Received: 30 April 2020 | Accepted: 3 May 2020

DOI: 10.1111/flan.12464

ORIGINAL ARTICLE

ACTFL AMERICAN COUNCIL ON THE TEACHING OF FOREIGN LANGUAGES WILEY

Corpora as open educational resources for language teaching

Nina Vyatkina

The Challenge

Language teaching with corpora, large electronic collections of texts, has been found effective and efficient. How can corpora be more broadly incorporated into teaching practice, especially for learners of languages other than English, at lower proficiency levels, and in secondary schools? What open educational resources can help teachers?

German and Applied Linguistics, University of Kansas, Lawrence, Kansas

Correspondence
Nina Vyatkina, 2080 Wescoe Hall, 1445 Jayhawk Blvd., Lawrence, KS 66045.
Email: vyatkina@ku.edu

Funding information
U.S. Department of Education, Grant/Award Number: P229A180008

Abstract

Corpora, large electronic collections of texts, have been used in language teaching for several decades. Also known as Data-Driven Learning (DDL), this method has been gaining popularity because empirical research has consistently shown its effectiveness for learning. However, corpora are still underutilized, especially with learners of languages other than English, at lower proficiency levels, and in non-university contexts. This is regrettable because DDL has a great potential for developing modular flipped content, especially for hybrid, remote, and online courses. This article first provides an overview of DDL applications and findings of empirical research. Next, it outlines obstacles to wider DDL implementation as well as available and possible solutions. Corpus user guides and exercise collections tied to specific corpora are discussed as one promising direction, and an example of such new open educational resources for teaching German is

presented. The article concludes with a discussion of implications and future directions.

KEYWORDS

all languages, authentic materials, computer-assisted language learning, corpora, data-driven learning

1 | INTRODUCTION

According to Szudarski (2018, p. 1, emphasis in the original), corpora are "large, principled and computer-readable collections of texts that allow analysis of *patterns* of language use across different *contexts*." The potential of corpora for language teaching has been recognized and utilized since their emergence back in the 1960s, first in indirect, and later in direct applications (Römer, 2011, p. 207). Corpus-based materials and activities can be a useful addition to the teachers' arsenal, especially in the era of hybrid, remote, and online teaching as they can supplement and enhance existing syllabi with digital, flipped, and modular content. In this article, we review different types of corpus-based applications to teaching and associated research. Then, we list persistent obstacles to a wide dissemination of corpora and propose a number of solutions, focusing on open educational resources in form of teacher guides and exercise collections. We conclude with a discussion of implications and future directions.

2 | INDIRECT APPLICATIONS: CORPORA AND LANGUAGE TEACHING MATERIALS

In indirect applications, researchers and materials writers have used the information derived from corpora to design syllabi, reference works, and teaching materials. The most important information collected for these purposes with automated corpus analysis tools has been word frequency, following the rationale that learning more frequently occurring words is more conducive to the development of L2 ability than learning rare words. Respectively, the most well-known corpus-informed syllabus was called the lexical syllabus (Willis, 1990). This research has also resulted in the publication of a number of word lists, such as the Academic Word List (Coxhead, 2000) that arranges English words typical of the academic register in frequency bands, so that teachers of English for Academic Purposes can select words from specific bands for teaching to learners depending on their language competency level. Furthermore, corpus researchers discovered that words had a tendency to occur in repeated lexical patterns (collocations) and lexico-grammatical patterns (colligations), which gave rise to new language acquisition theories that recognize the inseparability of grammar and the lexicon, such as usage-based approaches and construction grammar (see Ellis, 2017). A publication of several corpus-based pedagogical grammars and dictionaries followed (e.g., Carter & McCarthy, 2006; Rundell, 2007). However, recent developments in the field have mostly been connected with more direct applications, which are reviewed in the next section.

3 | DIRECT APPLICATIONS: DATA-DRIVEN LEARNING

In these applications, language teachers and/or learners interact with corpora directly (Römer, 2011, p. 207). Typically, such interaction involves searching corpora for specific words or phrases, with the resulting output in form of concordances, stacked truncated lines of text with the search item highlighted and centered in each line (Figure 1).

