33

TRAINING TEACHERS AND LEARNERS TO USE CORPORA

Agnieszka Leński-Szymańska

33.1 Introduction

As the present volume illustrates, corpus data are used in language education in a variety of ways and they are implemented by different parties involved in teaching and learning. Römer (2011) provided a useful review of the areas in which corpora are being exploited in language pedagogy. On the one hand, large text collections are very helpful to researchers and materials writers. They enable them to observe interesting but little-known facts about target language use, which later find their way into syllabi, reference works (dictionaries and grammars), and teaching materials (coursebooks and supplementary resource books). Corpora are also a limitless source of authentic citations featured in pedagogical resources. Römer (2011) labels these uses of corpora indirect applications. On the other hand, text collections can be consulted directly by language teachers and learners without the mediation of a researcher. This can take the form of students accessing a corpus or working with corpus data pre-selected and pre-processed by their teachers. Such interactions make it possible for teachers to prepare tasks and materials tailored precisely to their students’ specific needs, and for learners to study the language relevant to their interests. Direct corpus applications can also provide students with opportunities to become researchers themselves who are capable of observing, analysing, and drawing conclusions from examples of real language use. This mode of learning promotes a pro-active attitude, autonomy, and language awareness (Römer, 2011). The author calls such direct consultations of corpora by language teachers and students data-driven learning (DDL), echoing the term coined by Tim Johns (1991), which was originally reserved only for cases of students’ inductive and hands-on experience with text collections (Karlsen & Monsen, 2020).

In the last two decades, great progress has been made in the development of indirect applications of corpora. However, despite its advantages, as asserted by enthusiasts and confirmed by empirical evidence (Boulton & Cobb, 2017), data-driven learning has been scarcely implemented in language classrooms, a fact that is frequently lamented in the literature (cf. a recent review by Chambers, 2019). This scarcity has prompted researchers to, on the one hand, probe into the reasons for this lack of popularity, and, on the other hand, fervently advocate active promotion of direct applications of corpora in language pedagogy (Chambers, 2019; Römer, 2009). One important form of such promotion is effective teacher and student training in exploiting large text collections in order to support foreign language learning.
33.2 Core issues and topics in teacher and student training in DDL

33.2.1 Surveys of direct corpus applications

Over the last 20 years, a number of surveys have been conducted to query the implementation of corpora by language teachers. These surveys attempted to establish the extent to which language instructors make use of corpora. Some of these studies also attempted to establish the specific uses of corpora by educators in their practice as well as to uncover perceived obstacles in the application of corpora in language instruction.

One of the first such surveys was a questionnaire administered by Mukherjee (2004). He queried 248 English teachers in German secondary schools about their experience with large text collections. He discovered that almost 80% of them were unfamiliar with the term corpus linguistics and only 10% of the respondents had had some contact with corpus linguistics, although this did not necessarily entail having used corpus data in their teaching in any way. Almost 15 years later, the study was replicated in the same educational context by Callies (2019), who used a more detailed instrument to gather information on the instructors’ practices with corpora and their attitudes towards the benefits of corpus use for themselves and their students. Although only 26 teachers responded to his questionnaire, the results did demonstrate that almost half of the participants (42%) had not encountered the term corpus linguistics during their university studies and one quarter (23%) declared very little exposure to this methodology during their tertiary education. In addition, 80% of participants had no contact with corpora during their subsequent in-service teacher training. Nevertheless, a vast majority (70%) of the respondents declared having used corpora for various purposes, but frequently this use was very limited and their familiarity with corpus resources remained low. Among the respondents who acknowledged employing corpora in their teaching, the most frequent application was for reference (65%), that is, to look up individual words, phrases, or grammatical points. Fewer teachers used corpora for designing their own materials (46%), and even fewer encouraged their students to engage with corpus data (19%). Callies’ results, although not fully reliable due to the small sample size, illustrate that the direct applications of corpora in language instruction remain very scarce.

Tribble’s (2015) survey had a more international reach and included 560 respondents from 63 countries, with the largest numbers from the US, UK, and China. Three-quarters of respondents to his survey declared using corpora in their teaching. Although this figure appears encouraging, Tribble acknowledges that the channels selected for the distribution of the questionnaire might have resulted in a sample skewed towards corpus enthusiasts. It should also be noted that a large majority of Tribble’s respondents worked in a different context than the language instructors queried by Mukherjee (2004) and Callies (2019). Half of them were scholars (lecturers and researchers) and almost 80% taught languages at the tertiary level. The most frequent use of corpora reported in the questionnaire was the production of teaching materials in various forms: electronic (19%), paper-based (16%), and coursebooks (7%). Interestingly, the applications involving the instructor’s personal reference were almost equally frequent as students’ reference (both a little over 20%). However, corpus consultations in student-led DDL activities were rare (a little over 10%).

