

Data-driven Learning: A Student-centered Technique for Language Learning

Touraj Talai

English Department, Najafabad Branch, Islamic Azad University, Iran

Email: talai_najafabad@yahoo.com

Zahra Fotovatnia

English Department, Najafabad Branch, Islamic Azad University, Iran

Email: z.fotovatnia@iaun.ac.ir

Abstract—The use of student-centered methods in language teaching can be an alternative to be exerted by teachers in language classrooms. Such methods can create an atmosphere of excitement for the students because they themselves are supposed to discover word meanings, grammatical patterns, parts of speech and other aspects of language through receiving small tips from the teacher. In the present paper, data-driven learning (DDL) has been introduced as a student-centered technique in which samples of authentic language are taken from linguistic corpora to be presented to language learners for improving language proficiency and then different aspects of this technique have been elaborated on in order to make language teachers and students aware of its features and usefulness.

Index Terms—data-driven learning, corpus, concordancing

I. INTRODUCTION

DDL which was first coined by Johns in 1991 is a method in which learners read large amounts of authentic language and try to discover linguistic patterns and rules by themselves. DDL is famous because of its potential in language learning. It is a student-centered method; which exhorts rule and pattern discovery and learner autonomy. Johns (1988) expressed that DDL entails a shift in the role of teachers and students. In other words the teacher works as a research director and collaborator instead of transmitting information to the students directly and explicitly. DDL and Task-Based approach to language learning are sometimes taken as similar to each other because of the emphasis which is put on the role of students in both of them. Over the years, different investigators and language teachers have taken advantage of DDL for teaching different components of language such as collocations, grammatical points, affixes, etc. (Ball, 1996; Dyck, 1999; Kettemann, 1995; Tribble, 1997).

Samples of authentic language for preparing DDL exercises are usually taken from linguistic corpora. Corpus refers to a big collection of naturally occurring language produced by native speakers which is gathered from both spoken and written language. One of the advantages of corpus is its capability in familiarizing language learners with different patterns in the target language. With the use of corpus, we have the opportunity of analyzing natural language and finding the differences between formal and informal language. We can also learn about the features of spoken as compared with written language. At the present time corpora are mostly computerized but in the past they were produced in printed versions. The advent of computers has made us able to search for special words and patterns among corpora much faster and easier.

In DDL, a lot of emphasis is put on students' role. Teaching and learning are sometimes imagined as two complementary activities, but in reality, learning can occur without the existence of an instructor. Even in some situations, learning is more effective when there is no teacher. This is the basis of DDL in which the learners are exposed to authentic language whether in spoken or written form and are expected to discover patterns without the direct assistance of the teacher. It is believed that DDL has some advantages. For instance it improves learner independency and autonomy, enhances language awareness, and makes the learners able to cope with authentic language.

II. DIFFERENT ATTITUDES TOWARDS THE APPROPRIATENESS OF DDL FOR ADVANCED LEARNERS AND BEGINNERS

There is a controversy over a possible relationship between DDL and learners' language proficiency. The proponents of DDL take it as a synonym for task-based learning in which some activities are given to the learners who are asked to find patterns and rules of linguistic materials presented to them. Therefore, it is claimed that DDL is only useful for the learners at advanced levels and a lot of training is essential for its application. Johns (1986) said that DDL is appropriate for the learners who are adult, have enough motivation and, are sophisticated and intelligent.

A lot of literature on DDL shows that this type of learning has been used for the learners at advanced levels (Boulton, 2008a). The reason why in most of investigations about DDL the participants have been chosen from advanced level learners is that perhaps the researchers themselves have preferred this. In other words the researchers could choose their participants from among the learners at lower proficiency levels of language. Nevertheless there is a tendency towards thinking that DDL is only useful for the learners at advanced levels of language proficiency.