This visualization format makes patterns of use of the search item stand out and thus is conducive to inductive, analytical learning. For example, the viewer can infer from analyzing concordances in Figure 1 that in German, the noun *Internet* is frequently preceded by the preposition-article contraction *im* (*in* + *dem*). This pattern also points to a difference between German and English: although *Internet* is a borrowing from English, it is used in German with the preposition *in* and not *an*, which is the most direct equivalent of the English "on" (hence "on the internet" = *im Internet*). This method was named Data-Driven Learning (DDL) by Johns (1990), who adopted the term from computer science. DDL was defined as "the use in the classroom of computer-generated concordances to get students to explore the regularities of patterning in the target language, and the development of activities and exercises based on concordance output" (Johns & King, 1991, p. iii). This teaching method quickly started gaining popularity and the accompanying research followed, first in form of descriptive reports and then empirical studies (see Boulton, 2017, for a timeline). Typical DDL interventions focus on specific lexical and/or grammatical forms, with learners analyzing concordances and inferring patterns of use. In hands-on applications, learners search corpora directly on computers, independently or under the teacher's guidance, whereas in hands-off applications, the teacher prints concordances on paper beforehand or projects them on the big screen in class. In learning outcome studies, researchers compared the effectiveness of DDL to more traditional, usually deductive, teaching methods, or of different DDL methods (such as hands-on and hands-off). Learning gains have been tested via multiple-choice, gap-fill, or production tests (usually controlled or free writing). For example, Daskalovska (2015) found that Macedonian university students who explored concordances online learned English verb–adverb collocations better than their peers who used deductive textbook materials. Other types of DDL studies explored learner and teacher attitudes toward DDL (via questionnaires, interviews, etc.) or DDL processes (via observations, eye-tracking, screen captures, etc.). For example, Benavides (2015) found a variety of opinions of US university students about their experience using a large Spanish corpus for learning grammar that ranged from mostly positive to fairly negative. In what follows, we will only focus on direct corpus applications (i.e., DDL), to summarize existing

Damals hatte ich kein	**Internet**	, ich beschäftigte mich nur mit mir selbst.
Kurz darauf kommen im	**Internet**	Videos in Umlauf, die sich rasant im ganzen Land verbreiten.
…au und zur Sanierung der Kommunikationsnetze und des schnellen	**Internets**	vor allem auf dem Land, von Straßen, Brücken, Eisenbahnstrecken u…
…und eine Nummer eingestanzt, wer einen findet, soll den Fundort per	**Internet**	oder Handy-App der Uni melden.
Im	**Internet**	suchte der Bezirksverband Hessen-Süd jemanden für die Nachmittag…
Es gibt im	**Internet**	Musterbriefe für Klageschriften.
Entscheiden nicht Computer darüber, was wir im	**Internet**	zu sehen bekommen?
… in den Marktbewertungen der Leitunternehmen des kommerziellen	**Internets**	: Google/Alphabet schafft es auf 580 Milliarden Dollar, Apple auf soga…
Was kauft ein Mensch im	**Internet**	?
…ich vor allem daran, dass mit dem Gesetz die Rechtsdurchsetzung im	**Internet**	an private Firmen delegiert wird.

FIGURE 1 Concordances for *Internet* from the DWDS corpus (ZEIT newspaper subcorpus), Retrieved April 25, 2020 from https://www.dwds.de/r?q=Internet&corpus=zeit&date-start=1946&date-end=2018&format=kwic&sort=date_desc&limit=10

research and to explore ways in which corpora can be useful for language learning, especially in remote and online contexts.

