

# Should We Use It in Our Classrooms: An Analysis of Data-Driven Learning Research

Kai Bao

Correspondence: Kai Bao, Department of Foreign Languages and Literatures, Tsinghua University, Haidian District, Beijing, China.

Received: August 12, 2021

Accepted: August 23, 2021

Online Published: August 24, 2021

doi:10.5430/elr.v10n3p66

URL: <https://doi.org/10.5430/elr.v10n3p66>

## Abstract

Corpus linguistics has become increasingly important to both language researchers and teachers over the past three decades. As a popular practice of corpus linguistics, Data-Driven Learning (DDL) sees a rapidly growing body of research as well as instruction in the field. There is, however, a lack of comprehensive literature reviews that summarize the effectiveness, learners' perception, as well as factors affecting the success of DDL to guide its practices. In response, this study analyzes previous DDL research to show the feasibility of the activities in EFL classrooms. For the purpose, we collected and analyzed relevant research articles from 19 journals in the discipline of applied linguistics. Our analysis revealed that while DDL has been proved generally effective in improving learners' target language proficiency with respect to a variety of linguistic aspects, a set of its drawbacks have been elicited from the learners. The results indicate the instructors' need to take into account the learner as well as technique background before the introduction of DDL into their classrooms.

**Keywords:** Data-Driven Learning, review, effectiveness, perception

## 1. Introduction

Corpus linguistics analyzes collections of naturally occurring texts to approach the study of language (Bennett, 2010). It has become increasingly important to language research as well as pedagogy since 1990 due to the popularity of computers and concordance tools. Before DDL, corpora were primarily used as a source that provided realistic, rich, illustrative, and up-to-date language data for research and formulation of teaching materials (Braun 2005), which were then used by traditional instruction for language. Traditional instruction for language, either based on textbooks or dictionaries, provides learners with ready-made language patterns and thus prevents the direct contact between learners and texts that demonstrate these patterns. The instruction is therefore inconsistent with the pattern-finding acquisition process of language and criticized for reducing learning efficiency. In response, Johns (1986, 1988, 1991) proposed DDL that novelly advocates learners' direct use of corpora and has attracted considerable research interests and pedagogy practices.

A typical DDL activity consists of four steps: bringing up a question, drawing up a searching strategy, finding and observing the most relevant examples, and coming to conclusions (Kennedy & Miceli, 2001), based on a corpus and a concordance tool. Encouraging autonomous learning, DDL advocates role changes for both teachers and learners. Within the model, language teachers play roles of a trainer or a supervisor for corpus consultations. Accordingly, learners take on the responsibility of researchers who independently seek answers to their own language problems. Theoretically, DDL sees language acquisition as a process of discovering rules and principles underlying language in use, and is thus consistent with the usage-based theory and construction grammar as they all take pattern-finding as a major task and sign of language acquisition.

DDL is particularly beneficial to reading and writing pedagogy (Flowerdew, 2002), because both reading and writing are "text oriented and make use of words and word combinations, or lexical patterns, within the confines of discourse" (Jabbour, 2001, p. 294). DDL provides meaningful input for L2 (second language) writing instruction (Yoon, 2008), and its effectiveness has been illustrated by an increasing number of empirical studies that focus on different linguistic aspects or students of different L1 backgrounds as well as English proficiencies. Despite the rapidly growing body of studies, few research, however, has comprehensively summarized the effectiveness, learners' perceptions, and factors that affect the success of DDL activities to guide relevant practices.

## 2. Method

This study analyzes previous DDL research to summarize the effectiveness, learners' perceptions, and factors that affect the success of DDL activities. For the collection of DDL articles, we first retrieved target research articles on Google Scholar with keywords such as *Data-Driven Learning*, *concordance*, *corpus*, *consultation*, and *writing*. Upon the retrieval, we identified the origin journals and conducted searches within them. As a result, 19 journals were identified: *Journal of English for Academic Purposes*, *Journal of Language Teaching and Research*, *ReCALL (The Journal of the European Association for Computer Assisted Language Learning)*, *Journal of Second Language Writing*, *Language Learning & Technology*, *System*, *TESOL Quarterly*, *English Language Teaching Journal*, *English for Specific Purposes*, *Computer Assisted Language Learning*, *British Journal of Educational Technology*, *Eurasian Journal of Applied Linguistics*, *Corpora in the Foreign Language Classroom*, *Novitas-ROYAL*, *Asian EFL Journal*, *Journal of Computing in Higher Education*, *International Journal of Corpus Linguistics*, *International Journal of Assessment Tools in Education*, and *SAGE Open*, from which 79 research articles published between 2021 to 2000 were collected. This time frame on the one hand is highly relevant to the current practices and on the other involves the majority of research on DDL as a novel instruction approach.

To further enhance the relevance of the collected articles to our research purpose, we browsed the 79 research articles and screened out 26 pieces, which were not empirical studies that reported the DDL effectiveness, participants' perceptions, or factors of consultation success. As a result, 53 articles were selected for our analysis.

