

## **Data-driven Learning in Second Language Writing Class: A Survey of Empirical Studies**

<https://doi.org/10.3991/ijet.v12i03.6523>

Qinqin Luo

Southwest Petroleum University, Chengdu, China  
qinqinluo@126.com

Jie Zhou

Southwest Petroleum University, Chengdu, China  
15808263062@163.com

**Abstract**—Corpus technology is commonly used by researchers and professionals for language description; however it can also be employed by second or foreign language learners in what has come to be known as data-driven learning (DDL). DDL has been suggested as an effective approach to improve second language (L2) learners' writing competence. To popularize DDL approach among ordinary language teachers and learners, this paper offers an overview of empirical DDL research in writing published from 2010 to 2016, which can provide insights into how DDL approach is integrated into an actual writing classrooms and how much it can contribute to the development of L2 writing skill. The analysis of the surveyed studies reveals the great potentials of DDL activities in L2 writing class from different perspectives but it's also found that corpora are not superior to other traditional reference tools for some consultation purposes. For certain lexical problems, L2 learners are more willing to consult online dictionaries or other conventional tools. It is thus suggested to develop online platforms which could provide easy and free access to the user-friendly corpora along with other types of reference tools. The tendency of future studies is also predicted in the end.

**Keywords**—data-driven learning, second language writing, empirical studies  
How to work with this template

### **1 Introduction**

In recent decades, computer network as the core of modern information technology is developing rapidly, providing favorable conditions and vast space for language teaching and learning. The wide application of electronic corpora is a strong evidence for it. One application of corpora involving language learning through learners' direct interaction with language corpora is often referred to as data-driven learning (DDL). Such a term was firstly proposed by Johns & King [1] as an innovative approach to the implementation of concordancing materials in second language (L2) acquisition. This type of approach has been claimed to have the following obvious advantages.

Firstly, it exposes learners to massive authentic instances of a particular linguistic item, which can not only help learners to acquire common usages of lexis and grammar but also contribute to learners' increased awareness of language patterns [2]. Secondly, DDL approach fosters learner autonomy by encouraging learners to explore language features by themselves from corpus data [3]. Thirdly, DDL has an important corrective function [4]. By comparing their own writing with data produced by native writers, learners can get the help they need to revise the inappropriate expressions and ultimately improve their writing. Due to these advantages, DDL is suggested as an effective approach for the teaching and learning of L2 writing. This approach can not only be adopted before writing to help students brainstorm relevant ideas by searching the topic-related words or phrases, but also be used while writing and after writing to check whether an element of writing is correct or find alternatives to original simple words or inappropriate phrases[5] [6]. The past ten years has seen a gradual increase in empirical DDL studies in writing; however, more empirical studies are still in an urgent need to bring corpora to ordinary teachers and students. Therefore, it's necessary to review the existing studies so as to provide suggestions for further DDL studies in writing. The present study attempts to survey empirical DDL studies in L2 writing class, aiming to provide insights into how corpus technology can be integrated into an actual writing classroom and how much it can contribute to the development of L2 writing skill.

## **2 Theoretical background of DDL**

Since being proposed, DDL has been interpreted in various ways by different researchers. It seems extremely hard to give a watertight definition, thus a prototype of DDL was proposed by Boulton [7] as “the hands-on use of authentic corpus data by advanced foreign language or L2 learners in higher-education for inductive language learning of advanced usage.” However, those who declare to be implementing DDL involve learners of all levels of language proficiency including those lower-level or secondary school students and design activities based on printed corpus materials without requiring learners to consult corpora directly. Later, Smart [8] claims that it can be regarded as DDL as long as it meets the following two requirements: the use of authentic corpus data as learning materials or reference resources; the design of student-centered exploratory learning activities. So far, there is no consensus on the exact definition of DDL, and researchers name it in different ways such as corpus-based learning and learner concordancing, etc. In this paper, DDL is interpreted broadly referring to “any use of overt corpus data for foreign or second language learning or teaching” [9], whether the researchers claim it as DDL or not.

Earlier DDL studies are mainly theoretical, trying to analyze the language learning theories underpinning DDL. As stated in Ref. [10][11], Constructivism is an important theory that can support DDL. According to constructivist learning theories, knowledge acquisition is a dynamic process, which emphasizes learners' exploratory and discovery learning. This is much consistent with DDL in which learners are also encouraged to explore and discover linguistic regularities on their own through ana-

lyzing concordance lines. The knowledge about the usages of words is mainly acquired by learners' active involvement in tasks but not passively transmitted from teachers. However, constructivist learning was criticized by Kirschner et al. [12] for the fact that this discovery-based approach has high cognitive demands on the learner. Thus they argued for constructivist learning to adopt a "scaffolding" approach which was based on a socio-cultural theory of learning. This learning theory proposed by Vygotsky [13] holds that knowledge is constructed by collaborative dialogue and negotiation with guidance mediated by the teacher or student in the way of scaffolding. According to Vygotsky, scaffolding in teacher-student interaction is necessary to facilitate language development, which provides support for teacher guidance in DDL. To ensure the success of a DDL approach, learners usually need to get support or special training from teachers and then interact with corpus resources independently. In this process, they construct their knowledge by negotiation or mediation. Another theory underpinning DDL was learner autonomy, which was proposed by Henri Holec [14] who has illustrated that learners should be encouraged to take charge of their own learning and tackle problems independently. This has much in common with DDL for one of the aims of DDL is to make learners responsible for their own learning and become more autonomous language learners. In corpus investigation, students are usually instructed to follow these steps in solving language problems [15][16]: firstly, formulating and refining query terms; then observing concordance lines and selecting relevant examples; finally, evaluating the query results and drawing a conclusion. This process emphasizes learner autonomy by encouraging learners to act as researchers.

Besides the above theories, noticing hypothesis is another one that lends support to DDL. This theory demonstrates that learners' acquisition of linguistic input is more likely to increase if their attention is consciously drawn to linguistic features [17]. DDL tasks usually require students to concentrate on recurrent phrases, which would be an effective means to enhance learners' input through noticing. Schmidt [18] has come up with a related noticing hypothesis, demonstrating that "in order to overcome errors, learners must make conscious comparison between their own input and target language input". This has undoubtedly provided theoretical support for the adoption of DDL in error correction of writing. To L2 writers, there are usually many doubts and uncertainties about how to identify and correct errors in writing. In DDL, a corpus is regularly employed as reference resources to help learners discover the inappropriate expressions or correct the errors underlined by teachers or peers.

The language learning theories discussed above are actually not controversial and there is some overlap among them. However, they have, from different perspective, laid a solid foundation for the adoption of DDL approach in language learning, especially in L2 writing.

### **3 Methodology**

#### **3.1 Research questions:**

This study was guided by the following questions:

1. What corpora and corpus tools are used in these selected empirical studies?
2. What are the features of the research samples and designs in these studies?
3. What are the research focus and major findings of these recent DDL studies in writing?

#### **3.2 Research samples**

Since there is a lack of agreement on exactly what counts as empirical DDL studies, this paper adopts Boulton's [9, p.130] definition as " studies which subject some aspect of DDL to observation or experimentation with some kind of externally validated evaluation other than the researchers' own intuition".

Thus, the research samples have to meet the following criteria to be eligible for inclusion: (a) studies that provide information about the effects of corpus use by second/foreign language learners or learners' perceptions of corpus use; (b) the tasks assigned to learners are related to second/foreign language writing or as part of writing courses; (c) studies that offer information about data-collection or data-analysis; (d) studies that are published in SSCI (Social Science Citation Index) journals from 2010 to 2016. The following studies are excluded: (a) studies targeting native speakers of English. For instance, Friginal [19] provides statistical information about the effectiveness of corpus use in developing learners' research report writing skills but it focuses on native speakers, thus it's not included. (b) studies featuring corpora in languages other than English, such as Ref.[20] where an Italian corpus is utilized. (c) conference articles, reviews and editorials. (d) studies that are published in languages other than English.

#### **3.3 Research procedure**

To collect relevant studies on DDL, I searched Web of Science and Google Scholar using different combinations as search terms among data-driven learning, corpus-driven, corpus-based, corpus-assisted, concordancing, L2 writing, foreign language writing, EAP writing, ESP writing. I also conducted electronic searches of 10 journals in volumes covering the years 2010-2016. These journals were ReCALL, Journal of Second Language Writing, Language Learning & Technology, Computer Assisted language learning, Journal of English for Academic Purposes, English for Specific Purposes, ELT Journal, System, Applied linguistics, and Language learning. These are generally regarded as reputable SSCI journals which publish corpus-assisted studies and writing.

I firstly reviewed all the abstracts published in the SSCI journals, searching for information about research design, and excluded studies that don't meet the inclusion

criteria. If the abstract doesn't provide information about experimental design about corpus use in writing, the methodology section is examined. To avoid omitting some relevant studies, the reference lists of the selected papers are carefully read. Finally, 18 studies were selected based on the inclusion criteria described in the previous section. The 18 selected studies were coded for further investigation, initially including the following information: corpora and corpus tools used in the study, learners, duration of the study, interface (direct or indirect DDL), focus of investigation.