A particularly relevant aspect of Tribble’s study for the issues discussed in this chapter is the respondents’ answers concerning the reasons why some teachers did not exploit corpora in their practice. Two obstacles were most frequently selected among several explanations: a lack of time to develop corpus-based materials and a lack of knowledge of the potential of corpora
Training teachers and learners to use corpora

Because they don’t know how to use corpora in teaching
.... but don’t know how to put it to good use
... but we don’t know how to use them effectively

Figure 33.1 Selected concordances from the answers to an open question on the reasons for not using corpora in the classroom (Tribble, 2015, p. 56)

(both over 20%). A few concordance lines drawn from the answers to an open question in Figure 33.1 confirm the perceived lack of teachers’ competence related to the exploitation of corpora.

On the whole, Tribble’s results, although generally optimistic, reveal that, in spite of the increasing level of implementation of corpora in tertiary-level language education, instructors do not take advantage of the full potential of this medium for language learning. In addition, they believe that they do not have sufficient expertise in how to integrate corpus data into their practice.

No comparable wide-ranging surveys on corpus use have been conducted among language learners. If students are ever queried on their experience with corpora, this is usually in questionnaires distributed to control and experimental groups in studies into the effectiveness of different aspects of DDL. Such samples are usually small-scale and opportunistic, and not representative of larger student populations; nevertheless, their responses shed some light on learners’ familiarity with corpus resources and tools. For example, in a recent study set in the Iranian context (Saeedakhtar et al., 2020), almost none of 60 students declared previous contact with corpus resources. In a more comprehensive project, Karlsen and Monsen (2020) queried 154 Norwegian students in four secondary schools and taught by four teachers who had been trained in corpus linguistics during their university studies. In response to an open question about the digital resources consulted for language learning, none of the students mentioned a corpus-based tool or a text collection in any format. Only 6% of the students declared familiarity with the concept of a corpus and only two of them attempted to explain the notion. However, in a multiple-choice question, the respondents admitted having heard of the most popular corpus resources: BNC (15% of the students), COCA (11%), SKELL (8%), Just-the-word.com (6%), and AntConc (3%).

The responses to the surveys distributed among language instructors and students indicate that the information on corpora and their affordances in language education has had a limited reach. In addition, the respondents’ answers suggest that merely “spreading the word” about corpora, as recommended by Römer (2009, p. 84), may not be sufficient for their major uptake in education. Both teachers and learners need a systematic and dedicated training in DDL in order to take advantage of this medium in the classroom.

33.2.2 Skills needed for data-driven teaching and learning

Two types of skills are indispensable for language students to engage in data-driven learning: technical and corpus-analytical. First, it is necessary for learners to have technical skills of operating corpus tools efficiently, which involves being aware and capable of using different functionalities offered by language processing software – for example, wildcards,
Agnieszka Leńko-Szymańska

sorting concordance lines, or generating lists of collocates. Second, learners need to have acquired the skills of corpus analysis, even if only implicitly. This entails recognising the kind of information that can be searched in a corpus (e.g. contexts, patterns of use, variation), knowledge how to retrieve this information successfully (building effective queries), and an ability to interpret search results (by focusing on relevant information and identifying interesting patterns). The need for some of these skills can be reduced by teachers who pre-select appropriate citations from corpora and guide their students in the process of their analysis and interpretation. However, instructors have to be fully proficient in these competences, as well as in corpus-related pedagogical skills necessary for the creation of interesting, varied, relevant and instructionally-valid materials and tasks that are well integrated into lessons and syllabi (Callies, 2019; Leńko-Szymańska, 2017; Mukherjee, 2004). Both teachers and learners require assistance in developing these necessary sets of skills in courses, workshops, and seminars aimed at training teachers and students in DDL.

