

THE GOOD AND THE BAD OF THE CORPUS-BASED APPROACH (OR DATA-DRIVEN LEARNING) TO ESP TEACHING

Olivia CHIROBOCEA¹

¹Lecturer, Ph.D., Department of Modern Languages for Specific Purposes, Faculty of Letters, “Ovidius” University of Constanta, 1 Aleea Universitatii, Constanta, e-mail: olivia.ch31@gmail.com

Abstract: *This paper tackles a didactic method with a rising popularity among those involved, especially in English for Specific Purposes teaching. First proposed to the world in the 1980s by Tim Johns, it gained even more notoriety with its inventor's death in the late 2000s, a notoriety also helped by the technological advancement that would considerably ease the application of this method. With its specific terminology which includes terms such as corpus / corpora, data-driven, computer-assisted or concordance / concordancer, the data-driven learning or corpus-based approach to teaching a foreign language is especially useful to ESP teachers, the various activity domains of their learners making this method all the more useful, as corpora has specific importance in this field. Although heavily discussed and promoted, the method is still largely either misapplied or misunderstood by teachers and, apart from its obvious advantages, it has some important disadvantages which make it harder to use. This paper will present some of these issues in an attempt to make the method more familiar and applicable to the ESP teaching field. It will also emphasize some of the issues this problem may have in its application in Romanian contexts.*

Key words: *data driven learning, corpus, ESP, concordancer*

Introduction

The teaching of languages in general and of English in particular has always sparked new ideas and has opened the path for new methods, always refining methodologies, redefining terms and the actors involved (teachers and learners), in an attempt to keep up with the times. Data Driven Learning (DDL) is one such endeavor to update the methodology of teaching and learning a language by making use of technology and by bringing it closer to the learners' needs. This method also promises to teach the skills learners need in order to always be able to learn and understand language issues, even after class, in their future life. This is a commendable effort since language teaching and learning, regardless of methodology, is usually an activity revolving around a source of knowledge (the teacher) and a more or less passive receiver (the learner). DDL challenges this dynamics and turns the learner into a researcher, a language detective, terms coined by the initiator of this method, Tim Johns. Equipped with skills to look for themselves exactly what they need, learners will no longer be lost in useless activities that a regular English class contains regardless of individual learning speed.

DDL – the method

Data Driven Language Learning (or the Corpus-Based approach) was proposed in the 1980s, when personal computers were beginning to emerge as a source of knowledge and as a learning aid. The initiator was Tim Johns, a Birmingham University professor, who disseminated his ideas in a book co-authored with John Higgins, *Computers in Language Learning* (1984) and in various articles published in journals, collective volumes or conference proceedings, such as: “Micro-Concord: A

language learner's research tool” (1986), “Whence and whither classroom concordancing?” (1988), “From printout to handout: Grammar and vocabulary teaching in the context of data driven learning” (1990), “Should you be persuaded: two samples of data-driven learning materials” (1991), *Classroom Concordancing* (1991, editor), and “Data-driven Learning: The Perpetual Challenge” (2000). While some of the aspects Johns detailed in his early works may be outdated, especially those regarding the use of particular software which no longer exists or is much better and faster now, all other issues remain valid as they present the theory of the method and solutions to problems encountered by teachers while using it. The 1980s were also the age of the COBUILD project (Tribble and Jones, 1991; Johns, 2000; Gabrielatos, 2005), an effort by the Birmingham University to create the Collins Corpus, an electronic corpus of texts which later led to the production of the *Collins COBUILD English Language Dictionary*, the first corpus-based dictionary, and the Bank of English, a collection of over 600 million words compiled from the COBUILD corpus and taken from various sources including newspapers, books, informal conversations, radio and TV. Tim John was a contributor to this project (Boulton, 2012 : 23) working with the MicroConcord software in early CALL programs (Computer Assisted Language Learning).

