Corpus-Driven Instruction

Interest in using corpora, or large electronic collection of texts, in language instruction has been rapidly growing over the last few decades (see Römer, 2011, for an overview). Corpus-driven applications can be either indirect or direct. In indirect applications, instructors can use corpus-based reference grammars, textbooks, and dictionaries that include attested language samples instead of invented examples (e.g., Biber et al., 2002; Dodd et al., 2003). In direct applications, also called Data-Driven Learning, or DDL (Johns & King, 1991), teachers and learners explore corpora themselves.

DDL is grounded in usage-based theories, according to which language is learned inductively through repeated exposure to and practice with specific language models. Corpora can help teachers as a rich repository of such models, especially in foreign language teaching contexts, where authentic language samples are hard to come by. Furthermore, perceptual salience of corpus samples is enhanced by their graphic representation: corpus search results typically come in the form of concordances, or stacked text lines with the search words highlighted and placed in the middle (see Figure).

Figure. Concordances with the search word Deutsch ('German') from the DWDS corpus (http://dwds.de/).
Due to these characteristics, corpora lend themselves to an inductive approach to language teaching and learning, in which learners more or less independently engage in “pattern-hunting” and “pattern-defining” (Kennedy & Miceli, 2010, p. 31) with the teacher assisting them as a facilitator. DDL research has shown that corpus-driven instruction is at least equal to or more efficient than more traditional, deductive teaching methods for a number of instructional targets (Boulton & Pérez-Paredes, 2014; Cobb & Boulton, in press). However, for DDL to be successful, teachers should carefully guide the learners and sequence the tasks from less to more autonomous. More specifically, corpus-driven instruction at all language proficiency levels should start with teacher-designed corpus-based materials (e.g., concordance printouts) and progress toward independent learner corpus searches.

Teachers and learners can especially benefit from freely and publicly available language corpora, such as Corpus of Contemporary American English (http://corpus.byu.edu/coca/), Corpus del Español (http://www.corpusdelespanol.org/), Digital Dictionary of German (http://dwds.de/), or Russian National Corpus (http://ruscorpora.ru/en/index.html/). Notably, these corpora are equipped with various built-in search and analysis tools beyond simple concordancers, which allow teachers to design a wide variety of instructional tasks. Learners can be instructed to find, analyze, and compare selected words, phrases, or parts-of-speech. Other possible tasks include comparing word usage in different genres (e.g., fiction and journalism), exploring semantic prosody of a word (the tendency of near-synonyms to appear with different or similar attributes), and investigating waning and waxing of the popularity of a word over decades and centuries. For guides and models for designing specific DDL activities, please see Bennett (2010), O’Keeffe et al. (2007), and Reppen (2010), as well as CALPER Corpus Tutorial (http://calper.la.psu.edu/corpus_portal/tutorial_main.php).
References:


