Vocabulary learning through data-driven learning in the context of Spanish as a foreign language

Gang Yao
University of Murcia / Spain

Abstract – An increasing number of studies have shown the potential associations between corpus work and second language acquisition and teaching. Some research, for example, explores the effect of data-driven learning (DDL, Johns 1991) in the context of foreign language learning. Up till now, however, empirical quantitative studies on the topic have been limited, especially with respect to foreign languages other than English. In order to bridge this gap, a quasi-experimental longitudinal design was used in the present study to examine whether there is a statistically significant difference between the DDL approach to vocabulary learning and more traditional learning methods (e.g., dictionary approach) in the context of Spanish as a foreign language (SFL) by Chinese students. The study further gauged students' attitude towards DDL activities. The results of two post-tests revealed that the DDL group of students significantly outperformed the group of students following a traditional learning method. Furthermore, a questionnaire assessment collected from the experimental group showed that the respondents generally favored DDL and adopted a positive attitude towards its future application to Spanish learning.

Keywords – Data-driven learning; Spanish as a foreign language; vocabulary learning; empirical study

1. INTRODUCTION[1]

Vocabulary is considered to be one of the most important elements when learning a foreign language (Nation 2001). Lewis (1993: 89) comments that "lexis is the core or heart of language […]." Nation (2001) and Nation and Meara (2010) state that knowledge of vocabulary enables language use, which can be reflected in all language skills (i.e., listening, speaking, reading, and writing). Similarly, both Schmitt (2000) and

---

Jiménez-Calderón and Sánchez-Rufat (2017) agree that lexical knowledge is fundamental to communicative competence and second language acquisition. Barcroft (2005) highlights the importance of vocabulary in three aspects: communication, perception, and the way in which the knowledge of grammar is stored in the mind.

Although we are conscious of its importance, there exist many challenges and problems in second language vocabulary acquisition, a fact that is especially true in Chinese SFL context. Vocabulary teaching and learning procedures are relatively simple and old-fashioned in China (Guan 2013). For instance, the example sentences used for vocabulary teaching are often extracted from traditional monolingual or bilingual dictionaries or, in some cases, teachers themselves invent or compile those examples. Such sentences, which may have little authenticity and contextual adequacy, are unlikely to arouse learner's interest and attention. Meanwhile, students still draw heavily on teacher's explaining and rote learning (Chang 2001). This top-down learning process weakens their initiative and autonomous learning since they tend to receive linguistic input passively.

With the advent and development of computer technology, computer-assisted language learning (CALL) has become possible and is maturing. One thing that affects CALL significantly is the use of digitized corpora. From the very first Brown Corpus to large-scaled modern corpora (e.g., COCA, WebCorp), corpora have been exerting a considerable influence on language teaching and learning in many ways (O'Keeffe *et al.* 2007, Szudarski 2018). As freely available and easy-to-access corpus resources come into being, the above-mentioned problems and challenges in second language vocabulary can be addressed from a fresh angle, namely, corpus-aided vocabulary learning (Guan 2013; Yılmaz and Soruç 2015; Karras 2016). And one of the most significant representations of corpus-aided learning is data-driven learning (DDL), proposed by Johns (1991).

2. DATA-DRIVEN LEARNING: THEORETICAL BACKGROUND

DDL is a new approach to language learning and teaching in which students can inductively discover linguistic features and regularities by exploring "real and authentic language data" (Johns 1991). Essentially, DDL exploits the techniques of corpus use in contrast to traditional learning strategies, in which textbook learning and teacher's

explanations feature prominently. But the problem is that users cannot read a corpus directly; instead, they need to draw upon a program or software, namely, a concordancer. Using this interface, one can retrieve concordance lines after requesting a word, a phrase or a regular expression in the search bar. The concordance lines are typically presented in the form of Key Word in Context (known as KWIC), with the keyword displayed centrally and some words before and after (cf. Figure 1). Observing the concordance formatted as KWIC, the learner can, for instance, easily discover patterns, be sensitive to collocations, and enhance their learning strategy (Thurstun and Candlin 1998; Pérez-Paredes 2010). Take the example in Figure 1, i.e., the Spanish verb *infringir*. With the co-text around the search word, it would not be difficult for a learner to notice that the word usually takes nouns related to the law as objects. Now we can glimpse the main characteristic of DDL, namely that learners themselves discover linguistic regularities and make generalizations about linguistic phenomena based on observation, analysis, induction, and conclusion. This type of learning is also called *discovery learning* (Bernardini 2000, 2004). After an amount of training, *the learner as researcher* (Johns 1991; McEnery and Wilson 1997; Gavioli 2001) can engage in his own linguistic analyses and exploration.

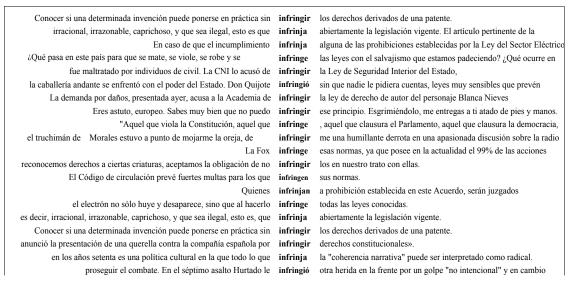|		
Conocer si una determinada invención puede ponerse en práctica sin	**infringir**	los derechos derivados de una patente.
irracional, irrazonable, caprichoso, y que sea ilegal, esto es que	**infrinja**	abiertamente la legislación vigente. El artículo pertinente de la
En caso de que el incumplimiento	**infrinja**	alguna de las prohibiciones establecidas por la Ley del Sector Eléctrico
¿Qué pasa en este país para que se mate, se viole, se robe y se	**infringe**	las leyes con el salvajismo que estamos padeciendo? ¿Qué ocurre en
fue maltratado por individuos de civil. La CNI lo acusó de	**infringir**	la Ley de Seguridad Interior del Estado,
la caballería andante se enfrentó con el poder del Estado. Don Quijote	**infringió**	sin que nadie le pidiera cuentas, leyes muy sensibles que prevén
La demanda por daños, presentada ayer, acusa a la Academia de	**infringir**	la ley de derecho de autor del personaje Blanca Nieves
Eres astuto, europeo. Sabes muy bien que no puedo	**infringir**	ese principio. Esgrimiéndolo, me entregas a ti atado de pies y manos.
"Aquel que viola la Constitución, aquel que	**infringe**	, aquel que clausura el Parlamento, aquel que clausura la democracia,
el truchimán de Morales estuvo a punto de mojarme la oreja, de	**infringir**	me una humillante derrota en una apasionada discusión sobre la radio
La Fox	**infringe**	esas normas, ya que posee en la actualidad el 99% de las acciones
reconocemos derechos a ciertas criaturas, aceptamos la obligación de no	**infringir**	los en nuestro trato con ellas.
El Código de circulación prevé fuertes multas para los que	**infringen**	sus normas.
Quienes	**infrinjan**	a prohibición establecida en este Acuerdo, serán juzgados
el electrón no sólo huye y desaparece, sino que al hacerlo	**infringe**	todas las leyes conocidas.
es decir, irracional, irrazonable, caprichoso, y que sea ilegal, esto es, que	**infrinja**	abiertamente la legislación vigente.
Conocer si una determinada invención puede ponerse en práctica sin	**infringir**	los derechos derivados de una patente.
anunció la presentación de una querella contra la compañía española por	**infringir**	derechos constitucionales».
en los años setenta es una política cultural en la que todo lo que	**infrinja**	la "coherencia narrativa" puede ser interpretado como radical.
proseguir el combate. En el séptimo asalto Hurtado le	**infringió**	otra herida en la frente por un golpe "no intencional" y en cambio

Figure 1: Concordance of 'infringir' retrieved from the 'Corpus de Referencia del Español Actual' (CREA) by the Royal Spanish Academy