## 3. Results

### 3.1 Effectiveness of DDL

The analysis showed that 40 of the 53 research articles explicitly reported the effectiveness of DDL activities, and the other 13 articles primarily concern participants' perceptions or factors affecting consultation successes. Table 1 shows that DDL activities have been proved generally effective for the teaching and learning of EFL writing with respect to a variety of language aspects.

Table 1. Effectiveness of DDL activities reported by previous research

Linguistic aspect	Research
general collocation	Altun, 2021; Çelik, 2011; Li, 2017; Rezaee et al., 2015; Sun & Wang, 2003
verb-noun collocation	Akıncı & Yıldız, 2017; Chan & Liou, 2005, Wu, 2021
verb-adverb collocation	Daskalovska, 2015
collocations of prepositions	Koosha & Jafarpour, 2006
abstract nouns	Huang, 2014; Yılmaz, 2017
linking adverbials	Boulton, 2009b; Cotos, 2014; Friginal, 2013
stance	Chang, 2012
rhetorical functions	Charles, 2007; Lee & Swales, 2006; Poole, 2016
logical connectors	Cresswell, 2007; Tseng & Liou, 2006
<i>be</i> usage	Moon & Oh, 2017
passive voice	Smart, 2014
general lexico-grammatical pattern	Tsai, 2021; Yoon, 2008
written error correction	Bridle, 2019; Chambers & O'Sullivan, 2004; O'Sullivan & Chambers, 2006; Charles, 2018; Crosthwaite, 2017; Dolgova & Mueller, 2019; Gaskell & Cobb, 2004; Gilmore, 2009; Kennedy & Miceli, 2001; Luo & Liao, 2015; Mueller & Jacobsen, 2015; Sun, 2003; Todd, 2001; Tono et al., 2014; Yoon & Jo, 2014

#### 3.1.1 Collocations

The studies concerning collocations fall into two strands: general collocations and specific collocations such as verb-noun collocations (Akıncı & Yıldız, 2017; Chan & Liou, 2005), verb-adverb collocations (Daskalovska, 2015), and collocations concerning prepositions (Koosha & Jafarpour, 2006). Despite the varied focuses, the collocation-oriented studies adopted a consistent research method to examine the effectiveness of DDL activities. In

general, the researchers set up an experimental group following DDL approach, and a control group following traditional instruction based on dictionaries or textbooks. The instruction session ranges between five to 15 weeks. Prior to the beginning, a pre-test is conducted to measure the target collocation competence; upon the end of the session, a post-test is carried out to compare the performance of the experimental and control group; after another two to eight weeks, a retention test is conducted to compare the long-term effects of corpus consultation and conventional instruction. Adopting the method, the studies revealed that their control groups were outperformed by the experimental groups in either post-tests or retention tests, or both.

### 3.1.2 Abstract Nouns

In terms of the studies concerning abstract nouns, Huang (2014) examined the effectiveness of paper-based concordance lines in improving the abstract noun use by L1 Chinese undergraduate students. An experimental group following DDL-based instruction and a control group following dictionary-based instruction were compared. A pre-test, a post-test, a retention test, and a writing task were conducted to measure the participants' lexico-grammatical competence regarding the target abstract nouns. After four weeks of instruction, it was found that the experimental group outperformed the control group with respect to both language accuracy and complexity. In the same manner, Yılmaz (2017) investigated the effectiveness of DDL in teaching abstract nouns to L1 Turkish EFL students. Undergoing roughly the same experimental procedures, the DDL group outperformed the dictionary group in terms of collocational patterns and error numbers.

### 3.1.3 Linking Adverbials

For the studies concerning linking adverbials, Boulton (2009b) examined the effectiveness of DDL approach in teaching English linking adverbials to intermediate and lower-level EFL learners in France. Similarly, a pre-test, a post-test, as well as a retention test were conducted to compare the performance of an experimental group using printed corpus data and a control group using traditional pedagogical resources including bilingual dictionaries and grammar books. The study revealed that even in the absence of pre-training, the experimental group outperformed the control group. Cotos (2014) used a combination of native-speaker and learner corpora to teach linking adverbials to international graduate students at a North American university. The researcher compared a group that used only native speaker corpus with another group that used both native speaker and learner corpus. The retention test showed that the group using both corpora outperformed the other by presenting 27.1% more correct use of target linking adverbials. Friginal (2013) explored the effectiveness of corpus instruction in teaching research report writing skills to college-level native speakers, and focused on four linguistic features: linking adverbials, reporting verbs, verb tenses, and passive sentence structures. After instruction of two weeks, the corpus-based group outperformed the other group following traditional instruction.

### 3.1.4 Rhetorical Functions

With regard to the studies concerning rhetorical functions, Charles (2007) investigated the effectiveness of combining printed materials and corpus-based activities to teach rhetorical functions to international students at Oxford University. She first used handouts to present target rhetorical functions, and then required the participants to identify according lexico-grammatical usages in a corpus. It was revealed that the activities were effective in promoting the learners' rhetorical skills as well as search techniques. Lee and Swales (2006) explored the effectiveness by using four L1 Chinese PhD students' own writings and a reference corpus to improve their rhetorical consciousness. The researchers required the PhD students to first compile a corpus with their own writings and then compare the corpus with a reference corpus consisted of published research articles. Likewise, the results showed a substantial increase of rhetorical consciousness and search techniques. Poole (2016) examined the effectiveness of DDL activities in teaching English rhetoric to undergraduate L2 writers, and revealed that 81% of the writers agreed that the activities were helpful in facilitating the identification and understanding of rhetorical strategies.

### 3.1.5 Logical Connectors

With respect to the research concerning logical connectors, Cresswell (2007) investigated the effectiveness of corpus consultations in enhancing learners' use of logical connectors. The researcher compared essays from two learner corpora: one written by DDL learners and the other by non-DDL learners. The results showed that DDL activities were moderately effective in enhancing the quantity and accuracy of logical connector usages. Tseng and Liou (2006) examined the effectiveness of corpus consultation in teaching English conjunctions to L1 Chinese undergraduate students in Taiwan. After corpus-based instruction of six weeks, a significant increase of conjunction competence was identified with tests and writing tasks.

### 3.1.6 Be Usage

Moon and Oh (2017) investigated the effectiveness of using learner and native speaker corpora in unlearning L1 Korean students' over-generated usage of *be*. The experimental group consulted a native speaker corpus and a learner corpus to identify error patterns, and the control group consulted textbooks for the correct usage. It was found that the experimental group outperformed the control group in a retention test.

### 3.1.7 Stance Expression

Chang (2012) examined the effectiveness of corpus consultations in improving the English stance use by L1 Chinese PhD students in America. The researcher first compiled a corpus consisted of fifteen well-written introductions of research articles, and then required the participants to consult the corpus when they were polishing their own introduction section. The results revealed a significant improvement in terms of more accurate stance use within the introduction section.

### 3.1.8 Passive Voice

Smart (2014) investigated the effectiveness of paper-based DDL activities in teaching passive voice to EFL students in an American university. The researcher set up a DDL group, a deductive group using corpus-informed teaching materials, and another deductive group using traditional materials. After instruction of two weeks, the DDL group significantly outperformed the two deductive groups.

### 3.1.9 Lexico-Grammatical Patterns

In terms of the research concerning general lexico-grammatical patterns, Yoon (2008) conducted a case study that evaluated the influence of DDL activities on the development of L1 Korean students' English writing skills. The author required six participants to search Collins COBUILD Corpus to address their writing problems and e-mail the teacher results weekly. After instruction of ten weeks, the participants' evaluations were gathered and showed that not only were most of their language problems effectively solved, but also their lexico-grammar knowledge and awareness enhanced. Tsai (2021) examined the effectiveness of corpus-based instruction in improving undergraduate business students' lexical and syntactic proficiency. By comparing the performance of 49 EFL students following corpus-based instruction and 58 EFL students following lecture-based instruction, the researcher found that the former group outperformed the latter in post-test with respect to lexical and syntactic complexity.

### 3.1.10 Written Error Correction

Research applying DDL to correct written errors falls into three stages. The research in the first stage mainly investigated the feasibility rather than the effectiveness with no control group involved. For instance, Todd (2001) explored the effectiveness of corpus consultation in correcting written errors made by 25 graduate students at King Mongkut's University of Technology Thonburi in Thailand. The method was typical: after pre-training, instructors collected participants' writing, coded errors, and then required the students to correct the errors using concordance tools and then submit the second draft. The analysis of the second draft showed that the participants achieved a 90% accuracy rate of correction. Chambers and O'Sullivan (2004, 2006) conducted a two-phase case study on Irish students learning French with corpora, and identified a 75% and 73% accuracy rate of correction in the first and second phase, respectively.

In the second stage of DDL-based written error correction research, control groups were involved. Luo and Liao (2015), as well as Mueller and Jacobsen (2015) compared the effectiveness of corpus- and dictionary-based consultation in revising English writings by L1 Chinese and L1 Thai students, respectively. The studies revealed the greater effectiveness of corpora than dictionaries. In the same manner, Yoon and Jo (2014) compared the effects of indirect and direct use of corpora to correct English written errors by L1 Korean students. Surprisingly, the indirect use group outperformed the other group, which indicated the importance of instructors' thorough interference.

In the third stage, researchers started investigating the factors that affect consultation successes in error correction. Bridle (2019) used Learner Styles Questionnaire to divide participants into four types: activists who favored novel experiences, pragmatists who favored practicing, reflectors who favored pondering, and theorists who acted logically. The results showed that the reflectors were least receptive to corpus consultations because they did not like ambiguity and large amounts of data. It was also noted that corpus consultation was more effective in correcting errors concerning synonyms. Similarly, Dolgova and Mueller (2019) found that local lexico-grammatical and register errors were more receptive to the approach; Crosthwaite (2017) found that errors of word choice, word form, collocations and phrasing were more readily corrected than errors of deletion or morphosyntax. Charles (2018) found a relevance between certain corpus functions and language problems rectified, and noted that N-grams, Word List and Keyword List functions of lexical analysis tools allowed the users to identify and revise unknown language problems.

### 3.2 Learners' Perceptions of DDL

#### 3.2.1 Positive Perceptions

Most participants noted that corpus consultation has a positive effect on their overall writing quality (Bridle, 2019; Chambers & O'Sullivan, 2004; Chang, 2014; Charles, 2014; Crosthwaite, 2017; Dolgova & Mueller, 2019; Luo & Liao, 2015; O'Sullivan & Chambers, 2006; Poole, 2016; Yoon, 2008).

The positive evaluations were primarily identified with vocabulary learning. Luo and Liao (2015) revealed that 93% of their participants noted that the corpus consultation assisted their learning of word usages; 73% considered the approach helpful in learning word meanings; 80% considered it useful in learning new words incidentally. In support, Yilmaz (2017) revealed that 80% of the participants considered corpus consultations helpful in general vocabulary learning, 93.4% in learning word meanings, and 73.3% in learning word usages. Huang (2014) reported that 95% of his students found DDL activities helpful in learning word meanings, and 90% in word usages. Slightly differently, Bridle (2019) found that 91% of the participants regarded corpus consultation helpful in learning word usages, but only 42% in learning word meanings, which is consistent with Yoon and Hirvela's (2004) 90% and 60% for word usages and word meanings, respectively.

In terms of collocation learning, Yilmaz (2017) found that 86.7% of the participants considered corpus consultation useful for collocation learning, while the percentage in Dolgova and Mueller's (2019) study was 69%, and in Huang's (2014) study 100%.

Corpus consultation was also generally perceived as a helpful tool in assisting grammatical pattern learning. Luo and Liao (2015) found that 60% of the participants emphasized the positive effects in this regard, in comparison with 46.7% according to Yilmaz (2017), 95% to Huang (2014), 33% to Bridle (2019), 88% to Yoon and Hirvela (2004), and 50% to Dolgova and Mueller (2019).

The other positive effects were perceived at the metalinguistic level. In Luo and Liao's (2015) research, 80% of the participants agreed that corpus-based activities increased their linguistic awareness, and 73% agreed that the approach increased their writing confidence. Yoon and Hirvela (2004) found that 88% of the participants reported an increase of writing confidence. Rezaee et al. (2015) reported that 100% of the participants found corpus treatments effective in raising their consciousness of word co-occurrences.

#### 3.2.2 Negative Perceptions

The most identified problem regarding corpus consultations is their time-consuming nature (Chamber & O'Sullivan, 2004; Chang, 2014; Charles, 2011; Friginal, 2013; Huang, 2014; Kennedy & Miceli, 2010; Luo & Liao, 2015; Yoon & Hirvela, 2004). Luo and Liao (2015) revealed that 87% of the respondents thought that the approach was time-consuming, and that percentage was 75% according to Huang (2014) and Yoon and Hirvela (2004).

Encountering unfamiliar vocabulary is another major difficulty reported by participants (Akıncı & Yıldız; 2017; Crosthwaite, 2017; Huang, 2014; Luo & Liao, 2015; Yilmaz, 2017; Yoon & Hirvela, 2004). In the study of Huang (2014), 80% of the participants reported having difficulty in corpus-based activities because of new words in concordance lines, and the percentage was 80% according to Luo and Liao (2015), 46.7% to Yilmaz (2017), and 38% to Yoon and Hirvela (2004).

Another reported difficulty is cut-off or incomplete sentences (Akıncı & Yıldız; 2017; Crosthwaite, 2017; Huang, 2014; Luo & Liao, 2015; Varley, 2009; Yoon & Hirvela, 2004). It was shown that 80% of the participants in Huang's (2014) study reported this problem, and the percentage was 73% in Luo and Liao's (2015) and 62% in Yoon and Hirvela's (2004) study.

The participants also complained about too many (Crosthwaite, 2017; Huang, 2014; Kennedy & Miceli, 2010; Luo & Liao, 2015; Varley, 2009; Yoon & Hirvela; 2004) or too few concordance lines (Crosthwaite, 2017; Huang, 2014; Luo & Liao, 2015; Yoon & Hirvela; 2004; Yoon & Jo, 2014). Kennedy and Miceli (2010) argued that students had problems in restricting the number of concordance lines displayed, and the students became overwhelmed with too many examples. In contrast, Varley (2009) reported that when the number of examples was limited, the students decided not to rely on corpora but resort to their own knowledge. Huang (2014) reported that 65% and 30% of the participants had too many and too few lines, respectively, and the percentage was 73% and 33% according to Luo and Liao (2015), and 25% and 62% to Yoon and Hirvela (2004).

The last common problem recognized is the lack of ability to induce rules (Crosthwaite, 2017; Huang, 2014; Luo & Liao, 2015; Yilmaz, 2017; Yoon & Hirvela, 2004). Huang reported that 45% of participants had this difficulty, Yilmaz reported 53.4%, and Yoon and Hirvela reported 50%.

### 3.3 Factors of Consultation Successes

#### 3.3.1 Selection of Corpus

There are two types of corpora used by DDL research: self-compiled specialized corpora (Chambers & O'Sullivan, 2004; Charles, 2018; Kennedy & Miceli, 2001; O'Sullivan & Chambers, 2006), and public available general corpora (Bridle, 2019; Crosthwaite, 2017; Dolgova & Mueller, 2019; Gaskell & Cobb, 2004; Gilmore, 2009; Luo & Liao, 2015; Mueller & Jacobsen, 2015; Tono et al., 2014). Self-compiled specialized corpora were either used in early studies or in EAP studies because of specific generic focuses. General corpora were employed by recent studies because of their greater availability (Chambers, 2007; Yoon, 2008) and size (Yoon, 2008). Overall, self-compiled specialized corpora have low availability but high representativeness to discipline- and genre-specific queries and quicker consultation thanks to their small size; general corpora, in contrast, have a limited relevance to specific disciplines and genres, longer consultation time, but a considerably greater amount of data and higher availability.

#### 3.3.2 Learners' L2 Proficiency

Although it is widely believed that DDL activities were more effective for intermediate and advanced L2 learners because corpus analyses demand intellectual capacities (Bennett, 2010; Mauranen, 2004), there are mixed results regarding the relevance between learners' L2 proficiency and their success of corpus consultation.

Lin (2021) compared the performance of 95 undergraduate students at three English proficiency levels, and found no statistically significant differences between the groups. Tono et al. (2014) compared the performance of 68 learners with upper-intermediate and 25 learners with low level of English proficiency in the same essay tasks. No significant difference, however, was identified. The low proficiency learners tended to correct more simple and similar errors, and the upper-intermediate learners tended to correct a greater variety of error types but with a lower accuracy rate. Boulton (2009b) compared students of low English proficiency in DDL activities and traditional pedagogical activities, and found that the participants benefited more from DDL approach even in the absence of pre-training. Boulton (2009a, p.3) therefore argued that the claim that students with low proficiency level were not likely to benefit from DDL activities seems "to be based on gut reaction rather than empirical evidence". In the same manner, Mueller and Jacobsen (2015) compared students of lower English proficiency in revising essays using corpora and dictionaries as reference tools, and found that the group of corpus consultation performed marginally better than the group using dictionaries. Yoon and Hirvela (2004) investigated intermediate and advanced students' perceptions of corpus consultation, and found that they all considered the approach helpful in promoting writing skills and confidence about L2 writing, and that the students of intermediate proficiency made even more positive evaluations.

#### 3.3.3 Target Error Types

A relevance between error types and the success of corpus treatment has been identified. Bridle (2019) found that corpora were only used for the correction of six error types when the students were offered alternative reference tools such as an electric dictionary. The six error types were: error of formal/informal usage, wrong word, article, grammar, missing words, and the unknown, among which formal/informal usage and wrong word took up 80% of the instances. Chambers and O'Sullivan (2004) discovered that DDL activities were most effective in correcting grammatical errors (gender and agreement, prepositions, verb forms/mood, negation and syntax), misspellings, accents and hyphens, lexico-grammatical patterning (native language interference, choice of verb and inappropriate vocabulary), and capitalization. Crosthwaite (2017) revealed that learners were most likely to use corpora to correct errors regarding word choice, word form, collocation and phrasing, but less likely to correct errors of deletion or morphosyntax. Dolgova and Mueller (2019) found that local lexicogrammatical and register errors were the most common errors in the learners' texts, but the accuracy rate of their correction was lower than that of register errors. Gaskell and Cobb (2004) examined ten error types and found that DDL activities were only effective in reducing three types of errors: the errors of capitals/punctuation, word order, as well as pronouns. Tono et al. (2014) revealed that the errors of omission and addition were more readily corrected than the errors of misformation.

## 4. Discussion

The studies reviewed indicate that DDL activities were generally effective in facilitating EFL learners' English learning. The effectiveness might result from the role changes advocated by the approach. The perspective of a proactive researcher instead of a passive learner impels students to pay attention to aspects rarely noticed. In addition, the learning shows high relevance to authentic language use because of the reliance of DDL activities on authentic language data that present practical language patterns in use. Those benefits enable learners to continue exploiting the resource even after the experimental session, thus allowing them to outperform those following traditional instruction particularly in retention tests.

