Corpus-based Instruction in the EFL Classroom: The Case of Data-driven Learning: 'Spreading the Word'

Mahfoudh BESSIDHOUM*
University of Oran -2-" Mohamed Ben Ahmed", Algeria, mbessidhoum@gmail.com
Ghania OUAHMICHE
University of Oran -2-" Mohamed Ben Ahmed", Algeria, ghaniaouahmiche@gmail.com

Received 20/12/2020 Accepted 02/04/2021 Published 27/09/2021

Abstract:

In line with the ubiquity of technological tools in education, corpus-based approaches and tools are making a paradigm shift in L2 instruction. Data- driven Learning is one approach that is increasingly gaining ground in the L2 classroom in many countries; however, in Algeria, it is nearly non-existent. Therefore this article aligns with efforts to 'spread the word' about corpus-based approaches to Second/Foreign language learning and teaching, particularly Data-driven Learning. The aim is to provide overview on the nature of this approach for Algerian EFL learners, teachers, program administrators as well as syllabus designers, in order to explore potential possibilities for its incorporation into the existing EFL curricula at the tertiary level.

Keywords: Corpus-based instruction, Corpus, Corpus work, Data-driven Learning, Spreading the word.

1.Introduction

Technology is transforming our lives by the day. The ubiquity of technical devices, such as personal computers, smart phones, tablets and other modern gadgetry is remarkable in both developed and developing countries. Likewise, in education, technology is omnipresent in many educational contexts around the world. In the realm of Second/Foreign Language learning and teaching, recent developments in Computer-assisted Language Learning (CALL) are transforming views about how languages are best taught and learned. As part of Computer-assisted Language Learning, "Corpus-based approaches to language learning and teaching are becoming increasingly common worldwide" (Bednarek, Crosthwaite, & Garcia, 2020, p.105). The high level of interest in corpus-based instruction in language learning and teaching is reflected in the abundant number of contributions in academic journals and conferences dedicated to the area. Journals, like CALICO, ReCALL, CALL, and research conferences, such as EUROCALL, as well as, Teaching and Language Corpora conference (TaLC) disseminate a huge amount of scholarship and debate in the field.

* Corresponding Author

One corpus-based approach in particular, called, Data-driven Learning (henceforth DDL) is making a huge impact in many educational settings, mainly in Europe, Asia and the USA. Since its advent about 4 decades ago, DDL has generated a great deal of research and scholarship. A relatively great number of seminal studies have been carried out so far (e.g. Johns,1991; Todd, 2001; Sun, 2003; O'Sullvian & Chambers, 2006; Gateskell & Cobb,2004; Yoon & Hirevela, 2004; Boulton:2010 a; Yoon & Jo, 2014, Chang, 2014, Charles,2014, Tono , Satake & Miura, 2014; Frankenberg & Garcia,2014; Croswaithe,2017, Mao & Croswaithe, 2019; Bridle,2019; Dolgova & Miueller,2019; Corino & Onesti,2019). Findings from most studies, demonstrate that generally DDL is a promising approach to L2 pedagogy.

Compared to other traditional approaches to language learning and instruction, DDL is an approach that gives learners opportunities to discover how language works through access to authentic language data stored electronically. DDL helps learners acquire linguistic knowledge that cannot be provided by other traditional teaching and learning tools and materials, such as, dictionaries, grammars and textbooks. The overall aim of this approach is to empower learners and make them more autonomous outside the classroom.

The recent surge in both research and application of DDL led researchers to talk about a 'corpus revolution' in L2 pedagogy (See Hyland & Wong, 2013; Boulton, 2017; Chambers, 2019). Godwin-Jones (2017) contends that this is mainly due to the fact that DDL aligns with contemporary trends and philosophies in Second Language Acquisition and also because of the availability of sophisticated corpus materials.

However, despite the incorporation of DDL materials into many educational contexts and the availability of technological resources in many developing countries, DDL remains a complete 'unchartered territory' in some educational contexts, like Algeria. Accordingly, calls have been made for outreach efforts; in this respect, Römer (2009, p. 84) expressed the need to "spread the word" to places where DDL is unheard of. Thus, this paper is meant to dovetail with such efforts, as it seeks to outline the DDL approach in order to explore potential possibilities for its integration into current higher education EFL curricula in Algeria.

This paper will start with defining DDL and exploring its roots and theoretical foundations. Then, discussion will move on to its pedagogical applications in the classroom, and the skills and language areas it is supposed to develop. Subsequently, its advantages and limitations will be discussed. Finally, the state of the art of the approach will be explored in the light of findings from empirical research.

2. What is DDL?

Data-driven Learning is the use of corpora for Second/foreign language learning. Corpora (singular corpus) are large collections of natural and authentic texts organised and stored electronically. The basic principle of DDL is that instead of relying on teachers or reference materials, such as dictionaries, grammars, or translation websites, learners access the language data through corpora "with varying degrees of guidance" (Boulton & Cobb, 2017,p. 349) from teachers, in order to find answers for their language questions and satisfy their language needs In that sense, DDL attempts to "cut out the middleman as far as possible" (Johns, 1991, p. 30); learners behave as researchers, and everyone is "a Sherlock Holmes" (Johns, 1997, p.101). Since its advent in the 1980's, several terms have been put forward to describe it. Some of the terms used in the literature include, 'classroom concordancing' (Johns, 1991); 'data-driven learning' (Johns & King, 1991; Boulton, 2016) 'corpus-based learning', 'learner concordancing', 'corpus-consultation' (Boulton, 2016), 'corpus-

aided language learning' (Huang, 2011), and recently in a more figurative terminology "A 24/7 native speaker assistant" (Meunier, 2019, p.430). Likewise, many definitions have been proposed for DDL, and a watertight definition is non-existent (Boulton, 2011). For our purposes, a broad definition of DDL that includes any use of authentic data for pedagogical purposes is accepted. Accordingly, DDL is defined as "any use of overt corpus data for foreign or second language learning or teaching" (Boulton, 2010 b, p.130). As will be shown later on in this paper, corpora can be used in a variety of ways, including, but not limited to, checking how words are used in context, and which words go together (e.g. collocations, chunks), as well as correcting errors, etc .

In attempt to capture the essence of DDL, Smart (2014,p.186) proposes that this approach has two main defining features: 1- real language data are used as sources of language learning materials or reference resources; and 2- learning activities are student-centred and focus on language discovery. A couple of implications concurring with current developments in L2 language pedagogy can be drawn from those two features: authenticity and autonomy. It is believed that corpora, being mostly native speaker textual data, provide more accurate and more practical descriptions of the language than any other ready-made learning and reference materials, such as grammar books, dictionaries, or textbooks (Johns, 1991; Flowerdew, 2009). Also, by placing students at the centre of the learning process, DDL stimulates them into taking charge of their own learning; a feature nearly absent in traditional methodology .

The student-centred nature of DDL might seem radical in educational contexts that prize the distinction between teachers' and students' roles, like Algeria; yet, DDL does not eliminate the need for instruction. But, the overreliance on formal teaching may lead to learners' excessive dependence on the teacher (Boulton, 2017), thereby depriving them from discovery-based knowledge, which is processed more deeply (Nunan, 2001) than transmitted knowledge. In DDL, corpus-based activities are normally blended with other course materials and resources, and the teacher is a facilitator of corpus consultation; he/she is "a learning expert rather than a language expert" (Bernardini, 2004, p. 28) giving learners opportunities to share responsibility for the learning process.