However, there arises another problem: How to train our learners to become familiar with DDL? Teachers play a role here: not a dominant one but as "director and coordinator" (Johns 1991: 3), designing concordance-based exercises to provide their students with practice. According to Boulton (2010a, 2010c), there are two types of

DDL exercises: one is hands-on DDL, which corresponds to "direct corpus consultation" (Chambers 2007: 4); the other one is hands-off DDL, which is equivalent to "indirect corpus consultation." In the first type of DDL, students are given considerable individual autonomy, which can be considered as a 'pure' DDL. However, in the case of inexpert learners, for instance, the extensive data retrieved from corpora may frustrate them because the data sometimes are "irrelevant", "incomprehensible", and "extremely chaotic" (Boulton 2010c: 6). Besides, corpus data are normally produced by native speakers, and thus the difficulty may lie outside the language competence of a learner. Lastly, direct corpus consultation requires access to a computer, which means learners have to know basic computer skills and schools may need to be equipped with multimedia rooms. But this condition is hard to achieve for some students and education centers (Pérez-Paredes 2005). Conversely, the hands-off DDL, a "soft version" (Gabrielatos 2005), is popularly introduced into regular classrooms as corpus resources. It is the teacher who consults the corpora directly, then selects appropriate language data depending on the learner's level, and finally prepares concordance-based exercises (e.g., handouts, worksheets). The advantages of this type of DDL may be easily noticed: students can still stand to benefit from direct access to authentic language data (instead of direct corpus consultation). At the same time, computer knowledge and competencies are not required. Thompson (2006) also points out that selected materials can help our learners turn their attention to the key elements, reduce confusion, and confine the range of possible answers (cf. also Stevens 1991). For newcomers to the corpus, preselected concordance lines make linguistic features more noticeable (cf. Sripicharn 2010). Moreover, prepared materials could reduce cognitive load at the beginning since learners only need to focus on a single new element (Boulton 2010a).

3. PREVIOUS STUDIES ON DDL

The use and effects of DDL for language learning and teaching have been studied thoroughly and systematically by many researchers. The emergence of several holistic surveys and syntheses suffices to show the trends and popularity of DDL, such as Chambers (2007), Boulton (2008, 2010b, 2017a), Boulton and Cobb (2017), Mizumoto and Chujo (2015), and Lee *et al.* (2019). All of them have given a fair summary of previous studies on DDL, and some of them have even conducted a meta-analysis. The

state-of-the-art review in this section will only focus on empirical studies which, in turn, can be divided into: qualitative analyses and quantitative analyses. It should be noted that there does not exist a strict boundary between these two categories. In practice, many studies adopted both methods to complement one another, i.e., mixed methods. Labeling a study qualitative only means the study has more qualitative characteristics than quantitative ones; and vice versa.

3.1. Qualitative studies

As stated in Boulton (2017b: 185), many initial publications related to DDL lie in "emic studies", with the goal of exploring what learners think about DDL; in other words, the evaluation of DDL. Specifically, there are three types of evaluation (Boulton 2008, as cited in Gilquin and Granger 2010: 365): evaluation of attitudes (what do learners think about DDL?), practices (how well do users work with DDL?), and efficiency (can learners really benefit from DDL?). In emic studies, information of this kind is usually collected through interviews, learning logs, and especially questionnaires.

The focus of this type of empirical study is usually on learners' written production with the aim of improving writing skills (e.g., Chambers and O'Sullivan 2004; O'Sullivan and Chambers 2006; Kennedy and Miceli 2010; Charles 2012; Chang 2014). Yoon and Hirvela (2004), for example, examined corpus use in students' L2 writing and their perceptions of it. The authors combined qualitative and quantitative analysis in their study. The feedback from the students was generally positive and most learners indeed favored corpus-assisted writing. In particular, corpus assistance was deemed beneficial in terms of learning common usage and collocates and boosting student's confidence in writing. There were, however, several problems or difficulties reported in the study, such as the time that was wasted on the corpus searches and students' proficiency level as an essential factor in corpus work.

Qualitative studies on vocabulary acquisition are relatively new. Jiao (2012) utilized corpora to help students ($N = 87$) learn English vocabulary. After a one-semester instruction with the aid of corpora, all participants were invited to take a survey. Again, the students generally acknowledged the merit of corpus work since it can contribute to autonomous learning and help the students grasp the correct usage of vocabulary (collocation, colligation, semantic prosody, etc.). Likewise, Tekin and Soruç

(2016) enabled 26 participants from a Turkish high school to use BNC to learn four target words and then received their reflective journals. The qualitative findings showed that the students considered corpus-assisted vocabulary learning activities interesting, innovative, practical but also complex. Aşık *et al.* (2016) reported 126 Turkish EFL learners' perceptions of DDL regarding lexical awareness and development. Although the data they collected was based on questionnaires and interviews, the authors quantified it to carry out a statistical analysis. The results revealed that the students held overall positive opinions about DDL tasks, although improvements can only be seen in certain aspects of lexical awareness, such as synonyms and collocations; while awareness concerning word frequency, idioms, and learning strategies did not achieve a satisfactory result.

Despite the fact that the majority of studies above lack quantitative data and statistical analyses, it does not mean they are of little value. On the contrary, during the initial phase of the development of DDL, qualitative research on learners' attitudes and behaviors is undoubtedly helpful for other researchers who want to know what has been done so far in this field and what are the advantages/problems of DDL (Chambers 2007).

3.2. Quantitative studies

Even though DDL has been developing and perfecting since the 1990s, we should address ourselves to some key questions: Does DDL indeed work for foreign language learning? To what extent it is effective? To answer these questions, it is necessary to provide more empirical evidence that focuses on measurable outcomes in order to shed light on the effectiveness and efficiency of DDL. That is, quantitative studies that observe DDL from a "more etic perspective" (Boulton 2017b: 186) are needed.

Within quantitative studies, according to Boulton (2017a, 2017b), two categories can be identified depending on the purpose of corpus use. But, again, there is no watertight delimitation between them. The first group aims to evaluate the effect of the corpus as a reference resource, particularly corpus use in practice exercise, translation or learner's written revision. Interestingly, most studies that fit into this category are qualitative (cf. Section 3.1). The reason behind this may be that it is difficult to quantify

the results of writing or translation evaluation. Notwithstanding, there are several quantitative studies that employ corpus as a reference resource.

For example, Gaskell and Cobb (2004) carried out a longitudinal experiment in an intermediate-low level English writing course. A total of 20 learners of English were involved in this experiment for over 15 weeks. The task consisted mainly of error identification, corpus consultation with instructor's aids, and independent searches in the corpus. The comparison of students' writings between the pre-test and post-test, as well as a questionnaire assessment, indicated that all students thought they had achieved improvements in grammar and error correction after the course. In Gilmore's (2009) short-term study, 45 intermediate-level Japanese learners of English first received a 90-minute training session to solve lexical and grammatical problems they had encountered in their writings. Their second compositions, which were graded by four native speakers of English, showed a significant improvement in terms of naturalness. Crosthwaite (2017) examined the adequacy of corpus use for student error correction in L2 writing during a series of DDL course. Teachers offered error feedback while students highlighted revisions made with corpus consultation or without it. The quantitative results revealed that with corpus-mediated correction, students can prevent lexical errors more successfully, but they were less likely to correct morphosyntactic errors. Students' feedback from the post-course questionnaires also confirmed the quantitative results. A similar writing enhancement experiment, conducted by Cotos *et al.* (2017), incorporated a corpus-based platform—Research Writing Tutor (RWT)—into a one-semester writing course. The RWT can automatically evaluate students' drafts and give them rhetorical feedback. Multiple comparisons in a mixed-methods design revealed that RWT-enabled DDL activities could improve the quality of students' writing, for instance in genre awareness.

