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/).

66 [2014]   Er mochte den Unterricht , war Klassensprecher , ehrgeizig , gut in **Deutsch** und Englisch , aber Mathe fiel ihm schwer .
67 [2014]   Die G-8-Schüler hatten schlechtere Mathematiknoten , in **Deutsch** gab es keinen Unterschied .
68 [2014]   Henning freut sich , mal wieder **Deutsch** reden zu können , aber manchmal kippt er in ein fast akzentfreies briti
69 [2014]   Er spricht über den kreativen Knast , den sein Deutschsein darstelle , **Deutsch** hier , **Deutsch** da .
70 [2014]   Mit Flash Boys , das am Donnerstag dieser Woche auch auf **Deutsch** erscheint , hat Michael Lewis voll ins Schwarze getroffen .
71 [2014]   So haben wir also vor allem die Texte – und auf **Deutsch** die Übertragungen von August Wilhelm Schlegel .
72 [2014]   Eine Gesamtnote 3 in **Deutsch** zum Beispiel sagt nichts darüber aus , wo ein Schüler seine Stärken u
73 [2014]   Dabei wird beispielsweise im Fach **Deutsch** unterschieden zwischen der verbalen Sprechfähigkeit und dem Sinn e
74 [2014]   Die Ausbildung ist komplett auf **Deutsch** – für Hang eine Herausforderung .
75 [2014]   Hang würde gerne besser **Deutsch** sprechen können , neben ihrer Vollzeitstelle besucht sie einen Deutsc
76 [2014]   Er ritzt sich die Arme , sein **Deutsch** bleibt schlecht , er fehlt in der Schule , arbeitet in einem Restaurant , b
77 [2014]   Jeden (Ab-)Satz habe ich in etwas moderneres **Deutsch** übersetzt , aber ich blieb sehr nah an Schleiermachers Übertragung .

1

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:

Bennett, G. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers.* Michigan: University of Michigan Press.

Biber, D., Conrad, S., & Leech, G. (2002). *The Longman student grammar of spoken and written English*. London: Longman.

Boulton, A., & Pérez-Paredes, P. (2014). Editorial: Researching new uses of corpora for language teaching and learning. *ReCALL*, *26*, 121-127.

Cobb, T., & Boulton, A. (in press). Classroom applications of corpus analysis. In D. Biber & R. Reppen (Eds.), *Cambridge handbook of corpus linguistics*. Cambridge: Cambridge University Press.

Dodd, B., Eckhard-Black, C., Klapper, J., & Whittle, R. (2003). *Modern German grammar: A practical guide (2nd edition).* London: Routledge.

Johns, T., & King, P. (Eds.) (1991). *Classroom Concordancing: English Language Research Journal, 4*. CELS: University of Birmingham.

Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning & Technology*, *14*, 28–44.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.

Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge: Cambridge University Press.

Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics, 31*, 205-225.