## **4 Results and discussion**

### **4.1 Corpora and corpus tools used in these studies**

Table 1 shows the detailed information about corpora and corpus tools used in individual studies. A common feature of most studies is the combined use of several corpora or several types of corpus tools, with only five studies employing a single corpus. Two types of corpora, including general corpora and specialized corpora, are most frequently employed by learners.

A large and general corpus, such as the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), is usually considered as a good reference resource for L2 writers since it is readily available and free to use with an Internet connection. Additionally, it can provide massive concordance lines about a linguistic item. Therefore, ten of the surveyed eighteen studies have selected general corpora as at least one type of corpus tools for learners. Among the ten, three studies employed a general corpus as the only tool and the other seven just included the general corpus as one type of reference tools for learners. For instance, in [16] [21], learners were allowed to access a variety of corpora and tools that were freely accessible on the web, including COCA and an 80-million-word subsection of the BNC, etc. In several studies, the general corpora were not provided directly to students but offered through an online link. Tono et al. [22] introduced a publicly available online corpus query system IntelliText in which the students could consult ready-made corpora including the BNC. Compared with the direct use of BNC, IntelliText provided a more user-friendly interface where students were limited to search merely word or phrase levels. In Ref [23], students were encouraged to exploit Lextutor (<http://www.Lextutor.ca>) which could provide an easy and quick access to BNC, Brown and other small-scale corpora.

Despite the obvious advantages of the large and general corpora, the specialized corpora, which are primarily compiled by researchers, also have their own special value. As stated by Kennedy & Miceli [15] and Yoon [24], a small but focused discipline-specific corpus provided writing models that were directly relevant to learners' needs and interests. In addition, the limited amount of concordance lines from the small and specialized corpora can prevent learners from getting lost and enabled them to interpret the results better. It can be found that three studies employed specialized corpus as the only corpus tool or as one type of tools to guide L2 learners' writing. Poole [25] presented the benefits of using a localized, specialized corpus to support

various corpus-aided activities in an L2 writing classroom. Chang [26] instructed students to use both a general corpus and a specialized corpus which included journal articles and conference papers in the participants' fields. The findings revealed that a specialized corpus is more helpful as reference sources for EAP students' academic English writing due to its direct relevance to their academic fields. It can be concluded that both the general and specialized corpora have their own merits and demerits, thus they can complement each other and be utilized for different purposes.

**Table 1.** Corpora and corpus tolls in these studies

Study	Description of corpora or corpus tools
Yoon, 2016a	the Corpus of Contemporary American English (COCA) (450 million-words) Google Web , Google Scholar ,Custom Search Engine JustTheWord (an 80 million-word subsection of the BNC)
Yoon, 2016b	the Corpus of Contemporary American English (COCA) (450 million-words) Google Web , Google Scholar ,Custom Search Engine JustTheWord (an 80 million-word subsection of the BNC)
Muller & Jacobsen, 2016	the British National Corpus, 2007 (100 million words) the Michigan Corpus of Academic Spoken English, 2002 the Corpus of Contemporary American English (450 million-words)
Poole, 2016	blog posts from The Rosemont Mine Truth (RMT) (70000words) press releases from Rosemont Copper Company (RCC) (17000 words)
Chen et al., 2015	PREF abricated Expression Recognizer (PREFER) A bilingual corpus-Hong Kong Parallel Text
Lai & Chen, 2015	Sinorama (1990–2000) Brown corpus, LOB, news articles from The Times on various topics, short stories The British National Corpus (BNC)
Yoon & Jo, 2014	Lextutor (accessible to BNC, Brown and other small-scale corpora)
Chang, 2014	the Corpus of Contemporary American English (400 million words) Michelangelo (self-compiled, 1352033 words)
Tono et al., 2014	Intelligent Tools for Creating and Analysing Electronic Text Corpora for Humanities Research ( IntelliText- accessible to BNC )
Cotos, 2014	a specialized corpus of research articles - 1322089 words The local learner corpus of writings produced by the participants
Huang, 2014	A topic-specific corpus consisting of texts from online websites and Louvain Corpus of Native English Essays (LOCNESS)- self-compiled a sub-corpus of opinion essays written by the British students in the LOCNESS corpus
Charles, 2014	Do-it-yourself corpora
Quinn, 2014	Collins WordBanks Online
Geluso, 2013	Google
Charles, 2012	Do-it-yourself corpora
Park, 2012	A corpus of academic texts and Google
Boulton, 2010	BYU-BNC (100 million words)
Conroy, 2010	Virtual Language Centre at Hong Kong Polytechnic University Lextutor ( <a href="http://www.lexutor.ca/">http://www.lexutor.ca/</a> ) Google ( <a href="http://www.google.com/">http://www.google.com/</a> )