33.2.3 Models of teacher and student training

Promoting the potential of data-driven learning must begin with teachers since it is from their instructors that learners are most likely to hear of, get exposure to, and learn to manipulate and interpret corpus data (Mauranen, 2004). Therefore, effective teacher training is the key element in popularising direct uses of corpora in language education. In the last decade, several books for language educators have been published showcasing multiple uses of corpora in the classroom (e.g. Flowerdew, 2012; Friginal, 2018; Poole, 2018; Reppen, 2010). In addition, a number of teacher training courses have been described in the literature (e.g. Leńko-Szymańska, 2014, 2015, 2017; Zareva, 2017). These courses targeted either practicing instructors or teacher trainees. Training in-service and pre-service teachers differs due to the instructional setting. The former is usually only available for short intensive courses. Seminars designed for the latter group are often interwoven with other teacher training classes and can span over a longer period.

Three approaches to designing teacher-training courses in data-driven learning can be identified in the literature. One type of training – offered to both pre-service and in-service teachers – are courses focusing on corpus linguistics (Abdel Latif, 2021; Ebrahimi & Faghih, 2017). Such classes start with introducing fundamental concepts in the field such as definition of a corpus, collocation, or concordance, and then participants are presented with different ways of querying text collections by using a range of tools for language analysis. These courses can supplement or expand on training in general linguistics or information-technology for language teachers. Their main objective is to help trainees become aware of the types of information that can be searched in a corpus and the types of analyses that can be performed with corpus tools. The syllabus is often constructed around available resources and software and emphasises the development of technical and corpus-analytical skills in language teachers. However, this approach does not always cater for the development of pedagogical skills related to designing data-driven activities.

Another approach to popularising data-driven learning among language teachers is making them use corpus resources and tools in their foreign language or language awareness classes (Callies, 2016; Farr, 2008; Heather & Helt, 2012). This option is mainly available to pre-service teachers for whom improving language proficiency or awareness is often an integral part of their teacher training programmes. For example, Zareva (2017) reported on a semester-long course in English grammar with a substantial corpus component offered to TESOL teacher trainees enrolled in an MA programme at an American university. As part of
Training teachers and learners to use corpora

the coursework, the participants completed a corpus-based project on a selected grammatical issue and prepared an in-class presentation as well as a written report on their findings.

The main premise behind this approach is that before language instructors can be expected to use corpora in their teaching, they need to have contact with data-driven education as language learners. The training follows the syllabus of a language course, i.e. trainees focus on issues related to various aspects of language such as grammar, vocabulary, or discourse structure. Yet, instead of using traditional materials, they are guided to consult corpora. This way pre-service teachers not only get acquainted with a variety of resources, tools, and their functions (technical skills), and learn to analyse and interpret corpus data (corpus-analytical skills) but they also get first-hand experience of the relevance of information derived from corpora for language learning. The benefits of such experience have been highlighted by Breyer (2011) who notes: “If teacher trainees can discover the potential of corpora for their own learning, then this may foster intrinsic motivation to make use of corpora in their profession as teachers. It also allows teacher trainees to explore and address the challenges that such an approach entails” (p. 230).

However, the efficiency of such training for language instructors has been questioned by some researchers due to its neglect of the DDL-related pedagogical skills (Leńko–Szymańska, 2017). As Breyer (2009) observes, “recognising that there is a significant difference between learning and teaching with corpora, as well as providing student teachers with the required skills, is of great importance” (p. 156). Thus, the third approach to teacher training in data-driven learning has emerged. It recognises the indispensability of corpus-related pedagogical skills in preparing teachers to exploit text collections in their classrooms (Breyer, 2009, 2011; Hüttner et al., 2009; Leńko–Szymańska, 2014, 2015, 2017). In these courses, teachers not only learn how to query corpora for relevant information and how to analyse and interpret search results, but also how to transform these results into pedagogically-sound and valuable language teaching materials and tasks. The syllabus covers typical language teaching problems and participants discover how corpora can assist learners in studying and mastering these points. An example of such a course addressed to pre-service teachers enrolled in a Master’s of Education programme at a German university was described by Le Foll (2020). The course was built around a collective project whose aim was to create A practical guide to using corpora for EFL teachers. The guide was meant to showcase various applications of corpora for teaching a range of language elements and skills to students at different educational levels. Each trainee (co-) worked on one chapter built around a topic of their choice. The chapters included step-by-step instructions on how to query a selected corpus or how to build and analyse a custom corpus for specific purposes, as well as a sample of corpus-related instructional materials related to the selected topic.

