
Much research in corpus-based language teaching and data-driven learning (DDL) has focused on the affordances of concordancing for language learning, albeit with a primary focus on English as the target language of investigation, corpus data unsuitable for classroom use, and an almost singular focus on writing. Moreover, as the functionality of corpus tools has progressed greatly over the past decade, the reliance on concordancing as the sole pedagogical approach available in DDL has given way to multiple textual, statistical and visual pathways into corpus data. Broadening the horizons of the field while harnessing the new options now available to us as DDL researchers is crucial if DDL is to continue to gain acceptance as a viable pedagogical approach in the language classroom and beyond. It is with great excitement, therefore, that the collated works of selected scholars featured in the *Teaching and Language Corpora (TaLC) 2018* conference are now available in book format, under the title *Beyond concordance lines*.

As Pascual Pérez-Paredes and Geraldine Mark point out in their editorial *Introduction*, “research in practice is at the heart of each chapter in this volume” (p. 6). In pursuit of this aim, the volume is divided into three parts, namely:

- Insights into the links between research and classroom uses of corpora
- Learner language research and its implications for teaching
- Applications of corpus use and DDL in diverse world contexts.

The editors note that what unifies the contributions is a “bird’s eye perspectives” (p. 2) on DDL, in terms of reflection on past practice, identifying strengths and best practice, while making predictions towards future advancement of the field. The purpose of the volume is therefore to provide a clearer understanding of what goes on before, during and following DDL, looking beyond concordancing in its current form and typical contexts to narrow the gap between language learning research and pedagogical practice.

Chapter 1 by Alex Boulton is an update of Boulton’s prior syntheses of DDL research (e.g. Boulton and Cobb 2014; Vyatkina and Boulton 2017), taking us up to 351 published studies up to and including 2018 (while also including a few in 2019 in the discussion). The difference between Boulton’s review and previous work is to focus on the evolution of the field over time while pointing out areas of future intervention. Boulton begins his review by challenging prior assumptions that concordancing is the sole means by which DDL takes place, pointing out that the boundaries of what constitutes DDL are, in fact, fuzzy. Next, Boulton rejects prior claims that DDL lacks empirical and theoretical support, pointing out a number of critical studies that have addressed either, or both. A survey-of-surveys of DDL is then
conducted, with Boulton critiquing both his and others’ approaches to such surveys while highlighting their relative contributions. The actual survey presented in the chapter points out a large increase in studies over the last decade, the most popular publication sources, the increase in other languages and language learning contexts studied (despite the continued dominance of English) and the diversity of learners and corpus tools and study designs. While the results are largely in line with those found in surveys and meta-analyses of other applied linguistics subfields (e.g., Hyland and Jiang 2021 for English for Academic Purposes (EAP); Crosthwaite et al. 2022 for corpus linguistics), the overall finding that DDL as a research area has come a long way in overcoming previous criticisms is very welcome. That said, Boulton laments a “lack of imagination” (p. 31) of DDL researchers in not going beyond well-trodden paths.

Chapter 2 by Anne O’Keefe continues her explorations into the intersection(s) between DDL pedagogy and theories of language learning. Echoing the comment of Boulton in the previous chapter, the lack of engagement with theoretical concerns in much DDL research is noted as “to the detriment of embedding DDL as a more mainstream pedagogy” (p. 36). In response, O’Keefe’s chapter suggests a link between DDL, usage-based language learning, and sociocultural theory, with the former embedding DDL into our present understanding of second language acquisition theory (in relation to the importance of concepts including frequency, constructions and chunks on SLA), and the latter focusing on how DDL may be enacted through peer- and teacher-scaffolded classroom activity. Particularly insightful is the statement that “we know a lot about the D for Data in DDL, but we need to think much more about the L for Learning” (p. 52). O’Keefe’s call for further experimentation in this area is particularly welcome.

Chapter 3 by Michael McCarthy, Tony McEnery, Geraldine Mark and Pascual Pérez-Paredes is a summary of conversations between these researchers looking back on 25 years of TaLC, presented in interview format. The topics covered include direct and indirect use of linguistic knowledge in the language classroom, the relative importance of linguistic terminology (e.g. collocation) for DDL, how learners’ implicit linguistic experience impacts doing DDL in practice, changes to learning contexts and methodologies over time, the role of frequency as a starting point for DDL discussions, and the challenges “of looking beyond the immediately discoverable” (p. 69), e.g., DDL and pragmatics. I actually found this interview format incredibly refreshing in its accessibility, and encourage more of this practice in edited volumes.

Chapter 4 by Hildegunn Dirdal kicks off the section on learner language research with a longitudinal study exploring L2 development of English -ing clauses by L1 Norwegian learners. Dirdal tracks the production of five secondary school students learning L2 English over four years, between the ages of 13 and 17 years. Data includes narrative and non-narrative texts taken at an average three data points per year, with analysis tracking the function of each -ing clause and its internal syntax development over the study period. The findings suggest a developmental pattern of -ing clauses from initial use with aspectual verbs only (e.g. start crying), to complements of prepositions (e.g. tired of doing the same thing…), post-modifiers of nouns (e.g. this is the pilot speaking) to subject clauses (e.g. Making friends is difficult). Individual differences across the five learners are also presented. The implications for teaching are that teachers need to be aware of the functional and syntactic complexity surrounding English -ing clauses, and that individual differences in acquisition may result in different developmental paths.