The effectiveness was better perceived by participants in learning word usages than in learning word meanings. The contrast can result from the absence of instructors' interference, without which the learners can hardly identify meanings of new words by contexts presented by concordance lines alone. To address the problem, annotated corpora and according concordance functions should be developed and used. The participants also expressed their preference for DDL activities in learning formulaic language, which can relate to the function of lexical analysis tools which are able to reveal word co-occurrence patterns within a few seconds. In addition to the benefits, a series of negative evaluations, such as time-consuming processes and incomplete sentences, were also reported by the learners. The causes underlying those negative evaluations, however, were not adequately accounted for by the studies.

The evaluations may result from the participants' lack of search technique. The current mainstream general corpora and concordance tools were primarily developed for research rather than pedagogical purposes, and they are thus not user-friendly enough even for many language instructors, let alone their students. As a consequence, many students may fail to fully utilize the functions of corpora and concordance tools. For instance, it was reported by Pérez-Paredes et al. (2012) that none of their participants, not even those who had received pre-training and regular guidances, used POS tags, regular expressions, and wildcards functions of concordance tools. Instead, it was found that the participants simply approached BNC search interface the same way they approached Google. For those participants, corpus consultations are more likely a one-shot deal where learners either achieve satisfied results in the first attempt or will never have them.

In addition to the major problems, the variations of facilities across different countries and regions should also be noted. It was reported by Yoon and Hirvela (2004) that 62% of the participants in their study complained about the speed of internet connection, and 12% about the access to computer and internet. We argue that students' access to the facilities should be taken into account before the introduction of any DDL activities. Moreover, as the majority of L1 English corpora are based in North America and Europe, the low speed of internet connection due to great geographical distances between Asia and the servers can considerably harm the Asian students' consultation experience as well as motivation. The selection of public corpora for consultations thus needs to be cautious.

## 5. Conclusion

A DDL activity "befits teaching second language reading and writing, since both activities are text oriented and make use of words and word combinations, or lexical patterns, within the confines of discourse" (Jabbour, 2001, p. 294). The approach, either directly employed as learning materials or indirectly used as the reference of lexicography and textbook development, is a primary language resource because of its "potential to make explicit the more common patterns of language use" (Tao, 2001, p. 116).

The implications of the argument are threefold. First, in terms of language knowledge, it is proved and acknowledged that corpus consultations bring significant developments to the lexico-grammatical and even rhetorical aspects of learners' writing. The corpus approach is based on the assumption that vocabulary and grammar (Halliday, 1992; Sinclair, 1991), form and meaning are inseparable (Ellis, 2008), and corpus-based activities function as an interface between those aspects and thus benefit the L2 writing teaching and learning (Gledhill, 2000; Hyland, 2002; Tribble, 1999, 2002). Second, DDL prompts learners to become a proactive researcher rather than a passive learner, thus is able to enhance their cognitive abilities as well as metalinguistic knowledge (Yoon & Jo, 2014). Third, corpora offer learners rich and authentic language use (Thurstun & Candlin, 1998). Different from other sources inevitably influenced by the subjectivity of editors and teachers, corpora present authentic language data, based on which learners find more practical lexico-grammatical patterns

However, with the effectiveness of DDL activities adequately demonstrated by the previous studies, its deficiencies need to be recognized, studied, and coped with. Its effects on learners of lower English proficiency is under-researched and thus still unclear, which can be explored by future research. Pedagogically, it is important for instructors to take into account the learner and technique backgrounds before the introduction of DDL activities into their classrooms. Technically, annotated corpora and learner-friendly concordance tools can be developed and used to thoroughly address the deficiencies of current DDL practices.

## References

- Akinci, M., & Yıldız, S. (2017). Effectiveness of corpus consultation in teaching verb+noun collocations to advanced ELT students. *Eurasian Journal of Applied Linguistics*, 3(1), 91-109. <https://doi.org/10.32601/ejal.461036>
- Altun, H. (2021). The learning effect of corpora on strong and weak collocations: Implications for corpus-based assessment of collocation competence. *International Journal of Assessment Tools in Education*, 8(3), 509-526.