4 | EMPIRICAL DDL RESEARCH

The number of empirical DDL studies has been growing rapidly, especially over the last 15 years, and has almost reached 500 by the end of 2019, only counting English-language journal articles, book chapters, and conference proceedings (Boulton, personal communication, May 19, 2020). This large accumulated body of knowledge has enabled evaluating the overall effectiveness of DDL via meta-analyses. The most comprehensive meta-analysis by Boulton and Cobb (2017) showed that, overall, DDL led to significant L2 learning gains and was more efficient than non-DDL teaching methods with a large effect size. Lee, Warschauer, and Lee (2018) meta-analyzed research on vocabulary learning and found that the knowledge gains were higher with DDL than with non-DDL methods. Furthermore, these two research syntheses showed that both the hands-on and hands-off types of DDL were effective methods that worked for different language targets (vocabulary, grammar, lexico-grammar, discourse), for learners at different proficiency levels, and in different contexts (second and foreign language teaching in different countries). These positive results have been attributed to a number of pedagogical and theoretical principles realized in DDL that have been long shown beneficial for language acquisition (see Boulton & Cobb, 2017, for a discussion): input flood (a large number of language use examples), input enhancement (graphics that highlight language use patterns), inductive pattern learning (learners' exploration of concordances), and deep processing (analysis, inferencing). Beyond objective learning outcomes, DDL has also been found to provide learner-centered experience, enhance learner autonomy, and be positively perceived by many learners (Pérez-Paredes, Ordoñana Guillamón, & Aguado Jiménez, 2018). Moreover, corpora can arguably relieve nonnative teachers from perceived responsibility vis-à-vis their learners as language experts and provide them with a plethora of authentic language use examples instead of or in addition to artificial textbook examples (Römer, 2011).

5 | OBSTACLES TO WIDER DDL IMPLEMENTATION

Despite these many empirically corroborated benefits of DDL, it remains surprisingly underutilized by language teachers. At the turn of the millennium, Conrad (2000) discussed the potential of corpora to revolutionize teaching, yet a decade later, Römer (2011) noted teachers' resistance toward corpora, and as recently as last year, Chambers (2019, p. 460) still discussed the need for "bridging the research-practice gap." Several reasons for this disconnect have been repeatedly identified (Ballance, 2017; Chambers, 2019; Pérez-Paredes et al., 2018; Wilson 2013). First, the overwhelming majority of DDL studies (more than 90%) have targeted university students of English as a Second or Foreign Language (ESL or EFL) with intermediate to advanced language competency (Boulton, personal communication, May 19, 2020). This fact has hindered dissemination of DDL practices to contexts that include learners of LOTEs (Languages Other Than English), lower competency learners, and young learners. Second, many corpora and corpus tools used in DDL studies are not freely and publicly available. For example, Sketch Engine (Kilgarriff et al., 2014) is a widely used corpus search and analysis tool associated with a large number of corpora yet free access to it is restricted only to researchers at

selected academic institutions. Third, most corpora have been developed by corpus linguists for corpus linguists with an interface too technical and difficult for non-specialist users like language teachers and learners. As Braun (2007, p. 308) noted: "Many of the widely accessible corpora were created as tools for linguistic research and not with pedagogical goals in mind." Although not all DDL researchers are corpus linguists, all of them had dedicated considerable time and effort to exploring corpora before designing and conducting their pedagogical experiments, which can hardly be expected from teachers who do not conduct DDL research. Additionally, due to space limitations, DDL studies have rarely provided many pedagogical details, which made them hard (or not appealing) to replicate for other teachers. Fortunately, the situation has recently begun changing with many DDL researchers, practitioners, and material developers working on solutions to the abovementioned issues. These emerging solutions are addressed in the next section.

6 | EMERGING SOLUTIONS

6.1 | Broadening the contexts for DDL applications

A growing number of publications have been expanding the DDL application contexts. Boulton and Cobb's (2017) meta-analysis demonstrated that DDL was as efficient for lower competency learners as for higher competency learners. Several studies showed that DDL worked well with younger learners, such as primary and secondary school students, given appropriate tasks and careful teacher guidance (e.g., Crosthwaite, 2019). What DDL researchers have recommended for these contexts is in line with ACTFL guiding principles (ACTFL, n.d.) for using authentic materials in language teaching in general: "Tailor the task to the proficiency level of the student (use the same text but change what you ask learners at each level to do with the text)." For example, verb–preposition collocations can be taught with DDL methods to both novice and more advanced learners but novice learners can work with printed concordances under the teacher's guidance, whereas more advanced learners can search for concordances directly in corpora (see Vyatkina, 2018, for more examples of tasks for different proficiency levels).

