

(Lexico)grammar and Data-Driven Learning

Gaëtanelle Gilquin
UCLouvain

Abstract

Among the different aspects of language that data-driven learning (DDL) can be used for, this entry focuses on lexicogrammar and grammar. It starts by showing that DDL has been applied more often to lexicogrammar than to grammar. It then presents the ways in which each of them can be approached through DDL, with some illustrations taken from the literature. It appears that DDL approaches to lexicogrammar and grammar bear some similarities, including exploratory and corrective functions, and that they each have unique contributions to make to language pedagogy.

Keywords: Collocates, Concordances, Constructions, Corrective data-driven learning, Learner corpus, Patterns, POS tagging, Word sketches

Introduction

Data-driven learning (DDL) is a pedagogical approach that involves the use of corpus data in the language classroom. The term also refers to the research field that explores this pedagogical approach and seeks to assess its effectiveness. In this entry, DDL is discussed, both as a pedagogical approach and a research field, in relation to two aspects of language that DDL can be applied to, namely, lexicogrammar (which straddles lexis and grammar) and grammar.

Research has shown that DDL is especially useful to cover lexicogrammatical aspects of language. According to Boulton and Cobb's (2017) meta-analysis, lexicogrammar is the language aspect that is most often dealt with in DDL publications. Their meta-analysis also reveals that, together with vocabulary, lexicogrammar is the aspect that yields the best results, as measured by effect size. Mizumoto and Chujo (2015) reach similar conclusions in the context of English language teaching in Japan.

Grammar, on the other hand, appears less often in DDL studies. In Boulton and Cobb's (2017) meta-analysis, DDL publications with grammar as their object of study are less frequent than those dealing with lexicogrammar and vocabulary – but more frequent than those dealing with discourse. The effect sizes are also generally smaller than for lexicogrammar and vocabulary, which suggests a lower degree of effectiveness.

In what follows, the focus is first on DDL approaches to lexicogrammar and then on DDL approaches to grammar. Illustrations from the literature are provided and, when evidence is available, the effectiveness of DDL in the reviewed studies is commented on. Although lexicogrammar and grammar are discussed in different sections, it should be emphasized that they do not represent clearly distinct categories. As its name indicates, lexicogrammar includes some grammatical aspects, which establishes a continuum between the two (as well as between lexicogrammar and vocabulary).

DDL Approaches to Lexicogrammar

Pearce (2007, p. 109) points out that “[t]he advent of corpus linguistics has made the identification of lexicogrammatical patterns much easier than it once was”. It is therefore not surprising that DDL should be particularly suitable to deal with such aspects. DDL has a unique contribution to make in this respect, since speakers’ intuitions tend to be rather poor when it comes to lexicogrammatical patterns (Hunston, 2002, pp. 20-21) and traditional pedagogical resources like dictionaries, grammar books or textbooks may not be readily amenable to the teaching of such patterns (O’Sullivan & Chambers, 2006, p. 50).

In DDL approaches to lexicogrammar, the starting point is usually a single word, but unlike DDL approaches to vocabulary (see *Vocabulary and DDL*), the focus is “less on meaning than on how the target word fits into its surroundings” (Boulton & Cobb, 2017, p. 380). Charles (2015), for example, describes how students were invited to use the Collocate tool of AntConc (Anthony, 2011) in order to examine the collocates of the verbs *show* and *suggest*, that is, the words occurring regularly in their close textual environment. The students first looked for the presence of *that* to the right of the verbs, which helped them identify the VERB + *that* pattern. They then considered noun and pronoun collocates to the left of the verbs, in an attempt to describe the nature of their subjects – either personal (e.g. *I, we*) or impersonal (e.g. *evidence, results*).

A DDL activity can also start from a pattern and encourage students to learn more about its use in context. In Vyatkina (2016), the DDL intervention centred around certain verb-preposition combinations in German (e.g. *warten auf*, “wait for”). Students were given worksheets with concordances that had been extracted from the DWDS corpus (www.dwds.de) and that showed occurrences of the combinations in context. On that basis, the students were able to observe how these combinations were used and, more particularly, what grammatical cases they took (e.g. accusative in the case of *warten auf*) and then to produce their own example sentences.