Learners also require special training in order to successfully engage in DDL. Language instructors frequently assume that students’ technical and corpus-analytical skills are gradually fostered whenever they come into contact with corpora in or outside the classroom. Instances of courses in which students are exposed to corpus data are abundant in the literature (e.g. Aston, 2015; Crosthwaite et al., 2019; Geluso & Yamaguchi, 2014; Vyatkina, 2016). In some DDL classes, students are presented with corpus materials offline – for example in the form of selected corpus citations or printed concordance lines, which they must examine and interpret (Boulton, 2010; Huang, 2014; Szudarski, 2019). In other courses, learners can directly consult a corpus but they are provided with specific instructions on how to build a query and what information to consider when scrutinising the results (Moles-Cases & Oster, 2015; Yoon & Hirvela, 2004). Such activities are claimed to have the benefit of exposing students to examples of authentic examples of language use to help them discover characteristics often disregarded in traditional teaching materials such as course books, dictionaries, or grammar
books. They are also believed to raise students’ language awareness (Gilquin & Granger, 2010; Römer, 2011).

Yet, it can be argued that exposing language learners to corpus data may be insufficient for the development of all the skills necessary for data-driven learning. Some educators claim that students should be explicitly taught these skills (Gilquin & Granger, 2010; Sripicharn, 2010). Bernardini (2002) summarised this approach to student training as follows “[i]n such frameworks, it is the rationale and/or methodology/ies of corpus linguistics that are put to didactic use, whilst the role of descriptively-adequate corpus-derived knowledge remains in the background” (p. 165). She calls this approach corpus-aided discovery learning and advocates giving control to students in the process of working with a corpus. The instructor creates a rich learning environment but has no influence over the language analysed and learned in class. Instead, their task is to accompany students “as a guide, not telling them what to do but advising them on how to pursue their own interests, suggesting alternative ways to proceed, other interpretations of the data or possible ways forward” (p. 166). In effect of such training, learners are expected not only to be capable of searching for, analysing, and interpreting corpus data on their own, but more importantly to “develop a frame of mind in which detours and distractions are not stigmatised, but valued as potential sources of unexpected serendipitous encounters” (p. 180).

Several such student-training courses have been reported in thematic volumes and academic journals. In the seminar offered by Bernardini (2002), students in English Studies worked with a variety of large reference corpora and corpus tools in order to discover the patterned quality of language. Doctoral students enrolled in the course described by Lee and Swales (2006) were encouraged to build their own domain-specific corpus of research articles and a corresponding corpus made of their own academic writing. The students subsequently examined the two collections to identify the differences related to a range of language use aspects. The majority of DDL courses are directed at advanced learners in the tertiary-level education. However, recent increased attempts to target younger learners include Crosthwaite and Stell (2019) who described the process of instructing two ten-year-old primary school students in the use of two corpus-based resources for revising lexical errors in their writing. All these student-training courses aimed not only at the improvement of the participants’ linguistic knowledge, but equally importantly at the development of their technical and corpus-analytical skills, thus enabling the students to apply data-driven learning techniques in their own learning outside the classroom.

33.3 Review of current state of research on the effectiveness of teacher and student training

33.3.1 Methodology for evaluating the effectiveness of teacher and learner training

The effectiveness of DDL teacher and learner training courses as well as more generally teachers’ and learners’ perceptions of DDL techniques have been scrutinised with a range of methods. The most popular instrument to gather relevant data has been the questionnaire. Questionnaires have been distributed before a course (e.g. Le Foll, 2020; Leńko-Szymańska, 2014; Mukherjee, 2004) in order to probe participants’ prior knowledge of and experience in corpus linguistics. They may also give an insight into reasons for selecting an elective course. Questionnaires administered directly after a training, on the other hand, give information on teachers’ as well as learners’ attitudes to corpus applications in language education.
Training teachers and learners to use corpora

(Leńko-Szymańska, 2014, 2015). Respondents may declare how difficult or time-consuming they found the new techniques. They may also express their opinions on the benefits and drawbacks of corpus-based teaching and learning. Finally, participants may also be queried about their intentions to use corpora in the future. Delayed questionnaires, circulated several months, or even years after the course offer an insight into teachers’ and students’ actual DDL practices (Charles, 2012; Farr, 2008). Respondents may also relate their unprompted and unguided corpus use or provide reasons for abandoning consultation of text collections.

