

CORPUS LINGUISTICS: POTENTIALS AND LIMITATIONS

Kamariah Yunus

Universiti Sultan Zainal Abidin (UniSZA), Malaysia.
kamariah@unisza.edu.my

Abstract

In this digital era, the role of computer technology as a resource for instruction of foreign language learners is increasing as educators recognise the ability of computer technology to produce both independent and collaborative learning environments. Computer technologies, for example the Internet, multimedia, and hypermedia have been introduced in English Language Learning and Teaching (ELLT) to foster language learning process, all of which fall under the category of Computer-Assisted Language Learning (CALL). Corpus linguistics is a systematic analysis of the actual (real) production of language (either spoken or written), in which texts are assembled using computer technology (concordancer) to form a large collection of authentic texts, called a corpus (plural-corpora) that comes in various sizes. Despite immense research on corpus linguistics in these recent decades, the potentials and limitations of Data-Driven Learning (DDL), the application of corpus linguistics in ELLT have not been widely discussed. Hence, this paper aims to review the potentials and limitations of DDL as a means of opening up opportunities for further studies. This insightful information is highlighted as a means of promoting DDL and producing independent learners in the 21st century classroom.

Keywords: corpus linguistics, Data-Driven Learning (DDL), corpora, concordancer

1. INTRODUCTION

In this new millennium, computer technology has revolutionised the fields of linguistic research (descriptive linguistics) and applied linguistics (language teaching and learning) with the advent of corpus linguistics. Corpus linguistics is defined as a systematic analysis of the actual (real) production of language (either spoken or written) as opposed to intuition. The texts (spoken or written production of language) are assembled to form a large collection of authentic texts called a corpus (plural-corpora) which comes in various sizes. The British National Corpus (BNC), for example, is a balanced synchronic text corpus comprising more than 100 million words. Language analysis is then performed using a tool called a concordancer, in which a large number of actual instances of the searched data, called patterns consisting of the Key-Word-in-Context (KWIC) or the nodes and their co-texts will be shown on the screen once typed.

The corpus-based approach to language learning is known as ‘classroom concordancing’ or DDL (Data-Driven Learning), an approach proposed originally by Johns (1991a) using the *Identify-Classify-Generalise* technique. This is an inductive approach, in which he put forth in his quote, “language-learner is also, essentially, a research worker whose *learning* needs to be driven by access to linguistic data” (Johns, 1991a, p.2). What he meant by the statement is that learners should be responsible for their own learning; that is, they should become learner-centred or autonomous learners through discovery learning. Learners should not leave it to teachers to help the process. This is an approach which has been proposed as striking the balance between the process and product approaches (Hadley, 1997), an approach which makes use of corpus technology (corpora and concordancers) to see the regularities of patterns of language use (Johns & King,

1991). This approach also suggests that grammar should consist largely of consciousness-raising activities rather than the teaching of rules (Rutherford, 1987). In other words, DDL has opened up a new model for ELLT in this century.

Classroom concordancing or DDL has contributed tremendously particularly in the teaching of English for Specific Purposes (ESP) courses. First, corpus-derived materials extracted from specialised purpose corpora enable ESP teachers in the creation of course syllabi and teaching materials for ESP courses (Bowker & Pearson, 2002; Gavioli, 2005; Nesselhauf, 2005; Römer, 2005a). Second, in the approaches known as corpus-driven or corpus-based, ESP learners, not the experts (teachers or researchers in the field), are given opportunities to experiment with the data derived from specialised corpora. ESP students are allowed to explore and work out with the concordance data (hands-on learning) to perform various language activities including, for instance, checking the correct usage of words and grammar of their written tasks, extending or deepening knowledge of existing language items, distinguishing close synonyms, detecting patterns of usage, collocation and colligation (phraseology), morphology, lexicography, sociolinguistics, and many others. Most importantly, this approach is very useful for the study of Language for Special Purposes (LSP), in which ESP learners acting as researchers will investigate the register and text type, discourse, and style of specialised languages, all of which fall under specialised phraseology of specialised disciplines such as medicine, law, and biology.

2. THEORETICAL PERSPECTIVES

There are two theories that frame DDL - Firth's (1957b) 'contextual theory of meaning' (a linguistic theory) and the socio-constructivist theory of learning (scaffolding) by Vygotsky (1978). Tognini-Bonelli (2001, p.14) claims that "When we bring corpus evidence into the classroom, it is important to understand the double role of corpus linguistics, entailing the methodological innovation and a theoretical one, because together they will account for a new way of teaching". Adopting the Firthian framework of the 'contextual theory of meaning', the central tenet of the theory is excerpted as follows:

We must take our facts from speech sequences, verbally complete in themselves and operating in contexts of situation which are typical, recurrent, and repeatedly observable. Such contexts of situation should themselves be placed in categories of some sort, sociological, and linguistic, within the wider context of culture.

(Firth, 1957b, p. 35)

Firth contends that the analysis of the meaning of utterances is the main goal of linguistics. He rejects any kind of distinction between 'langue' and 'parole' proposed by the father of modern linguistics (de Saussure, 1966) and Chomsky's (1965) ideas of 'competence' and 'performance' which considers language as a mental system, not as verbal behaviour. According to Firth, language is a set of events which speakers uttered, a mode of action and a way of 'doing things'. As utterances occur in real-life contexts, Firth argued that their meaning derived just as much from the particular situation in which they occurred ('extralinguistic') as from the string of sounds uttered (linguistic). This 'contextual theory of meaning' integrates language with the objects physically present during a conversation to ascertain the meaning involved.