In his 1991 article, in fact a chapter in *Classroom Concordancing*, Tim Johns explains the main innovation of the approach, namely the changing of the role of teacher and learner in a technology-based medium which allows for more freedom of understanding and learning. Thus, the typical method is given by Sinclair and Coulthard in 1975

and cited by Johns: “The teacher typically asks a question (answer already known) to check that learning has taken place: the learner attempts to answer that question: and the teacher gives feedback on whether the question has been successfully answered. Such is the I(nitiation)-R(esponse)-F(eedback) structure of the classroom exchange.” (Johns, 1991 : 1). This is where Johns introduces the idea of informant, that is passive, to replace the typical teacher. The informant in this case is a computer whose role is to provide enough information in authentic contexts to help the learners figure language problems on their own. The purpose is to “simply provide the evidence needed to answer the learner’s questions, and rely on the learner’s intelligence to find answers.” (Johns, 1991 : 2). And this requires two elements that will be discussed in the next section: corpus and concordancer.

As theorized, described and analyzed by many researchers and theorists (Scott, 2005; Boulton, 2009), DDL can be summarized in a few typical features: it is empirical and data-driven, as the very name suggests, it is not based on theory but generates it, it depends on corpora and on computer software, and it is discovery- and detection-oriented, attempting to make linguistic detectives out of learners, providing them with the skills for further, out-of-classroom learning. One of the main innovations of this method is the use of computers and software. While in the 1980s, when DDL was first proposed, such technology was in its infancy, it joggled the imagination of linguists as a golden door towards endless possibilities. Nowadays, the use of a computer is not a choice, but an imperative, an everyday necessity. The computer itself is not a unique medium for the use of software. People have technology in their pockets at all times as information can be accessed from mobile phones, tablets, notebooks, smart watches and other gadgets whose purpose is to shorten the distance between us and knowledge.

Thus, in theory, DDL should work wonders in this age where it seems to be most at home. However, that is not the case, and in spite of the countless studies and research documenting the possibilities of DDL, this method is not practiced in classroom as often as one might think. In my research for this article, I encountered numerous studies which debate the issues of the method, the solutions to typical problems and the many applications in the teaching of various aspects of language (vocabulary, grammar etc.). Although there are Romanian teachers and researchers that mention using corpora in their teaching practices or in their translation tasks, the method itself, as it was proposed, does not seem to have

much success or appeal in language classes in Romanian schools and academia. In any case, it is not heavily documented.

DDL – terminology

This method also comes with its own terminology as it introduces new concepts other methods do not have. Two of the most important such terms are corpus and concordancer. A corpus is a compilation of texts, it is “any collection of recorded instances of spoken or written language” (Gabrielatos, 2005), or simply, “any body of text” (McEnery and Wilson, 2001 : 197). Thus, any collection of texts makes up a corpus. It is not restricted to specialized texts as it may be understood. Even the assignments students have to prepare can make up a corpus (Gabrielatos, 2005), and this is an important one indeed, as a teacher draws conclusions after marking the papers and has a glimpse into his or her students’ knowledge at some particular moment in time. Each assignment is a new resource and evidence based on which progress can be assessed. While corpus is not an invention of the computer era, as such collections have existed for centuries, the compiling of a corpus is greatly eased by the assistance of software and the internet. With the help of technology, corpora are much easier to assemble and reference.

There are many types of corpus as there are many types of needs and purposes. Costas Gabrielatos speaks of various possibilities as corpora may be expandable or not, may contain just chunks of language or whole texts, may have a general or particular content and other traits: “Reference corpora have a fixed size; that is, they are not expandable (e.g., the British National Corpus), whereas monitor corpora are expandable; that is, texts are continuously being added (e.g., the Bank of English). Another design-related distinction is whether a corpus contains *whole texts*, or merely *samples* of a specified length. [...] In terms of content, corpora can be either *general*, that is, attempt to reflect a specific language or variety in all its contexts of use (e.g., the American National Corpus), or *specialised*, that is, aim to focus on specific contexts and users (e.g., Michigan Corpus of Academic Spoken English), and they can contain *written* or *spoken* language. [...] Finally, corpora can be monolingual (i.e., contain samples of only one language), or multilingual.” (Gabrielatos, 2005).