2.1. What are its roots?

A closer look at its history reveals that DDL stands at the intersection of Computer-assisted Language Learning (CALL) and Corpus Linguistics (CL). It is inextricably related to CALL in that it is a computer-based language learning and teaching approach; however, what sets DDL apart from the rest of CALL approaches is its use of electronic corpora. It is also solidly grounded in corpus linguistics, as it involves using its "tools and techniques for pedagogical purposes" (Gliquin & Granger, 2010, p.01). In the field of corpus linguistics, corpora are used by researchers and professionals for the purposes of language description and pedagogy; accordingly, linguists produce grammar guides, modern dictionaries, course books, testing designs and other materials. On the pedagogical front, attempts by the linguist Tim Johns - considered as the founding father of DDL- and his colleagues (e.g. Philip King), in the late 1980's and early 1990's to bring corpora to the L2 classroom gave birth to DDL, although earlier efforts to test their potential benefits in language education were made by other researchers, such as Peter Roe in 1969 (McEnery &Wilson, 1997) and later McKay in1980, as well as Ahmad, Corbett & Rogers in 1985.

2.2. What About its Theoretical Underpinnings?

Theoretically, DDL embraces contemporary philosophies and practices in Second Language Acquisition. DDL's popularity both in research and application "corresponds perfectly with a number of contemporary paradigms in research and practice in education in general, and in language teaching in particular" (Chambers, 2019, p. 05). It is based on the Noticing Hypothesis, Constructivism and Socio-cultural theories of language learning (Flowerdew, 2015). Besides, it reflects principles found in communicative language learning and usage-based theories, all of which will be explored below .

One outstanding theory supporting DDL is the Noticing Hypothesis, which claims that L2 Learning can be enhanced through both noticing and awareness (Schmidt, 2001). It is argued that Learners need opportunities to notice differences between their language and native speaker language (Swain & Lapkin, 1985). Corpus tools, such as concordancers (software tools that give access to corpora by organizing them and making them visible) provide opportunities for noticing the target language through two major teaching techniques, suggested by Sharwood Smith in 1993 and shown to be useful in language learning: "Input Flooding" and "Input Enhancement". The former pertains to frequent exposure to a wealth of the relevant target language structures (examples) provided by corpus tools that teachers want to focus on with their students (or learners themselves want to know how they work), and which can't be furnished by other learning materials, such as dictionaries and grammar reference guides. This exposure leads to awareness of language patterns as well as vocabulary development (Gabrielatos, 2005, p. 10). The latter concerns enhancing the visibility of the searched structures by highlighting them to learners through typographical tools, such as bolding and colouring (see figure 1 below). While it is true that exposure leads to noticing, problem-solving, as well as language awareness (Boulton, 2017), exposure alone is not sufficient for learning to take place. Corpus work involves additional components, such as cognitive processing, exploration of data through hypothesis-formulation and testing as well as drawing individual conclusions and generalizations on the basis of observed language patterns.

DDL is also informed by constructivism (e.g., John Dewey, Jean Piaget, George Kelly, Lev Vygotsky and Jerome Bruner, among others) of which the basic tenets are (inductive) discovery learning and learner autonomy. Constructivist learning theories consider learners as active participants in their learning process building knowledge through detection and exploration rather than depending solely on instructors. This kind of learning underpins DDL, as learners are supposed to be actively involved in investigating how language works through accessing corpora. The implications of constructivist learning are basically, induction (Bernaderni, 2002) and learner autonomy (Gavioli, 2009). In that sense DDL, especially in its 'hard' version, namely Direct, or Hands-on DDL endorses inductive learning in which learners ask questions about language and formulate hypotheses and test them arriving then at their own conclusions. It also encourages learners to behave as "researchers" (Johns, 1997, p. 101) taking responsibility for their own learning and dealing with their language problems independently. Notwithstanding, constructivism was not embraced by everyone. Some researchers (e.g. Kirschner, Sweller, & Clark, 2006) voiced objections to its heavy cognitive load on learners. Besides, traditionally-oriented teachers were concerned about their roles in the classroom as the main providers of knowledge, especially in direct DDL where the computer plays a major role, which in turn can be interpreted a "loss of expertise" by teachers (Hunston, 2002, p. 171).

Concerns about constructivism led socio-cultural theories of learning to call for the adoption of 'scaffolding' (Lev Vygotsky) to aid learners through the learning process. The Socio-cultural perspective holds that knowledge construction takes place through both teacher-student and student-peer collaboration and negotiation (Vygotsky, 1978), implying that without proper guidance and support, learners are less likely to successfully manipulate the target language and internalise information. This view is consistent with the less autonomous version of DDL, in which teachers lead learners through DDL activities using paper-based materials (Indirect DDL), as will be further explained in the next sections .

In addition to the above-and-often-cited theories underpinning DDL, other less-frequently-cited, albeit important theories include contemporary Communicative Approaches (see Godwin-Jones, 2017) to language teaching and learning and Usage-based Theories (Tomasello, 2005). DDL adopts important aspects of communicative language approaches, such as the use of authentic language (e.g. native speaker texts and conversations, podcasts, newspaper articles, television clips, songs, etc.), and the improvement of learners' metalinguistic knowledge and autonomy (Godwin-Jones, 2017). Further, usage-based theories to language learning maintain that most first language (mother tongue) acquisition happens inductively through massive exposure over the speakers' lifetime (Geluso, 2013). However, in L2 learning, "because learners usually do not have similar contact time and experience with the target language to fall back on, DDL can be a useful means of study" (Geluso, 2013, p.147). Thus, "Using corpora helps learners make decisions consciously that native speakers make subconsciously" (Thomas, 2020, p. 13). In that respect, DDL is in agreement with these theories. Moreover, it is important to note, that among the above-discussed theories, the noticing hypothesis "is referred to more often than either constructivist learning or Vygotskyan sociocultural theories in DDL pedagogic initiatives" (Flowerdew, 2015, p. 31), which means that induction is bedrock of DDL, although with varying degrees as we will see .

As a final point, some of the language learning theories discussed above do overlap to a certain extent; however, it is important to remember that they have laid a solid basis for the incorporation of DDL in language learning. Having dealt with theory, we now turn to application and show DDL types and how it actually works in the classroom.

2.3. How is DDL Operationalized?

By and large, two main pedagogical corpus applications are mentioned in the DDL literature: The 'Indirect applications' and the 'Direct applications' (Römer 2006, 2008, 2011). The former is research-based, reserved for materials writers and syllabus designers who research "how corpus work can contribute to an improvement of language teaching and help to make life easier for the learner"(Römer 2006, p. 125). The latter is classroom-oriented focusing on the direct manipulation of corpora by both teachers and learners. Discussion of the research-oriented applications is beyond the scope of this paper; for a detailed account, readers are directed to Römer (2006, 2008, and 2009). For our purposes, discussion and examples of classroom applications of DDL will follow .