Against the idea that emphasizes the mere appropriateness of DDL for the learners at advanced levels of language proficiency, some investigations present evidence for the efficiency of this type of learning for the students at lower levels. For example Sealy and Thompson (2007) proved that even primary school children can take advantage of corpora in their native language without a high level of sophistication. However, Braun (2007) claimed that when corpus is used for learning a foreign or second language, the situation is different and it can be less efficient in some environments such as the ones outside the university. This is why in most articles about DDL, the participants are from among the learners at high levels of language proficiency.

Regarding the appropriateness of DDL for the learners at very low levels, such as beginners, there are different opinions. As an example, Hadley (2002) said that all the efforts made by his Japanese colleagues to take advantage of DDL led to failure. Some other investigators, such as Boulton (2009) claimed that learners may benefit from being exposed to the complexity of corpus at early stages of learning. However this should be done with caution for lower level students and the learners should be trained about its correct use.

III. THE IMPORTANCE OF TRAINING IN DEALING WITH DDL

An important aspect in dealing with DDL is sufficient training which is necessary for the learners. Training is essential not only for the learners but also for the teachers. The learners should gain the skill of working with linguistic materials. Bernardini (2001) expressed that the complexity of the linguistic materials should not be overestimated and the learners should gain the ability to deal with them very quickly. Sinclair (2004) believed that every teacher and student can easily come to the corpus world and learn the language. When a teacher wants to take advantage of DDL in the environment of the classroom, he can give instructions to the students about how to deal with the data presented to them and how to extract rules and patterns without much help from the teacher.

IV. DIFFERENT ASPECTS OF DDL

DDL can be viewed from different angles. It involves some features such as using linguistic corpus, concordancing, being inductive, providing a continuum from product to process, raising learners' consciousness, using computer programs and linguistic corpus for learning activities. So in the following paragraphs these features are discussed and explained.

V. LINGUISTIC CORPUS AS A SOURCE OF INFORMATION FOR PREPARING DDL EXERCISES

Corpus linguistics has been introduced as a methodology in the field of linguistics. This methodology first appeared in 1964 when the first computerized corpus was produced. Leech (1997) said that the Brown Corpus compiled by Henry Kucera and W. Nelson Francis gathered in 1964 was the first computerized corpus. It was a collection of one million words and was a large corpus at that time. In computerized corpora the texts are stored electronically and are analyzed by means of computer programs which are called concordancers.

A. *Computerized Corpora*

Collecting large amounts of authentic language was not regarded as something new when corpus linguistics was introduced as a methodology for language teaching. Mayer (2009) expressed that early dictionaries were full of examples taken from naturally occurring language. Even concordance lines were used for a long time before the appearance of corpus linguistics. Tribble (1990) said that concordance lines were used for centuries from the Middle Ages.

Collecting large amounts of authentic language is a time consuming activity which has been simplified with the use of computer programs at the present time. Besides searching for keywords and their surrounding contexts, these programs can also calculate the frequency of vocabulary items. The ability of computer programs in the field of corpus linguistics is not limited to the ones mentioned above. These programs have other abilities such as finding collocations related to special vocabulary items. Since different wide corpora are available for teachers and researchers, there is no need for them to produce a corpus by themselves.

B. *The Direct and Indirect Use of Corpora*

Corpora can be used both indirectly and directly. If corpora are used indirectly, it means that they help us in making decisions about what materials to teach and when to teach them. Barlow (1996) notes that corpora can influence course design and determine the content of the materials which are going to be taught in language classrooms. On the other side, if they are used directly, it means that they are used directly by learners and teachers for the purpose of getting familiar with the use of language in the real world or as Fligelstone (1993) says they help us in the teaching process.

While the indirect use of corpus focuses on its influence on syllabus design and linguistic materials for the purpose of teaching, the direct use of corpus focuses more on the teachers and learners. In the direct approach instead of relying on researchers as the providers of linguistic materials, language learners and teachers themselves study the corpus in order to discover language patterns, word meanings, etc. (Bernardini, 2002). Tim Johns pioneered the direct use of corpus for learning grammar and vocabulary. Johns (2002) believed that the learners should encounter the linguistic data directly and tried to foster the role of learners as linguistic researchers. Johns (1997) also regarded the learner as “language detective” and called every learner as a Sherlock Holmes. This method, in which there is an interaction between corpora and students, is known as data-driven learning. The activities of DDL are usually taken from corpora.