Teachers can create their own corpus of texts, while large corpora are also available online or on other media for easy access. In the same vein, other authors advise to organize corpora using authentic and contemporary texts, and avoid literature, dialect, poetry or highly technical material (Tribble and Jones, 1991 : 13-15). In

what regards the avoidance of highly technical material, the opposite is true for ESP, which is mainly based on specialized language. In terms of how to collect corpora, the possibilities range from tedious methods such as keying (the actual typing of text by means of a word processor) to the contemporary use of online sources as the internet provides access to millions of examples of authentic texts and is able to cross-reference the learners' needs in a matter of seconds. The motivation behind the use of corpora is that the stock examples given by teachers in class to illustrate instances of grammar or vocabulary use are not relevant for real-life use of language (Boulton, 2009 : 10). A criticism to this very notion is given and maintained throughout the years by Henry Widdowson who argues that, on the contrary, these samples of language, as they are used in DDL, are taken out of context and thus lose this very authenticity the method claims (Widdowson, 1998, 2000). Another important reason for using corpora is the independence it affords learners as “students define their own tasks as they start noticing features of the data for themselves” (Johns, 1997 : 101), details that the teacher may not notice, or may not deem relevant, while for a learner it may be just so.

The second term to be defined is concordancer, which nowadays refers to a computer software or an online service. However, Tribble and Jones offer a more complex definition: “In its original sense a concordance is a reference book

containing all the words used in a particular text or in the works of a particular author (except, usually, the very common grammatical words such as articles and prepositions), together with a list of the contexts in which each word occurs. Each context may be indicated by means of a precise line reference, or by a short citation, or both.” (Tribble and Jones, 1991 : 1). The use of electronic concordancers offers obvious advantages such as interactivity, revealing of patterns of language otherwise less visible and the large amount of authentic contexts that offer a broader image of language.

As numerous authors illustrate (Johns, 1990, 1991; Gabrielatos, 2005; Boulton, 2009, 2012; Marzá, 2014), a concordancer is an efficient and easy tool to teach the use of any aspect of language, be it prepositions and conjunctions, collocations, *if*-clause, passive voice, word derivation, the use of articles, homonymy and, practically, anything, as utilizing a corpus is compatible with other approaches to teaching English as well (Gabrielatos, 2005). A corpus of texts from numerous instances of authentic use will, for example, provide learners with a better understanding of the multitude of uses for *as*. For example, inputting this keyword, *as*, into an online concordancer produces one thousand results, that is one thousand examples of *as* uses. Figure 1 is a print screen of only a fragment of this large result produced by an online concordancer (<http://lexutor.ca/conc/eng/>).

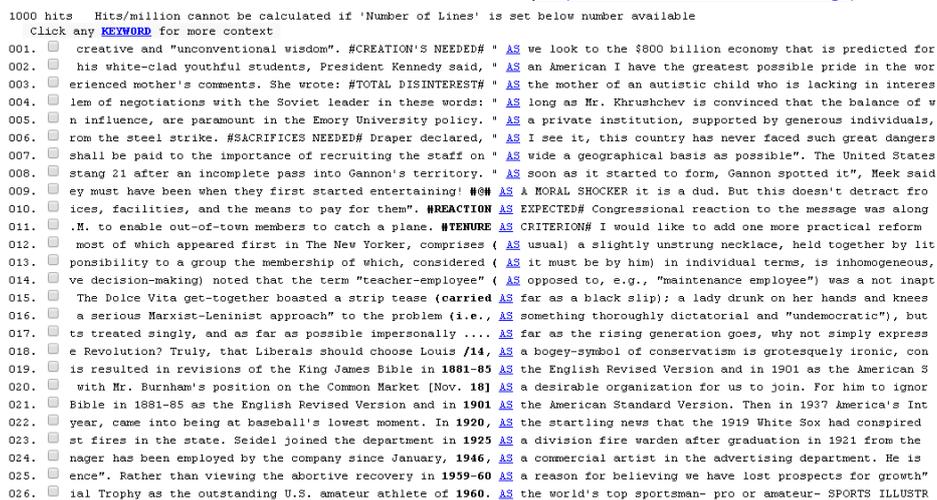


Figure 1

DDL – problems and solutions; applications to ESP

Like any other method of teaching a language, DDL has obvious advantages, but also less overt disadvantages that can arise from practice. Several theorists analyze these issues using research by other teachers, their classroom experience and statistics regarding the impact of this method on both teacher and learner. They

also offer solutions to ease the use of this method which seems somewhat paradoxical since, in spite of its obvious advantages and use of current technology, teachers are reluctant to employ it, especially in the Romanian education system. If they do use it, there is a shortage of studies to illustrate this.