Overall in a classroom setting, the simplest type of DDL activity is guiding learners to determine whether what they have written or spoken is used by native speakers (Forti & Spina, 2019). One of the pioneers of DDL: Geoffrey Leech (1997) proposes that DDL activities can be presented either directly (Direct DDL) or indirectly (Indirect DDL) in the language classroom. Over time, different terms have

been used to describe the two approaches (see table 1 below). The direct uses of corpora are mostly inductive-oriented and consist of learners exploiting computer corpus data first-hand to analyse and internalise the target language. Here, learners test their intuitions about the language, asking question then noticing various naturally-occurring texts in order to answer them and form their own conclusions. The indirect uses, on the other hand, are deductive-based and involve learners using corpus data second-hand with the help of teacher- prepared paper-based examples and tasks, such as gap-fills or quizzes (for a vivid illustration about paper-based materials in DDL, see Boulton, 2010 a; Flowerdew, 1996; Gabrielatos, 2005). Using this technique, learners attempt to "check the validity of rules from their grammar or textbook" (Gilquin & Granger, 2010, p.01), or to "exemplify" them (Lenko & Boulton, 2015, p.05). It is noteworthy that although the direct more inductive approach is "the ultimate goal of DDL educators as this promotes learner autonomy and allows for lifelong learning and using corpora as a reference resource beyond the classroom." (Vyatinka & Boulton, 2017, p.06), DDL tasks can also be deductive. In fact, the two approaches are not mutually exclusive but rather lie on a continuum (Gilquin & Granger, 2010), depending on several factors, including the availability of corpus materials (computers), leaning styles and preferences (deductive and inductive), learner's language proficiency and teacher's intervention (Yoon & Jo, 2014). They may also be affected by learners' cognitive styles (Field-dependent and field- independent) (Flowerdew, 2008), background culture (Boulton, 2009), as well as their corpus literacy (Mukherjee, 2002) and computer skills. Let's now turn to a concrete example of the direct, hands-on DDL technique.

DIRECT DDL	INDIRECT DDL
Direct use (Leech, 1997) Hands-on concordancing (Cobb, 1997; Boulton, 2012) Hard version (Gabrielatos, 2005) Direct consultation (Chambers, 2007) Inductive DDL (Creswell, 2007) Learner-corpus interaction (Römer, 2008), Learner-centred corpus-browsing projects (Mukherjee, 2006).	Indirect use (Leech, 1997) Hands-off concordancing (Boulton, 2012) Soft version (Gabrielatos, 2005) Indirect consultation (Chambers, 2007) Deductive DDL (Creswell, 2007) Teacher-corpus interaction (Römer, 2008), Teacher-led concordance-based activities (Mukherjee, 2006).

Table1: Different terms used to describe direct and indirect approaches to DDL: Adapted from Yoon & Jo (2014, p. 97).

As already stated, in DDL, learners are exposed to authentic language data usually produced by native speakers in order to notice how the target language works in lieu of relying completely on reference resources (e.g. Textbooks, dictionaries, grammar guides, etc.) or teachers. For that end, both a corpus (a large collection of texts) and a software tool to search the corpus, namely a concordancer, which displays occurrences and examples of words and expression in context, are needed. In a typical direct DDL activity, learners formulate a query of a word or a phrase on the computer

corpus and examine possible occurrences (otherwise called Concordance lines) of their use in different texts and contexts. Figure 1 below shows the results of a query for the use of the idiomatic expression "hold true" using an online concordancer called SkELL (Sketch Engine for Language Learning).The results are a collection of 20 sentences demonstrating the use of the phrase in a variety of contexts, which can be of use to language learners. The display helps the learner to focus attention on the target item which is bolded and heightened in red in every example provided. Activities like the one presented here can be prepared by the teacher with learners engaging with corpora; they can be integrated with other computer-assisted tasks, or chosen by learners with or without teachers' help (Gabrielatos, 2005, p. 12). Other features in SkELL and other online concordancers/ interfaces (e.g., the British National Corpus (BNC), the Corpus of Contemporary American English (COCA) and Flexible Language Acquisition (FLAX)) include synonyms, translation, frequency counts of words or phrases across a variety of registers (fiction, newspapers, Ted Talks and academic prose, etc.), and modes (spoken, written discourse) collocations exercises, lexicogrammatical patterning, examples of academic formulaic language, such as PhD abstracts of research articles across different disciplines (Law, Humanities and Social Sciences) and useful words for academic writing , and so on. All these features can be extremely valuable to L2 learners. Data used in the activities can be derived from different corpora, which fall into different categories and serve diverse purposes (see O'Keeffe & McCarthy, 2010; Wynne, 2005; Granger, 2008). These include: spoken and written or multimodal (with audio and video contents); monolingual, bilingual or multilingual; general (containing any type of text) or specialized (comprising only texts of certain subjects, topics, registers, genres, etc.); native speaker corpora and learner corpora; annotated (having interpretations of data, for example parts of speech, etc.) and unannotated corpora (having only plain texts). These electronic materials consist of millions of words and are updated around- the- clock to the point that some even contain a Covid-19 corpus (e.g. Sketch Engine).

A great deal of research and application of both direct DDL (e.g. Bernardini, 2000; Yoon & Hirevela, 2004; Gaskell & Cobb, 2004; Sun, 2003, 2007; Yoon, 2008; Gilmore, 2009; Tono et al, 2014; Cotos, 2014; Boulton, 2016; Croswaithe, 2017; 2019), and indirect DDL (e.g. Stevens, 1991; Tribble, 1997; Boulton,2009,2010 a ; Smart 2014; Huang, 2014; Forti 2019) have been conducted in the last forty years. While the availability of technological resources in many educational settings may lead to direct DDL being researched and used more compared to its counterpart, it (Direct approach) was not taken by everyone. The approach was blamed for overwhelming lower proficiency students with vast amounts of confusing and unfamiliar language, difficulties with corpus software manipulation (as many corpora are geared towards researchers not students) and the time-consuming nature of activities, etc. In response to these observed limitations, Boulton (2010 a) and other researchers (e.g. Johns, 1997; Gaskell & Cobb, 2004) called for 'taking the computer out of the equation' Boulton (2010 a) and using paper-based edited materials to scaffold learners into the more autonomous direct DDL. Nevertheless, this requires considerable time and preparation on the part of teachers (Meunier, 2002, p. 135), and may, thus, discourage them from trying DDL at all. Moreover, the indirect approach may help alleviate the above-mentioned concerns, as teachers try to calibrate corpus data to meet students' needs and levels. However, this approach was also criticized, because teacher editions and selections of corpus data to address students' errors on different levels may result in both a limitation of discovery learning (Bernardini, 2000) and a lack of representativeness of the data (Gabrielatos, 2005). Nonetheless,

the indirect approach can at least be valuable in contexts where technological resources, time, L2 proficiency, or motivation are lacking (Yoon & Jo 2014; Boulton, 2017).

In order to compare the efficacy of both approaches, a number of studies have been made recently. Yoon and Jo (2014) examined the effects of direct and indirect DDL on L2 writing error correction. The findings showed that while indirect DDL resulted in more error-corrections, especially for lower proficiency students, overall, most students "preferred the interactive aspect of direct corpora use" (Yoon and Jo,2014, p.112), yet, both approaches were useful in helping learners deal with a number of writing problems at the revision phase. The authors (Yoon & Jo) concluded that "Use of either approach makes for a win-win situation for teachers and students wherever corpus-based instruction is employed" (Yoon & Jo, 2014, p. 112). Another study made by Ackerley (2017), investigated collocation learning using both hands-on and hands-off DDL activities with Italian EFL learners. It was found that the indirect use was more beneficial to learners than the direct one, which "can be explained by the participants' proficiency level, since the intermediate students may have found the