Finally, and importantly, for the readership of this journal, there has been growing empirical evidence that DDL works not only for English but also for LOTEs: Catalan (Marco & van Lawick, 2015), Chinese (e.g., Chen, Wu, Yang, & Pan, 2016; Wong & Lee, 2016); Czech (Osolsobě & Vališová, 2012), French (e.g., Chambers & O'Sullivan, 2004), German (e.g., Ortner & Weber, 2018; Rets, 2017; Vyatkina, 2016a, 2016b), Italian (e.g., Kennedy & Miceli, 2001, 2010, 2017), and Spanish (e.g., Benavides, 2015). See the appendix for resources for working with DDL in these languages.

6.2 | Broadening access to DDL resources

Availability of DDL resources has been significantly broadened thanks to the growing number of open access corpora (see the appendix for references and links). These corpora contain large collections of texts (up to millions and even billions of words) that are designed to be representative of cross-sections of different registers and genres as well as historical periods (decades and even centuries). All these corpora are equipped with built-in search, analysis, and visualization tools, so that concordances, frequency lists, and other output forms can be created automatically for the search items typed in by the user. Importantly, such large national corpora

are not only free and publicly available but also relatively sustainable because their collection, maintenance, and expansion typically are handled by teams of corpus linguists under support of government research grants. Apparently, these advantages make open access corpora great candidates for becoming widely used Open Educational Resources (Pérez-Paredes et al., 2018). However, they still fall short of reaching this goal due to technical difficulties associated with the corpus-user interface. To obviate these difficulties and assist teachers in their corpus use, several solutions have been proposed, which are addressed in the next section.

6.3 | Scaffolding DDL for teachers

First, the necessity of DDL trainings for teachers has been repeatedly called for (e.g., Frankenberg-Garcia, 2012; Leńko-Szymańska, 2017). Free and open online teacher education courses in DDL would be a welcome development in this direction. Yet another promising solution is publication of corpus user guides for teachers and suggestions for DDL exercises. Such materials have been appearing throughout the DDL history. Although (in line with the trend in empirical DDL research) book-length teacher manuals have been restricted to English (e.g., Bennett, 2010; Friginal, 2018; Poole, 2019; Reppen, 2010), pedagogically oriented DDL articles in journals and edited volumes have also targeted LOTEs, including French (Kerr, 2009; Tyne, André, Benzitoun, Boulton, & Greub, 2014), Italian (Forti & Spina, 2019; Kennedy & Miceli, 2017), German (Neary-Sundquist, 2015; Schaeffer-Lacroix, 2016, 2020; Vyatkina, 2018, 2020a), and Spanish (Abad Castelló, 2019; González García, 2019).

Although these pedagogical materials go a long way toward engaging teachers in DDL, it is notable that they are stand-alone articles or books. Even teacher guides linked to specific corpora usually take the form of stand-alone pdf files (Kohn & Hoffstaedter, 2008; Shaw, 2011). In other words, teachers who want to use them for guidance in exploring corpora or creating corpus-based exercises must go back and forth between reading these materials and searching corpora online. What is still largely missing in the field (including EFL, ESL, and LOTEs) is online corpus user guides for teachers and exercises integrated with specific corpora.