Instead of predominantly quantitative data collected through questionnaires, reflective essays, and individual or focus group interviews may provide qualitative data on teachers’ or learners’ attitudes towards DDL as well as the effectiveness of teacher and learner training (Abdel Latif, 2021; Karlsen & Monsen, 2020; Le Foll, 2020). These instruments make it possible to examine teachers’ and learners’ perceptions and actual practices in a more detailed and nuanced way but, as for most qualitative methods, are limited to a smaller group of participants.

In addition to the methods based on teachers’ and learners’ reports and declarations, other, more evidence-based procedures of evaluating the effectiveness of instruction in DDL have been proposed. One of them is a systematic analysis of trainees’ and learners’ assignments or projects produced to meet course requirements (Breyer, 2011; Charles, 2015; Heather & Helt, 2012; Leńko-Szymańska, 2017). Such assignments are meant to explore the knowledge and skills acquired and developed by participants during the course and identify which sets of necessary competences are easier or more difficult to master. Other methodologies offering a great potential for researching actual use of corpus-based materials – in particular by language learners – are logs, screen captures, and use histories (Crosthwaite et al., 2019; Pérez-Paredes, this volume, Pérez-Paredes et al., 2011; Schaeffer-Lacroix, 2019). The application of eye-trackers has been advocated by O’Keeffe (2020), but so far no studies have exploited these instruments to collect data.

33.3.2 Results of the research into the effectiveness of teacher and learner training

Overall, the research into the effectiveness of teacher and student DDL training to date has produced ambiguous and difficult-to-interpret results. On the one hand, the questionnaires, reflective essays, focus groups, and interviews reveal generally positive reactions to corpora by pre-service teachers (Farr, 2008; Leńko-Szymańska, 2015), practicing instructors (Chen et al., 2019; Lin, 2016), and learners (Geluso & Yamaguchi, 2014; Yoon & Hirvela, 2004). The responses also confirm teachers’ and students’ recognition of the advantages of DDL such as access to examples of authentic language use and rich contextual information, fostering language awareness, and supporting autonomy (Szudarski, 2019; Zareva, 2017). Learners perceive the usefulness of DDL for studying formulaic language (Geluso & Yamaguchi, 2014), lexico-grammatical information (Szudarski, 2019); grammar (Zareva, 2017); and writing and reading (Charles, 2015; Cotos, 2014). However, at the same time, several drawbacks of this approach are frequently mentioned.

One group of these drawbacks pertain to technological challenges, such as limited access to computers and an adequate internet connection, and a lack of local IT support (Farr, 2008). In more recent surveys, these concerns have become less prominent; for example, technological difficulties were mentioned by less than 5% of respondents in Tribble’s (2015) survey. Nevertheless, hardware and software availability remains an obstacle in some educational
contexts (Ebrahimi & Faghih, 2017). The technical complexity of software and a lack of sufficient technical skills are also a challenge mentioned by respondents (Farr, 2008; Zareva, 2017). Another issue frequently raised is a dearth of suitable resources and tools meeting teachers’ and learners’ needs (Karlsen & Monsen, 2020; Schaeffer-Lacroix, 2019).

Another group of problems voiced by respondents is the high demands that corpus-based activities pose on time and cognitive involvement. Teachers complain that preparing suitable DDL materials and tasks is time-consuming (Tribble, 2015). Learners assert that the very process of analysing corpus data also takes a long time (Charles, 2012). Moreover, both instructors and students report that in spite of the received training, they lack appropriate corpus-analytical skills. Respondents frequently declare that corpus methods continue to be taxing as they require high cognitive involvement related to building appropriate queries, analysing the data, and interpreting the results (Zareva, 2017). In addition, corpora can be linguistically challenging even for advanced students (Charles, 2012, 2015) and the format of corpus results, i.e. truncated concordance lines, tables, and numbers, can be difficult for examination (Geluso & Yamaguchi, 2014; Zareva, 2017).

Finally, language teachers feel insecure about their DDL-related pedagogical skills such as designing corpus activities and integrating them into regular instructional practice. The responses to the questionnaires and interviews as well as the analyses of trainees’ assignments reveal that during the training teachers develop some idea of the ways in which corpora can be used for studying lexis and phraseology; however, designing DDL activities related to grammatical and pragmatic issues is more problematic, as such tasks require more abstract and complex queries and present more challenges in interpreting the data (Heyvaert & Laffut, 2008). Analyses of trainees’ assignments also highlight the need for more guidance in pedagogical aspects of DDL. This implies shifting the focus in training courses from corpus linguistics to such pedagogical aspects as lesson planning and materials design which would enable teachers to produce pedagogically-sound learning resources that are incorporated well into the syllabus (Le Foll, 2020; Leńko-Szymańska, 2017).