Some investigations have been done about the efficiency of using corpora directly for language learning. For example Watson (2001) conducted such an investigation. In his research, the students used internet as a source of information. In this study, the students were required to correct two mistakes they made in writing a report as their assignment. The mistakes were indicated by the teacher and the students were supposed to search the internet to find examples of use of the indicated words as a corpus and induce rules in order to self-correct their mistakes. The results showed that students could correct their mistakes in 78% of instances.

C. The Use of Linguistic Corpora in Language Classrooms

Corpora are not used only for the mere analysis of languages. Nowadays they also play the role of linguistic materials for language classes. Woolard (2000) expressed that corpora and the use of concordancing which were used only for the purpose of research about the language in the past, is being used as an important instrument for language teachers in their classes at the present time. Having access to corpora may be thought of as something difficult for language learners, but many authors who have done investigations about the use of corpora have published articles on the fact that corpora are easily accessible to the learners (e.g. Fox 1998; Kettemann, 1995; Stevens, 1995; Tribble, 1997; Wichmann, 1995).

Language teachers can play an important role in motivating the students to use corpora for language learning. This needs the teachers to believe in the usefulness of corpora for this purpose. Many articles have been published about the use of corpus in language classes. The teachers can study them and apply the methods to their classes. We cannot expect the learners to be aware of all ways for improving their language proficiency, so the teachers can assist them by introducing new methods. As corpora are produced by native speakers, the assistance of teacher as a guide in class can simplify the process of learning for the learners.

D. Linguistic Corpora for General and Specific Purposes

Corpora can be built for two different aims. Sometimes they are made for specific purposes and sometimes for general purposes. In order to build specific corpora the materials can be found on the internet. Such a kind of corpus needs not to be very large. It should only contain a reasonable number of examples in the target language. But in order to prepare a corpus for general purposes more problems and difficulties are encountered because the number of examples provided should be larger in number in order to be the representative of the target language (Watson, 2001).

Rizzo (2010) notes that a general corpus is gathered to be used for contrastive analysis or for providing a description of the general language. As a result, gathering linguistic materials involves collecting texts from a wide range of genres and topics to represent the usage of general language. On the contrary, in specialized corpora samples of a particular variety or register of the language are gathered in order to create a dictionary, analyzing the development of children language, studying the language when it is used in a specific topic area, etc. depending on the aims of the investigation.

VI. CONCORDANCING IN DDL

Language teachers can exert concordancing technique for presenting DDL exercises to the learners. This means that several sentence examples in which a special word or grammatical point has been used are gathered. After that the sentences are written under each other and may be cut from the beginnings and ends to be shortened for the purpose of centralizing the intended word or grammatical point in the middle of each line. This results in incomplete sentences, but it helps the teachers to sort the target word which is common in all the sentences exactly under each other. Through using this technique the attention of the learners is attracted to the intended word and its immediate context in different sentences. Then the learners can get familiar with the new word or grammatical point.

The presentation of such incomplete sentences may interrupt complete understanding. Hadley (1997) said that according to the results of the questionnaires given to the participants in this method, many participants believed that the concordance style of presenting part of the sentences was problematic for them. In this way, we cannot be sure about whether or not the learners can get the right meaning after reading such incomplete contexts. There are some tricks that can solve this problem. For example, a complete sentence can also be added to the end. Another solution is to present complete sentences instead of cut ones, however parts of the sentences may fall in the next line.