In order to compare the efficacy of both approaches, a number of studies have been made recently. Yoon and Jo (2014) examined the effects of direct and indirect DDL on L2 writing error correction. The findings showed that while indirect DDL resulted in more error-corrections, especially for lower proficiency students, overall, most students "preferred the interactive aspect of direct corpora use" (Yoon and Jo,2014, p.112), yet, both approaches were useful in helping learners deal with a number of writing problems at the revision phase. The authors (Yoon & Jo) concluded that "Use of either approach makes for a win-win situation for teachers and students wherever corpus-based instruction is employed" (Yoon & Jo, 2014, p. 112). Another study made by Ackerley (2017), investigated collocation learning using both hands-on and hands-off DDL activities with Italian EFL learners. It was found that the indirect use was more beneficial to learners than the direct one, which "can be explained by the participants' proficiency level, since the intermediate students may have found the

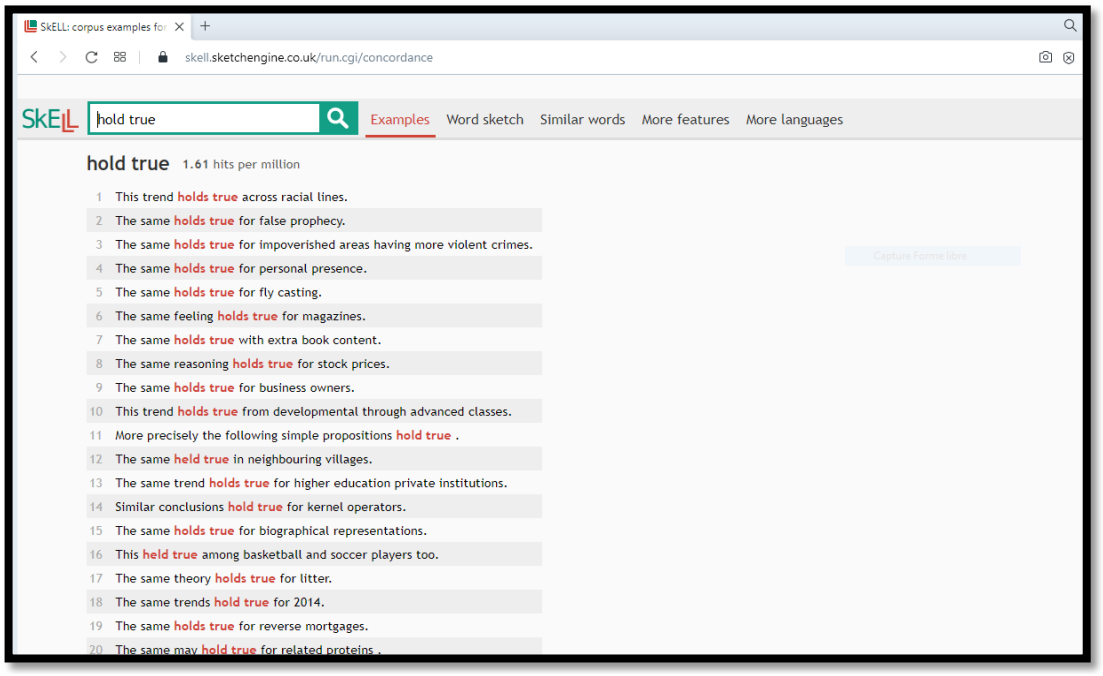

Figure1: Sample of concordance lines for "hold true" from the 'SkELL' corpus.

vast input resulting from direct corpus searches overwhelming." (Vyatkina & Boulton, 2017, p. 04). Finally, a recent narrative synthesis of studies have been completed by Vyatnika & Boulton (2017) found that hands-on DDL may be more appropriate for learning lexicogrammar, collocations, discourse moves and pragmatic routines, especially for students with high L2 proficiency. Conversely, hands-off DDL or traditional approaches may work more with more complex issues, like speech acts, and with lower L2 proficiency students. Although these studies show that L2 proficiency and type of instructional targets can play an important role in the effectiveness of the approach, they are not certainly the only factors. Other studies have indicated the impact of other variables related to logistics, leaners, as well as teachers (see, Mukherjee, 2002; Flowerdew, 2008; Boulton, 2009; Yoon & Jo 2014). Hence, the choice of either approach should carefully consider these issues.

2.4. What Skills and Features Does DDL Seek to Promote?

Whatever approach used, a reasonable question is in which areas of language learning is DDL being used. DDL has grabbed the interest of many researchers, practitioners and teachers worldwide; accordingly, "It has been applied in second and foreign language teaching contexts in many countries and at diverse educational institutions with learners at different proficiency levels" (Vyatkina & Boulton, 2017, p.01). The first DDL pedagogical initiatives focused on grammar (e.g. Johns, 1991; McKay, 1980; Ahmed et all, 1985); however, as the field grew rapidly in research and application and electronic corpora "developed in availability, speed, functionality, and visualization" (Croswaithe, 2020, p.01) , later developments expanded to cover a wide range of instructional targets and language skills. Recent research contributions and applications have centred around a number of areas where DDL can be used in the L2 classroom:1- Language learning, 2- skills development and, 3- translation training (see Lenko & Boulton, 2015), 4- pragramtics (Mizumoto & Chujo, 2015) and 5- discourse (Vyatkina & Boulton, 2017).

DDL was and is still being used for the teaching and learning of several language aspects, such as vocabulary (e.g. Shaw, 2011; Thurstun & Candlin, 1998; Frankenberg-Garcia, 2012; Geluso &Yamaguchi, 2014); collocations (e.g. Daskalovska, 2015; O'Keffe et al, 2007; Hyland, 2008; Huang, 2014; Li, 2017; Ackerley, 2017); phraseology (e.g. Römer, 2009; Adel, 2010; Geluso, 2013; Corino, 2014; Aston, 2015; Forti, 2019); grammar (e.g. Hadley, 2002; Hoey, 2009; Boulton, 2009; Cotos, 2014 Friginal, 2013; Mueller & Jacobsen, 2015); lexicogrammar (e.g. Allan, 2009; Götz, 2012). Moreover, it is used in teaching both productive and receptive skills, such as listening and speaking (e.g. Braun ,2007; Geluso & Yamaguch 2014; Kotani et al; 2016; Love, Dembry, Hardie, Brezina & McEnery, 2017); reading (e.g. Babych, 2015 ; Fuentes, 2015 ; Hadley & Charles , 2017), and writing (e.g. Todd, 2001; Kennedy & Miceli, 2001; Chambers & O'Sullivan, 2004; O'Sullivan & Chambers, 2006; Yoon & Hirevela, 2004; Sun, 2007; Yoon, 2008, 2011, 2014; Gilmore, 2009, Boulton, 2010 a, 2016 ; Charles, 2007, 2011, 2014; Tono et al; 2014; Yoon & Yo 2014; Quinn 2015, Mueller & Jackobsen, 2015; Dolgova & Mueller, 2019; Bridle, 2019; Croswaithe, 2017,2019). It is also used in translation pedagogy (e.g. Bernardini, 2002; Frankenberg-Garcia, 2014; Teresa Molés-Cases & Ulrike Oster, 2015; Marco & van Lawick, 2015; Sotelo, 2015). Last but not least, DDL is used in learning and teaching of pragmatics and discourse (e.g. Aijimer, 2020; Kübler, 2011; Bardovi-Harlig; Mossman & Su, 2017; Vyatkina, Nina, & Cunningham, 2015; Charles, 2014). It is mention worthy that although the examples above show a

considerable variety in research and application, the focus of research and practice until recently has been mainly on collocations and writing (Vyatinka & Boulton, 2017) in higher education contexts (Boulton, 2017).