Interestingly, certain corpus tools give direct access to lexicogrammatical patterns. This is the case of Sketch Engine and its “word sketches” (Kilgariff et al., 2010), which show lists of collocates organized according to their grammatical relationship with the search item. Figure 1 reproduces part of the word sketch of the noun *research* in the British Academic Written English (BAWE) corpus. It includes modifiers of *research* (e.g. *qualitative, recent*), nouns modified by *research* (e.g. *question, paper*), verbs used with *research* as a direct object (e.g. *conduct, undertake*) or as a subject (e.g. *show, indicate*), nouns followed by the preposition *of* and *research* (e.g. *piece, field*) and nouns coordinated with *research* (e.g. *development, theory*). Thomas (2015) refers to such information as “extended collocations”. He explains how students can group them into semantic groups so as to derive “word templates”, such as the following for the noun *allegation*: “[an injured party] makes an allegation about [a negative abstraction (potentially criminal, e.g. corruption) or a public figure]” (Thomas, 2015, p. 97). Thanks to these templates and through a process of “guided discovery learning” (Thomas, 2015, p. 102), students are expected to acquire productive knowledge of the lexicogrammatical uses of the target words.

modifier	modifies	object_of	subject_of	pp_obj_of	and/or
qualitative ... qualitative research	question ... research question	conduct ... conducting research	show ... research has shown	piece ... piece of research	development ... research and development
far ... further research	methods ... al . Research Methods in Politics	carry ... research carried	suggest ... research suggests that	lot ... a lot of research	extension ... benefit from formal research and extension
recent ... recent research	paper ... research paper	undertake ... research undertaken	indicate ... research indicates	purpose ... the purpose of the research	theory ... theory and research
market ... market research	study ... Home Office Research Study	need ... research is needed to	find ... research found that	method ... methods of research	study ... research and study
future ... future research	design ... research design	base ... based research	focus ... research focuses	basis ... the basis of further research	excavation ... excavation and research
quantitative ... of quantitative research	finding ... the research findings	focus ... research focussed on	reveal ... research revealed	field ... field of research	department ... the research and development department
previous ... previous research	project ... research project	complete ... research completed	aim ... this research aims to	area ... area of research	decade ... decades , research
empirical ... empirical research	agenda ... research agenda	review ... Evidence based research is critically reviewed to allow the	investigate ... research investigated	journal ... Underdevelopment " , Journal of Peace Research , Vol.	experience ... research and experience
scientific ... scientific research	method ... research method	publish ... research has been published	demonstrate ... research demonstrates	survey ... A Critical Survey of Recent Research in Chinese Economic	improvement ... research and improvement
office ... Home Office Research Study	centre ... research centres	require ... research is required	identify ... Research has identified that	majority ... the majority of research	literature ... literature and research
multi-strategy ... a multi-strategy research	laboratory ... research laboratories	begin ... began their research	centre ... research has centred	regulation ... The Regulation of Stem cell Research in the UK	research ... unlike quantitative research , qualitative research does not have
home ... Home Office Research Study	institute ... Research Institute	perform ... research performed by the	conduct ... research conducted	aim ... aim of this research	question ... questions and previous research

Figure 1. Word sketch of the noun *research* extracted by means of Sketch Engine from the BAWE corpus

The above examples are instances of what Kennedy and Miceli (2017) call “pattern hunting”, which involves letting students explore a corpus to find language patterns and information about them. DDL can also serve the purpose of “pattern refining”, which provides students with “ways of editing their text for lexico-grammatical accuracy” (Kennedy & Miceli, 2017, p. 94). Li (2023) illustrates pattern refining, with students using the Corpus of Contemporary American English (COCA) to correct the collocation errors underlined in their essays. The study demonstrates the superiority of the DDL approach over the use of a dictionary and indicates a sustained learning effect.