Another type of corpora is do-it-yourself corpus, which are mainly compiled by learners. There are two studies encouraging students to construct and consult this type of corpora. In Charles's [27] study, students were instructed to build their own individual, discipline-specific corpora and use them as reference resources for academic writing. To explore whether students truly use their own corpus later, Charles [28] conducted a further study to investigate their corpus use in their writing course by questionnaires and interviews one year after the completion of the corpus. The findings were positive; nevertheless, compiling a corpus is really time-consuming and requires much specialized corpus knowledge and skills, which may pose great challenges to students, especially those who are weak in using computers and those lower-level language learners.

Besides the above-mentioned corpora, Google as a corpus have attracted more and more attention from researchers. Among the previous 12 studies reviewed by Yoon [24], only one study exploited search engine as a corpus tool while there are five studies exploiting the search engine Google as the single tool or as one type of the tools. Strictly speaking, Google is not a real corpus, but it can be used as a dynamic corpus due to the following two reasons: the Internet can be considered as the body of machine-readable text and the search engine can be used as the concordancer [29]. Compared with the traditional corpora, Google has its own advantages. Firstly, it is more flexible than most concordancers in retrieving relevant sentences based on multiple-word queries [30]. Secondly, it is easy to use without high demand for learners to master complex corpus consultation skills and familiar to most learners while offering massive authentic language data from billions of web pages. In Ref. [31] [32], most students showed positive attitudes toward Google as a reference resource in L2 writing. They realized the usefulness of Google in improving the naturalness and accuracy of their writing and expressed their willingness to continue using it in the future writing.

## **4.2 Research samples and designs**

It can be seen clearly from Table 2 that the sample sizes of these studies are generally small. Four case studies featured less than 10 participants, four studies involved participants between 10 and 30 and ten studies targeted over 30 participants. There is Ref. [31] involving more than 100 participants. In this study, four groups of students with the total number of 165 took part in the experiment.

As for language proficiency about the research samples, all the studies involved learners in university settings since DDL was usually regarded as more appropriate for advanced, sophisticated foreign or second language learners in higher education. In addition, a vast majority of studies targeted advanced or upper-intermediate learners, with only two [3][33] featuring lower-level learners. In Ref. [33] study, the lower-level language learners generally showed positive attitudes toward corpus consultation for error correction, but they found it difficult to use an online corpus. In their second experiment which mainly focused on basic-level learners, the participants performed marginally better while using the online corpus to deal with language issues that have been identified as problems. Ref. [3] showed similar positive results

that lower-level learners could also gain more benefits while adopting a DDL approach in tackling certain language problems. The two studies both proved the effectiveness of DDL for lower-level learners from different angles; nonetheless, researchers or teachers may have to invest more time in guiding them or make more careful designs about the tasks.

Furthermore, it can be revealed from Table 2 that most of these studies are just based on short-time data and there is no study lasting for more than one year. In these studies, the duration of experiment is reported in different ways, which makes the comparison difficult. Most studies introduced it by giving weeks but the second experiment in Muller & Jacobsen [33] is reported by showing months. Besides that, there were two studies [30] [34] describing the duration of DDL by showing semester. In 12 of the studies, the exposure to DDL was more than 5 weeks. It seems that studies by Chang [26] and Muller & Jacobsen [33] lasted for the longest time, which was 22 weeks and 6 months respectively. The studies based on longer-time experiment usually started with a course or project and then presented the research questions to evaluate some aspects of DDL, while the shorter-time study often began with the research question and then create an experimental situation mainly in the form of one-time tests or questionnaires. This finding is similar to the results reported in Ref. [9] which summarized 27 DDL studies from 1996 to 2009.

When it comes to the implementation of DDL, the majority of these studies have adopted the direct DDL approach, with only three Ref. [3] [25] [35] employing the indirect DDL approach. Yoon & Jo [23] applied both direct and indirect approach, aiming to compare the effects of corpus use on error correction, error correction patterns, and learning strategy use between the two approaches. It was concluded that both approaches were effective in assisting learners to tackle certain writing problems, especially in the revising stage of writing. Nevertheless, the indirect use of corpora seemed to have greater effectiveness in error correction for most learners. Thus, the types of tasks are significant factors to determine whether to take indirect or direct DDL. Although both the two ways of DDL could bring encouraging results, Ref. [3][23] still found that direct corpus use was more suitable for higher-level students and indirect corpus use might bring more benefits for lower-level students since the limited number of examples on paper-based materials could reduce their cognitive burden.