Chapter 5 by Rolf Kreyer analyses verb-noun collocations with do, give, make and take in another longitudinal learner corpus study, this time of L2 English writing produced by 83 secondary school age L1 German speakers taken from the Marburg corpus of Intermediate Learner English (MILE). Focusing on the productive appropriate and erroneous use of these collocates, the study is particularly interesting for its focus on low/intermediate level L2 learners between A2/B2 CEFR. The findings suggest an overall decrease in the use of these particular collocates over time as their vocabulary increases, variation in the per cent of inappropriate attempts to produce individual collocates modified by L1 influence, and the persistence of certain erroneous collocates over time. The study should be commended for being one of the largest ‘long’ longitudinal corpus studies, despite the author’s modest claim that it is still “too small” (p. 116).

Chapter 6 by Martin Weissner presents a proof-of-concept study into analysing spoken learner language through pragmatically and error annotated corpora, using the Dialogue Annotation and
Research Tool (DART). Exploring the data of 20 L1 Chinese/L2 English interviews from the LINDSEI corpus at the speech act level, the results suggest that errors in interlanguage production may affect the coherence of the dialogue produced, while the L2 learners’ use of emphatic discourse markers is not reflective of genuine attempts at interaction. Having more learner corpora annotated for error and pragmatic features is always a good thing, and I certainly agree that the approach taken in this chapter represents a “more communicative perspective” (p. 144) than more traditional error analyses allow.

Chapter 7 by Gregory Hadley and Hiromi Hadley explores the impact of DDL on extensive reading, in a rare study of the affordances of DDL on receptive skills. This study followed 71 university level L2 learners of English from a range of L1 backgrounds (including Japanese, Chinese, French and more) through a sixteen-week applied English course where DDL tools were used both in and out of class on a corpus of graded reading materials across seven proficiency levels. The findings suggest that students using DDL read faster and read more than learners studying through extended reading alone. The authors are to be commended for their detail in explaining the research design, the pedagogical approaches taken in class, and particularly the difficulties they encountered along the way. This level of honesty and reflection on DDL is refreshing and very welcome.

Chapter 8 by Tanjun Liu tests the affordances of #Lancsbox for academic collocation learning (using the BAWE corpus), comparing its use against an online collocations dictionary with English major undergraduates in China. Students took a pre- and post-test of receptive and productive collocation knowledge before taking two hours of training on #Lancsbox or the dictionary, followed by four practice sessions involving written argumentative essays. The findings suggest some degree of growth of collocation knowledge in the DDL group, albeit there was no statistical significance in this growth pre- to post-test, nor between DDL and dictionary groups. In qualitative comments, the learners outlined differences in what they would use either tool for, e.g., DDL is useful for ‘collocation learning’, the dictionary is useful for ‘writing’. This detailed information is arguably more important for teachers than the comparison of the two tools itself. The study is also notable as a relatively rare case where students were found to use DDL outside of the classroom.

Chapter 9 by Clare Wolfarth, Claude Ponton and Catherine Brissaud presents Scoledit, an NLP tool used to analyse a longitudinal corpus of L1 French learner writing by younger learners aged between 6 and 11 years. The authors outline the challenges involved in building a corpus composed of original handwritten texts (which anyone with young children can attest are very hard to read!) before describing the corpus itself, the NLP tools used to annotate it, and the annotated error categories including spelling and verb forms/morphology. A range of teaching recommendations are provided for verb tense issues and morphology/word segmentation. The study is one of the few longitudinal corpus studies on younger learners involving languages other than English and represents a useful addition to the field.

Chapter 10 by Yukio Tono explores the creation of a range of multilingual resources and e-learning systems for 28 languages involving functional descriptors aligned to the CEFR. Following a description of the CEFR-J project where Japanese language resources have been created to help with English language teaching in Japan, the chapter then outlines how the use of machine translation and multilingual corpora can help convert these resources into a range of other languages. The resources presented are an absolute goldmine for language teachers and learners regardless of knowledge of Japanese, and the presented links between these resources and their applications for teaching materials and task design are at the cutting edge of what is possible in DDL ‘beyond’ the analysis of concordance lines.

Overall, the volume is a very strong contribution to the field of corpora for language learning, and the editors and contributors are to be commended for broadening the scope of corpus and DDL research for education, while also pointing out key areas for future research. As a reviewer, I have only two relatively minor concerns with the volume. First, a number of the chapters have already been superseded by later publications by the same authors (e.g. Boulton and Vyatkina 2021; O’Keeffe 2021). Granted, edited volumes take longer to produce, and COVID will have certainly played a part in the delay to this volumes’ publication following the 2018 conference. Second (and this is, I feel, an issue with the TaLC format itself, rather than the editors here) is that certain chapters are less teaching-focused, and despite the best efforts of the authors to tie in their findings with suggestions for pedagogy, there is sometimes little
presented beyond the typical ‘the findings will help raise awareness of [linguistic feature X]’ that we often see in published education-focused CL research. That said, I greatly enjoyed reading this accessible, polished volume on corpora for language education, and strongly believe the book would make excellent recommended reading on CL course reading lists.

References