<https://doi.org/10.21449/ijate.845051>

- Bennett, G. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Ann Arbor, MI: University of Michigan Press. <https://doi.org/10.3998/mpub.371534>
- Boulton, A. (2009a). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, 35(1), 81-106.
- Boulton, A. (2009b). Testing the limits of data-driving learning: language proficiency and training. *ReCALL*, 21(1), 37-54. <https://doi.org/10.1017/s0958344009000068>
- Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, 17(1), 47-64. <https://doi.org/10.1017/s0958344005000510>
- Bridle, M. (2019). Learner use of a corpus as a reference tool in error correction: Factors influencing consultation and success. *Journal of English for Academic Purposes*, 37(2009), 52-69. <https://doi.org/10.1016/j.jeap.2018.11.003>
- Çelik, S. (2011). Developing collocational competence through web based concordance activities. *Novitas-ROYAL*, 5(2), 273-286.
- Chambers, A. (2007). Popularising corpus consultation by language learners and teachers. In E. Hidalgo, L. Quereda & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 3-16). Amsterdam: Rodopi. [https://doi.org/10.1163/9789401203906\\_002](https://doi.org/10.1163/9789401203906_002)
- 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>
- Chan, T., & Liou, H. (2005). Effects of web-based concordancing Instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning*, 18(3), 231-250. <https://doi.org/10.1080/09588220500185769>
- 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>
- Chang, P. (2012). Using a stance corpus to learn about effective authorial stance-taking: A textlinguistic approach. *ReCALL*, 24(2), 209-236. <https://doi.org/10.1017/s0958344012000079>
- Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes*, 6(4), 289-302. <https://doi.org/10.1016/j.jeap.2007.09.009>
- Charles, M. (2011). Using hands-on concordancing to teach rhetorical functions: Evaluation and implications for EAP. In A. Frankenburg-Garcia, L. Flowerdew & G. Aston (Eds.), *New trends in corpora and language learning* (pp. 26-43). London: Continuum. <https://doi.org/10.5040/9781474211925.ch-002>
- Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes*, 35(7), 30-40. <https://doi.org/10.1016/j.esp.2013.11.004>
- Charles, M. (2018). Corpus-assisted editing for doctoral students: More than just concordancing. *Journal of English for Academic Purposes*, 36(11), 15-25. <https://doi.org/10.1016/j.jeap.2018.08.003>
- Cotos, E. (2014). Enhancing writing pedagogy with learner corpus data. *ReCALL*, 26(2), 202-224. <https://doi.org/10.1017/s0958344014000019>
- Cresswell, A. (2007). Getting to 'know' connectors? Evaluating data-driven learning in a writing skills course. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 267-287). Amsterdam: Rodopi.
- Crosthwaite, P. (2017). Retesting the limits of data-driven learning: feedback and error correction. *Computer Assisted Language Learning*, 30(6), 447-473. <https://doi.org/10.1080/09588221.2017.1312462>
- Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, 28(2), 130-144. <https://doi.org/10.1080/09588221.2013.803982>
- Dolgova, N., & Mueller, C. (2019). How useful are corpus tools for error correction? Insights from learner data. *Journal of English for Academic Purposes*, 39(5), 97-108. <https://doi.org/10.1016/j.jeap.2019.03.007>
- Ellis, N. (2008). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J.

- Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 52-83). Malden, MA: Wiley-Blackwell. <https://doi.org/10.1002/9780470756492.ch4>
- Flowerdew, L. (2002). Corpus-based analyses in EAP. In J. Flowerdew (Ed.), *Academic discourse* (pp. 95-114). Harlow, UK: Longman. <https://doi.org/10.4324/9781315838069-14>
- Friginal, E. (2013). Developing research report skills using corpora. *English for Specific Purposes*, 32(2013), 208-220. <https://doi.org/10.1016/j.esp.2013.06.001>
- 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>
- Gledhill, C. (2000). The discourse function of collocation in research article introductions. *English for Specific Purposes*, 19(2), 115-135. [https://doi.org/10.1016/s0889-4906\(98\)00015-5](https://doi.org/10.1016/s0889-4906(98)00015-5)
- Halliday, M. A. K. (1992). Language as system and language as instance: The corpus as a theoretical construct. In J. Svartvik (Ed.), *Proceedings of Nobel Symposium 82 on Directions in corpus linguistics, Stockholm, 4-8 August 1991* (pp. 61-77). Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110867275.61>
- Huang, Z. (2014). The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL*, 26(2), 163-183. <https://doi.org/10.1017/s0958344014000020>
- Hyland, K. (2002). Activity and evaluation: Reporting practices in academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp. 115-130). Harlow, England: Longman. <https://doi.org/10.4324/9781315838069-15>
- Jabbour, G. (2001). Lexis and grammar in second language reading and writing. In D. Belcher & A. Hirvela (Eds.), *Linking literacies: Perspectives on L2 reading-writing connections* (pp. 291-308). Ann Arbor: University of Michigan Press.
- Johns, T. (1986). Micro-concord: A language learner's research tool. *System*, 14(2), 151-62. [https://doi.org/10.1016/0346-251x\(86\)90004-7](https://doi.org/10.1016/0346-251x(86)90004-7)
- Johns, T. (1988). Whence and whither classroom concordancing. In T. Bongaerts, P. De Haan, S. Lobbe & H. Wekker (Eds), *Computer applications in language learning* (pp. 9-27). Dordrecht: Foris. <https://doi.org/10.1515/9783110884876-003>
- Johns, T. (1991). Should You Be Persuaded: Two Examples of Data-driven Learning. in T. Johns & P. King (Eds), *Classroom Concordancing, ELR Journal 4* (pp. 1-16). Birmingham: Centre for English Language Studies, University of Birmingham.
- 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.
- Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology*, 5(3), 77-90.
- Koosha, M., & Jafarpour, A. A. (2006). Data-driven learning and teaching collocations of prepositions: The case of Iranian EFL adult learners. *Asian EFL Journal*, 8(4), 192-209.
- Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1), 56-75. <https://doi.org/10.1016/j.esp.2005.02.010>
- Li, S. (2017). Using corpora to develop learners' collocational competence. *Language Learning & Technology*, 21(3), 153-171.
- Lin, M. H. (2021). Effects of data-driven learning on college students of different grammar proficiencies: A preliminary empirical assessment in EFL classes. *SAGE Open*, 11(3), 21582440211029936. <https://doi.org/10.1177/21582440211029936>
- Luo, Q., & Liao, Y. (2015). Using corpora for error correction in EFL Learners' writing. *Journal of Language Teaching and Research*, 6(6), 1333-1342. <https://doi.org/10.17507/jltr.0606.22>
- Mauranen, A. (2004). Spoken corpus for an ordinary learner. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 89-105). Amsterdam, the Netherlands: John Benjamins.