As regards the research tools, most studies used more than one research instruments, including questionnaires, classroom observation, recalling, tracking, verbal reports and most frequently questionnaires. Six studies adopted a pre-test/post-test format, including only two also with the delayed post-test. It can be found that most studies just provide qualitative data about DDL and the studies that offer quantitative statistics showing significant difference are still in an urgent need.

**Table 2.** Research designs of DDL in writing

Study	Learners	time	interface	Data-collection
Yoon,2016a	6 advanced	7-20 weeks	direct	Survey, interview, screen recording, recall
Yoon,2016b	2 advanced	7-20 weeks	direct	Survey, interview, screen recording, recall
Muller, Jacob-sen,2016	Exp1:78; Exp 2:39 lower level	Exp1: 3 weeks Exp2: 6 months	direct	Survey, gap-fill test
Poole, 2016	21 intermediate	16 week	indirect	surveys
Chen et al., 2015	55 intermediate	16 weeks	direct	single-group pre/post-test, questionnaires
Lai, Chen, 2015	14 intermediate	16 weeks	direct	stimulus recalls, interviews, class observations, research notes
Yoon, Jo, 2014	4 freshmen	10 weeks	Indirect vs direct	Interview, pre /post writing think-aloud protocols, learning journals, and observation notes
Chang, 2014	10 high intermediate	22 weeks	direct	Interviews, surveys, search logs and feedback logs, writing productions
Tono et al., 2014	93 upper intermediate	4 weeks	direct	Two sets of timed essay, error-correction results
Cotos, 2014	32 intermediate advanced	17 weeks	direct	Pre- test, immediate and delayed post-test, questionnaires
Huang, 2014	40 upper- intermediate	4 weeks	indirect	Pre/post test + delayed, questionnaires
Charles, 2014	40 advanced	6 weeks	direct	Questionnaires, interviews
Quinn, 2014	58 intermediate	a semester	direct	questionnaires
Geluso, 2013	25 lower- intermediate	14 weeks	direct	Pre- and post-versions of phrases
Charles, 2012	50 advanced	6 weeks	direct	questionnaires
Park, 2012	3 upper intermediate	a semester	direct	screen recordings, corpus search queries, oral and written reflections, the student's essays
Boulton, 2010	71 lower-level learners	5 weeks	indirect	Pre- and post-tests, questionnaire
Conroy, 2010	165 advanced	6 weeks	direct	Survey, interview, writing productions

### 4.3 Research focus and major findings

These studies can be classified roughly into three categories according to their research focuses, though inevitably they overlap. The studies in the first group were primarily about the effects of DDL on writing. These studies can be further divided into two streams. The first stream of these studies showed the effectiveness of DDL by comparing writing with corpora and other types of reference resources such as dictionaries. Boulton [3] compared the usefulness of paper-based corpus materials and traditional teaching materials in helping L2 learners to deal with typical problems in their written productions, finding that learners using corpus materials performed much

better than those using dictionary entries and other traditional materials. Huang [35] made the similar comparison but mainly focused on L2 learners' lexico-grammatical use of abstract nouns in their writing. In this study, the experimental group was encouraged to use paper-based concordance lines while the control group was offered dictionaries to study the usage of the words. The results revealed the obvious advantages of paper-based or indirect DDL in improving the accuracy and complexity of students' writing. Unlike the two studies, Muller & Jacobsen [33] adopted a direct DDL approach, trying to figure out whether EFL learners achieved greater accuracy in tackling the identified language problems when using an electronic dictionary or COCA. As a result, learners performed marginally better while using COCA for the enhancement of revision skills in writing. Chen et al [36] also used the direct DDL approach, aiming to compare the effectiveness of a corpus-based paraphrase tool PREFER and dictionaries in improving students' writing performance. These studies all confirmed the advantages of corpora as reference resources in handling learners' writing problems, but they could not completely replace the dictionaries. It is thus concluded that corpus tools and dictionaries are complementary to each other. To testify the effectiveness, Yoon [16] explored the processes and outcomes of the combined use of corpus tools and dictionaries by L2 learners and their evaluations of the reference suite as a problem solving tool that supported the completion of their writing tasks. The findings of this small-scale study demonstrated the positive results about the use of multiple reference resources in a single interface in advanced L2 learners' writing process. The second stream of studies explored the effects of using different types of corpora on writing. Chang [26] examined the usefulness of general corpora and specialized corpora as reference resources for academic English writing, while Cotos [37] compared the effectiveness of combining native-speaker and learner corpora and exposing students only to native-speaker data in improving L2 writers' knowledge of linking adverbials. These studies were mainly conducted by comparing the usefulness of different reference tools; however, Geluso [32] made a different comparison, which contrasted learners' initial Google-informed phrases vs final non-Google-informed phrases, suggesting that Google could be an effective reference tool to improve the naturalness of their productive language use. Unlike the above two streams of studies, Tono, et al [22] concerned about the effectiveness of DDL in correcting different types of errors in writing. In this study, students were required to consult corpora for correcting three types of lexico-grammatical errors in writing. A significant difference was found in the accuracy rate of error correction, indicating that omission and addition errors could be more easily identified and corrected than misformation errors.

