 425–432.

Kilgarriff, A., Marcowitz, F., Smith, S. & Thomas, J. (2015) Corpora and language learning with the Sketch Engine and SKELL. *Revue française de linguistique appliquée*, 10(1): 61–80. https://doi.org/10.3917/rfla.201.0061

Kirk, J. M. (2002) Teaching critical skills in corpus linguistics using the BNC. In Kettemann, B. & Marko, G. (eds.), *Teaching and learning by doing corpus analysis*. Amsterdam: Rodopi, 155–163. https://doi.org/10.1163/9789004334236_014

Lai, S.-L. & Chen, H.-J. H. (2015) Dictionaries vs concordancers: Actual practice of the two different tools in EFL writing. *Computer Assisted Language Learning*, 28(4): 341–363, https://doi.org/10.1080/09588221.2013.839567

Laufer, B. & Levitzky-Aviad, T. (2006) Examining the effectiveness of 'Bilingual Dictionary Plus': A dictionary for production in a foreign language. *International Journal of Lexicography*, 19(2): 135–155. https://doi.org/10.1093/ijl/eck006

Laufer, B. & Waldman, T. (2011) Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, 61(2): 647–672. https://doi.org/10.1111/j.1467-9922.2010.00621.x

Li, S. (2017) Using corpora to develop learners' collocational competence. *Language Learning & Technology*, 21(3): 153–171. https://doi.org/10125/44625

Liou, H.-C. (2019) Learner concordancing for EFL college writing accuracy. *English Teaching & Learning*, 43: 165–188. https://doi.org/10.1007/S42321-019-00022-5

Liou, H.-C., Chang, J. S., Chen, H.-J., Lin, C.-C., Liaw, M.-L., Gao, Z.-M., Jang, J.-S. R., Yeh, Y., Chuang, T. C. & You, G.-N. (2006) Corpora processing and computational scaffolding for a web-based English learning environment: The CANDLE project. *CALICO Journal*, 24(1): 77–95. https://doi.org/10.1558/cj.v24i1.77-95

Liu, D. & Jiang, P. (2009) Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal*, 93(1): 61–78. https://doi.org/10.1111/j.1540-4781.2009.00828.x

Mizumoto, A. (2017) Initial evaluation of AWSuM: A pilot study. *Vocabulary Learning and Instruction*, 6(2): 46–51.

Nation, I. S. P. (2003) *Learning vocabulary in another language* (4th ed.). New York: Cambridge University Press.

Nesi, H. & Haill, R. (2002) A study of dictionary use by international students at a British university. *International Journal of Lexicography*, 15(4): 277–305. https://doi.org/10.1093/ijl/15.4.277

O'Sullivan, Í. (2007) Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL*, 19(3): 269–286. https://doi.org/10.1017/S095834400700033X

O'Sullivan, Í. & Chambers, A. (2006) Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, 15(1): 49–68. https://doi.org/10.1016/j.jslw.2006.01.002

Oxford, R. L. (1990) *Language learning strategies: What every teacher should know*. New York: Newbury House.

Park, K. (2012) Learner–corpus interaction: A locus of microgenesis in corpus-assisted L2 writing. *Applied Linguistics*, 33(4): 361–385. https://doi.org/10.1093/applin/ams012

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press

Pérez-Paredes, P., Sánchez-Tornel, M., Alcaraz Calero, J. M. & Aguado Jiménez, P. (2011) Tracking learners' actual uses of corpora: Guided vs non-guided corpus consultation. *Computer Assisted Language Learning*, 24(3): 233–253. https://doi.org/10.1080/09588221.2010.539978

Quinn, C. (2015) Training L2 writers to reference corpora as a self-correction tool. *ELT Journal*, 69(2): 165–177. https://doi.org/10.1093/elt/ccu062

Schmidt, R. (2001) Attention. In Robinson, P. (ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press, 3–32. https://doi.org/10.1017/CBO9781139524780.003

Sinclair, J. (2004) New evidence, new priorities, new attitudes. In Sinclair, J. M. (ed.), *How to use corpora in language teaching*. Amsterdam: John Benjamins, 271–299. https://doi.org/10.1075/scl.12.20sin

Sun, Y.-C. (2003) Learning process, strategies and web-based concordancers: A case study. *British Journal of Educational Technology*, 34(5): 601–613. https://doi.org/10.1046/j.0007-1013.2003.00353.x

Tarp, S., Fisker, K. & Sepstrup, P. (2017) L2 writing assistants and context-aware dictionaries: New challenges to lexicography. *Lexikos*, 27: 494–521. https://doi.org/10.5788/27-1-1412

Tono, Y. (2000) On the effects of different types of electronic dictionary interfaces on L2 learners' reference behavior in productive/receptive tasks. In Heid, U., Evert, S., Lehmann, E. & Rohrer, C. (eds.), *Proceedings of the 9th EURALEX International Congress, EURALEX 2000*. Stuttgart: Institut fur Maschinelle Sprachverarbeitung, Universität Stuttgart, 855–862.

Tono, Y., Satake, Y. & Miura, A. (2014) The effects of using corpora on revision tasks in L2 writing with coded error feedback. *ReCALL*, 26(2): 147–162. https://doi.org/10.1017/S095834401400007X

Tribble, C. (2002) Corpora and corpus analysis: New windows on academic writing. In Flowerdew, J. (ed.), *Academic discourse*. Harlow: Longman, 131–149.

Wolter, B. & Gyllstad, H. (2011) Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, 32(4): 430–449. https://doi.org/10.1093/applin/amr011

Yoon, C. (2016) Concordancers and dictionaries as problem-solving tools for ESL academic writing. *Language Learning & Technology*, 20(1): 209–229. https://doi.org/10125/44453

Yoon, H. (2008) More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology*, 12(2): 31–48. https://doi.org/10125/44142

Yoon, H. & Hirvela, A. (2004) ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13(4): 257–283. https://doi.org/10.1016/j.jslw.2004.06.002

About the authors

Shu-Li Lai is an assistant professor in the General Education Center at National Taipei University of Business, Taiwan. Her research interests include computer-assisted language learning, second language acquisition, and EFL writing.

Jason S. Chang is a professor at the Department of Computer Science at National Tsing Hua University, Taiwan, where he directs a Natural Language Processing Group. His research interests include natural language processing, digital learning, and machine translation. The NLP lab has developed a battery of tools for language learners.

Author ORCiD. Shu-Li Lai, https://orcid.org/0000-0002-9976-9279
Author ORCiD. Jason S. Chang, https://orcid.org/0000-0002-8227-7382

https://doi.org/10.1017/S0958344020000105 Published online by Cambridge University Press