The other group of studies that focus on the effect of the corpus as a learning aid shows an interest in examining how the corpus can assist learners in linguistic elements of language learning such as vocabulary and grammar. Stevens' (1991) study could be considered a pioneering work in this sense. He innovatively used concordance-based exercises, instead of conventional gap-fillers, to aid students in vocabulary learning. The result suggested that this new type of exercises was easier and more useful for learners and that it can become a viable alternative to the traditional exercises. Cobb (1997, 1999) put DDL into practice in a stricter sense by designing and introducing a

well-known web-based platform—Compleat Lexical Tutor.[2] In his experiments, one group of students were asked to learn 240 English words based on that interactive platform during one semester, while the other group of students followed a traditional teaching method with a dictionary. The findings suggested that both treatments were effective for the acquisition of word meaning in a short period, but only the experimental group performed significantly well on the retention of vocabulary for an extended period.

Entering the 21st century, studies on vocabulary learning through DDL start to appear. For example, Allan (2006) carried out an experiment in which 18 advanced learners of English were engaged. The experimental group ($N = 13$) was given the concordance-based task to learn vocabulary for over 12 weeks. A quantitative analysis of the results indicated that the DDL group outperformed the other group, although the conditions of the two groups during the experiment were not entirely comparable. A more in-depth study was carried out by Anani-Sarab and Kardoust (2014), who investigated the potential and implication of corpus in the context of English as a Foreign Language (EFL) in an experiment with 34 Iranian students who were preparing for an English test. The experimental group adopted DDL activities to learn phrasal verbs, while the control group did the same through dictionary-based activities. After 14 sessions of instruction, the results from the immediate and delayed post-tests showed that the DDL group achieved greater improvements. However, this study did not include students' assessment of DDL activities in the experiment, which is popularly considered an important aspect in research of this kind.

Yılmaz and Soruç (2015) and Soruç and Tekin (2017) examined the effectiveness of DDL on vocabulary learning and teaching by contrasting the concordance-based vocabulary instruction with the traditional instructions (such as dictionary definitions, synonyms, fill-in-the-blank exercises). Despite the difference in settings, both experiments reported that DDL vocabulary learning activity yielded better results after comparing the pre-test and the post-test. Soruç and Tekin (2017) integrated a further delayed post-test to compare it with the immediate post-test. The result also supported the superiority of DDL. Interviews from both experiments also recorded students' positive attitudes towards it. Nonetheless, the procedure of Yılmaz and Soruç's (2015) study was not entirely clear. For example, the duration of the DDL activity and its retention effect

---

[2] Compleat Lexical Tutor: https://www.lextutor.ca/.

were unknown. Differing from other research on vocabulary learning through DDL, Karras (2016) conducted a more large-scale study, in which 100 international students of a Vietnam secondary school participated over eight weeks. In this study, both the experimental and control groups had online dictionary learning activities, but the first group received an extra DDL training. Based on the weekly results, the author reported that both groups achieved improvements but the DDL group obtained significantly higher scores than the other group. It is worth noting that the factor of grade level influenced the effect of DDL. However, since the two groups received a different amount of treatment, it cannot be said that they were comparable.

In addition to vocabulary learning, corpus work has been proved helpful in learning other linguistic features. For instance, corpus-driven lexico-grammatical learning has been shown to provide favorable outcomes, primarily on collocation and colligation (Chan and Liou 2005; Koosha and Jafarpour 2006; Huang 2014; Daskalovaska 2015; Vyatkina 2016; Li 2017). Boulton (2009), Smart (2014), and Moon and Oh (2017), in turn, paid more attention to aspects of grammar such as the passive voice, the overuse of *be*. All these studies indicate that DDL activities significantly improve learners' grammatical capacity.

More recently, the application of DDL is diversifying; in other words, DDL research does not only cover linguistic features (e.g., vocabulary, grammar, writing), but also shifts attention towards language comprehension and production (Frankenberg-Gacia 2014), reading (Hadley and Charles 2017), and pragmatic routines (Bardovi-Harlig *et al.* 2017).

To conclude, previous studies, both qualitative and quantitative, have shed light on the procedures and effects of learning a foreign language through corpora. While corpus use has been applied for diverse purposes in second language learning and teaching, there is still a particular pedagogical and research interest in vocabulary learning, as can be seen above. Though the majority of studies on the topic claim that the DDL approach contributes greatly to language learning, there are some limitations we should not ignore. The first and most evident problem is that most previous studies take English as the target language (Römer 2011; Vyatkina 2016), as also evidenced in a series of comprehensive surveys and syntheses (Chambers 2007; Boulton 2008, 2010b; Boulton and Cobb 2017). Consequently, the effect of DDL on other languages is inadequately tested and even remains unstudied. Secondly, according to Asención-

Delaney *et al.* (2015), there is a lack of systematicity across studies, and it is hard to make generalizations based on the findings. Lastly, some studies fail to satisfy the conventional norms of empirical research, which makes them difficult to replicate.

Given that the said challenges and issues would, to some extent, diminish the effect of DDL, more methodologically sound studies seem necessary. Therefore, the purpose of the present study is to critically and systematically examine the effect of DDL on Spanish vocabulary learning and contrast it with traditional dictionary-based activity learning. These are two research questions in the study:

1. Are DDL activities more effective than dictionary-based exercises in Spanish vocabulary learning? If so, to what extent?
2. What were the reactions of our learners of Spanish in the experimental group towards DDL activities?

4. METHODOLOGY

This experiment was designed to compare the efficacy of learning Spanish vocabulary through a DDL approach and a traditional dictionary-based approach after a longitudinal observation. To this end, two groups of students were involved in the study: one group dealt with paper-based DDL materials (hands-off DDL) while the other group worked with dictionary-based materials. Over the three-week experiment, the subjects were asked to participate successively in the pre-test, the immediate post-test, and the delayed post-test. A questionnaire was targeted at assessing how learners in the first group perceived DDL. All the collected data underwent a quantitative and qualitative analysis.

4.1. Participants

A total of 34 conveniently available university students were recruited for this experiment, but only 32 of them completed the whole procedure. All participants were in the 20–23 age range. Most of them were female (30) while there were only two males. For the sake of comparison, 32 participants were divided into two groups: the experimental group and the control group (henceforth referred to as EG and CG respectively). Each group was made up of 16 subjects.

All participants had a similar language background and foreign language learning experience. They were all L1 Chinese undergraduates majoring in Spanish Language and Literature. They had received two or three years of formal instruction in Spanish. In general, their Spanish proficiency level was upper-intermediate.[3] All of them reported that they had not had any prior experience on corpus work.

4.2. Materials

In order to analyze if vocabulary learning through DDL is more effective than learning it through dictionary consultation, a total of 38 Spanish words were selected as candidates for target items. The 38 words were extracted from the *Curricular Plan for Advanced Courses of the Specialty of Spanish* and considered advanced vocabulary, which was supposed to be unfamiliar to our participants.[4] After carrying out a pre-test in order to remove items already known by the participants, only 10 of the 38 initial words were used, namely, *clandestino, esporádico/ca, exponencial, inverosímil, latente, palpable, proliferación, rehusar, tajante,* and *vehemente* (cf. Section 4.3 for details).

Further, 30 words grouped in pairs of three were selected in order to examine if DDL is more effective than traditional learning methods, i.e. dictionary use, in allowing a distinction between confusing synonyms. The words in question were *someter, obligar, imponer, optativo, opcional, selectivo, énfasis, hincapié, relieve, proveer, suministrar, proporcionar, asignar, designar, resignar, hostil, adverso, opuesto, intenso, intensivo, tenso, colmar, apartar, involucrar, diametralmente, integralmente, sospechosamente, respetuoso/sa, respetable,* and *respetado*. Their selection was based on the suggestion, by several experienced teachers of Spanish, that these were still problematic for students of Spanish despite being familiar with them.