All these challenges result in declarations of some instructors in their post-course questionnaires and interviews that despite the training they have received, they do not plan to incorporate DDL techniques in their classrooms (e.g. 30% of respondents in the study by Le Foll, 2020). The inquiry into instructors’ actual practices carried out several years after completing a corpus training confirms that this is frequently the case (Karlsen & Monsen, 2020).

### 33.4 Current contributions to the research on teacher and student training in DDL

This section will present in more detail two DDL training courses discussed in the literature: one designed for pre-service teachers and one offered to advanced learners in an academic context. The courses’ content and methods were described exhaustively in a series of publications which also included thorough analyses of the effects of the instruction examined with a range of instruments. Both cases can serve as typical examples of the kinds of issues addressed in teacher and student training and the methodology employed for determining its effectiveness.

#### 33.4.1 An example of a teacher-training course

The course designed by Leńko-Szymańska (2014, 2015, 2017) is an elective offered to students enrolled in a Master’s programme in applied linguistics at the University of Warsaw. The
Training teachers and learners to use corpora

course has been running annually in either a winter or summer semester for the last eleven years. It consists of 13–15 90-minute sessions divided into three non-consecutive modules. The first module introduces the participants to the fundamental concepts of corpus linguistics and presents samples of research in this area. The next module focuses on the exploitation of large general L1 corpora for teaching language elements and skills. The third module presents the implementation of small purpose-built collections for instruction in language for special purposes. The majority of sessions adopt a workshop format.

During the workshops, the participants perform guided corpus queries and analyse their results. However, these corpus consultations are frequently accompanied by lead-in or follow-up language activities which do not involve interactions with a corpus, such as reading a text, dictionary searches, gap-filling exercises, or group discussions. An example of such a task is provided in Figure 33.2. In addition, trainees are requested to evaluate the corpus-based activities through individual guided reflection, class discussions or short writing assignments, focusing on the instructional value of the completed tasks. The combination of these assignments ensures that the participants learn to use different corpus resources and tools, while discovering how to analyse corpus data and interpret the results. They can also

1. **Read the text at** [http://www.guardian.co.uk/international/story/0,,1182857,00.html](http://www.guardian.co.uk/international/story/0,,1182857,00.html). There are a number of events mentioned in the text. Arrange them in chronological order.

2. **What are the synonyms of the word coup.** Go to the following address to find more: [http://www.wordandphrase.info/frequencyList.asp](http://www.wordandphrase.info/frequencyList.asp)

3. **Study the information (frequency in different genres, collocates and word patterns of each of the synonyms) available in different widows on the page.** Pay particular attention to frequent preposition following the synonyms.

4. **Which of the synonyms found in the corpus were used in the text?**

5. **Find words and expressions in the text that are related to putsches and revolutions.**

6. **Which nouns are followed by the preposition against?** List as many as you can find. Now check in a corpus by following the procedure described below. Study the first 30 items on the list. Can you see any similarities in meaning between these nouns? Can you group them in semantic fields? If necessary, study these words in context.

7. **Work in pairs.** Tell your partner about any familiar putsch or revolution you are familiar with. If you cannot think of any historical or political event, recount the putsch described in the text you have just read.

*Figure 33.2 A DDL vocabulary task*
observe how to merge corpus explorations with other types of classroom activities and develop critical evaluation skills of the merits and drawbacks of the proposed DDL techniques. In this manner, the course caters for the development of the three types of skills needed for the implementation of DDL in language teaching: technical, corpus-analytical, and pedagogical. The trainees are also provided with an opportunity to put the newly acquired competence and skills into practice. For the final course assignment, students are requested to compile a small (ca. about 30K word) corpus of specialised language texts, analyse it for teaching purposes and build a coherent lesson around their findings (see a sample DDL vocabulary task in Figure 33.2). In addition to corpus documentation and the results of an unassisted exploration of the data, the assignment must include a detailed lesson plan and all teaching materials.