While a linguistic unit (formal item) relies on its linguistic environments (contexts) in order to make meanings, meanings are further derived from extralinguistic contexts, contexts of situations and a much wider context of situations – culture. The sets of speech events are communicative events (functions) which are spoken and used by a society or discourse community (a group that share the same discourses, see Swales, 1990) in a given culture. These speech events make up a restricted language called a dialect or register (variation according to the use of language).

This present study employs the scaffolded DDL approach. This approach lends support from the social constructivist theory (scaffolding) introduced by Vygotsky (1978). Vygotsky defines scaffolding instruction as the “role of teachers and others in supporting the learners’ development and providing support structures to get to that next stage or level” (Raymond, 2000, p.176). This theory postulates that learners would reach the mastery level if they are scaffolded at the ZPD (Zone of Proximal Development). ZPD is the area between what a learner can do independently (mastery level) and what can be accomplished with the assistance of a competent adult (teacher) or peer (van Der Stuyf, 2002). Vygotsky believes that any child (learner) could be taught any subject effectively using scaffolding techniques by applying the scaffolds at the ZPD. This scaffolding strategy helps learners reduce the cognitive workload at the initial stages.

Scaffolding instruction is temporary and as the learners’ abilities increase, the scaffolding provided by the more knowledgeable other is progressively withdrawn. In scaffolded DDL classrooms, the teacher would scaffold learners in drawing conclusions at grammar rules or word meanings by providing printout concordance materials and guided DDL tasks before learners are left alone to work independently after they have mastered the skill. This may reduce the cognitive workload among learners when they have to use higher order learning skills such as generalising and formulating, the skills which might be foreign to Asian students and to those who are used to the deductive learning approach for so long (Smith, 2009).

According to O’Keefe, McCarthy, and Carter (2007), the main focus of socio-cultural theories is the social nature of classroom interaction. Learners “collectively construct their own knowledge and understanding by making connections, building mental schemata and concepts through collaborative meaning-making” (Walsh, 2006, as cited in O’ Keefe et al., 2007, p.228). Scaffolding is realised in dialogues (between a teacher and learners or within learners themselves in the form of self-dialogue (manifested in ‘private speech’) to comprehend the meaning, for example the content of a subject under study. Scaffolding is also a teacher strategy to assist learners to make sense of difficult tasks. The strategy comes in the forms of challenge and support (Walsh, 2006). A teacher provides the amount of challenge to maintain learner interest, motivation, and involvement, whereas the support is given to ensure students’ understanding of tasks. Scaffolded support recedes once a learner “can internalise external knowledge and convert it into a tool for conscious control” (Bruner, 1990, p.25).

In DDL context, scaffolding consists of problem-solving tasks (constructive) used to scaffold concordancing which are “provided to students in the form of questions termed as ‘question prompts’ or ‘scaffolding prompts’” (King, 1991, 1992; King & Rosenshine, 1993; Lin & Lehman, 1999; Scardamalia et al., 1994, as cited in Ha Le, 2010, p.19), ‘guided tasks’ (Boulton, 2010a), or “search skills for students to ‘discover’ collocations by themselves” (Woolard, 2000, p.33) in the

study of collocations. These scaffolding prompts would benefit learners cognitively by eliciting “learners’ self-explanation, self-questioning, self-monitoring, and self-reflection during their learning processes”, guiding “students in their knowledge construction, knowledge integration, and knowledge representation during their work on complex learning tasks”, linking “their arguments or explanations with their existing knowledge”, and finally, making students’ thinking more apparent and explicit, in which they are “better able to recognise areas in which their own understanding is lacking and to engage in knowledge integration” (Chang & Sun, 2009, as cited in Ha Le, 2010, p.20). Since the prompts are the problem-solving tasks, thus to scaffold concordancing is to scaffold a problem-solving process.

3. POTENTIALS AND LIMITATIONS OF DDL

Up to now, DDL has been accepted by some with open arms, but there are still others who are rather skeptical. Many have treated DDL as direct corpus consultation, not as the scaffolded DDL approach which makes use of concordance printouts, the approach which was originally proposed by the proponent of DDL - Tim Johns (1991a, 1991b). There are arguments that DDL would work best if learners are given help by instructors at the initial stages prior to allowing them to work with DDL independently by making use of concordance printouts (see Boulton, 2008a, 2009a, 2009c, 2010), a scaffolded DDL approach that takes the middle-ground position. This section reviews the DDL potentials and limitations of DDL.

3.1. The Potentials of DDL

One of the greatest contributions of DDL is to the teaching and learning of English for specific purposes (ESP). In fact, according to Aston (1998), Belcher (2006), Bernardini (2004), Conrad (2005), Gabrielatos (2005), Gavioli (2005), Pearson (1998), Sinclair (1991, 2004a), and Tognini-Bonelli (1993, 2001), the most accepted contributions of corpus linguistics have been in the descriptions of language for specific purposes, in which the language structure and use with emphasis on lexico-grammatical patterns or collocations are investigated. According to Conrad (2005, p.399), “teachers and students of a specialised variety want to know the characteristics of that variety and, therefore, analysis of a corpus of that variety is clearly useful”.

Using concordances may also help ESP learners grasp the lexis, concepts, usages, and pragmatics of specialised languages (Nolte, Ancarno, & Jones, 2018; Yunus, Mohamad, Ab Rashid, & Wahab, 2016). Many other ESP researchers and practitioners such as Aston (1996), Fuentes (2001), Gavioli (2005), Jabbour (2001), Pearson (1998), and Sinclair (1991, 2004a) suggested to make use of specialised corpora in teaching ESP since this type of corpora would be more representative of the needs of a small group in terms of developing both declarative and procedural knowledge. Moreover, the generic knowledge can be greatly facilitated if one understands how word combinations are structured and how they operate in textual environment. Gavioli (2005) and Hunston (2002) contend that to ESP practitioners, the issue of “what to teach” is very significant and crucial as opposed to “how to teach”.