6.4 | Open digital DDL materials integrated with open corpora

A new project that addresses this gap is *Incorporating Corpora*, conducted at the University of Kansas Open Language Resource Center (OLRC; http://olrc.ku.edu). The Center focuses on the creation of Open Educational Resources (OER) for teaching and learning a variety of LOTEs at the secondary and post-secondary level. The DDL project website (https://corpora.ku.edu) contains free and open materials for teaching and learning German with a free and open German corpus DWDS (a German acronym for the Digital Dictionary of German; http://dwds.de). The materials include a brief introduction to DDL and DWDS, a corpus user guide, and a series of exercise modules that focus on selected aspects of German grammar and vocabulary (Vyatkina, 2020b). What sets these materials apart from other DDL handbooks and exercise collections, besides their free and open nature, is their integration with a specific corpus. In both the user guide and the exercises, offline text and DWDS corpus snapshots alternate with links that open in a separate tab and take the user directly to the corpus. Nonspecialist explanations for how to conduct some basic corpus searches are interspersed with online quizzes, in which the user conducts these searches using the provided links, answers multiple choice or short answer

questions, and gets immediate automated feedback. The exercises are structured following the pedagogical principle of guided induction that has proven to be beneficial in DDL (see Vyatkina, 2018 for details): progressing from corpus example exploration to controlled practice to metalinguistic pattern induction to communicative free practice. These exercises can be tailored by the teacher to specific needs and the level of language and corpus competence of their learners. Teachers may opt for preprinting selected exercises on paper, projecting them on a big screen for the whole class, or assigning them for individual online completion in class or at home.

Selected exercises from *Incorporating Corpora* were piloted with students in second-year and fourth-year German classes at the University of Kansas, a large public US university. They focused on verb-preposition collocations, a lexico-grammar area that has been notoriously difficult for learners of German (Vyatkina, 2016a, 2016b). After a brief introduction of the DWDS corpus in class, exercises were assigned for independent online completion at home and learner opinions were collected via an online questionnaire. The results were very encouraging as students in both groups found the explanations clear and the exercises helpful (more advanced students for review and less advanced students for initial learning). Learners praised the modules for easy navigation, engaging nature, gradual progression, immediate feedback, and ability to work through the exercises at their own pace and to retry. Furthermore, the second-year group performed on a subsequent test on a par with their peers in another section of the same course who were taught with a deductive non-DDL method (a 10-page-long paper handout with teacher explanations followed by exercises).

7 | CONCLUSION AND IMPLICATIONS

The above overview shows that corpora have been successfully used in language teaching for decades. Corpus-based word frequency lists have informed teaching syllabi and materials, and both teachers and learners have used corpora directly to search for language use examples and explore patterns. Such inductive DDL applications have led to significant learning gains, especially in vocabulary and grammar knowledge, and have been frequently more efficient than non-DDL teaching methods while at the same time enhancing learner autonomy and providing individualized learning experiences. These aspects of DDL make it especially suitable for applications in hybrid, remote, and online courses. Corpus-based modules can be developed to supplement and enhance existing syllabi with digital and flipped content. Students can conduct corpus searches individually on their own computers and at their own pace, and report the results to the teachers via worksheets or other conventional media. This type of work would also contribute to larger educational goals such as developing students' critical thinking, analytical ability, and digital literacy.

Although DDL has predominantly targeted EFL and ESL university students, there is no inherent reason for why it should be restricted to these contexts. The field has recently been expanding, and the readers of this journal can take heart in the fact that DDL can also work with LOTEs, primary and secondary schools, as well as beginning learners. There is still a great potential for growth in these areas. Publication of teacher guides and DDL exercise collections integrated with specific corpora would be especially helpful to teachers. *Incorporating Corpora*, introduced above, is an exemplar of such a project that presents an alternative "third way" to hands-on and hands-off DDL, a middle ground "between the polished, albeit limited, linguistic information neatly systematized in dictionaries and the countless other linguistic facts that can

be gleaned from corpora, but which only experienced corpus users are able to access" (Frankenberg-Garcia, 2014, p. 141). The pilot study conducted with the project's materials showed that even lower-competency learners were capable of autonomous DDL when it was scaffolded through an online guiding interface. It is planned to continuously maintain, update, and expand the project's modules as well as to test them with other teachers and students, including those in secondary schools. Although *Incorporating Corpora* is focused on a specific German corpus, it can serve as a model for creating similar materials for other languages and corpora, and it is hoped that other DDL researchers and practitioners will follow suit.