The studies in the second category investigated learners' search behavior or strategy use while using corpora. Park [30] explored the process in which students interacted dynamically with a corpus system by documenting interactions as evidencing each student's efforts to tackle lexico-grammatical issues by retrieving, evaluating, and appropriating search results. This study collected four kinds of data including screen recordings, corpus search queries, oral and written reflections, and the student's essays. It was found that learners used the corpus most frequently for lexical issues, including collocation, word choice, choice of chunks and then prepositions. However,

they rarely consult corpus for the issues regarding content, organization, or rhetorical strategies. To examine the learning strategies learners employed in direct and indirect corpus use, Yoon & Jo [23] investigated four Korean EFL students' correction behaviors in concordance analysis, finding that cognitive strategy and social strategy were the first and second most frequently used strategies in both the corpus use settings. Unlike the above studies, Charles [28] investigated from a different perspective, aiming to find out participants' corpus use one year after the completion of their own corpus. As indicated by the findings, 30% of students never used it for writing due to the small size of the corpus and its lack of reliability and convenience, whereas most participants consulted the corpus for checking grammar and lexis while composing and revising essay. In addition, 93% of them believed that the use of their own corpus has positive effects on their academic writing. In order to figure out learners' search behavior while facing different types of reference resources, Lai & Chen [38] explored how EFL learners used the dictionaries and corpus tools to solve writing problems by tracking learners' tool consultation process. As demonstrated by the findings, corpus tools were mainly exploited for determining word usages, collocation information, and grammar patterns, whereas a bilingual dictionary was primarily used for finding out information about word form and word meaning. Focusing on more various types of reference resources, Yoon [21] investigated individual differences in using a variety of resources which were put within a single interface and the main factors that are potentially related to specific resource choices and referencing patterns. In this small-scale study, learners performed differently in various ways including the frequency of use, problem types, query purposes, resources consulted, and their perceptions of using such resources as an aid for writing. However, corpora were neither the most preferred resource according to the frequency of use nor the most appropriate for some consultation purposes, and it was most frequently consulted for some confirmatory problems and preposition errors or idiomatic expressions [21][38].

The studies in the third category focused on learners' evaluation of corpus use in L2 writing class. This type of study is in the minority. Some studies [16][21][26][33] which fall into the first two groups have also reported students' attitudes, but the investigation of attitudes is only one minor part of their research thus they are excluded from this type. In the studies that fall into the last group, data are mainly collected by questionnaires or surveys and interviews. In [27], initial and final questionnaires were collected to explore advanced-level students' attitudes towards do-it-yourself corpus-building in academic writing class. It was found that 90% students believed in the usefulness of corpus in improving their writing and they were willing to use it in the future although there were still some problems concerning do-it-yourself corpus-building especially for students without specialized computer skills. In another study conducted by Conroy [31], data were primarily collected by surveys and semi-structured interviews, investigating university students' use of, and attitudes towards, Internet based resources which facilitate independent language learning. The results were also positive since 89% of respondents expressed using corpora was either "very rewarding" or "somewhat useful" while only 9% reported some frustration with corpus consultation. To help learners overcome frustrations and guide teachers to prepare L2 writers for corpus consultation, Quinn [34] implemented a corpus training module

into EFL writing course and trained novice users to exploit the corpus for self-correcting their writing; then learners' reaction to the corpus use was investigated through questionnaires at the end of the course. As revealed by the study, learners were generally satisfied with the class training although a few students reported difficulties in searching the all-English corpus interface. In this study, most learners expressed that using corpus for error-correction could help them improve the naturalness of their writing and increase their confidence in lexical choices. Besides error-correction in writing, a more recent study conducted by Poole [25] reported students' perceptions of corpus use in the teaching and learning of rhetoric in an undergraduate writing course, and the participants responded positively to the corpus use in helping them understand the concepts of rhetoric in academic writing.