In English for Legal Purposes (ELP), the establishment of English for Academic Legal Purposes (EALP) courses to international students is still in its infancy (Master, 2005). According to Pérez-González (1999c), legal English is a newcomer to ESP in higher education. Knowing the fact that legal language is a formulaic language, how could those theories about legal language be applied in

classroom setting? The fleeting of research which was carried out so far in this field (English for Specific Purposes) has suggested the use of corpora for solution. According to Master (2005), the macro-linguistic concerns in EALP are legal writing and the use of computer-mediated materials. Bruce's study (2002) on an EALP course in Hong Kong is one of a few publications available, and he asserts that EALP teachers should put an emphasis on the rhetorical aspects of arguments in legal problem answer writing. This involves the inclusion of conventionalised or formulaic legal expressions. Candlin, Bhatia, and Jensen (2002) also justify the need for a computer-mediated resource bank of English and discourse-based materials for teaching EALP because they found that the 37 EALP books currently available to be too context-specific and of little use outside that context.

Research on legal phraseology has probably become active since the inception of corpus linguistics. There are a few specialised legal contract corpora designed so far which can be used for language descriptions. They include the AARHUS corpus (Danish-French-English) corpus in contract law by the Business School of AARHUS, developed 20 years ago, the academic collocation corpus by Durrant (2009) on academic legal writing including legal writing articles, legal contract corpora by Gozdz-Roszkowski (2003, 2004) on the analysis of lexical bundles and contract terminology, and the Memorandum of Understanding (MoU) contract corpus designed by Awab (1999) on the analysis of modals. Durrant (2009), for example, produced a list of semi-technical (academic) collocations collecting from research articles written by several faculties and schools at Nottingham University including those from the law school. His research is very valuable, probably for two reasons. First, it confirms that colligations of preposition are the most frequent patterns found in all academic writing genres. Second, it shows that it is not the technical or specialised collocations that are frequent in academic writing but academic collocations. Colligations of prepositions characterise the dominant aspects of legal language which are frequently apparent across legal genres. DDL can be used to teach collocations of specialised courses, for example, legal discourse (see González, 1999c). We may notice that ESP/EALP, collocation, and DDL approach are inter-related. Their interrelationship is illustrated in Figure 1.

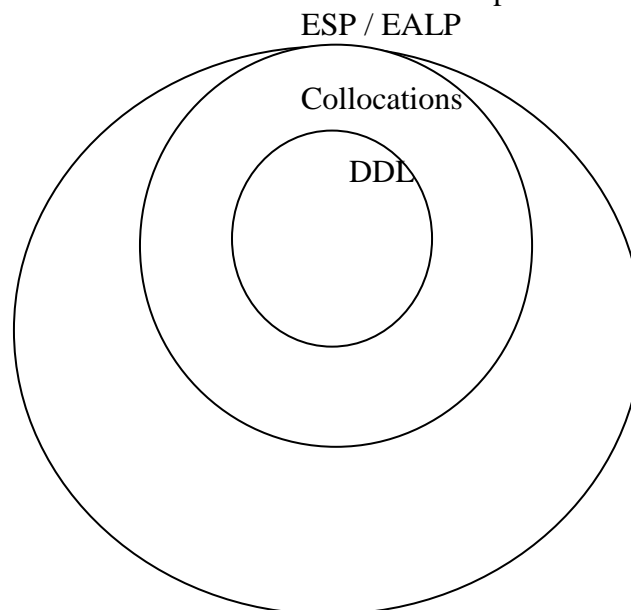


Figure 1: The Inter-relationship between ESP/EALP, Collocation, and DDL

Besides, DDL is also claimed as a communicative approach (Lewis, 1993, 1997, 2000) in a sense that it does not only engage learners with language facts (focusing their attention on language rules) but also engage learners in a communicative atmosphere with problem-solving tasks (discovery learning). In classrooms, learners need to interact with other group members in their groups to finish the tasks. This is in contrast to the students in the deductive or conventional approach where they become passive due to teacher-led approach.

DDL approach may also claim its advantages over other language learning methods which originated in the past few decades, for example Direct Method (DM), Audio-Lingual Method (ALM), and Grammar-Translation Method (GTM). Brown (2000), Lewis (1993), and Richards and Rodgers (2001) claim that the premises of the three approaches are basically flawed. They reasoned that many of the grammar rules taught in ESL classrooms are inaccurate or plain wrong; that is, they are not based on current usage. They also pinpointed that many of the grammar rules taught are frequently incomprehensible to students, for instance the aspect of *voice* in English. Because of the difficulty, learners often fail to understand abstract meta-language or the discourse function of grammar. Besides, they also claimed that there has been very little research evidence indicating that explicit knowledge of grammar aids acquisition of the grammatical system.

Moreover, many linguists have argued that grammar is not the only basis of language acquisition but should include fluency of language use in meaningful contexts (Lewis, 2000; Partington, 1998; Stubbs, 1996; Yunus, Mohamad, & Waelateh, 2016). Johns (1991b) mentioned that teaching grammar as a product cannot provide a full description of the complexity of the language. They are the products of “intuition-based armchair linguistics” (p.30) as evident from dictionaries, grammar books, and course books (traditional ELT materials).

Communicative Language Teaching (CLT), however, is a process approach that encourages creativity and self-discovery by students as they experiment with the language. A genuine CLT approach such as the task-based approach (Nunan, 1995, 1999) does not focus on forms (grammar and vocabulary) since its main principle in most classroom activities is to get the meaning across. Nattinger (1984) describe some major characteristics of the CLT syllabus, for instance an emphasis on meaning-making, language use in contexts, and a stress on fluency rather than accuracy in language learning.