2.5. What are the Potential Advantages of DDL?

The literature shows that DDL is suggested as an effective approach in promoting L2 learning. Several advantages were enumerated by researchers and practitioners (e.g. Aston, 2001; Gabrielatos, 2005; Frankenberg-Garcia, 2005; Clifton & Phillips, 2006; Römer, 2008; Flowerdew, 2008, 2015; Sonbull & Schmitt, 2010; Gilquin & Granger 2010; Boulton, 2011; Boulton & Cobb 2017). Some of these are discussed below.

One advantage of corpus work and DDL activities is that they expose learners to authentic language data (Johns, 1991; Gabrielatos, 2005; Flowerdew, 2015; Chen, 2011; Boulton, 2009; Gilquin & Granger, 2010; Clifton & Phillips, 2006; Römer, 2008), something which other ELT materials cannot afford (e.g. Dictionaries, grammars and textbooks). Corpus materials provide both accurate and practical data as they are extracted from real texts and are directly relevant to learners (when they search for them) (Johns, 1991).They also improve learners' ability to handle authentic language (Boulton, 2009) in real-life communication, which reflects current ideas espoused by communicative language approaches.

A second benefit of corpus materials and DDL is raising learners' awareness to frequency information of words, expressions and structures of the target language (Flowerdew, 2015; Boulton, 2009; Boulton & Cobb, 2017; Meunier, 2019). Concordancers provide statistical information about the occurrences of language items across registers (e.g. academic prose, fiction, newspapers, etc.) and modes (written or spoken discourse). This improves genre learning and awareness (McEnery & Wilson, 2001; Boulton, 2009), because it gives learners opportunities to compare their knowledge and intuitions about language with those of native speakers, leading ultimately to improve their understanding about how language works.

Further, corpora and DDL have an advantage over traditional instructional approaches, which is discovery learning (Gilquin & Granger, 2010; Cobb & Boulton, 2017). DDL tasks allow learners to be actively involved in the learning process (Gabrialatos, 2005), as they learn inductively how the target language operates. Rather than learning abstract rules from textbooks, learners notice patterns of the target language from numerous examples. According to Boulton & Cobb (2017) this reflects contemporary learning theory which contends that "Rules are hard, patterns are easy" (p.149). Discovery learning leads to increased language awareness (Gabrielatos, 2005; Boulton, 2009; Yoon, 2011), as self-discovered knowledge is processed more deeply (Nunan, 2001).

Another advantage is error-correction, especially in writing (Gilquin & Granger, 2010; Crosthwaite, Storch & Schweinberger, 2020). Corpus data helps students compare their writing production with native speakers or experts' texts and correct their errors, hence improve their writing. Several DDL studies have used corpus-based written-corrective feedback to guide learners to correct grammatical and/or lexical errors (particularly at the revising stage), and reported positive results (e.g. Todd, 2001; Gilmore, 2009; Tono et al 2014; Quinn, 2015; Croswaithe, 2017; Crosthwaite e al; 2020) although correction accuracy depended on the types of errors.

In addition to all the above benefits, corpus-based instruction and DDL help learners develop general learning skills and abilities, such as analytical and problem-solving skills (Vyatinka & Boulton, 2017), retention and recall (Cobb, 1999; Sonbull

& Schmitt, 2010, as cited in Cotos , Link, & Huffman 2017), as well as cognitive and metacognitive skills (Aston, 2001; O'Sullivan, 2007), such as "predicting, observing, noticing, thinking, reasoning, analysing, interpreting, reflecting, exploring, making inferences (inductively or deductively), focusing, guessing, comparing, differentiating, theorising, hypothesising, and verifying" (O'Sullivan, 2007, p. 277). DDL also proved good for developing learning strategies (Kennedy & Miceli, 2001; Sun, 2003), and for learning collocations, phraseology, vocabulary and morphology, (O'Keeffe, McCarthy & Carter, 2007; Römer, 2009; Boulton, 2009, Corino, 2014; Meunier, 2019). All the aforementioned benefits are believed to foster learner autonomy and life-long learning (Boulton, 2009, 2010; Gilquin & Granger, 2010; Lin & Lee, 2015; Yoon, 2011, as cited in Mizumoto, & Chujo, 2015).

To conclude, DDL can, therefore, present potential opportunities for both learners and teachers to improve Second/Foreign language learning and instruction. That being said, however, as any other approach or method in language learning and teaching, this approach is far from being a panacea.

2.6. What About its Limitations and Problems?

DDL has attracted a lot of interest among researchers and practitioners; nonetheless this interest has not been transformed into widespread practice. Chambers (2019, p. 05) maintains that there is a gap between research and practice in DDL despite its "excellent fit with current linguistic and educational paradigms". In a similar vein, Boulton (2019, p.05) believes that " the main consumers of research are other researchers rather than teachers or decision-makers". This reluctance about adopting DDL in the language classroom is largely due to its problems and limitations.

One the most common objections to corpus work and DDL is the reliance on induction. In DDL-based activities, especially hands-on concordancing, learners are supposed to engage with corpora asking questions, making hypotheses, confirming or disconfirming them and drawing conclusions. This discovery-based learning approach, which is based on "observation-hypothesize-experimentation" (Huang, 2017,p. 05) is different from the traditional deductive-based instruction approach of 'presentation-practice-production '; thus it can be incompatible with some students learning styles and preferences, as well as, with educational institutions' practices.

However, critics of induction do not consider the fact that DDL is not entirely induction-oriented because tasks "can be plotted on a cline of learner autonomy, ranging from teacher-led and relatively closed concordance-based activities to entirely learner-centred corpus-browsing projects" (Mukherjee, 2006,p. 12). Therefore, the more demanding inductive-based tasks can be conducive to more advanced learners, whereas the less challenging deductive ones may be more appropriate for lower proficiency learners, and "In-between totally teacher-led DDL and totally learner-led DDL, there is a whole range of activities, with various types of 'filters' exercised by the teacher" (Gilquin & Granger, 2010, p. 05). Further, tasks can even be entirely deductive when learners check corpus data and examples to uncover the validity of rules they have learned from grammar books or textbooks and illustrate them. Finally pure induction can be avoided altogether by using a proposed model called 'Guided Induction' (GI) (see Carter & McCarthy, 1995; Flowerdew, 2009). Under this approach, examples are carefully handpicked by the teacher from the corpus in order to illustrate uses of particular language items. Students then interact with the data in groups or peers, and then they are guided by the teacher to explore other examples.

After that, students are left to their own devices to elicit rules based on the data studied.

Another perceived problem is fragmented sentences presented by some concordance lines in online concordancers. These sentences can be confusing and frustrating to lower language proficiency learners, as they contain incomplete data. But researchers believe that this issue can be solved by teachers through selecting only complete sentences or completing unfinished ones and presenting them through paper-based tasks.

Furthermore, doubts have been raised about the benefits of authentic data. Widdowson (2000) claims that authentic samples of the target language provided by corpora may not be authentic because "they are taken out of a larger context and often come from texts that are of little interest or relevance to learners" (Vyatnika & Boulton, 2017, p. 02). Also Cook (1995) highlighted the significance of context, particularly in spoken language:

Whereas writing encourages the notion that language may be divorced from the circumstances of its production, and can often be understood without reference to them, speech is as often inseparable from these circumstances, and can only be apprehended in the context of the knowledge of the participants, their paralanguage and the situation. (P.42)

Nevertheless, this need for authentication of DDL materials could be addressed by choosing texts which are specific to learners 'disciplines, interests or even relying on their own texts (Charles, 2012). Meanwhile, "corpora of simplified data (e.g. of textbooks, exam papers or simplified novels) might still be used" (Boulton, 2011, p. 570) to account for difficulties in understanding samples from native speaker texts.