- Moon, S., & Oh, S. (2017). Unlearning overgenerated be through data-driven learning in the secondary EFL classroom. *ReCALL*, 30(1), 48-67. <https://doi.org/10.1017/s0958344017000246>
- Mueller, C. M., & Jacobson, N. D. (2015). A comparison of the effectiveness of EFL students' use of dictionaries and an online corpus for the enhancement of revision skills. *ReCALL*, 28(1), 3-21. <https://doi.org/10.1017/s0958344015000142>
- 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>
- Pérez-Paredes, P., Sánchez-Tornel, M., & Calero, J. M. A. (2012). Learners' search patterns during corpus-based focus-on-form activities: A study on hands-on concordancing. *International Journal of Corpus Linguistics*, 17(4), 482-515. <https://doi.org/10.1075/ijcl.17.4.02par>
- Poole, J. (2006). E-learning and learning styles: Students' reactions to web-based Language and Style at Blackpool and the Fylde College. *Language and Literature*, 15(3), 307-320. <https://doi.org/10.1177/0963947006066129>
- Rezaee, A. A., Marefat, H., & Saeedakhtar, A. (2015). Symmetrical and asymmetrical scaffolding of L2 collocations in the context of concordancing. *Computer Assisted Language Learning*, 28(6), 532-549. <https://doi.org/10.1080/09588221.2014.889712>
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press. <https://doi.org/10.2307/330144>
- Smart, J. (2014) The role of guided induction in paper-based data-driven learning. *ReCALL*, 26(2), 184-201. <https://doi.org/10.1017/s0958344014000081>
- Sun, Y. C. (2003). Learning processes 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>
- Sun, Y. C., & Wang, L. Y. (2003). Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning*, 16(1), 83-94. <https://doi.org/10.1076/call.16.1.83.15528>
- Tao, H. (2001). Discovering the usual with corpora: The case of remember. In R. Simpson & J. Swales (Eds.), *Corpus linguistics in North America: Selections from the 1999 symposium* (pp. 116-144). Ann Arbor: University of Michigan Press. <https://doi.org/10.1016/j.knosys.2009.05.009>
- Thurstun, J., & Candlin, C. (1997). *Exploring academic English: A workbook for student essay writing*. Sydney: NCELTR.
- Todd, R. W. (2001). Induction from self-selected concordances and self-correction. *System*, 29(1), 91-102. [https://doi.org/10.1016/s0346-251x\(00\)00047-6](https://doi.org/10.1016/s0346-251x(00)00047-6)
- 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. (1999). *Writing difficult texts (doctoral dissertation)*. Lancaster University, UK.
- Tribble, C. (2002). Corpora and corpus analysis: New windows on academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp. 131-149). Harlow, England: Longman. <https://doi.org/10.4324/9781315838069-16>
- Tsai, Y. R. (2021). Exploring the effects of corpus-based business English writing instruction on EFL learners' writing proficiency and perception. *Journal of Computing in Higher Education*, 33, 475-498. <https://doi.org/10.1007/s12528-021-09272-4>
- Tseng, Y., & Liou, H. (2006). The effects of online conjunction materials on college EFL students' writing. *System*, 34, 270-283. <https://doi.org/10.1016/j.system.2006.01.006>
- Varley, S. (2009). I'll just look that up in the concordancer: Integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning*, 22(2), 133-152. <https://doi.org/10.1080/09588220902778294>
- Wu, Y. J. (2021). Discovering collocations via data-driven learning in L2 writing. *Language Learning & Technology*, 25(2), 192-214.
- Yılmaz, M. (2017). The effect of Data-Driven Learning on EFL students' acquisition of lexico-grammatical patterns in EFL writing. *Eurasian Journal of Applied Linguistics*, 3(2), 75-88. <https://doi.org/10.32601/ejal.460966>
- 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>

Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology*, 12(2), 31-48.

Yoon, H., & Jo, J. W. (2014). Direct and indirect access to corpora: An exploratory case study comparing students' error correction and learning strategy use in L2 writing. *Language Learning & Technology*, 18(1), 96-117.

### **Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).