In the field of English for Specific Purposes, DDL can be a reliable and useful teaching method,

especially if the learners are particularly motivated to understand the subtleties of the specialized language, be it technical, scientific, legal, medical or otherwise. This is one important aspect that can make the method successful or not: the motivation of the learners, especially if corpora is introduced to them directly, which may be baffling. Tim Johns himself admits this methodology was designed for a particular type of student: “(adult: well motivated: a sophisticated learner with experience of research methods in his subject area) with particular needs (fairly closely specifiable in terms of target texts) in a particular learning/teaching situation (in which a great deal of emphasis is placed on developing students’ learning strategies and on their responsibility for their own learning)” (Johns, 1986 : 161). Given this, Johns expressed his concerns in 1986 that time would tell whether the method worked for other types of learners.

Motivation comes from need and often our students are not aware of their needs. In my experience teaching specialized English to science students at university (in specialties such as Biology, Ecology, Agriculture or Horticulture), they are not always highly motivated to learn English. While most of them acknowledge the importance of being able to speak English, others dismiss it entirely, even though the course is designed using corpora and authentic information specific to their respective domains, and based on needs analysis and the range of possible jobs specific to the field of activity they train for. They doubt whether they will be able to find these jobs at the end of school and question how relevant scientific English will be for their future needs. Also, some are not aware that scientific research nowadays is only published in English. They may simply be uninterested in research, for which English is an indispensable tool.

However, motivation, albeit in small amounts and adapted to each group, may be increased by using user-friendly software, by allowing learners to be involved in the decisions related to what to learn, according to their lacks and needs (Boulton, 2009 : 9), as this will help them see the relevance of their work. They should also be encouraged to search on their own, to look for answers to their own inquiries, something that, again, many students are reluctant to do and prefer to be given the information directly by the teacher (Boulton, 2009 : 6, 10).

Apart from students’ motivation, another important element that may deter many teachers from using this method is the lack of equipment. In the Romanian academic context, faculties within a university usually outsource the English classes to a specialized department, independent or within a philology faculty, which provides English teachers

for technical, medical, scientific, business and other specialties. These faculties do not usually offer a classroom equipped with the necessary devices, computers in this case, with appropriate software, or even connection to the internet. Busy timetables and a shortage of space usually means the English class will be provided with a regular or specialized classroom (laboratory), whatever is available in that particular time interval.

Apart from the likely lack of equipment, there is also the question of time. Most universities allot a two-hour English course per week, sometimes even less, for two years (normally) or one year. Various objective or subjective factors may contribute to this situation such as curriculum design, ministerial impositions, and other administrative issues, and sometimes these numbers fluctuate from one year to the next depending on the number of students that enlist each year and the factors already mentioned above. DDL requires time, first to familiarize oneself to the method itself, to the use of corpora and concordance, to the independence that some learners do not prefer and then to the actual work and inquiries one must accomplish (Marzá, 2014 : 130-131). However, once the initial problems are overcome (learning how to use the software does not take long, certainly no longer than one sitting), time is no longer an issue and concord activities can proceed efficiently.

If time and space are not the issue, then the methodology is the one that makes the difference. As specified above, the method was thought to make learners language detectives and while that may work for students in humanities, ESP students are not linguists, nor are they very interested in the linguistic spectacle and diversity. Even so, they are not an amorphous mass. Some are interested in knowing more and question why their version for an exercise, for example, is not the right one, demanding more explanations. Others, on the contrary, are not interested at all and usually require simple solutions, one answer to one question, which may not always be possible. It depends on their mindset.