Leńko-Szymańska (2014, 2015, 2017) evaluated the effectiveness of the course through two questionnaires administered at the end of two editions of the course as well as through an analysis of 53 end-of-the-semester projects collected from five trainee cohorts. The two questionnaires showed somewhat conflicting results. The first group of respondents had a less positive judgment of their skills in manipulating corpora, as well as of their ability to use corpora for their own language study and for teaching. One year later, a different cohort showed more confidence in these respects, but mainly in corpus-based analyses and teaching of lexis and phraseology, and less so in relation to grammar and discourse problems. More importantly, most of the trainees also declared recognition of the benefits of corpora in language teaching, especially of general corpora but also, to a lesser extent, of small purpose-built specialised corpora. The results revealed that the respondents saw a greater value of corpora as a tool for teachers in creating general language and ESP teaching materials than in having language learners engage directly in data-driven activities. However, the study also showed that a few participants had developed negative attitudes to the value of corpus-based activities in language teaching.

A subsequent analysis of the trainees’ final course projects (Leńko-Szymańska, 2017) produced more disappointing results. It showed that the participants did not develop the expected level of expertise in direct applications of corpora for language teaching. The projects contained only the simplest analyses prescribed in the instructions (selection of specialised terminology and interesting clusters), which demonstrated that in unguided contact with a corpus the trainees lacked intuitions about the selection of pedagogically-valid language points suitable for corpus-based explorations. Corpus data was integrated into the projects’ teaching materials chiefly as a source of examples for presentation and controlled practice activities such as gap-filling and matching. The lesson plans contained almost no hands-on activities engaging potential learners in the analysis of corpus data in a printed or electronic format. Leńko-Szymańska (2017) concluded that if teacher-trainees contact with corpora is limited to one course, the pre-service teachers are not likely to develop a sufficient expertise in the three types of skills indispensable for implementing DDL techniques in their classrooms.

33.4.2 An example of a student-training course

Charles (2012, 2014, 2015) provided a description of a DDL student-training course offered for several years at the Oxford University Language Centre. The course was addressed to graduate students in multiple disciplines and focused on the development of their skills in academic writing through investigating grammar and rhetorical functions in discipline-specific texts. It consisted of six weekly two-hour optional and non-assessed sessions. In the more recent editions of the course, the first class served as an introduction to the concept of
Training teachers and learners to use corpora

Carry out the following tasks and make notes of the information.

a. How frequently does ‘that’ collocate on the RIGHT of the verb? FREQUENCY =

b. Is ‘show* that’ a frequent pattern of this verb in your corpus? YES/NO

c. Which nouns and pronouns collocate most frequently on the LEFT of the verb? Are any of these likely to construct claims?

d. Look at the concordances and original files to check the function of the collocates and fill in the table below.

e. Write down one typical example for each noun/pronoun. Choose examples which will be useful in your own writing. An example is given below.

f. Compare your findings with those of another student. Discuss any similarities or differences you see and try to explain them.

Figure 33.3 Example of a guided discovery task on making and modifying claims (Charles, 2015, p. 154)

hands-on concordancing and its usefulness for improving the quality of learners’ writing, as well as to the relevance of small self-compiled corpora for investigating rhetorical conventions used in domain-specific academic texts. In the subsequent sessions, students explored individual functionalities offered by text analysis packages — represented by AntConc in the course — for searching information on selected language points. In this way, students were simultaneously practising the technical skills of operating the software and the corpus-analytical skills of examining and interpreting retrieved data. Each class started with the instructor’s short presentation of a selected type of analysis based on a small corpus of theses. Subsequently, all students received identical tasks which they completed by querying their own specialised collections. Finally, the participants were given an opportunity to compare their results with their fellow students from other disciplines and discuss the significance of their findings for their own writing practices. Figure 33.3 presents an example of such a task, focusing on the behaviour of the verb show to perform the rhetorical function of making and modifying claims and studied with the help of the Collocates function.

The effects of the course were examined with several instruments. Immediate course evaluation questionnaires were completed by the participants after each of the four editions of the course. The overall results showed an overwhelmingly positive response. The students perceived the main strength of the course in working with discipline-related corpora, which ensured the relevance of the findings for their own particular needs. They also felt that corpus consultations helped them improve their writing. Interestingly, 9 out 42 students (Charles, 2015) mentioned the benefit of discovering discourse studies as a field of academic inquiry. Some reservations were voiced about the size of the self-compiled corpora, which were too small to return a sufficient amount of data to some queries and about their inadequately cleaned format which obscured the results. The answers also indicated that non-native authorship of the texts selected for the personal collections undermined the trust in the obtained results. In addition, several students
reported some linguistic problems with reading concordance lines and complained about the time necessary to process large amounts of language data.