In addition, while corpora provide learners with huge opportunities to discover the workings of the target language through the massive amount of data they contain, this wealth of data itself is believed to be confusing and overwhelming to lower proficiency learners in particular. The corollary of this is that the search process can be both arduous and time-consuming.

In order to deal with this problem, it was suggested that learners should be introduced to smaller, more user-friendly corpora. Examples of small corpora include 'The Flax' (Flexible Language Acquisition) and 'Just The Word'. Search options in these tools are both restricted and intuitive. Through 'Jut the Word' corpus, for example, learners can work on collocations through the part-of-speech option (POS), hence concentrating only on one of type of activity. Moreover, it is recommended to transition learners to independent concordancing by starting with corpus printouts and hand-outs (Boulton, 2010 a; Gaskell & Cobb, 2004).

What is more, from the teachers' standpoint, it is always argued that preparing DDL tasks is time-consuming and "ready-made teaching materials are few and far between" (Vyatnika & Boulton, 2017, p. 02). Besides, some teachers complain about their lack of knowledge and expertise to deal with technical issues arising in corpus work. However, many free interfaces and tools are available today via the Internet, and many do not entail corpus literacy or computer skills. Also there are numerous guidelines helping teachers understand corpus-based approaches to teaching (e.g. Bennett, 2010; Flowerdew, 2012; Gilquin & Granger, 2010; Reppen, 2010; Shaw, 2011). Further, the technical issues involved in corpus manipulation should not be overestimated (Bernardini, 2001), as they could be reduced by using paper-based DDL tasks (Boulton, 2017).

Other voiced concerns were about the nature corpus materials. Many corpora and concordancing tools are thought to be specifically made for corpus linguists, not

for teachers and learners. Thus, they require a great deal of technical knowhow. This in turn could deter both teachers and learners from using them.

Recognizing this problem, many corpus linguists and researchers started lately to design tailor-made materials to cater for the needs of both learners and teachers. As results, many manageable and subject-specific interfaces are available today. Examples of such materials include the 'Oxford Online Collocations Dictionary', the 'Hong Kong Engineering Corpus', the 'Sentence Corpus of Remedial English (SCoRE) and countless others. In addition to that, training both learners and teachers on corpus work can alleviate these problems. In this respect, various training programs have been introduced so far (e.g. Carter & McCarthy, 1995; Flowerdew, 2009; Kennedy & Miceli, 2010, 2016). However, calls for training do not mean that it is a requirement or that training should be extensive. Chambers (2005) argued against training students as if preparing them to be future corpus linguists. Besides, many researchers have noted successful corpus work with students with brief training sessions or no training at all (e.g. Boulton, 2010 a , 2016; Gilmore, 2009; Mueller & Jacobsen, 2015; Tono et al; 2014).

To conclude, there seems to be a heated debate about the merits and demerits of DDL; however hitherto, much of what has been claimed remains largely anecdotal in the absence of rigorous empirical evidence. Hence, some research syntheses have been made to determine whether the approach works or not, some of which will be outlined in the next section.

2.7. What is the State of the Art of DDL?

Over the last three decades, a great body of empirical research has emerged to investigate the state of the work of DDL as an approach to L2 instruction. About 300 empirical studies have been made since the inception of DDL (Croswaithe, 2020, p.01). DDL empirical studies are defined 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" (Boulton, 2010b, p.130). Empirical contributions fall into two broad categories: emic studies and etic studies, with two research areas for each (Boulton, 2017). Emic studies investigate corpus-based approaches from an insider's perspective involving learners': a- attitudes about and reactions to using corpora through data-elicitation tools, such as questionnaires, interviews and logs, and b- their practices and behaviours when using it. Etic studies on the other hand, pertain to an outsider's (researcher's) objective evaluations of the approach, through experimental designs, scrutinizing its efficiency both as: 1- a learning utility for some specific instructional targets and skills, and 2- a reference resource, especially for writing and translation.

In order to draw conclusions from this large body of research about whether DDL is a viable approach to L2 instruction, both qualitative and quantitative syntheses and evaluations of primary studies of both emic and etic perspectives have been conducted. Qualitative evaluations generally take the form of surveys, narrative syntheses and reviews and seek to track the evolution of the field by investigating methodologies, topics and learning outcomes, etc. These include Chambers (2007), Boulton (2010), Boulton (2017a), Boulton (2017 b), Vyatnika & Boulton (2017), Cheng (2010), Römer (2011), Yoon, Choongil (2011), Flowerdew (2015), Pérez-Paredes (2019), Tribble (2015) and Luo & Zhou (2017). Quantitative syntheses on the other hand, consist mainly of meta-analyses, which are secondary quantitative evaluations and analysis of effect size in (quasi) experimental studies. These assess whether DDL actually works in terms of both effectiveness and efficiency. Hence,

"effectiveness studies look at DDL's ability to increase learners' skills or knowledge through a pre/posttest design (P/P); efficiency studies through a control/experimental group comparison (C/E) of different ways of covering the same content"(Cobb & Boulton, 2017, p. 27). DDL Meta-analyses encompass Muzomito & Chujo (2015), Cobb & Boulton (2015), Boulton & Cobb (2017), as well as Lee, Warschawer & Lee (2018). To the best of our knowledge, from 2007 to 2020, a total 16 of both categories of research synthesis (narrative surveys and meta-analyses) were executed. Findings gleaned from these will be discussed below.

In terms of attitudes, both Chambers' (2007) review of 10 influential studies along with Boulton's (2017) survey of longitudinal studies, demonstrated that overall, a variety of students held positive attitudes towards DDL activities, especially in terms of the nature of some of its features, such as authenticity and discovery learning. However, some studies, reported negative reactions, with subjects describing the tasks as hard and tedious. This raises questions about which language targets and skills is DDL most amenable to; how much guidance learners need to navigate corpora, and what types of corpora should be used. Moreover, although the results from both surveys mean that DDL can be appealing to learners, the evidence should be interpreted cautiously. Boulton (2017) warns of the 'novelty factor' on the part of researchers and 'the Hawthorne effect' on the part of students in some primary studies (e.g. Yoon & Hirevela, 2004). He observes that researchers themselves were enthusiastic and students may have thus responded positively. Also, given that most primary studies have been made by researchers in the field, it is possible that only positive findings are reported (Mizumoto & Chujo. 2015) leading to the 'file drawer problem' (where negative results are not disseminated; see Cobb & Boulton, 2017), and putting in turn the validity of conclusions into question. Accordingly, concerns like these can be addressed if more studies are conducted by teachers who have no stake in DDL.