ACKNOWLEDGMENTS

Many thanks go to my University of Kansas colleagues Jonathan Perkins, Keah Cunningham, and Schirin Kourehpaz for their help in developing, designing, and publishing the OERs for *Incorporating Corpora*. I would also like to thank Alex Boulton and my fellow participants of the EuroCALL 2020 symposium *Data-driven learning for languages other than English* Magdalena Abad Castelló, Luciana Forti, and Eva Schaeffer-Lacroix for their many suggestions for corpus resources and DDL publications on LOTEs. This study was supported in part by the U.S. Department of Education grant P229A180008.

ORCID

Nina Vyatkina http://orcid.org/0000-0002-2778-8016

REFERENCES

Abad Castelló, M. (2019). Uso de corpus lingüísticos por y para profesores de español como lengua extranjera [Uses of corpora in Spanish language teaching]. *Revista Electrónica de Didáctica del Español Lengua Extranjera*, *31*, 1–20. https://www.culturaydeporte.gob.es/dam/jcr:39778ef0-dddf-422f-b3c2-a8f08730225b/07-uso-de-corpus-linguisticos.pdf

ACTFL. (n.d.). *Use of authentic texts in language learning* [Electronic version]. Retrieved from https://www.actfl.org/guiding-principles/use-authentic-texts-language-learning

Ballance, O. (2017). Pedagogical models of concordance use: Correlations between concordance user preferences. *Computer Assisted Language Learning*, *30*, 259–283.

Benavides, C. (2015). Using a corpus in a 300-level Spanish grammar course. *Foreign Language Annals*, *48*(2), 218–235.

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

Boulton, A. (2017). Research timeline: Corpora in language teaching and learning. *Language Teaching*, *50*(4), 483–506.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language learning*, *67*(2), 348–393.

Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL*, *19*(3), 307–328.

Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English*. Cambridge, UK: Cambridge University Press.

Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching*, *52*(4), 460–475.

Chambers, A., & O'Sullivan, Í. (2004). Corpus consultation and advanced learners' writing skills in French. *ReCALL*, *16*(1), 158–172.

Chen, H. H.-J., Wu, J.-C., Yang, C. T.-Y., & Pan, I. (2016). Developing and evaluating a Chinese collocation retrieval tool for CFL students and teachers. *Computer-Assisted Language Learning*, *29*(1), 21–39.

Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly*, *34*(3), 548–560.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238.

Crosthwaite, P. (Ed.). (2019). *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*. New York, NY: Routledge.

Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, *28*(2), 130–144.

Ellis, N. C. (2017). Cognition, corpora, and computing: Triangulating research in usage-based language learning. *Language learning*, *67*(S1), 40–65.

Forti, L., & Spina, S. (2019). Corpora for linguists vs. corpora for learners: Bridging the gap in Italian L2 learning and teaching. *EL.LE: Educazione Linguistica. Language Education*, *8*(2), 349–362. https://edizionicafoscari.unive.it/en/edizioni4/riviste/elle/2019/2/

Frankenberg-Garcia, A. (2012). Raising teachers' awareness of corpora. *Language Teaching*, *45*(4), 475–489.

Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. *ReCALL*, *26*(2), 128–146.

Friginal, E. (2018). *Corpus linguistics for English teachers*. New York, NY: Routledge.

González García, V. (2019, March 4). *Los corpus discursivos en la clase de ELE: atrévete a usarlos y aprovéchate de todo lo que te ofrecen* [The discursive corpora in the SFL class: Dare to use them and take advantage of everything they offer you]. Retrieved from https://www.difusion.com/en/corpus-discursivos-en-clase/

Johns, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, *10*, 14–34.

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

Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology*, *5*(3), 77–90.

Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning & Technology*, *14*(1), 28–44.

Kennedy, C., & Miceli, T. (2017). Cultivating effective corpus use by language learners. *Computer Assisted Language Learning*, *30*(1-2), 91–114.

Kerr, B. J. (2009). Applications of corpus-based linguistics to second language instruction: Lexical grammar and data-driven learning. In S. L. Katz & J. Watzinger-Tharp (Eds.), *Conceptions of L2 grammar: Theoretical approaches and their application in the L2 classroom* (pp. 128–150). Boston, MA: Heinle Cengage Learning.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The sketch engine: Ten years on. *Lexicography*, *1*, 7–36.