According to Hadley (2002), DDL is a more preferable approach since it “appears to utilise the strengths of both product and process approaches to teaching grammar successfully” (p.106). Lewis (1993, 1997, 2000) also support this view when they claim that DDL is a communicative approach in a sense that it does not only engage learners with language facts by focusing their attention on language rules (form-focused instruction) (Ellis, 2005), but it also engages learners in a communicative atmosphere with problem-solving tasks (discovery learning). For example, in DDL classrooms, learners are expected to interact with other group members in their groups to finish the tasks. This is in contrast to students in the product or deductive approach, in which they become passive due to teacher-led approach. Besides, in DDL classrooms also, the role of a teacher has also changed from a teacher as a knowledge provider to a teacher as a ‘facilitator’. In the deductive language and grammar learning, the teacher is the driver and the students the passengers. In contrast, in DDL language learning, teacher plays more of the role of a co-pilot and navigator while the students take control of their own learning (Johns, 1991b).

Besides that, DDL works with authentic and genuine data as compared to the made up linguistic instances written in the structural grammar textbooks. Learners will be presented with the concrete facts of language, showing evidence of the contexts of situation of the text. Woolard (2004, p.40) asserts that concordances “provide much richer sources of co-textual information than dictionaries, and they can lead to a more exploration of the collocates of a word”. This simply means that DDL provides students the opportunity to observe a grammatical phenomenon of the language, to make hypothesis of how grammatical rules work, and to experiment to see if their hypothesis is correct (Payne, 2008). As opposed to DDL, in the traditional grammar learning, the teaching of grammar is conducted through the process of presentation of information done by the teacher. The students then practise with this information, and later they produce new contents.

In other words, DDL approach exposes learners to multiple instances of linguistic examples. They would not rely on textbooks anymore, the main companions to instruction in a traditional classroom. They will get the opportunity to discover language rules on their own through *Identify-Classify-Generalise* technique (Johns, 1991a) or *Observe-Hypothesise-Experiment* technique (Lewis, 2000). In the deductive approach, on the other hand, language items are presented through the PPP technique - presentation, practice, and production. In this approach, teacher intervention dominates the whole lesson. Flowerdew (2009) even suggests an improvement to Johns’ (1991a) DDL by adding a much more ‘soft’ DDL. She suggested a more ‘pedagogic-processing’ technique namely *Illustration*, *Interaction*, *Intervention* (optional), and *Induction*. This technique, she claims, is a middle-ground between the prescriptive and descriptive grammars.

DDL encourages learners to use their intuition based on corpus evidence and derive at grammatical rules through hypotheses-making processes. Though learners in the conventional method may also use their intuition to guess the rules and practise with language, Francis (1993) proposes that such a practice is unreliable because “there is often a difference between what they think they say and they actually do say” (p.86).

The DDL type of learning also trains learners to be independent, self-corrective, constructive, and autonomous especially in finishing the communicative tasks. According to Lee and Liou (2003, p.49), “the main advantage of the DDL approach is that it encourages students to take responsibility for their language learning”. This is because students have become more liberated from teacher-directed learning. Moreover, this kind of problem-solving approach benefits learners since it “exploits the learners’ natural tendency to work things out” (Bourke 1996, p.14). This approach is also more advantageous and “ensuring motivation” (Bernardini, 2004, p.106).

Moreover, it is argued that DDL approach could increase learners’ awareness of the facts or rules of language through consciousness-raising activities or tasks. Consciousness-raising is defined as deliberate attempts to draw learners’ attention specifically to the formal properties of the target language (Rutherford, 1987). Odlin (1994, p.14) claims that “consciousness-raising can succeed in changing interlanguage competence”. Ellis (1994, p.643) informs that “in consciousness-raising activities the learners are not expected to produce the target structure, only to understand it by formulating some kind of cognitive representation of how it works”. In the structural approach, consciousness-raising to grammatical rules is increased at developing implicit knowledge of the rule only through form-focused instruction (Rutherford, 1987). In contrast, DDL raises learners’ awareness of the

convention of a specific genre or register both through discovery learning (inductive) and form-focused instruction (Ellis, 2001, 2005). Granger and Tribble (1998) stress the importance of form-focused instruction, especially for adult learners since it is argued that incidental learning is not very effective with them.

Relevant to the concept of consciousness-raising is scaffolding instruction. Through scaffolding, learners' awareness of the target language is raised through consciousness-raising tasks, prepared by teachers in advance. Ha Le (2010) found in her study that her subjects in the experimental group who were treated with both concordancing and scaffolding (in the form of question prompts) scored significantly better than those in the control group who were treated with concordancing only in the posttest and delayed posttest. This finding was in line with Boulton's study findings (2008d, 2010a) indicating that even lower proficiency students can work better with DDL given scaffolding DDL instruction. This type of instruction, according to him, may reduce some of the difficulties associated with 'hands-on' work.

Another beneficial effect of DDL is that learner motivation can be raised via the use of technology as teaching aids (Boulton, 2008a; Chambers, 2005; Leńko-Szymańska & Boulton, 2015). Learners of today are the virtual students who prefer technological learning and teaching aids like computers and other multimedia in comparison to teachers' lecture and traditional books (Boulton, 2009b; Gavioli, 2001; Kern, 2006). According to Boulton (2008c), the current research on DDL as a whole has been reported positive with participating learners enjoyed DDL work because of this very nature.