Forty students from one cohort of the course (Charles, 2012) answered a delayed questionnaire probing their long-term use of the self-compiled corpora. Seventy percent of the participants declared some use of their collections in the period of 12 months after the course. Interestingly, five of the declared non-users stated their intentions to consult their corpora once they started working on their dissertations and two other non-users queried large general resources (the BNC) instead. One-third of the students used their corpora at least once a week, with 10% declaring consultations several times a day. In addition, 30% of the respondents made modifications to their personal collections after the course, which involved adding and/or deleting files.

Finally, Charles (2015) conducted a qualitative analysis of a sample of students’ written answers on the worksheets submitted after each class. The results demonstrated that the guided discovery techniques used in the course enabled the students to discover important and relevant facts about grammar and discourse conventions used in their own disciplines. In addition, comparisons with fellow peers as well as class discussions helped them understand the specificity of the writing in their domains and reinforced their confidence as autonomous language investigators, the abilities which – according to Charles – would not only foster the quality of their written language but also their general intellectual development as researchers. Charles (2015) summarised the effects of the tasks in the following words:

The ‘same task, different corpus’ approach allows students to check the findings of research against data from a specialised corpus in their own field. This problematizes established accounts of academic writing, which paves the way for students to adopt a more nuanced and critical attitude towards such descriptions. Being able to refer to highly specific corpus data not only builds students’ knowledge of the norms of usage in their own field, but enables them to back up their intuitions with evidence and thus increases their confidence in their own judgements. (p. 139)

Charles (2012, 2015) results confirm that her students acquired the sufficient levels of technical and corpus-analytical skills which enabled them to become independent corpus users.

### 33.5 Future directions of research

Although direct applications of corpora by teachers and students remain scarce, we can observe, over the last two decades, a growing awareness of the availability and expediency of corpora in language education. An increasing number of instructors apply corpus resources in their everyday practice. Yet, a closer look into the responses from various surveys among language teachers and learners indicates that data-driven learning is not equally popular in all educational settings. It is primarily used by instructors who are corpus researchers themselves and who work in academic contexts, particularly in language-related programmes such as translation. This observation has led researchers to voice the need for promoting DDL, particularly among ‘ordinary’ teachers working in pre-tertiary settings. A number of accounts of teacher- and learner-training courses have indeed appeared in the literature. However, the research into their effectiveness has produced somewhat discouraging results. Although both instructors and students develop positive attitudes to corpora and recognise their benefits in language education, neither party acquires sufficient skills to become fully autonomous corpus users.
Three principal problems hindering successful implementation of DDL have been identified: high time demands, complexity of the required skills, and a lack of pedagogic integration. Future developments in the area should address these specific concerns. First, teachers and learners need pedagogically-relevant resources and tools as well as off-the-shelf DDL materials and tasks (Chambers, 2019; Gilquin & Granger, 2010). Such solutions contradict some of the argued advantages of corpora – authenticity and relevance to students’ specific needs. However, it seems that without some compromises instructors and students will continue to turn towards conventional resources such as coursebooks, dictionaries, and reference grammars, which do not require onerous and time-consuming searches. Second, beyond highly specialised corpus linguistics or DDL courses, corpus use must have greater representation in language, linguistics, and general teaching methodology classes. Only then will DDL likely reach a tipping point resulting in the mastery and constant reinforcement of the complex technical and corpus-analytical skills necessary for querying and analysing corpus data. Finally, more research and training should support the successful integration of DDL into the language learning process and the development of teachers’ corpus-related pedagogical skills. Consequently, instructors and students will become aware of how to successfully combine corpus activities with other effective pedagogic techniques and learning strategies. Meeting these conditions – admittedly not unproblematic – will finally bring about a change not just in teachers’ and learners’ perceptions and attitudes, but in their routine DDL practice.

Further reading


Wicher, O. (2019). Data-driven learning in the secondary classroom. A critical evaluation from the perspective of foreign language didactics. In P. Crosthwaite (Ed.), Data-driven learning for the next generation. Corpora and DDL for pre-tertiary learners (pp. 31–46). Routledge. https://doi.org/10.4324/9780429425899-3. Although this paper does not deal with teacher and learner training per se, it is recommended to instructors as it stresses the importance of implementing DDL in language education in accordance with the current theories of language instruction and the models of lesson stages.
Agnieszka Leńko-Szymańska

References


522
Training teachers and learners to use corpora