The learning materials consisted of two worksheets covering the same 20 target items for both groups but with different contents. Materials for the EG were based on concordance lines extracted from the annotated version of the *Corpus de Referencia del Español Actual* (CREA). The concordance lines were selected carefully according to Gilquin and Granger's (2010: 362) criteria: 'readability', 'frequency', and 'usefulness'.

---

[3] We chose upper-intermediate learners as participants since they were believed to have sufficient language proficiency to read high-level learning materials (such as corpus concordance and monolingual dictionaries) without the need to consult other reference books.

[4] This curricular plan is used to guide teachers and students of Spanish through the teaching and learning of Spanish in advanced courses.

Therefore, all selected concordance lines tried to avoid cut-off sentences as much as possible in order to "enhance familiarity and comprehensibility" (Moon and Oh 2017: 6). Besides, only frequent and common usages (i.e., collocation, colligation) of the target items were included so that learners could guess the word meaning and find the collocational patterns (Sripicharn 2003). To reduce participants' reading burden, manageable quantities of lines (3–5 lines in our case) were more appropriate (cf. Cobb 1997). A DDL learning material sample is given below (Figure 2).

<u>Infringir</u>

1. ¿Qué pasa en este país para que se mate, se viole, se robe y se   **<infrinjan>** las leyes con el salvajismo que estamos padeciendo? ¿Qué ocurre en la juventud?

2. ¿Le pongo algunos ejemplos? El delincuente habitual es un hombre que ha decidido **<infringir>** las leyes para vivir. Comprende la necesidad de las leyes, no las discute, pero se las salta.

3. Notificamos al administrador del sitio de que estaban **<infringiendo>** las leyes electorales de California y que era necesario parar la actividad del sitio Web,

4. si es que el fiscal determina, como lo hizo con los otros dos detenidos, que **<infringieron>** la ley y por lo tanto pueden ser incluso sentenciados.

5. Si se trata de señalar a los muchachos cuando usan pelo largo, arete o drogas, de alcoholizarse, de **<infringir>** normas y causar problemas, tendríamos primero que cuestionarnos a nosotros mismos

(*fuente: Corpus de Referencia del Español Actual*)

Figure 2: Sample of DDL learning materials

Dictionaries usually offer a distinct point of comparison (Cobb 1997; Yoon and Hirvela 2004; Boulton 2010a; Anani-Sarab and Kardoust 2014, Karras 2016) since they are one of the most common resources of foreign language learning and teaching. In the worksheet of the CG, dictionary definitions or example sentences were taken from two authoritative Spanish monolingual dictionaries: 1) *Diccionario de Uso del Español* by Moliner (2007), and 2) *Diccionario de la Lengua Española* by the Royal Spanish Academy. Note that for those polysemous target items, only the dictionary meanings that matched the learning materials of DDL group were used. A sample of dictionary-based learning materials is offered below (Figure 3).

infringir

Infringir la ley.

Infringir las disposiciones sobre abastos.

(fuente: Diccionario de uso del español)

Figure 3: Sample of dictionary-based learning material

4.3. Instruments

All test instruments were identical for all participants; namely, a pre-test, an immediate post-test, and a delayed post-test, except that the learning materials for the two groups in the immediate post-test were different, as pointed out before. Thus, any difference that the tests would produce can be attributed to the aid of corpus resources (Cobb 1997).

The aim of the pre-test was twofold: 1) to discard advanced words that were familiar to the subjects (cf. Anani-Sarab and Kardoust 2014; Yılmaz and Soruç 2015), and 2) to prove the homogeneity of the two groups in terms of language competence before the post-tests (Johns *et al.* 2008). For the first aim, all participants were asked to take a small and straightforward word recognition test. They had to select the words they recognized and give their corresponding definitions in L1 or L2. Note that some non-target words were added to the pre-test to disguise the target words (cf. Yılmaz and Soruç 2015). For the second aim, the scores obtained by the participants in the EEE-4 were used since they were conveniently available.[5]

The immediate post-test aimed to examine the performance of two groups after using differential learning materials. The test consisted of two parts and each part contained 10 familiar multiple-choice questions that covered all target items, with different learning materials displayed. Questions in the first part had been adapted directly from Cobb's (1997) work while the last ten questions were inspired by Boulton's (2010a) study. Each question was worth 5 points. Correct answers for each scored 5 points and incorrect answers, in turn, 0 points. The maximum score of the immediate post-test was thus 100.

Regarding the delayed post-test, its implementation was mainly targeted at comparing the effects of two differential treatments on the retention of vocabulary knowledge over a long period. The design of this test was practically identical to the

---

[5] EEE-4 is a Chinese national exam for undergraduate students learning Spanish, which is usually organized during the final semester of the second academic year.

former one, except that no learning materials were provided to our participants this time. Thus, the results obtained from the test were based entirely on the recall of those target items that the participants learned from the last test. Another noteworthy aspect is that the grading system for the first part of the delayed post-test was different from the previous one. The first ten questions were scored using an adapted version of the Vocabulary Knowledge Scale (Wesche and Paribakht 1996), which was a self-report scale that allowed our participants to assess how well they knew the items. Specifically, there were three scales and each scale represented different scores (cf. Table 1). In the last ten questions, the format and grading were the same as before. However, the order of questions was altered to avoid" practice effectu (cf. Anani-Sarab and Kardoust 2014).

Scale (translated from the Chinese version)	Score
I know this word and I'm sure it means_____	5
I know this word but I'm not sure what it means. It could mean_____	2
I've seen this word before, but I don't know what it means.	0

Table 1: Adapted vocabulary knowledge scale for the first ten questions of the delayed post-test

After the immediate post-test, the participants of the EG were asked to fill out a questionnaire about their perceptions of the previously performed test. The questionnaire, written in L1, consisted of two types of questions. The first one consisted of ten questions with a 5-point Likert scale. With this scale, respondents were requested to rate each question according to their agreement or satisfaction. After they had answered, scores were assigned to each response. For instance, strongly agree/very satisfied = 5; agree/satisfied = 4; neither agree/satisfied nor disagree/dissatisfied = 3; disagree/dissatisfied = 2; strongly disagree/very dissatisfied = 1. Note that to compensate for the acquiescence response bias (Lavrakas 2008), i.e., a tendency for people to agree rather than disagree in statements, some items in the questionnaire were purposely constructed, in other words, negatively worded questions, such as item 5 and 9. For example, item 9 is translated as "Do you think reading concordance lines wasted your time?" The second type was made up of five open-ended questions, in which our respondents had a chance to comment on the advantages and weaknesses of the DDL

approach. Some question samples are translated as follows: "In addition to the understanding of words' meaning, what else did you learn from the concordance lines?" or "In comparison with traditional dictionary-based learning strategy, what are the advantages and shortcomings of concordance-based learning strategy?"

4.4. Procedure

Due to limited resources and logistic problems, the experiment could not be carried out in an ordinary classroom. All test materials were posted on an online platform.[6] Admittedly, it is hard to reach the same conditions as normative tests do, but this form had several advantages: 1) subjects can freely access to the tests regardless of their location, 2) participants may not experience the same pressure as under examination conditions, and 3) the online platform can automatically grade the test and collect valuable data such as the time that each participant spends on the tests.

Before each test, a reminder was given to all participants. Firstly, the aim of the experiment was to test whether they could infer the meaning of those unfamiliar words and distinguish the synonyms based on the provided learning materials. Secondly, during the test consulting any reference books (e.g., dictionary, textbook) or online tools was not allowed; thirdly, there was no time limit for the test so they had sufficient time to read the materials and finally choose the most appropriate answers according to their level of knowledge.

The pre-test was performed once the 32 subjects had been recruited. The primary purpose of this test, as mentioned before, was to select the words that our participants could not recognize. In the end, 10 words that fulfilled the requirement were chosen. Having done this, these 10 target items, along with 10 other pairs of synonyms, were used to design materials for the later post-tests.