Most importantly, DDL benefits learners in the study of lexico-grammatical patterns (collocations, colligations, and particularly colligations of prepositions). Many ESP researchers and practitioners including Celce-Murcia and Larsen-Freeman (1999) and Yunus, Awab, and Ab Rashid (2016) propose that colligations of prepositions be taught in context, for example, through corpora. This is because collocation errors are not easy to be explained except in a large number of contexts (Lewis, 1997, 2000). VanPatten, Williams, and Rott (2004, as cited in Ellis, 2006, p.87) also assert that "establishing connections between form and meaning is a fundamental aspect of language acquisition". Therefore, any grammar teaching that fails to describe the form-meaning connections of the target language must necessarily be inadequate.

3.2. The Limitations of DDL

While the literature has shown many great potentials of DDL, we cannot ignore the fact that DDL, as does any other approach or technology, has its many limitations as well. This section will also review the shortcomings of DDL in the light of the traditional approach.

The first attack on DDL is the data itself. DDL is a data-driven approach, in which data have become very important and need to be authentic. However, authentic data are sometimes rather daunting to be interpreted especially for lower proficient learners (Balunda, 2009; Boulton, 2009c; Gavioli, 2005; Hadley, 1997; Koosha & Jafarpour, 2006; Yoon & Hirvela, 2004). Another limitation of DDL is concerned with the term 'authentic'. Widdowson (1996) has constantly held a negative view towards the use of authentic data in classrooms. To him, authentic does not suggest meaningful in sociolinguistic sense, thus rejecting the notion of culture and society in texts as mentioned by Firth (1957b).

In addition, many scholars have doubted the practicality and efficacy of DDL as a teaching method that can improve learning (see Boulton, 2010a; Chambers,

2005; Gaskel & Cobb, 2004; Kern, 2006). Salaberry (2001) argues that the use of ICT in classroom allows 'technology-driven instruction' to take over from a 'pedagogically-driven approach'. And this, according to him, is a permanent danger.

Meanwhile, Jarvis (2004) expressed doubts whether DDL can guarantee an improved learning or motivation. Chambers and Kelly (2004, p.1) also felt the same thing when they asked others to think whether DDL is "a good thing pedagogically". Boulton (2010c) made a similar claim that "DDL is certainly no panacea to language learning, as is any other approach or technology" (p.14).

Furthermore, success with DDL in language learning does take into consideration of learner language learning styles and motivation (Boulton, 2009b). Many researchers have claimed that DDL may not be suitable for all learner profiles (see Boulton, 2009b; Cresswell, 2007; Chambers, 2005; Flowerdew, 2009; Tyne, 2009). Kaszubski (2008, p.174) found that his students fall into three categories in doing corpus consultation - "adopters, minimal users, and refusers", and this was presumably due to their learning style preferences. Some of his subjects were found to adapt to DDL more quickly ('adopters'), while others were found hard to adapt to it ('refusers').

Similarly, in Boulton's (2009b) study, he reported that there was some correlation between learners' receptivity to DDL and learning style preferences. He thus concluded that DDL seemed to appeal to those with the strongest visual preference. Yoon (2008, p.45) also reported that "a wide variety of individual experiences and learning contexts were involved in deciding the level of the students' willingness and their degree of success in using corpora". Meanwhile, Yoon and Jo (2014) also assert that students can be motivated further to adopt corpora as their learning tools if their needs and wants for using corpus as assistant to language learning are examined. Chambers (2005, p.119) also suggested that "differences in motivation or learning styles may explain the considerable variation in the success of the [DDL] activity".

Besides, the use of hands-on concordancing (direct application of corpora in classrooms) has left learners to ponder at large data (Hafner & Chandlin, 2007; Todd, 2000). Many students are incapable of or cannot endure learning without teacher supervision. To some of them, free or 'serendipitous' corpus exploration (Bernardini 2000) requires training or previous experience. And according to Mukherjee (2006, p.14), "it is doubtful... whether this extremely autonomous corpus-based activity can be fruitfully put into practice in the reality of ELT classrooms". Students are 'technophobic' to direct application of DDL (see Bernardini, 2002; Mukherjee, 2004; Seidlhofer 2000). And even if they are not perhaps as 'technophobic' as the respondents in Bernardini's (2002), Mukherjee's (2004), and Seidlhofer's (2000) studies, students are more comfortable with the traditional roles of teacher as knower and learner as the recipient of knowledge, the roles which was claimed by Boulton (2009b) to be stronger in France than in some other cultures.

DDL also challenges the language teaching approach which has been a tradition for so many decades in Asian context. While DDL has been proven to work in Europe (Boulton, 2008c, 2008d, 2009c, 2010; Johns, 1991a, 1991b; St. John, 2001), DDL has not been accepted with open arms in Asian countries, for instance in Japan, Taiwan, China, and Malaysia. This is due to the fact that ESL Asian learners have been exposed to the deductive (traditional) approach, for example Audio Lingual Method (ALM), for decades despite the introduction of

more inductive (non-traditional) approaches to English language learning. The deductive approach emphasises the role of teachers as sole knowledge providers. In this approach, learners become the recipient of knowledge, taking more of a passive approach to language learning. This type of learning contradicts the one proposed by DDL; that is, to take an active role in the process of learning by hypothesising and formulating rules. This learning approach “does not seem to fit too comfortably into the received model of Asian pedagogy” (Smith 2009, p.2). Yeh, Liou, and Li (2007) and Kenzhen (2015) also claim that the educational system and general background culture in Asian setting, for example in Taiwan, encourages more of the deductive approach.