However, while the rule-based approach may be too abstract or not relevant to actual use of language and frequencies, with rules that are difficult to understand or remember and apply (Boulton, 2009 : 4-5), inductive learning may prove more useful in the long run, more motivating and relevant: “Detecting patterns and regularities also allow learners to realize that much of language use is highly fuzzy, with typical or frequent uses rather than rules and exceptions.” (Boulton, 2009 : 5). An important criticism to this method is related to this very issue: “One criticism that has been leveled against analysis based on concordance output is that the very methodology

itself, in the form of concordance and keyword searches, limits the analysis to a somewhat atomized, bottom-up type of investigation of the corpus data” (Flowerdew, 2005 : 324).

The fact is that learners are diverse and teachers learn from experience that no one method works every time with everybody. Some may find comfort in the knowledge of rules and in being told what to learn while others may appreciate the freedom and flexibility of DDL, the diversity of contexts and the authenticity of use provided by this method. This fact is proven by the previous studies that show encouraging results, but “not world-shattering” (Boulton, 2009 : 6) as learners have varying reactions to this method. Statistics also show that while learners are generally enthusiastic about DDL, there is always some degree of negativity towards the method which refers to the mechanical or tedious activities, and the overwhelming corpora (Boulton, 2009 : 8-9).

The data to be used is also part of the challenges this methodology has to face. The debate usually includes topics such as how much corpora, what kind of corpora to offer the learners in order to maintain a balanced approach and avoid overloading or discouraging them (Johns, 1990 : 294), as well as why use corpora and not stick to the invented but simplified contexts. There are, of course, opinions for and against, but regardless of criticism, any teacher that wants to employ this method will use experience and knowledge of their own students to settle these issues. Thus, by constantly assessing learners, methodology will adapt to each group. It is what most teachers normally do, as practice and results demand it.

In terms of what data to use, the resource most utilized by virtually anybody in this day and age is the internet, the World Wide Web, no matter how vilified it may be by serious researchers who do not trust the free and unchecked information that floats in cyberspace. Whether to use the internet as a learning tool or not is highly debated in didactic circles, as Alex Boulton compiles such contradictory opinions in his 2012 article. His conclusion is that, in spite of academics’ rejection of the internet as an unreliable source with too many unknowns or variables, the undisputed reality is that learners do use it, to an ever increasing rate: “The main point here is that if even linguists can overcome qualms about using web data, then it would seem unreasonable to prevent others from using it, especially perhaps language learners who do not need to be as scrupulous in their requirements as researchers: the decision should be *pedagogically* driven rather than based on non-pertinent *research* criteria.” (Boulton, 2012 : 24). In fact, not just learners use it, but academics as well, though still not an easy thing to admit. Since knowledge is but a click

away, sitting in a pocket and waiting to be accessed from the ever present gadgets everyone possesses, nobody can stop learners from using this resource and it should be acknowledged as such. It may be unchecked and unreliable from an academic point of view, but it is what most people use whenever they need an answer because it is the most easily accessible. The internet itself is a corpus, for if one inputs a term into a browser, they will obtain a large number of websites with contexts in which that term is used, and that is exactly what corpus is. In fact, Boulton names Google a concordancer (Boulton, 2012 : 24).

For ESP classes, if adapted to the needs and requirements of the class (in terms of time, space, type of learners), DDL would be highly successful. Corpora is nowhere more important than in the teaching and learning of specialized language. Each domain has its own specificities and legal English, for example, is different from scientific English or business English or medical English and so on. My own experience with teaching scientific language has led me to characterize this type of language as accurate, simple, lacking ornaments and involving an obligatory quality, objectivity. Unlike other domains like law or business, where local culture, traditions and customs influence the language to a large extent and involve a lot of subjectivity, science is the same everywhere in the world. The language of science is universal: a researcher in China has to understand the same thing a researcher in Brazil does because science has to be reproducible, the experiments must be explained so that any scientist in the world can repeat them with the same results. This very simplicity and accuracy makes it one of the best candidates for the use of DDL.