DDL Approaches to Grammar

The relatively low number of DDL studies that have dealt with grammar may be due to the impression that DDL is “difficult to reconcile with the ‘big themes’ of language teaching, such as ‘tenses’ or ‘articles’” (Hunston, 2002, p. 184). Misguided perceptions may also explain some of the less positive evaluations of DDL approaches to grammar. In Yoon and Hirvela (2004), students having experienced DDL evaluated it as somewhat less helpful for grammar than for most other aspects (usage of vocabulary/phrases, writing skills, academic writing ability, etc.). This, however, is arguably linked to the fact that “grammar as presented through a corpus orientation did not conform to their notion of learning about grammar” (Yoon & Hirvela, 2004, p. 269).

DDL may also seem less crucial for grammar than for lexicogrammar because many pedagogical resources include materials for grammar teaching. However, it should be emphasized that DDL approaches to grammar are quite different from traditional approaches. Traditionally, grammar is taught by means of abstract rules that learners have to study. The problem is that learners may master the rules without being able to apply them (Gaskell & Cobb, 2004). DDL, by contrast, encourages learners to observe the grammatical regularities

found in corpus data, which makes them more active in the process and hence more likely to remember and reproduce the discovered patterns. In Boulton (2007), students proved able to detect patterns on the basis of concordances from the British National Corpus (BNC) illustrating the future uses of *will* and *going to* and to apply these patterns to new contexts. While the results were similar to those of students using traditional grammatical rules, Boulton (2007, p. 10) argues that they would have been better with “more appropriate corpora, more carefully selected concordances, and especially learner training in interpreting the concordances”.

As is the case with lexicogrammar, DDL can also start from a pattern, for example, a grammatical construction whose features are examined in the context of corpus data. In Gilquin (2021), three constructions were the focus of a DDL intervention: the *MAKE* causative construction (e.g. *She tried to make him look at her*), the *way* construction (e.g. *She elbowed her way through the crowd*) and the *into* causative construction (e.g. *He had blackmailed her into marrying him*). Students received a printed handout with concordance lines from the BNC in which the constructions had been highlighted. They were also given a worksheet with some tasks to carry out with the help of the concordance, for example, determining the syntactic category of the infinitive in instances of the *MAKE* causative construction (transitive, intransitive or copular) and identifying the prepositional phrase in instances of the *way* construction. As a reflection of the overlap between grammar and lexicogrammar, the students also had to discover the words typically occurring in some slots of the constructions (e.g. the main verb in the *into* causative construction).

Certain types of corpora are particularly useful for DDL activities centring around grammatical aspects. This includes part-of-speech (POS) tagged corpora, in which each word has been assigned a label indicating its word class. Friginal (2018, p. 32), for instance, explains how students can be asked to define a noun on the basis of the POS tags of the words found before and after nouns in a POS tagged corpus (e.g. preceded by an article, followed by another noun). In Hegelheimer (2007), it is a different type of corpus that is used to help students improve their grammatical accuracy, namely, an error-tagged learner corpus, that is, a corpus of texts produced by language learners in which errors have been identified and corrected. Hegelheimer (2007) describes how, through the iWRITE program, students can select an error type, such as tense inconsistency, and see examples of this error type from the corpus as well as corrected versions.

It has been suggested that DDL approaches to grammar may be less suited to “present sequences of new structures that learners are hypothesized to be ready for” than to “shed light on structures they have already shown a need for” (Gaskell & Cobb, 2004, p. 303). This points to the relevance of the corrective function of DDL for grammar, that is, the use of corpus data to make one’s texts grammatically more accurate. In O’Sullivan and Chambers (2006), learners of French were asked to correct and improve their essays with the help of a French corpus and WordSmith Tools (Scott, 1996). The analysis of the changes made after consulting the corpus revealed that, of all errors found in the essays, it was the category of grammatical errors that led to the highest number of revisions.

Conclusion

DDL has gained respectability as a pedagogical approach to lexicogrammar, which it is ideally suited to deal with, since corpus techniques help make lexicogrammatical patterns more

noticeable. While DDL has been less often used to cover grammatical aspects of language, it has been shown to be suitable for this purpose too.