The learners should get familiar with this type of technique to cope with the sentence examples. When the learners encounter such examples taken from the corpus, they should discover the patterns and rules by themselves. So they should be familiar with inductive learning. Sun (2003) said that the learners who are only familiar with deductive language learning need to receive a lot of guidance to deal with concordancing in linguistic corpus. The use of

concordancing in linguistic corpora can be used as well as using dictionary for improving vocabulary knowledge. Frankenberg-Garcia (2005) came to the conclusion that with a little training about coping with concordances, corpus can be as efficient as dictionary for the purpose of improving language proficiency.

VII. DDL AS AN INDUCTIVE APPROACH

DDL is an inductive approach as it does not present word meanings and grammatical rules directly. As an example if a teacher wants to teach a point of grammar to the students in an inductive way, he gives a text full of that grammatical point to the learners, and in this way, the students themselves should try to discover the structure of the sentences and the grammatical points of the text. Researchers believe that an inductive approach is more useful for acquiring simple rules and patterns (Reber, 1989; Robinson, 1997). In a study done by Lee & Liou (2003), they came to the conclusion that the students who liked inductive learning were more successful when they were given a DDL lesson. In DDL the data can be taken from linguistic corpora and the learners can learn the word meanings and grammatical rules indirectly and in an inductive way.

VIII. A CONTINUUM FROM PRODUCT TO PROCESS IN DDL

Bastone (1995) believed that DDL is a method which is a continuum from product to process. It has the features of product approach since specific linguistic aspects are given to the learners in a lot of sentences of the target language. It also has the features of process approach because it improves creativity and self-discovery in language learners. There was a disadvantage for product approaches in the past. The disadvantage was that in earlier product approaches, idealized sentences were extracted for the purpose of language teaching, but in DDL, lots of authentic sentences produced by the native speakers are exerted. The researchers in favor of DDL as a product approach are preparing increasing proofs that show how the non-idealized information taken from corpora and the use of concordances can be exploited in language classes (Tribble 1996, Kettemann 1995, Tribble & Jones 1990).

IX. CONSCIOUSNESS RAISING IN DDL

The DDL approach in teaching vocabulary and grammar leads to consciousness-raising of the learners. Rutherford and Smith (1998) defined consciousness-raising as “the deliberate attempt to draw the learner’s attention specifically to the formal properties of the target language”. In DDL, learners are not regarded as the mere recipients of linguistic data, but as researchers who study the regularity of language. The duty of teachers is only to encourage the learners without directly assisting them in the process of language learning and discovering linguistic rules. After receiving some tips from the teacher, the learners’ consciousness regarding an intended language component increases and in the next step the learners will be able to produce language using what they have learned in DDL exercises.

The use of computers and computer programs can be part of DDL. Boulton (2008a) came to the conclusion that from 50 studies about DDL, all but eight involved the use of computer by learners. Besides the use of computers, DDL can be used through printed materials instead of computer programs for the presentation of data to the learners. Some students may be technophobic (Bernardini, 2002). So the use of printed papers can be more effective for them as compared to the use of computers. Using papers has some advantages. For instance it is more economic for the researchers because they don’t need laboratories equipped with computers for doing their investigations about DDL. In other words printed materials are more accessible than computer programs. In DDL, linguistic materials which are taken from linguistic corpora can also be presented to the learners with the use of computers. So it can be regarded as a combination of corpus and CALL.

X. CORPUS AS A SOURCE OF INFORMATION FOR DDL ACTIVITIES

The materials in DDL are usually taken from different types of corpora. In a corpus, the type of language which is used is not simplified as the phrases and sentences of learners’ dictionaries. In other words they are full of expressions produced by native speakers of the target language. As a result, there might be more than one difficult word in the sentences taken from corpus. Boulton (2010) believed that DDL seems almost a difficult approach for language learning, because it involves the use of corpora and computer programs combined with a new approach. In DDL, the learners are not taught explicit rules, but read corpora to recognize patterns among a lot of examples which are given to them; consequently it is like natural approach to language learning (Gaskell & Cobb, 2004p. 304; Scott & Tribble, 2006, p. 6; Boulton, 2009).