The immediate post-test was conducted one week after the first test. Before the participants started to answer the questions, two practice items were added so that they could familiarize themselves with the two different types of questions (cf. Stevens 1991). There were three short pauses over the test with the purpose of mitigating the fatigue effects. When the immediate post-test finished, the 16 subjects of the EG were asked to complete the questionnaire.

---

[6] Wenjuanxing: https://www.wjx.cn/

The last test was performed two weeks later. The participants were not informed about the form of the test so that they were not able to prepare for it or consult dictionaries beforehand. This time no practice items were provided since they were already familiar with the question types. Test breaks, however, remained unchanged.

5. RESULTS

5.1. Test results

The first step was to ensure the homogeneity of the two groups before two post-tests. As mentioned before, the scores that our participants obtained from the EEE-4 were conveniently utilized. The descriptive statistics of the scores roughly indicated similar language competence between the EG and the CG (M_{EG} = 79.56, SD_{EG} = 6.18; M_{CG} = 77.31, SD_{CG} = 6.77). A later t-test confirmed that the two groups did not differ significantly (p = .334), thereby suggesting that any diverging outcome would be due to different experimental treatments, i.e., DDL approach and traditional method.

Based on the data collected in the immediate post-test, it can be clearly noticed that the EG (M = 78.75, SD = 12.58) achieved greater performance than the CG (M = 50.00, SD = 13.66). However, this did not show whether the difference between them was statistically significant or not. An independent samples t-test based on groups' mean score seemed appropriate. But before testing this hypothesis, it was necessary to check whether the collected data satisfies the assumptions of the t-test, namely, õthe normality assumptionö (Brezina 2018: 13). The type of normality tests to be chosen depends on the sample size. If the sample size is big, Kolmogorov-Smirnov test is used and, if it is not, Shapiro-Wilk test is used instead.[7] Given that our sample size is small (N = 32), the Shapiro-Wilk normality test was run. The result demonstrated that the data of the two groups did conform to a normal distribution (p_{EG} = .078, p_{CG} = .577). The next step was to carry out the t-test, whose result (cf. Table 2) indicated that there was a statistically significant difference between the DDL group and the dictionary-based group in the immediate test [t(30) = 6.191, p < .001, 95% CI (19.27, 38.23)]. However, what the t-test was unable to tell is "how large this difference is and whether it is

---

[7] There is no accurate standard reference regarding if the sample size is big enough for a Kolmogorov-Smirnov test, but the conventional cut-off size is 50 (https://statistics.laerd.com/spss-tutorials/testing-for-normality-using-spss-statistics.php). Recent research, however, shows that t-test can be robust to the violation of the normality assumption, which means that even if the data follow an abnormal distribution that does not interfere with valid results of the t-test (cf. Brezina 2018).

practically important" (Brezina 2018: 14). Therefore, it is necessary to report effect sizes here. In our case, the well-known Cohen's *d* was used to measure the effect size of the difference. The result ($d = 2.19$) suggested that it was much larger than Plonsky and Oswald's (2014) L2 field-specific criterion for a large effect size ($d = 1.00$).

Groups	*N*	*M*	*SD*	*t*-value	df	*p*
Experimental group	16	78.75	12.58	6.191	30	< .001
Control group	16	50.00	13.66			

Table 2: Result of *t*-test for the two groups in the immediate post-test

The same statisticcn tests were then performed in the delayed post-test. First, the Shapiro-Wilk normality test showed that the data collected from the two groups were also normally distributed ($p_{EG} = .364$, $p_{CG} = .069$). Subsequently, the results of the *t*-test are presented in Table 3. As can be noticed, there was a statistically significant difference between the two groups [$t(30) = 2.600$, $p = .014$, 95% CI (3.54, 29.46)]. Moreover, the effect size of this difference was $d = 0.92$, which was close to the L2 field-specific benchmark for a large size (cf. Plonsky and Oswald 2014).

Groups	*N*	*M*	*SD*	*t*-value	df	*p*
Experimental group	16	49.19	18.29	2.600	30	.014
Control group	16	32.69	17.61			

Table 3: Result of *t*-test for the two groups in the delayed post-test

Based on the results of the between-group comparisons in two post-tests, the positive effect of DDL approach on Spanish vocabulary learning is quite clear and statistically meaningful. It seems that learners in the EG performed better in terms of lexical awareness and word recall with the aid of corpus resources, i.e., concordance-based materials. The CG, in contrast, did not achieve satisfying results using only dictionary-based materials. Nevertheless, it is worth noticing that there was a marked decrease in the mean score of both groups between the two post-tests, as illustrated in Table 4.

	Experimental group	Control group
Immediate post-test	78.75	50.00
Delayed post-test	49.19	32.69
Difference	-29.56	-17.31
Change (% of the immediate post-test)	-37.54%	-34.62%

Table 4: Result of within-group comparison between the immediate post-test and delayed post-test

As can be seen in Table 4 above, both groups experienced a decrease of nearly 35% in the mean score, which means that all participants had difficulties in recalling words learned two weeks before. Further examination revealed that the first part of the test materials (i.e., advanced vocabulary) contributed to such a significant decrease: the average score in this part declined by 57.48% (-55.90% for the EG and -59.06% for the CG). In the synonym discrimination part, however, some students showed better performance in the delayed post-test than the immediate one. Possible explanations for this will be discussed in due course.

5.2. Questionnaire results

The ten Likert-type questions were analyzed first. According to their attitudes and beliefs, 15 out of 16 participants were generally satisfied with the concordance-based activity, 13 students liked the sentences displayed in KWIC format and 15 agreed that the concordance lines were readily comprehensible. All participants agreed that concordance lines provided a rich context for the target vocabulary and 15 considered that concordance lines helped them differentiate those synonyms. The encouraging feedback from the participants was enhanced when they were asked whether they would support the application of corpus resources to Spanish learning: 14 participants gave an affirmative answer.

Using a 5-point Likert scale, participants' perceptions were quantifiable. The mean score and standard deviation for each question are displayed below (cf. Table 5).

Item	Mean	*SD*
1	4.37	0.619
2	4.13	0.806
3	4.25	0.775
4	4.31	0.602
5	2.94	1.340
6	4.56	0.512
7	4.44	0.629
8	4.62	0.719
9	2.50	1.155
10	4.44	0.727

Table 5: Descriptive statistics for the first ten Likert-type questions

Overall, our participants rated the concordance-based activities highly, except for item 5 and 9. As mentioned in Section 4.3, these two items were asked from a negative perspective of DDL. A larger standard deviation of the two items suggested an evident variation in our participants' perception. In other words, some students indeed thought reading concordance lines wasted their time while others did not agree.

Regarding the highly-rated scores, if these questions *per se* were poorly designed, the result would not be as reliable as it looked. Therefore, Cronbach's alpha was utilized to measure the internal consistency estimate of the reliability of the rated scores. The first step was to eliminate item 5 and 9 since they were out of line with the other eight questions. After this, the reliability analysis was performed. The result is provided in Table 6: it indicates that the coefficient alpha for the eight *pro-DDL* questions is statistically acceptable (α = .74), and it confirms that the majority of our respondents had a positive attitude towards DDL.[8]

---

[8] According to the rules of thumb, $0.7 \leq \alpha < 0.8$ means the internal consistency is acceptable.

Cronbach's alpha	Number of items
0.74	8

Table 6: Cronbach's alpha reliability coefficient for 8 Likert-scale questions

With regard to the five open-ended questions, the participants had an ambivalent reaction to DDL activities. Most respondents considered the concordance lines clear, and they provided rich contexts for target vocabulary and synonym distinction. However, many expressed their dissatisfaction with the quantity and length of those concordance lines. Some unfamiliar words also affected the understanding of the whole sentence. When they were asked whether they had obtained more valuable information based on the concordance lines, 13 respondents answered yes. Among them eight mentioned that the collocations helped them recognize words and answer the questions, two said that they knew the common usage of the words thanks to concordance lines, and one pointed out that concordance lines of each synonym pair helped discriminate their meanings.