DDL approaches to grammar and lexicogrammar actually bear some similarities. They can both involve exploring the textual environment of a linguistic item or taking a whole pattern as a starting point. They can also fulfil a corrective function, when learners use a corpus as a source of information to correct and improve their texts. In addition, both have unique contributions to make to language pedagogy: DDL approaches to lexicogrammar provide insights into an aspect of language whose treatment in traditional pedagogical materials tends to leave much to be desired, while DDL approaches to grammar give learners an active role of pattern discoverers rather than rule memorizers.

Although both types of approaches inherit the challenges associated with DDL, they also offer considerable opportunities (see *Future Challenges and Opportunities for the Field of DDL*), which teachers and students alike should seize to help with the development of (lexico)grammatical competence.

References

- Anthony, L. (2011). AntConc 3.2.4 [Computer Software]. Waseda University. <https://www.laurenceanthony.net/software>
- Boulton, A. (2007). 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 2007*. University of Birmingham. https://ucrel.lancs.ac.uk/publications/cl2007/paper/126_Paper.pdf
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393. <https://doi.org/10.1111/lang.12224>
- Charles, M. (2015). Same task, different corpus: The role of personal corpora in EAP classes. In A. Leńko-Szymańska, & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 131–154). John Benjamins. <https://doi.org/10.1075/scl.69.07cha>
- Friginal, E. (2018). *Corpus linguistics for English teachers: New tools, online resources, and classroom activities*. Routledge.
- Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32(3), 301–319. <https://doi.org/10.1016/j.system.2004.04.001>
- Gilquin, G. (2021). Using corpora to foster L2 construction learning: A data-driven learning experiment. *International Journal of Applied Linguistics*, 31(2), 229–247. <https://doi.org/10.1111/ijal.12317>
- Hegelheimer, V. (2007). Helping ESL writers through a multimodal, corpus-based, online grammar resource. *CALICO Journal*, 24(1), 5–32. <https://doi.org/10.1558/cj.v24i1.5-31>
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.
- Kennedy, C., & Miceli, T. (2017). Cultivating effective corpus use by language learners. *Computer Assisted Language Learning*, 30(1-2), 91–114. <https://doi.org/10.1080/09588221.2016.1264427>
- Kilgarrieff, A., Kovář, V., Krek, S., Srdanović, I., & Tiberius, C. (2010). A quantitative evaluation of word sketches. In A. Dykstra, & T. Schoonheim (Eds.), *Proceedings of the 14th EURALEX international congress* (pp. 372–379). Fryske Akademy.

- Li, L. X. (2023). Promoting accuracy of collocation use in L2 writing: The role of data-driven learning in indirect corrective feedback. *Computer Assisted Language Learning*, Online first. <https://doi.org/10.1080/09588221.2023.2292554>
- Mizumoto, A., & Chujo, K. (2015). A meta-analysis of data-driven learning approach in the Japanese EFL classroom. *English Corpus Studies*, 22, 1–18. https://jaecs.com/jnl/jnl_22.pdf
- O'Sullivan, Í., & Chambers, A. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, 15(1), 49–68. <https://doi.org/10.1016/j.jslw.2006.01.002>
- Pearce, M. (2007). *The Routledge dictionary of English language studies*. Routledge.
- Scott, M. (1996). WordSmith Tools [Computer Software]. Oxford University Press. <https://lexically.net/wordsmith>
- Thomas, J. (2015). Stealing a march on collocation: Deriving extended collocations from full text for student analysis and synthesis. In A. Leńko-Szymańska, & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 85–108). John Benjamins. <https://doi.org/10.1075/scl.69.05tho>
- Vyatkina, N. (2016). Data-driven learning for beginners: The case of German verb-preposition collocations. *ReCALL*, 28(2), 207–226. <https://doi.org/10.1017/S0958344015000269>
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13(4), 257–283. <https://doi.org/10.1016/j.jslw.2004.06.002>