## **5 Conclusion**

The studies surveyed above have generally revealed that DDL in L2 writing class is effective in the following respects: increasing the accuracy rate of correcting linguistic errors, improving the productive use of language in writing, promoting learner autonomy, fostering different cognitive skills, enhancing language awareness and noticing skills, etc. Nonetheless, some problems reported by students can't be ignored; for instance, being time-consuming in searching and interpreting massive concordance lines, frustrations from failing to get what they want. To overcome these problems, these factors such as learners' language proficiency, individual learning styles, the choice of corpus tools, task designs should be taken into consideration. For novice or lower-level learners, the user-friendly corpora or corpus tools and the indirect DDL approach could be a good choice; meanwhile, the DDL tasks shouldn't be too complex at the beginning, which could be about tackling the confirmatory problems in writing. Additionally, proper learner training about corpus use is still necessary before learners' independent use of corpora especially for the technically less competent learners. Furthermore, just as suggested by Yoon [24], continued efforts for teacher training will be also beneficial to the application of corpus technology in L2 writing classroom.

Some new tendencies can be seen from these studies. To begin with, web as a corpus or search engine-based corpus such as Google would be more popular due to its accessibility and flexibility. One of the main obstacles to corpus use is learners' lack of specialized knowledge or skills about corpus technology. However, just by conducting some simple searches in Google, learners can quickly and accurately find out the frequency and occurrence of linguistic patterns, which may help improve learners' naturalness of language use [32]. Furthermore, the combined use of different types of reference tools would be more helpful to ensure the effectiveness of DDL activities. A corpus has been proved to be useful as a means of writing assistance in improving learners' writing competence; nonetheless it couldn't completely take the place of other traditional reference resources. As Yoon [16][21] indicates, different types of reference tools are complementary to each other and they could be put within a single interface for learners to consult for different purposes. Thus, to popularize DDL ap-

proach, more online platforms, which could provide easy and quick access to user-friendly corpora as well as other types of reference tools should be developed and open to public.

There are still limitations of these surveyed studies, which may provide insight into the future DDL studies. In terms of research design, many studies are based on a very small sample of participants and report only learners' subjective impression on corpus use without providing quantitative data. Thus, further study should involve larger samples offering more quantitative data to ensure the objectivity. In addition, a majority of studies evaluate the short-term benefits of DDL activities in writing thus there is in need of longitudinal studies to complement the existing research.

## 6 Acknowledgment

This work was supported by the National University Foreign Language Teaching and Research Project (No. 2016SC0021A), the Southwest Petroleum University Teaching and Research Project (No. 2016JXYJ-39) and the 16<sup>th</sup> SWPU College Students Open Experiment Project (No. KSZ16182). Additionally, I am very grateful to the reviewers' insightful comments and recommendations.

## 7 References

- [1] Johns, T., & King, P. (Eds.). (1991). Classroom concordancing (pp. 1-16). Centre for English Language Studies, the University of Birmingham.
- [2] Gabrielatos, C. (2005). Corpora and Language Teaching: Just a Fling or Wedding Bells? *TESL-EJ*, 8 (4):1-35.
- [3] 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>
- [4] Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In A. O'Keeffe and M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. London: Routledge, pp. 359-369. <https://doi.org/10.4324/9780203856949.ch26>
- [5] Luo, Q. (2015). Application of data-driven learning in EFL writing instructions for engineering students. *World Transactions on Engineering and Technology Education*, 13 (3): 238-244.
- [6] Luo, Q. (2016). The effects of data-driven learning activities on EFL learners' writing development. *SpringerPlus*, 5 (1): 1-13. <https://doi.org/10.1186/s40064-016-2935-5>
- [7] Boulton, A. (2011). Data-driven learning: The perpetual enigma. In S. Goźdź-Roszkowski (Ed.), *Explorations across Languages and Corpora*. Frankfurt: Peter Lang, pp. 563-580.
- [8] 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>
- [9] Boulton, A. (2010). "Learning outcomes from corpus consultation," in M. Moreno Jaen, F. Serrano Valverde & M. Calzada Perez (Eds.), *Exploring New Paths in Language Pedagogy: Lexis and Corpus-Based Language Teaching*, Equinox, pp.129-144.
- [10] Chambers, A., & Kelly, V. (2004). Semi-specialised corpora of written French as a resource in language teaching and learning. *Teanga*, 21: 114-133.