As for learners' practices when engaged in DDL tasks, Boulton (2017) reviewed some seminal studies (e.g. Pérez-Paredes et al; 2012; Kennedy & Miceli, 2010; Charles, 2014) and found generally that the majority of students could use hands-on concordancing (Direct DDL); however, the amount of training learners should receive was still unresolved. The author concluded that the paper-based approach (Indirect DDL) could be an alternative to direct DDL when material and linguistic resources or motivation are lacking (Boulton, 2017). In another narrative review made by Vyatnika & Boulton (2017), direct DDL was found to be more conducive to students with high L2 proficiency. Even though findings from both surveys may be understood in a way that hands-on concordancing may be challenging without training, or linguistic resources and proficiency, a number of studies have been carried out where subjects received brief training sessions ranging from 05 minutes to 90 minutes (e.g. Boulton & Wilhelm, 2006; Boulton, 2010; Gilmore, 2009; Mueller & Jacobson, 2015; Tono et al; 2014) or no training at all, especially when hyperlinks are provided to direct them to correct their errors (e.g. Gatskell & Cobb, 2004; Vincent, Nesi & Quinn, 2015; Boulton 2016; Mathew, Nesi, & Vincent2018). Some of these studies have reported positive findings. For example, Boulton (2010) found that lower proficiency learners successfully detected patterns only after a five-minute training session. Also, Tono et al (2014) reported good levels of accuracy in error-correction with intermediate students after a 20-minutes introduction, although not for all types of errors.

While uncovering learners' reactions to and behaviours when performing DDL-based activities is important, it does not tell us about the viability of the

approach, because as Boulton (2020) puts it "What if they like it but benefit very little or not at all… or even if it has a negative effect on language use, or if they don't like it but do benefit? "(Boulton, personal communication, March, 17, 2020). Therefore, measurements of learning outcomes from DDL are required- that is- assessing its effects when it is used both in assisting the learning of certain language targets (as a learning aid), and checking production especially in writing and translation (as a reference resource). One of these is Boulton's (2010 b) survey of 27 L2 studies, which reported promising results for most of them, although the results lacked statistical significance on some research questions. Nevertheless, Boulton (2010b) reports that generally studies found that:

DDL can be usefully employed for learners of many different language backgrounds and in different situations when appropriately adapted, whether using sophisticated equipment or the simplest of materials, in pursuing individual language interests or in tightly controlled activities, for high and low levels alike. (p. 17)

Further, in his qualitative narrative synthesis, Cheng's (2010) concluded that "DDL has been found to be a useful language learning methodology, and that there is evidence that learners can indeed benefit from being both language learners and language researchers." (p. 320). Taken together, the results from both reviews (Boulton's and Cheng's), paint a positive picture about learning outcomes from corpus consultation. However, because narrative reviews like these may report biased evidence, as findings are based mostly on self-reportage, more reliable evidence from the more objective and rigorous meta-analyses was called for by researchers (e.g. Muzomito and Chujo, 2015).

In response to calls for more objective measurements of the effectiveness of DDL, the authors of the most recent general meta-analysis, Cobb and Boulton (2017), evaluated quantitatively the effect size of a large number of empirical studies. The authors reported great effect sizes and claimed that they " reach the somewhat surprising and possibly encouraging conclusion that DDL works pretty well in almost any context where it has been extensively tried" (p.386), and that it "is better than many traditional teaching methods for various instructional targets." (Vyatkina & Boulton, 2017, p.01). That being said, however, studies reviewed were largely confined to English, and did not address other issues and areas, like speaking skills, long-term effects and applications of DDL at the professional level (Cobb & Boulton, 2017). Pulled together, the meta-analyses carried out by Chujo and Muzomito (2015) (14 studies), Cobb & Boulton (2015) (116 studies), Boulton & Cobb (2017) (64 studies) and Lee et al (2019) (29 studies), demonstrate that " DDL is most effective in the following cases: with vocabulary, in within-groups designs, in a foreign language context, at higher proficiency levels, with mixed paper/computer-based modalities, for in-depth knowledge of vocabulary and with more than 10 sessions" (Forti, 2019.p 365). Some caveats are worth-mentioning about these meta-analyses though. For example, Chujo and Muzomito's (2015) was restricted only to Japanese primary studies, and Lee et al (2018) dealt exclusively with DDL studies for vocabulary, which does not mean necessarily that it does not work for other language areas. This leaves us with two general and comprehensive meta-analyses made by Boulton & Cobb (2015, 2017), which are in fact the most-widely cited in the literature. Therefore, it can be assumed that at least evidence from these two meta-analyses should be considered seriously.

Finally, to answer the question raised in this section, "empirical studies of DDL have generally provided positive results" (Chambers, 2019, p. 04). From the above research synthesis, it seems reasonable to conclude that despite some

reservations about the data reported in qualitative narrative reviews, quantitative scientific evidence proves that DDL is by and large a viable approach to L2 language instruction.

3. Conclusion

This paper has briefly overviewed the area of corpus-based L 2 instruction, particularly the DDL approach. It may seem clear to both learners and teachers coming from knowledge-transmission and deductive-oriented educational contexts, like Algeria, that corpus-based methodology and techniques could be extreme and challenging. However, it is both note-worthy and unquestionable that improved learning does not rely solely on instruction in its pure traditional sense. In line with the modern day 'zeitgeist', current trends in L2 education endorse learner autonomy and life-long learning, both of which are basic tenets of corpus-based approaches and techniques to L2 instruction. While it may be hard to draw firm conclusions from the debates raised about the advantages and limitations of corpus-based approaches, Ivor Timmis (2015) forcefully reminds us that "Both the empirical and theoretical arguments for DDL suggest that DDL should be part of the repertoire of teachers and materials writers" (p.141). Nonetheless, it must pointed out that these approaches and techniques are not a replacement of older methods, materials and activities but a supplement to them (Boulton, 2009; Johns, 1991; Meunier, 2002). Therefore it is high time for teachers, practitioners, program administrators and syllabus designers alike to critically evaluate potential possibilities of integrating corpus-based methodologies and techniques, such as DDL with existing approaches to L2 learning and instruction in order to improve present-day practices, respond to the demands of current times and emulate other successful educational systems around the globe. In closing, perhaps it is proper to conclude with James Wilson's statement who observes that "Surely it is more appropriate to consider the benefits or advantages of combining tradition with technology and introducing corpora to traditional teaching practices and/or to other innovative modes of delivery" (Wilson, 2013, p. 63).

References

Ackerley, K (2017). Effects of corpus-based instruction on phraseology in learner *English. Language Learning & Technology, 21*(3), 195–216. http://llt.msu.edu/issues/october2017/ackerley.pdf

Aijmer, K. (2020).Contrastive Pragmatics and Corpora. *Contrastive Pragmatics* (1), 28–57. Retrieved from https://brill.com/view/journals/jocp/1/1/article-p28_3.xml?language=en&body=citedBy-35532

Aston, G. (2015). Learning phraseology from speech corpora. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning (1st ed., pp. 63–84).*Amsterdam: John Benjamins.

Babych, S. (2015). Textual cohesion patterns for developing reading skills. In Leńko-Szymańska & Boulton A (Eds.*), Multiple Affordances of Language Corpora for Data-driven Learning(1st ed., 155-176).*Amsterdam: John Benjamins.

Bardovi-Harlig, K; Mossman; Su & Y. (2017). The effect of corpus-based instruction on pragmatic routines. *Language Learning & Technology, 21* (3), 76–10. Retrieved from https://eric.ed.gov/?id=EJ1156121

Bednarek, M; Crosthwaite, P; & Garcia, A. (2020). Corpus linguistics and education in Australia. *Australian Review of Applied Linguistics, 43*(2), 105-116. Retrieved from https://doi.org/10.1075/aral.00029.edi

Bernardini, S. (2004). Corpora in the classroom. In J. M. Sinclair (Ed.), *How to Use Corpora in Language Teaching. (1st ed., 15–36).* Amsterdam; Philadelphia, PA: John Benjamins.

Boulton, A. (2009). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics, 35*(1), 81-106.

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

Boulton, A. (2010b). 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,* (*1st ed., 129-144).* Equinox.