Thus, given its major importance in ESP, proper corpora with a reasonable amount of authentic texts and a good concordancer may be of considerable use in ESP classes. Such tools are especially useful for teaching collocations specific to a particular domain. For example, a simple inputting of the word *environmental* into a concordancer (<http://lertextutor.ca/conc/eng/>) produces over 500 results (Figure 2). Further on, the same concordancer will produce a thousand results when asked to find contexts with words ending in the suffix *-al*, a derivation exercise stemming from the use of *environmental* previously. These results, however, are offered by an online concordancer with its own corpora, which range from Disney scripts to presidential speeches, therefore the sheer number of results for a single search is overwhelming and not very useful for class activities. By working with specific corpora, for example, a small number of authentic texts regarding ecology or environmental issues

introduced into a licensed software, the results will be fewer but more focused on the task at hand. Such useful online engines for concord and language patterns are: Word Smith Tools (<http://www.lexically.net/wordsmith/>), Concordance

(<http://www.concordancesoftware.co.uk/>), AntConc (<http://www.laurenceanthony.net/software/antconc/>), Sketch Engine (<https://www.sketchengine.co.uk/>) and others.

531 hits Standardized to 88 per million (hits/corpus size x 1,000,000)
 Click any **KEYWORD** for more context

001. he results of their analysis strengthens their belief that " **ENVIRONMENTAL** scarcity does appear to be related to a range

002. ks. The arguments about the importance of architecture and " **ENVIRONMENTAL** design" in preventing crime that have develop

003. s. They received £15 million from Operation Clean - Up (an **ENVIRONMENTAL** improvement project) and about £70 million fr

004. ely to be performed for a longer period of time (**duration**). **ENVIRONMENTAL** events which are pleasurable will, if they con

005. . A Variety of Issues The variety of record categories (e.g **ENVIRONMENTAL** records, EPOS data, health records) is quite 1

006. ental monitors. The Wireless Integrated MicroSystems (**WIMS**) **ENVIRONMENTAL** monitor testbed (EMT) is a multicomponent micr

007. mentally sensitive or inaccessible (see also Stewart 1987). **ENVIRONMENTAL** monitoring of sites for the disposal of toxic

008. diverse as business - development initiatives (DoE, 1988b) **ENVIRONMENTAL** improvements (JURUE, 1986b) and some employem

009. onentially as genetic engineering develops (section 8.4.3). **ENVIRONMENTAL** BIOTECHNOLOGY II: OTHER APPLICATIONS There is

010. idual species response. Ecological processes and **Aboriginal** **ENVIRONMENTAL** interactions from the Keep River region of the

011. ecosystem benefits, services, and resources. Concern **about** **ENVIRONMENTAL** quality and the long-term livability of urban

012. form a template or schema with which the information **about** **ENVIRONMENTAL** events and self is processed. If subsequently

013. of 1060 million) such developments have also brought **about** **ENVIRONMENTAL** degradation. For the purpose of discussion the

014. ed with an increased demand for fuelwood are bringing **about** **ENVIRONMENTAL** change at an unprecedented rate. As Ives (1987

015. g events, such as the Winter Olympics, can also bring **about** **ENVIRONMENTAL** change which is not just a consequence of the

016. nd biogeochemical cycles which, when disrupted, bring **about** **ENVIRONMENTAL** degradation. In the developing nations, which

017. nd the attendant services that they require can bring **about** **ENVIRONMENTAL** change. The very landscape that creates the ma

018. ology and uncertainty, as well as CEO need for **achievement**, **ENVIRONMENTAL** constraints and selected Chinese characteri

019. dock closures inevitably led to the visible scars of **acute** **ENVIRONMENTAL** deterioration and abject physical dereliction.

020. le - class households to live in east Glasgow. In **addition**, **ENVIRONMENTAL** standards have improved - sometimes dramatical

021. 1990 and £13m. in the two succeeding years for **additional** **ENVIRONMENTAL** research. Perhaps the most significant develop

022. Such developments in South America are not without **adverse** **ENVIRONMENTAL** consequences, though they are more limited tha

023. 6.4 per cent of export earnings and while there are **adverse** **ENVIRONMENTAL** effects of this development, as will be discus

024. socio - economic status, poor quality of parenting, **adverse** **ENVIRONMENTAL** conditions have all been associated with otiti

Figure 2

There are numerous possibilities of using corpora in the classroom, but either one of two methods is usually employed (Cortes, 2013 : 1), the soft one or the hard one. The teacher either uses corpora to design materials or introduces corpora directly to learners in order to teach them how to use this resource and develop skills for further learning. Each of these methods has its own advantages and disadvantages.