- [11] Flowerdew, L. (2015). Data-Driven learning and language learning theories. Multiple Affordances of Language Corpora for Data-Driven Learning, 69: 15-36. <https://doi.org/10.1075/scl.69.02flo>
- [12] Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, 41(2): 75-86. [https://doi.org/10.1207/s15326985ep4102\\_1](https://doi.org/10.1207/s15326985ep4102_1)
- [13] Vygotsky, L. (1986). *Thought and Language*. Cambridge MA: The MIT Press.
- [14] Holec, H. (1981). *Autonomy and foreign language learning*. Oxford: Pergamon.
- [15] Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language learning & technology*, 5(3): 77-90.
- [16] Yoon, C. (2016). Concordancers and dictionaries as problem-solving tools for ESL academic writing. *Language Learning & Technology*, 20 (1): 209-229.
- [17] Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction*. Cambridge: CUP, pp. 3-32. <https://doi.org/10.1017/CBO9781139524780.003>
- [18] Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, et al. (Eds.), *Proceedings of CLaSIC 2010*, Singapore, December 2-4 (pp. 721-737). Singapore: National University of Singapore, Centre for Language Studies.
- [19] Friginal, E. (2013). Developing research report writing skills using corpora. *English for Specific Purposes*, 32 (4): 208-220. <https://doi.org/10.1016/j.esp.2013.06.001>
- [20] 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.
- [21] Yoon, C. (2016). Individual differences in online reference resource consultation: Case studies of Korean ESL graduate writers. *Journal of Second Language Writing*, 32: 67-80. <https://doi.org/10.1016/j.jslw.2016.04.002>
- [22] 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>
- [23] 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.
- [24] Yoon, C. (2011). Concordancing in L2 writing class: An overview of research and issues. *Journal of English for Academic Purposes*, 10(3):130-139. <https://doi.org/10.1016/j.jeap.2011.03.003>
- [25] Poole, R. (2016). A corpus-aided approach for the teaching and learning of rhetoric in an undergraduate composition course for L2 writers. *Journal of English for Academic Purposes*, 21: 99-109. <https://doi.org/10.1016/j.jeap.2015.12.003>
- [26] 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>
- [27] Charles, M. (2012). 'Proper vocabulary and juicy collocations': EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes*, 31(2): 93-102. <https://doi.org/10.1016/j.esp.2011.12.003>
- [28] Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes*, 35: 30-40. <https://doi.org/10.1016/j.esp.2013.11.004>

- [29] Sha, G. (2010). Using Google as a super corpus to drive written language learning: a comparison with the British National Corpus. *Computer Assisted Language Learning*, 23(5): 377-393. <https://doi.org/10.1080/09588221.2010.514576>
- [30] Park, K. (2012). Learner–corpus interaction: A locus of microgenesis in corpus-assisted L2 writing. *Applied Linguistics*, ams012: 1-26.
- [31] Conroy, M. A. (2010). Internet tools for language learning: University students taking control of their writing. *Australasian Journal of educational technology*, 26(6): 861-882. <https://doi.org/10.14742/ajet.1047>
- [32] Geluso, J. (2013). Phraseology and frequency of occurrence on the web: native speakers' perceptions of Google-informed second language writing. *Computer Assisted Language Learning*, 26(2): 144-157. <https://doi.org/10.1080/09588221.2011.639786>
- [33] Mueller, C. M., & Jacobsen, N. D. (2016). 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>
- [34] Quinn, C. (2014). Training L2 writers to reference corpora as a self-correction tool. *ELT Journal*, ccu062:1-13.
- [35] Huang, Z. P. (2014). The effects of paper-based DDL on the acquisition of lexicogrammatical patterns in L2 writing. *ReCALL*, 26(2): 163-183. <https://doi.org/10.1017/S0958344014000020>
- [36] Chen, M. H., Huang, S. T., Chang, J. S., & Liou, H. C. (2015). Developing a corpus-based paraphrase tool to improve EFL learners' writing skills. *Computer Assisted Language Learning*, 28(1): 22-40. <https://doi.org/10.1080/09588221.2013.783873>
- [37] Cotos, E. (2014). Enhancing writing pedagogy with learner corpus data. *ReCALL*, 26(2): 202-224. <https://doi.org/10.1017/S0958344014000019>
- [38] 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>

## 8 Authors

**Qinqin Luo** is currently a lecturer in School of Foreign Languages, Southwest Petroleum University, Chengdu, China. Her research interests include computer-assisted language teaching, second language writing and corpus linguistics. (Email: qinqinluo@126.com).

**Jie Zhou** is a student in School of Foreign Languages, Southwest Petroleum University. Her research area includes applied linguistics and language pedagogy.

Article submitted 15 December 2016. Published as resubmitted by the authors 21 February 2017.