Besides students, DDL is also a challenge for teachers. The application of DDL requires teachers to give high commitment, and they need to be technology savvy. Some should be threatened by technology and even some become computer phobia or ‘technophobics’. Teachers’ resentment to DDL is partly due to their “resentment of new technology and the time spent mastering it, as well as the risk to face in front of learners who are possibly more literate than the teachers in ICT (Information and Communication Technology)” (Boulton 2010c,p.3). DDL is indeed a ‘perpetual challenge’ (Johns, 2002) both for teachers and learners. Direct application of DDL is not the true spirit of DDL (Boulton, 2010c). This is in contrast to the original motivation of using the data as suggested by Johns (1991b) with learners; that is, the use of handouts or printed concordance outputs.

4. CONCLUSION

To conclude, DDL approach has many benefits. It sheds light on the importance of lexis and grammar (lexicogrammar) and specialised phraseology in ESP. The neo-Firthians like Sinclair, Halliday, and Hoey view grammar and vocabulary as ‘complementary’ units rather than as separate entities. Moreover, DDL approach also enhances language learning through multiple contexts and rejects the deductive approach in vocabulary teaching which emphasises ‘single words out of context’ (McCarthy, 2001, p.63). Multiple exposures to language samples or contexts through technology instead of a handful of made-up samples in textbooks give the opportunity for learners to lengthen the memory retention of the patterns (Cobb, 1997; Nation, 2001) especially where few learners have time to do reading for natural, multi-textual lexical acquisition (Cobb, 1997; O’Keefe, McCarthy, & Carter, 2007). Finally, DDL has become a stepping stone for learners to try out their potentials as ‘travelers’ or ‘language researchers’ compared to the traditional role (Johns, 1991a, 1991b).

On the other hand, DDL has many limitations as well. The limitations of DDL are due to many factors, among which include its practicality as a new approach replacing the traditional approach, different learners’ learning styles, technophobic students, teachers’ resentment of new technology, and some other barriers for instance technical and logistic aspects (Johns, 2002). These limitations, as lamented by Boulton (2009a), Leech (1991, 1997), and Thompson (2002), have resulted in lack of research interest and application of DDL in classrooms despite more DDL resources available online. However, due to the fact that DDL is a ground-breaking approach that affords teaching and learning in various ways (Rapti, 2013,) particularly in ESP context, and that DDL enables ESP teachers in the creation of course syllabi and teaching materials for ESP courses (Bowker & Pearson, 2002; Chambers & Kelly, 2004; Gavioli, 2001, 2005; Nesselhauf, 2005; Römer 2005a), its limitations should be minimised for deeper and further research in this area.

5. REFERENCES

- Awab, S. (1999). Multi-word units in a corpus-based study of memoranda of understanding: Modal multi-word units. Unpublished doctoral thesis, University of Lancaster, UK.
- Balunda, S.A. (2009). Teaching academic vocabulary with corpora: Student perceptions of data-driven learning. Unpublished master's dissertation, Indiana University, Indiana.
- Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis: Proceedings of the Fourth International Conference on Teaching and Language Corpora* (pp. 165-182). Amsterdam: Rodopi.
- Boulton, A. (2007a). DDL is in the details... and in the big themes. In M. Davies, P. Rayson, S. Hunston, & P. Danielsson (Eds.), *Proceedings of the Corpus Linguistics Conference (CL2007)* (pp.1-13), held 27th-30th July, 2007 at the University of Birmingham, UK.
- Boulton, A. (2007c). But where's the proof? The need for empirical evidence for data-driven learning. In M. Edwards (Eds.), *Proceedings of the BAAL annual conference* (pp 13-16). London: Scitsiugnil Press.
- Boulton, A. (2008a). *Evaluating corpus use in language learning: State of play and future directions*. Paper presented at AACL 2008 (American Association for Corpus Linguistics), Brigham Young University, Provo, Utah, USA.
- Boulton, A. (2008b). *Off-the-peg materials for DDL*. Paper presented at the New Trends in Corpus Linguistics for Language Teaching and Translation Studies: In Honour of John Sinclair, University of Granada/University Jaume I, Granada, Spain.
- Boulton, A. (2008c). Looking (for) empirical evidence for DDL at lower levels. In B. Lewandowska-Tomaszczyk (Eds.), *Corpus linguistics, computer tools, and applications* (pp. 581-598). Frankfurt am Main: Peter Lang.
- Boulton, A. (2008d). DDL: Reaching the parts other teaching can't reach? In A. Frankenburg Garcia (Eds.), *Proceedings of the Eighth Teaching and Language Corpora Conference* (pp. 38-44). Lisbon, Portugal: Associação de Estudos e de Investogação Científica do ISLA-Lisboa.
- Boulton, A. (2009a). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, 35(1), 81-106.
- Boulton, A. (2009b). *Corpora for all? Learning styles and data-driven learning*. Paper presented at the Fifth Corpus Linguistics Conference, University of Liverpool, and Liverpool, UK.
- Boulton, A. (2009c). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, 21(1), 37-51.
- Boulton, A. (2010a). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60(3), 534-572.
- Boulton, A. (2010c). Data-driven learning: On paper, in practice. In T. Harris & M. Moreno Jaén (Eds.), *Corpus linguistics in language teaching* (pp 17-52). Bern: Peter Lang.
- Boulton, A. (2011a). Data-driven learning: The perpetual enigma. In Gozdz-Roszkowski & B. Lewandowska-Tomaszczyk (Eds.), *Explorations across languages and corpora* (pp. 563-580). Frankfurt: Peter Lang.