Kohn, K., & Hoffstaedter, P. (2008). SACODEYL corpus-based language learning: Guidelines for teachers and learners. Retrieved from https://www.um.es/sacodeyl/data/publications/SACODEYL_guidelines_for_teachers.pdf

Lee, H., Warschauer, M., & Lee, J. H. (2018). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, *40*(5), 721–753.

Leńko-Szymańska, A. (2017). Training teachers in data driven learning: Tackling the challenge. *Language Learning & Technology*, *21*(3), 217–241.

Marco, J., & van Lawick, H. (2015). Enhancing translator trainees' awareness of source text interference through the use of comparable corpora. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 225–244). Amsterdam, the Netherlands: John Benjamins.

Neary-Sundquist, C. (2015). A corpus-based pedagogy for German vocabulary. In A. J. Moeller (Ed.), *Learn languages, explore cultures, transform lives* (pp. 201–215). Report of the Central States Conference on the Teaching of Foreign Languages, University of Nebraska–Lincoln. Retrieved from http://www.csctfl.org/documents/2015Report/CSCTFL%20Report_2015.pdf

Ortner, G. J., & Weber, U. S. (2018). Using corpora to inform teaching practice in German studies. *Per Linguam*, *34*(2), 69–83.

Osolsobě, K., & Vališová, P. (2012). Using data-driven methods in teaching Czech as a foreign language. In J. Thomas & A. Boulton (Eds.), *Input, process and product: Developments in teaching and language corpora* (pp. 184–195). Brno, Czech Republic: Masaryk University Press.

Poole, R. (2019). *A guide to using corpora for English language learners*. Edinburgh, UK: Edinburgh University Press.

Pérez-Paredes, P., Ordoñana Guillamón, C., & Aguado Jiménez, P. (2018). Language teachers' perceptions on the use of OER language processing technologies in MALL. *Computer Assisted Language Learning*, *31*(5-6), 522–545.

Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge, UK: Cambridge University Press.
Rets, I. (2017). Vocabulary retention and concordance-based learning in L3 acquisition. *Eurasian Journal of Applied Linguistics*, *3*(2), 313–324.
Rundell, M. (Ed.). (2007). *Macmillan English dictionary for advanced learners* (2nd ed.). Oxford, UK: Macmillan Education.
Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, *31*, 205–225.
Schaeffer-Lacroix, E. (2016). Talking about German verb particles identified in concordance lines: From spontaneous to expert-like metatalk. *Language Awareness*, *25*(1–2), 127–143.
Schaeffer-Lacroix, E. (2020). Integrating corpus-based audio description tasks into an intermediate-level German course. *International Journal of Applied Linguistics*, 1–20. https://doi.org/10.1111/ijal.12294
Shaw, E. M. (2011). *Teaching vocabulary through data-driven learning*. Provo, UT: Brigham Young University. Retrieved from https://corpus.byu.edu/coca/files/Teaching_Vocabulary_Through_DDL.pdf
Szudarski, P. (2018). *Corpus linguistics for vocabulary: A guide for research*. New York, NY: Routledge.
Tyne, H., André, V., Benzitoun, C., Boulton, A., & Greub, Y. (2014). *French through corpora: Ecological and data-driven perspectives in French language studies*. Newcastle Upon Tyne, UK: Cambridge Scholars Press Ltd.
Vyatkina, N. (2016a). Data-driven learning for beginners: The case of German verb-preposition collocations. *ReCALL*, *28*(2), 207–226.
Vyatkina, N. (2016b). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology*, *20*(3), 159–179.
Vyatkina, N. (2018). Language corpora for L2 vocabulary learning: Data-driven learning across the curriculum. In P. Ecke & S. Rott (Eds.), *Understanding vocabulary learning and teaching: Implications for language program development* (pp. 121–145). Boston, MA: Cengage Learning.
Vyatkina, N. (2020a). Corpus-informed pedagogy in a language course: Design, implementation, and evaluation. In M. Kruk & M. Peterson (Eds.), *New technological applications for foreign and second language learning and teaching* (pp. 306–335). Hershey, PA: IGI Global.
Vyatkina, N. (Ed.). (2020b). *Incorporating corpora: Using corpora to teach German to English-speaking learners* [Online instructional materials]. Lawrence, KS: KU Open Language Resource Center. Retrieved from https://corpora.ku.edu
Willis, D. (1990). *The lexical syllabus*. London, UK: Harper Collins.
Wilson, J. (2013). *How can we make the most of corpora and data-driven learning (DDL) in language learning and teaching?* York, UK: The Higher Education Academy.
Wong, T.-S., & Lee, J. S. Y. (2016). Corpus-based learning of Cantonese for Mandarin speakers. *ReCALL*, *28*(2), 187–206.