XI. INADEQUATE INVESTIGATIONS ABOUT DDL

Whenever a lot of investigations are done about the usefulness of a new method in the field of language teaching and learning, it is expected that many language teachers find it useful and efficient to be used in their classes. Unfortunately, because of inadequate investigations about the efficiency of DDL, many language teachers are unaware of its application. DDL has not become a major approach for learning languages in the world and this is why only few DDL

materials are found to be used by teachers in language classes (Boulton, 2008b). More investigations about the efficiency of this approach can lead to its introduction to the world of academic language teaching in the coming years.

XII. CONCLUSION

DDL is a student-centered method in which natural instances of language produced by the native speakers are gathered and presented to the learners for the purpose of improving language proficiency. DDL exercises can be used to learn new items of vocabulary, grammatical points, collocations, parts of speech, etc. As DDL exercises are extracted from linguistic corpora, some believe that this technique is more useful for the learners at advanced levels, so that they can cope with samples of natural language which have not been simplified. The familiarity of language teachers and learners with DDL can give them the opportunity of exerting this technique to increase learner autonomy. This emphasis on the role of the learners can make language learning more exciting for them since they are active during the process of learning and should discover language rules, word meanings, etc. by themselves.

REFERENCES

- [1] Ball, C. N. (1996). Tutorial notes: Concordances and corpora. Retrieved March 24, 2003 from <http://www.georgetown.edu/cball/corpora/tutorial.html>.
- [2] Barlow, Michael. (1996). Corpora for Theory and Practice. In: *International Journal of Corpus Linguistics* 1(1), 1-37.
- [3] Bastone, R. (1995). Product and process: Grammar in the second language classroom. In M. Bygate, A. Tonkyn, and E. Williams, (Eds.), *Grammar and the language teacher*. (Pp. 224- 236). London: Prentice Hall.
- [4] Bernardini, S. (2001). Spoilt for choice: A learner explores general language corpora. In G. Aston (Ed.), *Learning with corpora* (pp. 220-249). Houston: Athelstan.
- [5] Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp.165-182). Amsterdam, New York: Rodopi.
- [6] Boulton, A (2008a). But Where's the Proof? The Need for Empirical Evidence for Data-Driven Learning. In M. Edwardes (ed.) *Proceedings of the BAAL Annual Conference 2007*. London: Scitsiugnil Press, 13-16.
- [7] Boulton, A (2009b). Testing the Limits of Data-Driven Learning: Language Proficiency and Training. *ReCALL*. 21/1, 37-54.
- [8] Boulton, A. (2010). Data-driven learning: on paper, in practice. In T. Harris & M. Moreno Jaen (eds.) *Corpus linguistics in language teaching*. Bern: Peter Lang, p. 17-52. [pre-publication version]
- [9] Braun, S. (2007). Integrating corpus work into secondary education: from data-driven learning to needs-driven corpora. *ReCALL*, 19(3): 307-328.
- [10] Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, 25(3),301-315.
- [11] Dyck, G. N. (1999). Concordancing for English teachers. Paper presented at TESL Manitoba, Canada. Retrieved March 21, 2003 from <http://home.cc.umanitoba.ca/~gdyck/conc.html>
- [12] Fox, G. (1998). Using corpus data in the classroom. In Tomlinson, B. (ed.) (1998) *Materials Development in Language Teaching*. Cambridge: Cambridge University Press. pp. 25-43
- [13] Frankenberg-Garcia, A. (2005a). A peek into what today's language learners as researchers actually do. *International journal of lexicography*, 18(3), 335-355.
- [14] Fligelstone, S. (1993). Some reflections on the question of teaching, from a corpus linguistics perspective. *ICAME Journal* 17, 97-109.
- [15] Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32(3), 301-319.
- [16] Hadley, G. (1997). A survey of cultural influence in Japanese ELT. *Buletin of Keiwa College* 6: 61-87.
- [17] Hadley, G. (2002). Sensing the winds of change: An introduction to data-driven learning. *RELC journal*, 33(2), 99-124.
- [18] Johns, T (1986). Micro-Concord: A Language Learner's Research Tool. *System*. 14/2, 151-162.
- [19] Johns, T. (1988). Whence and Whither Classroom Concordancing? In P. Bongearnts/ P. de Hann/ S. Lobbe/ H. Wekker (eds) *Computer applications in language learning*. Dordrecht: Foris, 9-27.
- [20] Johns, T (1997), Contexts: The Background, Development and Trialling of a Concordance-based CALL Program. In: Wichmann et al. 1997, 100 115.
- [21] Johns, T. (2002). Data-driven learning: The perpetual challenge. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 107-117).
- [22] Kettemann, B. (1995). On the use of concordancing in ELT. *TELL&CALL* 4: 4-15.
- [23] 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.
- [24] 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). London: Longman.
- [25] Meyer, C.F. (2009). In the profession: The "empirical tradition" in linguistics. *Journal of English Linguistics*, 37, 208-213.
- [26] Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of experimental psychology: General*, 118, 219-35.
- [27] Rizzo, C. R. (2010). Getting on with corpus compilation: From theory to practice. *ESP World*. Issue 1 (27), Vol 9, pp. 1-23.
- [28] Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in second language acquisition*, 19, 223-47.
- [29] Rutherford, W. & Smith, M. (1988). (Eds.) *Grammar and Second Language Teacher: A Book of Readings*. In Sinclair, John (ed.), 1987, *Looking Up: An account of the COBUILD project* (pp. 107- 116). Boston: Heinle and Heinle.
- [30] Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. Amsterdam: John Benjamins.
- [31] Sinclair, J (2004). *How to use corpora in language teaching*. Amesterda: John Bejamins.