Producing endless results with endless contexts for various inquiries is indeed a tedious activity, preferred for individual work and relevant to those who are searching for a particular answer. As classroom activity in the current context of many schools and academic programs, the most efficient solution is the soft method. Thus, the teacher has access to the corpora, uses the concordancer, devises tasks relevant to the specific group of learners and prints them as handouts to be used in class (Gabrielatos, 2005). The alternative, as mentioned previously in this article, would be to expose the learners directly to the corpora and the software, provided they have the skills to tackle this task. Dedicating the entire class to working with the hard version of this method is also probably not a feasible solution. The actual work with a concordancer may be left as homework or as solution for individual inquiries for further clarification of language items. In introducing these two approaches to DDL, Costas Gabrielatos distinguishes between “*text-based* and *corpus-based* approaches to data-driven learning” as “a data-driven, awareness-raising approach is not necessarily linked to the use of corpora”, meaning “teachers can use texts containing the target language features and,

through awareness-raising tasks, guide learners to discover the behavior of lexical, grammatical or discourse elements” (Gabrielatos, 2005).

The specialized literature offers examples of DDL-based course design for various needs to teach specific aspects of vocabulary or grammar (Johns, 1990, 1991, 2000; Gabrielatos, 2005; Boulton, 2009; Kilgarriff et al, 2015; Al Saeed & Waly, 2009; Aston, Bernardini and Stewart, 2004; Gavioli, 2005; Guan, 2013; Marinov, 2013; Marzá, 2014). A possible example of ESP class could see activities starting from a small number of authentic specialized texts or fragments of text (text-based approach) which can be thoroughly exploited into emphasizing a myriad of issues, from irregular plurals (which occur often in scientific contexts as many specialized terms originate from Latin or Greek) to the use of prepositions and the change in meaning they incur (phrasal verbs), the difference between adjectives and adverbs, or word derivation. All these make up a lexicogrammatical approach. Each issue raised is better understood by using concord activities in order to let the learners observe the varying contexts in which they occur within corpora specific to the learners’ domain (science, law, business, medicine etc.). Arguably, the chosen texts may not offer enough instances of the problem in need of illustration, but the alternative is to use the classical method, texts written for pedagogical use, which are no longer authentic. A solution would be to further use an online concordancer with a wider corpora for additional examples and contexts like the ones given example in Figures 1 and 2.

By using DDL, it is also easy to observe certain language patterns which are specific to English in general (word order, sequence of tenses) and also to particular domains (collocations, words with different meanings according to field, and others), an approach acknowledged by Costas Gabrielatos as well: “Language learners in countries where the target language is not widely spoken often lack opportunities for the rich

language exposure that is essential for developing the ability to recognize patterns. [...] Representative corpora can offer condensed exposure to language patterns.” (Gabrielatos, 2005). A well-designed course will motivate learners and help them experiment and better understand the importance of lexico-grammatical patternings.

Conclusion

The ideal factors required for DDL to work as intended are: highly motivated learners aware of their needs, access to computers equipped with licensed software and internet connection, and a rich corpus of texts specific to the learners' domain. This is true both for EFL and ESP. However, in order for this method to work and have positive results in terms of knowledge acquired, but also of learner satisfaction, DDL can be adapted to the classroom conditions by keeping some elements and forgoing others, depending on the actual conditions and on the type of learners. DDL is also particularly adaptable to ESP and vice-versa since they share certain common features: neither rely entirely on a textbook, they are both innovative and rather new in the context of English teaching, and both examine data and make use of corpora. The method is indeed challenging both for teacher and learners, but, if adapted properly, it can surely yield the best results. And one of the most important positive aspects of the method is that it teaches skills for constant learning.