Given that all our participants had previous experience in learning Spanish through dictionary-based activities, the next question was about the advantages and shortcomings of concordance-based activities (i.e., DDL) when compared to traditional learning strategies. Half of the participants believed that they understood the meaning of the words more precisely with the help of contexts. Three of them stated they had a more profound impression of those words because of their repeated occurrences in the context. Four held the opinion that concordancing was useful in terms of synonym discrimination. Two learners gained benefit from the collocations. As regards the disadvantages, many complained about the time they spent on the test. This feeling was also evidenced by the time recorder of the online platform: the average duration of the immediate post-test for the EG was more than 30 minutes, contrasting with around 20 minutes for the CG. Four participants mentioned that the concordance lines were "too lengthy and too many" to have the patience to read them all. Another four claimed that, in this sense, the dictionary was more convenient and straightforward to consult the meaning of a word. An interesting comment by one of the participants was that "with this method [concordance-based approach], it is hard to understand the meaning of the

word, while a bilingual dictionary can provide more precise definitions and more common collocations," while another comment pointed out that "the vocabulary learned through concordance will be easily forgotten: concordance lines will only help memorize the meaning of the word and usages during a short run."

To summarize, most comments from our participants were considered positive. They generally agreed that the context was rich enough to learn words, collocations were easy to detect, and concordance lines helped the distinction of synonym pairs. At the same time, some problems of DDL should not be neglected, for instance, the time-consuming process and the interference from unfamiliar words. These drawbacks are also reported in several previous studies (cf. Cheng *et al.* 2003; Yoon and Hirvela 2004; Chambers 2005; Chambers 2007; Boulton 2010a; Geluso and Yamaguchi 2014).

6. DISCUSSION AND CONCLUSIONS

Based on the quantitative and qualitative analyses of the two post-tests and a questionnaire assessment, DDL activities have been found to be more effective and efficient than the traditional dictionary-based activities in terms of vocabulary learning. And the difference between the two treatments was statistically significant with a large effect size (research question 1). Moreover, most of our participants took a positive attitude towards DDL activities and its future application (research question 2).

The study has provided empirical evidence for the effectiveness of DDL activities on a language other than English, which is meaningful for the popularization and acceptance of DDL. Also, the study has followed the procedures of empirical research by drawing on previous pioneering studies (e.g., Cobb 1997; Boulton 2010a; Anani-Sarab and Kardoust 2014). Hence, the research design of the present experiment is methodologically and statistically sound. Besides, the details and procedure of the experiment have been reported as much detail as possible so that it would be feasible to replicate it in the future.

Although the findings of the present study are in line with the reassuring conclusions from prior research, such as Cobb (1997, 1999), Chambers (2005), Boulton (2007, 2010a), Soruç and Tekin (2017), the experimental results of our particular case should be interpreted cautiously. The first point to note is that two learning materials in the immediate post-test (cf. Figure 2 and Figure 3) presented an unbalanced amount of

information. The DDL learning materials had more textual input than the dictionary-based ones, a feature that also characterized Cobb's (1997) experiment. In our case, since less informative definitions and example sentences are one of the major flaws of two Spanish dictionaries (cf. Calderón-Campos 1994 for discussion), the learning materials for the control group inevitably offered less textual input. The contents of the concordance-based materials, on the other hand, may be problematic as well. For the good of our participants, only a manageable number of concordance lines (3–5 lines in our case) were selected. However, the cherry-picking lines may expose a bias in favor of researchers or teachers instead of catering for learners. For example, sometimes the concordance lines may have contained vocabulary that is beyond the proficiency level of learners, but the researcher may not have realized it. With a few lines, moreover, it is easy to go to extremes: at one end, concordance provides too many similar examples of one specific usage, and learners will get bored easily and, at the other end, concordance contains too little or no data that students would like to learn, and they probably will get frustrated (cf. Flowerdew 1996).

Secondly, from the immediate post-test to the delayed post-test, both groups underwent a substantial drop in performance (a loss of nearly 35% in the mean score), a result that merits further discussion. The primary factor could be that the forgetting curve was at work. Since the interval between the two post-tests was two weeks, and during that period our subjects did not receive any similar treatments, forgetting the knowledge learned before seems a logical outcome. Another explanation for this is that the students did not receive any feedback or correction after the immediate post-test. Thus, there was no way for them to know what the correct answers were, let alone to learn from the potential errors. In this situation, the participants would answer the questions of the last post-test based merely on what they learned during the immediate post-test. Theoretically, it seems unlikely that the performance in the delayed post-test reaches the same level as in the former post-test. Indeed, except for two subjects from the EG and another three from the CG who succeeded in keeping the same scores as the previous test, the remainder obtained lower scores in the last test.

Another potential issue relates to the results emerging from the questionnaire assessment. In both Likert-type and open-ended questions, the majority of our participants gave a positive evaluation of this brand-new learning method, a result in line with many other pertinent studies (e.g., Yoon and Hirvela 2004; Boulton 2010a;

Yılmaz and Soruç 2015; Moon and Oh 2017). Undoubtedly, participants' positive reaction is highly significant for the public acceptability of DDL and its future application to foreign language learning and teaching, but "their [participants'] subjective appreciations of their own learning may not be reflected in actual learning" (Boulton 2010b: 3), because the overwhelmingly positive reactions may be ascribable to 'the novelty factor' and 'the Hawthorne effect' (Boulton 2017b: 185). In addition, the acquiescence bias (Lavrakas 2008) in those responses is hard to avoid even with negatively worded items involved (cf. Section 4.3), and this is especially true when questions belong to the agree/disagree type. In other words, participants tend to put in a good word for the formulated statements because they would like to show their politeness and respect. Another issue to address here is that the coefficient alpha for the first part of the questionnaire was statistically acceptable but did not reach the benchmark of good or excellent. The reason could be due to the small number of Likert-type questions. Typically, a larger number of items lead to a larger α.

Despite the fact that the present study aims to examine the effectiveness and efficiency of DDL, there is one caveat: its objective does not lie in demonstrating that traditional methods like dictionary-based activities are not effective in foreign language learning and teaching. In reality, as one of the most traditional and classical study tools, the dictionary is and will play a fundamental role in language learning and teaching (Anani-Sarab and Kardoust 2014). Interestingly, nowadays corpora exert considerable influence on the publishing work and has hastened the birth of corpus-based dictionaries (McEnery and Wilson 1997; Römer 2011). In this sense, the new generation of dictionaries shows the corpus to good advantage. Returning to our study, it is noteworthy that the participants of both groups did not recognize the target items at first. But with the aid of both learning materials, all of them have acquired vocabulary knowledge. However, as Chambers (2010) indicates, in working with concordance, the learner can check and confirm whether one particular use is correct or not, thereby reinforcing the learning process. The longitudinal experiment conducted by Cobb (1999), as mentioned before, also suggests that both concordance and dictionary information bring benefits in the short run, but only the concordance group retains knowledge for an extended period. From the students' point of view, DDL also outweighs dictionary learning in terms of the benefits. As Yoon and Hirvela (2004: 277) reported, "they [participants] agreed that a dictionary is useful for acquiring the

meaning of words, but a corpus is more useful for learning how and where to put words in context."

We should also acknowledge there are several obstacles on the way to implementing DDL, which is the reason DDL is still placed outside the mainstream of foreign language learning and teaching, or 'marginal practice' in Boulton's term (2017a: 483). There are several possible explanations for this situation. First of all, infrastructure or logistics is one of the biggest problems for DDL (Gilquin and Granger 2010). If it is the case of hands-on DDL, its implementation means that schools or universities have to be well equipped with computers and servers, which will increase the equipment budget (Pérez-Paredes 2005). Moreover, if corpus resources are not freely available, buying licenses also costs a large amount of money. Secondly, corpus linguistics is a relatively new field whereas the techniques of corpus are developing at a dizzy speed. Many foreign language teachers have never received any relevant training and probably are reluctant to adopt a new method like DDL (Pérez-Paredes 2005). Thus, how to make teachers "corpus literate and comfortable with mechanics of corpus analysis" (Szudarski 2018: 106) is a problem. Another issue that possibly emerges is the dilemma of the teacher's role. DDL places more emphasis on the central role of the learner during the learning process, well known as 'learner-centred' (Mukherjee 2006: 12). On the one hand, learners' autonomy reaches the maximum with DDL; teachers play a less central role and have less control over the class, on the other hand (Gilquin and Granger 2010). How to balance teacher-led against learner-led awaits more discussion. Thirdly, from learner's perspective, they need to overcome technophobia if they are to search a corpus personally. Besides, they also should ideally know some basic corpus query skills. Even with the soft version of DDL, some learners still think reading paper-based concordance is time-consuming and difficult. In this respect, Gilquin and Granger (2010) argue that DDL may only be suitable for certain learners, depending on their learning style. All these issues above are challenging the development and application of DDL. As Römer (2011: 206) pointed out, there is still "much work to be done in bridging the gap between research and practice."

The scope of this study is limited regarding the sample size and the representativeness. Only 32 subjects were involved in the experiment. All of them were from universities and with an upper-intermediate level of Spanish. Thus, the results may not apply to other Spanish learners with different proficiency levels. Future work needs

to increase the size and variety of sample. It is recommended to use power analysis (Faul *et al.* 2007) to determine the required sample size based on a fixed α, power, and effect size. More studies are needed to examine the effectiveness of DDL on other levels of language proficiency (e.g., lower level, intermediate level), especially in the SFL context. It would also be profitable to include more linguistic features in future research instead of only focusing on one or two since language learners are supposed to master a language as a whole (Boulton 2017b). Lastly, future empirical studies could extend the experiment to see whether the DDL activities are more beneficial following the continuum from teacher-led to learner-led. By making these endeavors, we have reason to believe that DDL will have a vast potential for its future application and popularization.

REFERENCES

Allan, Rachel. 2006. *Data-driven Learning and Vocabulary: Investigating the Use of Concordances with Advanced Learners of English*. Dublin: Trinity College Dublin.

Anani-Sarab, Mohammad R. and Amir Kardoust. 2014. Concordance-based data-driven learning activities and learning English phrasal verbs in EFL classrooms. *Issues in Language Teaching* 3/1: 89–112.

Asención-Delaney, Yuly, Joseph G. Collentine, Karina Collentine, Jersus Colmenares and Luke Plonsky. El potencial de la enseñanza del vocabulario basada en corpus: Optimismo con precaución. *Journal of Spanish Language Teaching* 2/2: 140–151.

Aşık, Asuman, Arzu Sarlanoglu Vural and Kadriye Dilek Akpınar. 2016. Lexical awareness and development through data driven learning: Attitudes and beliefs of EFL learners. *Journal of Education and Training Studies* 4/3: 87–96.

Barcroft, Joe. 2005. La enseñanza del vocabulario en español como segunda lengua. *Hispania* 88/3: 568–582.

Bardovi-Harlig, Kathleen, Sabrina Mossman and Yunwen Su. 2017. The effect of corpus-based instruction on pragmatic routines. *Language Learning and Technology* 21/3: 76–103.

Bernardini, Silvia. 2000. Systematising serendipity: Proposals for concordancing large corpora with language learners. In Lou Burnard and Tony McEnery eds. *Rethinking Language Pedagogy from a Corpus Perspective*. Hamburg: Peter Lang, 225–234.

Bernardini, Silvia. 2004. Corpora in the classroom: An overview and some reflections on future developments. In John Sinclair ed. *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins, 5–36.

Boulton, Alex. 2007. But where's the proof? The need for empirical evidence for data-driven learning. In Michael Edwardes ed. *Proceedings of the BAAL Annual Conference 2007*. London: Scitsiugnil Press, 13–16.

Boulton, Alex. 2008. Esprit de corpus: Promouvoir l'exploitation de corpus en apprentissage des langues. *Texte et Corpus* 3: 37–46.

Boulton, Alex. 2009. Testing the limits of data-driven learning: Language proficiency and training. *ReCALL* 21/1: 37–54.

Boulton, Alex. 2010a. Data-driven learning: Taking the computer out of the equation. *Language Learning* 60/3: 534–572.

Boulton, Alex. 2010b. Learning outcomes from corpus consultation. In María Moreno-Jaén, Fernando Serrano Valverde and María Calzada Pérez eds. *Exploring New Paths in Language Pedagogy: Lexis and Corpus-Based Language Teaching*. London: Equinox, 129–144.

Boulton, Alex. 2010c. Data-driven learning: On paper, in practice. In Tony Harris and María Moreno-Jaén eds. *Corpus Linguistics in Language Teaching*. Bern: Peter Lang, 17–52.

Boulton, Alex. 2017a. Research timeline: Corpora in language teaching and learning. *Language Teaching* 50/4: 483–506.

Boulton, Alex. 2017b. Data-driven Learning and Language Pedagogy. In Steven L. Thorne and Stephen May eds. *Language, Education and Technology*. New York: Springer, 181–192.

Boulton, Alex and Tom Cobb. 2017. Corpus use in language learning: A meta-analysis. *Language Learning* 67/2: 1–46.

Brezina, Vaclav. 2018. *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press.

Calderón-Campos, Miguel. 1994. Sobre la elaboración de diccionarios monolingües de producción: Las definiciones, los ejemplos y las colocaciones léxicas. In Peter Jan Slagter ed. *Aproximaciones a Cuestiones de Adquisición y Aprendizaje del Español como Lengua Extranjera o Lengua Segunda*. Amsterdam: Rodopi, 105–119.

Chambers, Angela and Íde O'Sullivan. 2004. Corpus consultation and advanced learners' writing skills in French. *ReCALL* 16/1: 158–172.

Chambers, Angela. 2005. Integrating corpus consultation in language studies. *Language Learning and Technology* 9/2: 111–125.

Chambers, Angela. 2007. Popularising corpus consultation by language learners and teachers. In Encarnación Hidalgo, Luis Quereda and Juan Santana eds. *Corpora in the Foreign Language Classroom*. Amsterdam: Rodopi, 3–16.

Chambers, Angela. 2010. What is data-driven learning? In Anne O'Keeffe and Michael McCarthy eds. *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 345–358.

Chan, Tun-pei and Hsien-Chin Liou. 2005. Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning* 18/3: 231–251.

Chang, Ji-Yeon. 2014. The use of general and specialized corpora as reference sources for academic English writing: A case study. *ReCALL* 26/2: 243–259.

Chang, Jung. 2001. Chinese speakers. In Michael Swan and Bernard Smith eds. *Learner English: A Teacher's Guide to Interference and other Problems*. Cambridge: Cambridge University Press, 310–324.

Charles, Maggie. 2012. Proper vocabulary and juicy collocations: EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes* 31/2: 93–102.

Cheng, Winnie, Martin Warren and Xun-feng, Xu. 2003. The language learner as language researcher: Putting corpus linguistics on the timetable. *System* 31/2: 173–186.

Cobb, Tom. 1997. Is there any measurable learning from hands-on concordancing? *System* 25/3: 301–315.

Cobb, Tom. 1999. Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning* 12/4: 345–360.

Cotos, Elena, Stephanie Link and Sarah Huffman. 2017. Effects of DDL technology on genre learning. *Language Learning and Technology* 21/3: 104–130.

Crosthwaite, Peter. 2017. Retesting the limits of data-driven learning: Feedback and error correction. *Computer Assisted Language Learning* 30/6: 447–473.

Daskalovska, Nina. 2015. Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning* 28/2: 130–144.

Faul, Franz, Edgar Erdfelder, Albert-Georg Lang and Axel Buchner. 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39/2: 175–191.