How to cite this article: Vyatkina N. Corpora as open educational resources for language teaching. *Foreign Language Annals*. 2020;53:359–370. https://doi.org/10.1111/flan.12464

APPENDIX: REFERENCES TO SELECTED CORPORA

Most corpora listed below are free and open to the public; although, some of them require registration, especially for access to extended resources and functionalities.

English

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 600 million words, 1990–present*. Retrieved from https://www.english-corpora.org/coca/

Davies, M. (2004). *British National Corpus* (from Oxford University Press). Retrieved from https://www.english-corpora.org/bnc/

French

Praxiling: UMR 5267. (2019). *Corpus 14*. Constance, Germany: University of Constance, ORTOLANG (Open Resources and TOols for LANGuage). Retrieved from https://hdl.handle.net/11403/corpus14/v2

Reinhardt, J. (2019). *Les interrogatives directes tirées de dix romans policier* [Interrogative sentences from ten detective novels]. Constance, Germany: University of Constance, ORTOLANG (Open Resources and TOols for LANGuage). Retrieved from https://hdl.handle.net/11403/interrogatives-in-novels/v1

Tutin, A., & Hartwell, L. (n.d.). *Scientext. A French and English Corpus of Scientific Texts*. Retrieved from https://scientext.hypotheses.org/homepage

German

COSMAS I/II: Corpus Search, Management and Analysis System. (1991). Mannheim, Germany: Leibniz-Institut für Deutsche Sprache. Retrieved from http://www.ids-mannheim.de/cosmas2/

DWDS: Digitales Wörterbuch der deutschen Sprache (Digital Dictionary of German). (n.d.). Berlin, Germany: Berlin-Brandenburg Academy of Sciences. Retrieved from https://www.dwds.de/

Italian

Spina, S. (2015). *Corpora Di Italiano. Learner Corpora Di Italiano*. Perugia, Italy: University for Foreigners of Perugia. Retrieved from https://www.unistrapg.it/cqpwebnew/

Russian

Russian National Corpus. Moscow, Russia: Institute of the Russian Language, Russian Academy of Sciences. Retrieved from http://ruscorpora.ru/old/en/index.html

Spanish

Davies, M. (2002). *Corpus del Español: 100 million words, 1200s–1900s (Historical/Genres)*. Retrieved from http://www.corpusdelespanol.org/hist-gen/

Davies, M. (2016). *Corpus del Español: Two billion words, 21 countries (Web/Dialects)*. Retrieved from http://www.corpusdelespanol.org/web-dial/

Real Academia Española corpora. (2020). Madrid, Spain: Real Academia Española. Retrieved from https://www.rae.es/recursos/banco-de-datos

Other Resources

European Commission (2008). *SACODEYL: European youth language*. Retrieved from https://www.um.es/sacodeyl/

European Commission (2009–2011). *BACKBONE: Pedagogic corpora for content and language integrated learning*. Retrieved from http://projects.ael.uni-tuebingen.de/backbone/moodle/

Weisser, M. (2016). *Non-English, parallel & multilingual corpora: a selection*. Retrieved from http://martinweisser.org/corpora_site/corpora2.html