Bibliography

- [1] Al Saeed, N. and Waly, S. (2009). “Corpus for Classrooms: Ideas for Material Design”, Proceedings of the 10th METU ELT Convention.
- [2] Aston, G., Bernardini, S. and Stewart, D. (2004). (eds.) *Corpora and Language Learners*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- [3] Boulton, A. (2009). “Data-driven learning: reasonable fears and rational reassurance”, *Indian Journal of Applied Linguistics*, vol. 35, no. 1. Available online: https://hal.archives-ouvertes.fr/file/index/docid/326990/filename/2009_boulton_IJOAL_reassurance.pdf
- [4] Boulton, A. (2012). “What data for data-driven learning?”, *The EUROCALL Review. Proceedings of the EUROCALL 2011 Conference*, Vol. 20, pp. 23-27.
- [5] Cortes, V. (2013). “Corpora in the Teaching of Language for Specific Purposes”, Chapelle C.A. (ed.) *The Encyclopedia of Applied Linguistics*, Blackwell Publishing Ltd. Published, pp 1-7.
- [6] Flowerdew, L. (2005). “An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies”, *English for Specific Purposes*, 24, pp. 321–332.
- [7] Gabrielatos, C. (2005). “Corpora and language teaching: Just a fling or wedding bells?”, *Teaching English as a Second Language – Electronic Journal*, 8/4. Available online: <http://tesl-ej.org/ej32/a1.html>
- [8] Gavioli, L. (2005). *Exploring Corpora for ESP Learning*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- [9] Guan, X. (2013). “A Study on the Application of Data-driven Learning in Vocabulary Teaching and Learning in China’s EFL Class”. *Journal of Language Teaching and Research*, Vol. 4, No. 1, pp. 105-112.
- [10] Johns, T. (1986). “Micro-Concord: A language learner’s research tool”, *System*, vol. 14, no. 2, 151-162.
- [11] Johns, T. (1990). “From printout to handout: Grammar and vocabulary teaching in the context of data driven learning”, Odlin T. (ed.), *Perspectives on Pedagogical Grammar*. Cambridge University Press, pp 293-313.
- [12] Johns, T. (1991). “Should You Be Persuaded - Two Samples of Data-Driven Learning Materials”, Johns T. and King P. (eds.) *Classroom Concordancing*, ELR Journal Vol. 4, pp. 1-16. Scanned 2010, M. Scott
- [13] Johns, T. (1997). “Contexts: The background, development and trialling of a concordance-based CALL program”, Wichmann A., Fligelstone S., McEnery T. & Knowles G. (Eds.), *Teaching and language corpora*. New York: Addison Wesley Longman, pp. 100-115.
- [14] Johns, T. (2000). “Data-driven Learning: The Perpetual Challenge”, Kettemann B. and Marko G. (eds.) *Teaching and Learning by Doing Corpus Analysis: Proceedings of the Fourth International Conference on Teaching and Language Corpora*, Graz 19-24 July, Vol. 1, pp. 107-117.
- [15] Kilgarriff. A., Marcowitz, F., Smith, S., Thomas, J. (2015). “Corpora and Language Learning with the Sketch Engine and SKELL”. *Revue Française de Linguistique Appliquée* 20(1), pp. 61-80.

- [16] Marinov, S. (2013). “Training ESP Students in Corpus Use – Challenges of Using Corpus-Based Exercises with Students of Non-Philological Studies”, *Teaching English with Technology*, 13 (4), 49-76.
<http://www.tewtjournal.org>
- [17] Marzá, N. E. (2014). “A Practical Corpus-based Approach to Teaching English for Tourism”, *International Journal of Applied Linguistics & English Literature*. Vol. 3, No. 1, pp. 129-136.
- [18] McEnery, T. and Wilson, A. (2001). (eds.) *Corpus linguistics*. Edinburgh University Press.
- [19] Scott, M. (2005). “Corpus Linguistics and ESP – is there a link?”, Power Point presentation (at XIX Seminário Nacional de Inglês Instrumental), available at
http://webcache.googleusercontent.com/search?q=cache:zCM66d7Eb5QJ:www.lexically.net/downloads/corpus_linguistics/CL_ESP_PUC_2005.ppt+&cd=1&hl=ro&ct=clnk&gl=ro
- [20] Tribble, C. and Jones, G. (1991). *Concordances in the classroom*. London: Longman.
- [21] Widdowson, H. (1998). “Context, community and authentic language”. *TESOL Quarterly*, 32(4), 705–716.
- [22] Widdowson, H. (2000). “On the limitations of linguistics applied”, *Applied Linguistics*, 21/1, 3-25.