- Boulton, A. (2011d). Language awareness and medium-term benefits of corpus consultation. In A. Gimeno Sanz (Eds.), *New trends in computer-assisted language learning: Working together* (pp.39-46). Madrid: Macmillan ELT.
- Boulton, A. (2012c). Beyond concordancing: Multiple affordances of corpora in university language degrees. In D. Macaire & A. Boulton (Eds.), *Languages, cultures and virtual communities. Procedia: Social and Behavioral Sciences*, 34, 33-38.
- Boulton, A. (2016). Integrating corpus tools and techniques in ESP courses. *ASp*, 69, 111-135.
- 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
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67 (2), 348-393.
- Bourke, J. (1996). In praise of linguistic problem-solving. *RELC Journal*, 27, 12-29.
- Bowker, L. & Pearson, J. (2002). *Working with specialised language: A practical guide to using corpora*. London: Routledge.
- Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course*. Boston: Heinle and Heinle Publishing Company.
- Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning and Technology*, 9(2), 111–125.
- Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning and Technology*, 9(2), 111–125.
- Chambers, A., & Kelly, V. (2004). Corpora and concordancing: Changing the paradigm in language learning and teaching. In A. Chambers, J.E. Conacher, & J.M. Littlemore (Eds.), *ICT and language learning: Integrating pedagogy and practice* (pp. 183-202). Birmingham: Birmingham University Press.
- Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, 25(3), 301–315.
- Cresswell, A. (2007). Getting to 'know' connectors? Evaluating data-driven learning in a writing skills course. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the Foreign Language Classroom* (pp. 267-287). Amsterdam: Rodopi.
- Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, 28, 157–169.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. (2001). Introduction: Investigating form-focused instruction. In R. Ellis (Eds.), *Form-focused instruction and second language learning* (pp. 1-46). Oxford: Blackwell.
- Ellis, R. (2005). Instructed language learning and task-based teaching. In E. Hinkel (Eds.), *Handbook of research in second language teaching and learning* (pp. 713-728). Mahwah, New Jersey and London: Lawrence Erlbaum Associates.

- Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, 40(1), 83-107.
- Firth, J. R. (1957a). Modes of meaning. In J. R. Firth (Eds.), *Papers in linguistics, 1934-1951* (pp. 190-215). London: Oxford University Press.
- Firth, J.R. (1957b). A synopsis of linguistic theory, 1930-1955. In J. R. Firth (Eds.), *Studies in linguistic analysis* (pp. 1-32). Oxford: Philological Society.
- Firth, J. R. (1968). Descriptive linguistics and the study of English. In F. R. Palmer (Eds.), *Selected papers of J. R. Firth 1952-59* (pp. 96-113). London and Harlow: Longman.
- Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistic*, 14(3), 393-417.
- Francis, G. (1993). A corpus-driven approach to grammar: Principles, methods and examples. In M. Baker, G. Francis, & E. Tognelli-Bonelli (Eds.), *Text and technology: In Honour of John Sinclair* (pp. 137-156). Amsterdam: John Benjamins.
- Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing Errors? *System*, 32, 301-319.
- Gavioli, L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. In Aston, G. (Eds.), *Learning with corpora* (pp. 108-137). Houston, Texas: Athelstan.
- Gavioli, L. (2005). *Exploring corpora for ESP learning*. Amsterdam: John Benjamins.
- Granger, S., & Tribble, C. (1998). Learner corpus data in the foreign language classroom: Form-focused instruction and data-driven learning. In S. Granger (Eds.), *Learner English on computer* (pp. 199-209). London: Longman.
- Hadley, G. (1997). *Concordancing in Japanese TEFL: Unlocking the power of data-driven learning*. Retrieved 22 October 1998, from <http://www.nuis.ac.jp/~hadley/publication/>
- Hafner, C.A., & Chandlin, N. (2007). Corpus tools as an affordance to learning in professional legal education. *Journal of English for Academic Purposes*, 6, 303-318.
- Ha Le, Thanh. (2010). *Learning lexical collocations with concordancing and scaffolding*. Unpublished master's dissertation, University of Groningen, the Netherlands. Retrieved 13 March 2011, from scripties.let.eldoc.ub.rug.nl/FILES/.../Le/MA_1938851_T_H_Le.pdf
- Jarvis, H. (2004). Investigating the classroom applications of computers on EFL courses at higher education institutions in UK. *Journal of English for Academic Purposes*, 3, 111-137.
- Johns, T. (1991a). Should you be persuaded: Two samples of data-driven learning materials. *English Language Research Journal (New Series)*, 4, 1-16.
- Johns, T. (1991b). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *English Language Research Journal (New Series)*, 4, 27-45.
- Johns, T. (1993). Data-driven learning. An update. *TELL & CALL*, 2, 4-10.
- Johns, T. (1997a). Contexts: The background, development and trialling of a concordance-based CALL Program. In A. Wichmann, S. Fligelstone, T.