Flowerdew, John. 1996. Concordancing in language learning. In Martha C. Pennington ed. *The Power of CALL*. Houston: Athelstan, 97–113.

Frankenberg-Garcia, Ana. 2014. The use of corpus examples for language comprehension and production. *ReCALL* 26/2: 128–146.

Gabrielatos, Costas. 2005. Corpora and language teaching: Just a fling or wedding bells? *The Electronic Journal for Teaching English* 8/4: 1–32.

Gaskell, Delian and Thomas Cobb. 2004. Can learners use concordance feedback for writing errors? *System* 32/3: 301–319.

Gavioli, Laura. 2001. The learner as researcher: Introducing corpus concordancing in the classroom. In Guy Aston ed. *Learning with Corpora*. Bologna: Athelstan, 108–137.

Geluso, Joe and Atsumi Yamaguchi. 2014. Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL* 26/2: 225–242.

Gilmore, Alex. 2009. Using online corpora to develop students' writing skills. *ELT Journal* 63/4: 363–372.

Gilquin, Gaëtanelle and Sylviane Granger. 2010. How can data-driven learning be used in language teaching? In Anne O'Keeffe and Michael McCarthy eds. *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 359–370.

Guan, Xiaowei. 2013. A study on the application of data-driven learning in vocabulary teaching and learning in China's EFL class. *Journal of Language Teaching and Research* 4/1: 105–112.

Hadley, Gregory and Maggie Charles. 2017. Enhancing extensive reading with data-driven learning. *Language Learning and Technology* 21/3: 131–152.

Huang, Zeping. 2014. The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL* 26/2: 163–183.

Jiao, Binkai. 2012. An empirical study on corpus-driven English vocabulary learning in China. *English Language Teaching* 5/4: 131–137.

Jiménez-Calderón, Francisco and Ana Sánchez-Rufat. 2017. Posibilidades de aplicación de un enfoque léxico a la enseñanza comunicativa del español. In Guadalupe Nieto Caballero ed. *Nuevas Aportaciones al Estudio de la Enseñanza y Aprendizaje de Lenguas*. Cáceres: Universidad de Extremadura, 11–23.

Johns, Tim. 1991. Should you be persuaded: Two examples of data-driven learning materials. *ELR Journal* 4: 1–16.

Johns, Tim, Lee Hsingchin and Wang Lixun. 2008. Integrating corpus-based CALL programs in teaching English through children's literature. *Computer Assisted Language Learning* 21/5: 483–506.

Karras, Jacob N. 2016. The effects of data-driven learning upon vocabulary acquisition for secondary international school students in Vietnam. *ReCALL* 28/2: 166–186.

Kennedy, Claire and Tiziana Miceli. 2010. Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning and Technology* 14/1: 28–44.

Koosha, Mansour and Ali A. Jafarpour. 2006. Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL Journal* 8/4: 192–209.

Lavrakas, Paul J. 2008. *Encyclopedia of Survey Research Methods*. California: SAGE Publications.

Lee, Hansol, Mark Warschauer and Jang H. Lee. 2019. The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics* 40/5: 721–753.

Lewis, Michael. 1993. *The Lexical Approach: The State of ELT and a Way Forward*. Boston: Heinle.

Li, Shuangling. 2017. Using corpora to develop learners' collocational competence. *Language Learning and Technology* 21/3: 153–171.

McEnery, Tony and Andrew Wilson. 1997. Teaching and language corpora. *ReCALL* 9/1: 5–14.

Mizumoto, Atsushi and Kiyomi Chujo. 2015. A meta-analysis of data-driven learning approach in the Japanese EFL classroom. *English Corpus Studies* 22: 1–18.

Moliner, María. 2007. *Diccionario del Uso del Español*. Gredos: Madrid.

Moon, Soyeon and Sun-Young Oh. 2017. Unlearning overgenerated *be* through data-driven learning in the secondary EFL classroom. *ReCALL* 30/1: 48–67.

Mukherjee, Joybrato. 2006. Corpus linguistics and language pedagogy: The state of the art – and beyond. In Sabine Braun, Kurt Kohn and Joybrato Mukherjee eds. *Corpus Technology and Language Pedagogy*. Bern: Peter Lang, 5–24.

Nation, Paul and Paul Meara. 2010. Vocabulary. In Norbert Schmitt ed. *An Introduction to Applied Linguistics*. London: Hodder Education, 34–51.

Nation, Paul. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

O'Keeffe, Anne, Michael McCarthy and Ronald Carter. 2007. *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.

O'Sullivan, Íde and Angela Chambers. 2006. Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing* 15/1: 49–68.

Pérez-Paredes, Pascual. 2005. Data-driven learning y el aprendizaje de idiomas. *Greta: Revista para Profesores de Inglés* 13/1–2: 5–10.

Pérez-Paredes, Pascual. 2010. Appropriation and integration issues in corpus methods and mainstream language education. In Tony Harris and María Moreno-Jaén eds. *Corpus Linguistics in Language Teaching*. Bern: Peter Lang, 53–73.

Plonsky, Luke, and Frederick L. Oswald. 2014. How big is 'big'? Interpreting effect sizes in L2 research. *Language Learning* 64/4: 878–912.

Römer, Ute. 2011. Corpus research applications in second language teaching. *Annual Review of Applied Linguistics* 31: 205–225.

Royal Spanish Academy. Online. *Diccionario de la Lengua Española* (22. ed.). At https://dle.rae.es/?w=diccionario. Accessed on 25/09/2019

Royal Spanish Academy. Online. Banco de Datos CREA: *Corpus de Referencia del Español Actual*. At http://www.rae.es. Accessed on 25/09/2019

Schmitt, Norbert. 2000. *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

Smart, Jonathan. 2014. The role of guided induction in paper-based data-driven learning. *ReCALL* 26/2: 184–201.

Soruç, Adem and Bilal Tekin. 2017. Vocabulary learning through data-driven learning in an English as a second language setting. *Educational Sciences: Theory and Practice* 17/6: 1811–1832.

Sripicharn, Passapong. 2003. Evaluating classroom concordancing: The use of concordance-based materials by a group of Thai students. *Thammasat Review* 1: 203–236.

Sripicharn, Passapong. 2010. How can we prepare learners for using language corpora? In Anne O'Keeffe and Michael McCarthy eds. *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 371–384.

Stevens, Vance. 1991. Concordance-based vocabulary exercises: A viable alternative to gap-filling. *English Language Research Journal* 4: 47–61.

Szudarski, Paweł. 2018. *Corpus Linguistics for Vocabulary: A Guide for Research*. London: Routledge.

Tekin, Bilal and Adem Soruç. 2016. Using corpus-assisted learning activities to assist vocabulary development in English. *The Turkish Online Journal of Educational Technology* 1270–1283.

Thompson, Paul. 2006. Assessing the contribution of corpora to EAP practice. In Zoe Kantaridou, Iris Papadopoulou and Ifigenia Mahili eds. *Motivation in Learning Language for Specific and Academic Purposes*. Macedonia: University of Macedonia.

Thurstun, Jennifer and Christopher N. Candlin. 1998. Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes* 17/3: 267–280.

Vyatkina, Nina. 2016. Data-driven learning for beginners: The case of German verb-preposition collocations. *ReCALL* 28/2: 207–226.

Wesche, Marjorie and T. Sima Paribakht. 1996. Assessing Second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review* 53/1: 13–40.

Yılmaz Enes and Adem Soruç. 2015. The use of concordance for teaching vocabulary: A data-driven learning approach. *Procedia-Social and Behavioral Sciences* 191: 2626–2630.

Yoon Hyunsook and Alan Hirvela. 2004. ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing* 13/4: 257–283.

Corresponding author
Gang Yao
University of Murcia
Department of English Philology
30001 Calle Santo Cristo 1
Spain
e-mail: gang.yao@um.es