- [32] Sealey, A. & Thompson, P. (2007) Corpus, concordance, classification: young learners in the L1 classroom. *Language awareness*, 16(3): 208-216.
- [33] Stevens, V. (1995) Concordancing with language learners: Why? When? What? *CAELL Journal* vol. 6 no. 2 pp. 2-10.
- [34] Sun, Y-C. (2003). Learning process, strategies and Web-based concordancers: A case-study. *British Journal of Educational Technology*, 34(5), 601-613.
- [35] Tribble, C. & Jones, G. (1990). Concordances in the classroom: A resource book for teachers. Essex, U.K.: Longman.
- [36] Tribble, C. (1996). Concordances in the classroom. London: Longman.
- [37] Tribble, C. (1997). Improvising corpora for ELT: Quick and dirty ways of developing corpora for language teaching. Available at: <http://web.bham.ac.uk/johnstf/palc.htm>.
- [38] Tribble, C. & Jones, G. (1990). Concordances in the classroom: A resource book for teachers. Essex, U.K.: Longman.
- [39] Tribble, C. (1997). Improvising corpora for ELT: Quick and dirty ways of developing corpora for language teaching. Available at: <http://web.bham.ac.uk/johnstf/palc.htm>.
- [40] Watson Todd, R. (2001). Building and using your own corpus and concordance. *ThaiTESOL bulletin* vol. 14 no. 2 pp.1-9.
- [41] Wichmann, A. (1995). Using concordances for the teaching of modern languages in higher education. *Language Learning Journal* no. 11 pp. 61-63.
- [42] Woolard, G. (2000). Collocation - encouraging learner independence. In Lewis, M. (ed.) (2000) *Teaching Collocation: Further Developments in the Lexical Approach*. Hove: Language Teaching Publications. pp. 28-46.

Touraj Talai received his M.A. in TEFL from Islamic Azad University, Najafabad Branch. His areas of interest are Applied Linguistics, Psycholinguistics and Phonetics.

Zahra Fotovatnia has a PhD. In TEFL from Islamic Azad University, Science and Research Branch. She is an assistant professor at the English department of Islamic Azad University, Najafabad Branch. She has presented and published papers in international conferences and journals.