- McEnery & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100-115). Harlow: Addison Wesley Longman.
- Johns, T. (2002). Data-driven learning: The perpetual challenge. In B. Kettermann and G. Marko (Eds.), *Teaching and learning by doing corpus linguistics* (pp. 107-117). Amsterdam: Rodopi.
- Johns, T., & King, P. (1991). Classroom concordancing. In T. Johns & P. King (Eds.), *English Language Research Journal*, 4, 17-25.
- Kaszubski, P. (2000). Selected aspect of lexicon, phraseology, and style in the writing of Polish advanced learners of English: A contrastive, corpus-based. Unpublished doctoral thesis, Adam Mickiewicz University, Poznań, Poland.
- Kern, R. (2006). Perspectives on technology in learning and teaching languages. *TESOL Quarterly*, 40(1), 183-210.
- Kezhen, L. I. (2015). The use of concordance programs in English lexical teaching in high school. *Higher Education of Social Science*, 8(1), 60-65.
- Koosha, M., & Jafarpour, A. A. (2006). Data- driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL Journal*, 8, 1-13.
- Lee, C-Y., & Liou, H-C. (2003). A study of using web concordancing for English vocabulary learning in a Taiwanese high school context. *English Teaching and Learning*, 27(3), 35-56.
- Leech, G. (1991). The state of the art in corpus linguistics. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics* (pp. 8-29). London: Longman.
- Leech, G. (1997). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and Language Corpora* (pp. 1-23). Harlow: Addison Wesley Longman.
- Leńko-Szymańska, A., & Boulton, A. (2015). *Multiple affordances of language corpora for data driven learning*. Amsterdam: John Benjamins Publishing Company.
- Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. London: Language Teaching Productions.
- Lewis, M. (1997). *Implementing the Lexical Approach: Putting theory into practice*. Hove: Language Teaching Productions.
- Lewis, M. (2000). Language in the Lexical Approach. In M. Lewis (Eds.), *Teaching collocation: Further developments in the Lexical Approach* (pp.126-154). Hove: Language Teaching Productions.
- McCarthy, M. (2001). *Issues in applied linguistics*. Cambridge: Cambridge University Press.
- Mukherjee, J. (2004). Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In U. Connor & T. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective* (pp. 239-250). Amsterdam: Rodopi.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nattinger, J. R. (1984). Communicative language teaching: A new metaphor. *TESOL Quarterly*, 18, 391-407. Doi: 10.2307/3586711
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.

- Nolte, I., Ancarno, C., & Jones, R. (2018). Using corpus methods to explore a survey on inter-religious relations in Yorubaland, southwest Nigeria. *Corpora*.
- Nunan, D. (1995). *Atlas: Learning-centered communication: Student's Book 1*. Boston: Heinle & Heinle.
- Nunan, D. (1999). Second language teaching and learning. *Journal of Languages and Linguistics*, 20(2), 66-99.
- Odlin, T. (1994). Introduction. In T. Odlin (Eds.), *Perspectives on pedagogical grammar* (pp. 1-22). Cambridge: Cambridge University Press.
- O'Keefe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- Partington, A. (1998). *Patterns and meaning: Using corpora for English language research*. Amsterdam: John Benjamin Publishing Company.
- Payne, J. S. (2008). Data-driven South Asian language learning. Retrieved 30 April 2010, from <http://salrc.uchicago.edu/workshops/sponsored/061005/DDL.ppt>.
- Rapti, N. (2013). Data-driven grammar teaching and adolescent EFL learners in Greece. *Corpus Linguistics and Variation in English: Focus on Non-Native Englishes*, 13, Retrieved 21 February 2012, from <http://www.helsinki.fi/varieng/series/volumes/13/rapti/>
- Richards, J. C. & Rogers, T. S. (1986). *Approaches and methods in language teaching: A description and analysis*. Cambridge, UK: Cambridge University Press.
- Römer, U. (2005a). *Progressives, patterns, pedagogy. A corpus-driven approach to English progressive forms, functions, contexts and didactics*. Amsterdam: John Benjamins.
- Rutherford, W. E. (1987). *Second language grammar: Learning and teaching*. New York: Longman.
- Seidlhofer, B. (2000). Going non-native? Do we need a new model for teaching EFL? *ELT News*, 42, 61-68.
- Smith, S. (2009). *Corpora in the classroom: Data-driven learning for freshman English*. Retrieved 1 January 2010, from http://www.nccu.edu.tw/_smithsgj/Corpora.
- St. John, E. (2001). A case for using a parallel corpus and concordancer for beginners of a foreign language. *Language Learning and Technology*, 5(3), 185-203.
- Thompson, P. (2002). What use are corpora in the teaching of EAP. In A. Cresswell, S. Bernardini, & G. Aston (Eds.), *Proceedings of the Fifth Teaching and Language Corpora Conference* (5th ed.) (pp. 71-72). Bertinoro, Italy: University of Bologna.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at work* (vol. 6). Amsterdam: John Benjamins.
- VanPatten, B., Williams, J., & Rott, S. (2004). *Form-meaning connections in second language acquisition*. Mahwah, New Jersey: Lawrence Erlbaum.
- Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, Massachusetts: Harvard University Press.
- Walsh, S. (2006). *Investigating classroom discourse*. London: Routledge.
- Widdowson, H. G. (1996). *Linguistics*. USA: Oxford University Press.

- Woolard, G. (2000). *Collocation - encouraging learner independence*. In M. Lewis (Eds.), *Teaching collocation: Further developments in the lexical approach* (pp. 28-46). Hove: Language Teaching Publications.
- Yeh, Y., Liou, H-C., & Li, Y-H. (2007). Online synonym materials and concordancing for EFL college writing. *Computer Assisted Language Learning*, 20(2), 131-52.
- Yoon, H. (2008). More than a linguistic reference: the influence of corpus technology on L2 academic writing. *Language Learning & Technology*, 12(2), 31-49.
- Yoon, H., & Jo, J. W. (2014). Direct and indirect access to corpora: An exploratory case study comparing students' error correction and learning strategy use in L2 writing. *Language Learning & Technology* 18(1), 96-117. Retrieved from <http://lt.msu.edu/issues/february2014/yoontojo.pdf>
- Yunus, K., Mohamad, M., Ab Rashid, R., & Wahab, Z. (2016). Corpus analysis of the multi - Level marketing persuasive phraseology. *International Journal of Applied Business and Economic Research*, 14 (14), 1233-1238.
- Yunus, K., Mohamad, M., & Waelateh, B. (2016). The breadth of receptive vocabulary knowledge among English major university students, *Journal of Nusantara Studies*, 1(1), 7-17.
- Yunus, K., Su'ad Awab., & Ab Rashid, R. (2016). Colligations of prepositions: Essential properties of legal phraseology. *International Journal of Applied Linguistics and Literature*, 5(6), 199-208.