Boulton, A. (2011). Data-driven learning: the perpetual enigma. In Goźdź-Roszkowski S (Ed.), (1st ed., 563-580). *Explorations across languages and corpora.* Peter Lang, Frankfurt.

Boulton, A. (2017). Data-driven learning and language pedagogy. In S. Thorne & S. May (Eds.), Language, Education and Technology: Encyclopedia of Language and Education. New York: Springer. DOI 10.1007/978-3-319-02328-1_15-1. 182-191.

Boulton, A. (2019). Data-driven learning for younger learners: Obstacles and optimism. Foreward to P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners.* Abingdon: Routledge.

Boulton, A & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348-393. https://www.researchgate.net/publication/313729217

Bridle M, (2019). Learner use of a corpus as a reference tool in error correction: Factors influencing consultation and success. *Journal of English for Academic Purposes.37*, 52-69. Retrieved from https://doi.org/10.1016/j.jeap.2018.11.003

Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching.52(4), 460-475.* https://www.cambridge.org/core

Cheng, W. (2010). What can a corpus tell us about language teaching. In M. O'Keefe, Anne & Mc Carthy (Eds.), *The Routledge Handbook of Corpus Linguistics, (1st ed., 319-332).* Abingdon: Routledge.

Cook, G. (1995). Theoretical issues: transcribing the untranscribable. In, G. Leech, G. Myers & J. Thomas (Eds.), *Spoken English on Computer.Transcription, markup, and application (1st ed., 35-53).* Harlow, England: Longman.

Cotos, E, Link, S, Huffman, S. (2017). Effects of DDL technology on genre learning. *Language Learning & Technology, 21* (3), 104–130. https://scholarspace.manoa.hawaii.edu/bitstream/10125/44623/1/21_03_cotoslinkhuffman.pdf

Crosthwaite, P. (2020). Taking DDL online: Designing, implementing and evaluating a SPOC on data-driven learning for tertiary L2 writing. *Australian Review of Applied Linguistics, 43*(2), 169-195. https://doi.org/10.1075/aral.00031.cro.

Crosthwaite, P; Storch, N; & Schweinberger M. (2020). Less is more? The impact of written corrective feedback on corpus-assisted L2 error resolution. *Journal of Second Language Writing 49,* https://doi.org/10.1016/j.jslw.2020.100729

Dolgova, N., & Mueller, C. (2019). How useful are corpus tools for error correction? Insights from learner data. *Journal of English for Academic Purposes,(39),* 97–108. https://www.researchgate.net/publication/332041278_How_useful_are_corpus_tools_for_error_correction_Insights_from_learner_data

Flowerdew, L. (2015). Data-driven learning and language learning theories. *Multiple Affordances Language Corpora for Data-Driven Learning*,(1st ed., 15–36.).Amsterdam: John Benjamins.

Forti, L., & Spina, S. (2019). Corpora for linguists vs. corpora for learners: Bridging the gap in Italian L2 learning and teaching. *EL.LE: Educazione Linguistica. Language Education, 8*(2), 349– 362. https://edizionicafoscari. unive.it/en/ edizioni4/ riviste/ elle/2019/2/

Gabrielatos, C. (2005). Corpora and language teaching: just a fling or wedding bells? *The Electronic Journal for English as a Second Language 8*(4), 1-35. http://www-writing.berkeley.edu/TESL-EJ/ej32/a1.html

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://www.tandfonline.com/doi/abs/10.1080/09588221.2011.639786

Gilquin, G.; Granger, S. (2010). How can Data-Driven Learning be used in Language Teaching? O'Keeffe, In A.; McCarthy, M. (Eds.), *The Routledge Handbook of Corpus Linguistics, (1st ed., 319-332).* Routledge: London.

Godwin-Jones, R. (2017). Data-informed language learning. *Language Learning & Technology, 21*(3), 9–27. http://llt.msu.edu/issues/october2017/emerging.pdf

Huang, L.-S. (2017). Has corpus-based instruction reached a tipping point? Practical applications and pointers for teachers. *TESOL Journal, 8*(2), 295–313. https:// doi.org/10.1002/tesj.271

Hunston, S. (2002). *Corpora in Applied Linguistics.* Cambridge: Cambridge University Press.

Johns T. (1991). From printout to hand -out: grammar and vocabulary teaching in the context of data-driven learning. In Johns T. & King P. (dir.) Classroom concordancing. *English language research journal ,(4),* 27-45.

Johns T. (1997). Contexts: the background, development and trialling of a concordance-based CALL program. In Wichmann A., Fligelstone S., McEnery T. & Knowles G. (dir.) Teaching and language corpora, Harlow: Addison Wesley Longman. 100-115.

Lee, H; Warschawer, M; & Lee J.H. (2018). The Effects of Corpus Use on Second Language Vocabulary Learning: A Multilevel Meta-analysis. *Applied Linguistics 40*(5), 721–753. https://academic.oup.com/applij/advance-article-abstract/doi/10.1093/applin/amy012/4953772

Leńko-Szymańska, A & Boulton, A. (2015). *Multiple Affordances of Language Corpora for Data-driven Learning.* Amsterdam: John Benjamins.

Meunier, F. (2002). The pedagogic value of native and learner corpora in EFL grammar teaching. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching. (1st ed., 119–141).* Amsterdam/Philadelphia: John Benjamins.

Meunier, F (2019). Data-Driven Learning: From Classroom Scaffolding to Sustainable Practices. *Educazione Linguistica. Language Education, 8*(2), 423-434. DOI 10.30687/ELLE/2280-6792/2019/02/010

Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese EFL classroom. *English Corpus Studies, (22),* 1–18. https://mizumot.com/files/ecs2015.pdf

Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art – and beyond. In S, Braun; K, Kohn; & J, Mukherjee. (Eds.), *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods (1st ed., 5-24)* Frankfurt: Peter Lang.

Nunan, D. (2001). *Second language teaching and learning.* Boston: Heinle & Heinle Publishers.

O'Sullivan, I. (2007). Enhancing a process-oriented approach to literacy and language learning: the role of corpus consultation literacy. *ReCALL 19* (3), 269-286.

Römer, U. (2006). Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik, 54*(2), 121–134.

Römer, U. (2009). Corpus research and practice: What help do teachers need and what can we offer? In K. Aijmer (Ed.) *Corpora and language teaching, (1st ed., 83-101)* Amsterdam. The Netherlands: John Benjamins.

Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL, 26*(02), 184-201.

Thomas, J. E. (2020). *Two web-based tools for learning language from language.* Manuscript submitted for publication. http://www.versatile.pub/versatext-info.html

Timmis, I. (2015). *Corpus linguistics for ELT.* Abingdon: Routledge.

Vyatkina, N., & Boulton, A. (2017). Corpora in language learning and teaching. *Language Learning & Technology, 21*(3), 1–8. http://llt.msu.edu/issues/october2017/commentary.pdf

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Widdowson, H. (2000). On the limitations of linguistics applied. *Applied Linguistics,* (21) 1, 3-25.

Wilson, J. (2013). *Technology, Pedagogy and Promotion. How Can We Make the Most of Corpora and Data-Driven Learning (DDL) in Language Learning and Teaching?* The Higher Education Academy. https://www.heacademy.ac.uk/system/files/Corpus_Technology_pedagogy_promotion2.pdf

Yoon, H & Jo, 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. https://scholarspace.manoa.hawaii.edu/bitstream/10125/44356/1/18_01_yoonjo.pdf