Chapter 7

Language Corpora for L2 Vocabulary Learning: Data-Driven Learning Across the Curriculum

Nina Vyatkina, University of Kansas

Introduction

Empirical instructed second language acquisition (ISLA) research on second language (L2) vocabulary has shown that data-driven learning (DDL), or teaching and learning languages with the help of corpora (large, structured electronic collections of texts), is beneficial for L2 vocabulary acquisition. Nevertheless, it is still far from becoming a common pedagogical practice, not least because few pedagogical manuals and user-friendly corpus tutorials have been published to date.

This chapter describes how DDL with an open-access German language corpus has been used across the curriculum in a German Studies program at a North American university. I report empirical results and present specific pedagogical suggestions and activities for using a corpus to enhance L2 vocabulary knowledge at all proficiency levels and show how DDL can help learners improve not only the breadth of their L2 vocabulary knowledge (the number of words the basic meaning of which the learner knows) but also the depth of this knowledge (Nation, 1990), including collocations, frequency, and grammatical patterns. Although this chapter uses a German program as a case study, its pedagogical suggestions can be applied to teaching any language for which open-access electronic corpora are available.

In what follows, I provide a brief overview of intersections between corpus linguistics and L2 vocabulary research and then focus in on ISLA and DDL. Next, I describe empirical DDL research results and how to connect it to teaching practice for L2 vocabulary. The rest of the chapter is devoted to a curricular proposal for DDL activities to teach vocabulary across an L2 curriculum, concluding with a brief summary and outlook.

Corpus Linguistics and L2 Vocabulary Research

There are numerous intersections between corpus linguistics and L2 vocabulary teaching and research. This is not surprising as corpus linguistics is predicated upon the postulate of the primacy of lexis (words and coselection of words) over

grammar in terms of meaning creation (Sinclair, 1987, 1991), and thus both disciplines share their major object of study. First, corpus analysis results have been used in L2 vocabulary research for reference purposes. Word frequency lists have been extracted from native corpora of different languages and used as a baseline for measuring L2 learners' vocabulary size (Laufer & Nation, 1995). Second, L2 textbook writers have used corpus-derived information for making decisions on what vocabulary to include and in what curricular progression. Third, more direct applications of corpora in teaching languages in general and L2 vocabulary in particular have been expanding since the 1980s. This teaching method and the associated research field became known as data-driven learning, or DDL (Johns, 1990). According to Boulton and Cobb (2017), DDL is academic inquiry "into the effectiveness of using the tools and techniques of corpus linguistics for second language learning or use" (p. 348).

It is also of note that the concept of lexis has been developing along similar lines in both corpus linguistics and L2 vocabulary research. Corpus linguists have argued against a strict separation of lexis and grammar and, instead, referred in their work to "lexico-grammar" (Sinclair, 1991), that is, tendencies of certain words to occur in certain grammatical patterns. This notion agrees with the concept of "depth" of L2 vocabulary knowledge, first put forth by Nation (1990, 2001). Whereas much of earlier research explored only the "size" of L2 vocabulary knowledge at the level of minimal form–meaning mapping (L1–L2 word translation), Nation has argued that many additional aspects must be included, such as the word's grammatical functions, usage frequency, register, semantic associations, and collocations. On a theoretical level, these approaches clearly align themselves with usage-based language learning theories (e.g., Ellis, 2014), which posit that languages are learned through repeated exposure to usage examples, progressing from individual items (words) to lexico-grammatical patterns to generalizations about abstract categories and principles (grammatical rules).

ISLA and DDL Principles Relevant to L2 Vocabulary Learning

Specific pedagogical principles that are beneficial for L2 vocabulary acquisition in instructed settings have been explored in empirical ISLA studies. Laufer (2017) summarizes these main principles as "The Three 'I's" necessary for successful vocabulary learning: *Input*, *Instruction*, and *Involvement*. More specifically, the focus of attention has been shown to be critical to what is learned (Barcroft, 2003) and form-focused instruction has been found to be effective (Peters, 2014; Webb & Kagimoto, 2011). Next, studies have shown that repeated exposure (Laufer & Rozovski-Roitblat, 2011), visual input enhancement (Peters, 2012; Sharwood Smith, 1993; Sonbul & Schmitt, 2013), and high involvement load (Kim, 2011; Laufer & Hulstijn, 2001) all lead to more L2 vocabulary learning. As such, these research results lend credence to Schmidt's (1990, 2001) noticing hypothesis,

which states that L2 units and features are better learned if they are attended to and noticed at some level of awareness.

Empirical DDL research has been developing mostly in parallel to mainstream ISLA research with few DDL scholars explicitly situating their studies within SLA frameworks (see Flowerdew, 2015, for a discussion). Nevertheless, many of DDL research designs and findings can be explained with a reference to the aforementioned theoretical and pedagogical principles. Typically, learners in DDL get exposed to Laufer's (2017) first "I" condition for successful L2 vocabulary learning, *Input*, through so-called concordances: stacked lines of text with the search item highlighted (e.g., bolded or colored) and centered (Figure 7.1). Concordances provide learners with several to hundreds to millions of examples (depending on how representative of the search item the chosen corpus is). This input is thus rich, repeated, concentrated, and graphically enhanced, which draws the learner's attention to the language patterns (Schmidt, 2001). Corpus examples also have the added benefit of representing naturally occurring language (vs. artificially created textbook examples). As Boulton and Tyne (2015) remark, "corpora bring to the fore a distilled set of authentic uses that the individual would be hard pressed to tease out of the data manually or based on occasional incidental encounters" (p. 303).

Laufer's second "I," *Instruction*, is primarily realized in DDL through the method of guided induction (Herron & Tomasello, 1992), "in which teachers help learners co-construct rules by directing their attention to relevant aspects in the input and asking guiding questions" (Cerezo, Caras, & Leow, 2016, pp. 265–266). Guided induction is operationalized in DDL as another set of "The Three 'I's" (Carter & McCarthy, 1995)—*Illustration* (learners are being exposed to corpus examples), *Interaction* (learners analyze the data and discuss the patterns), and *Induction* (learners induce rules from data analysis)—as well as the fourth "I" of the teacher's *Intervention* (Flowerdew, 2009). Several recent ISLA studies

Ein Extraleben bekommen gute Spieler vom	**Computer**	geschenkt, wenn sie eine besonders hohe Punk
Ich kann nur im	**Computer**	nachsehen, wenn Sie mir die Nummern sagen.
Er setzte sich wieder vor den	**Computer**	, legte eine neue Seite unter den Deckel und st
so nicht wegen des Computers, sondern trotz des	**Computers**	.
Verzweifelt versucht jeder zweite Brite, mit dem	**Computer**	zu sprechen - so eine Microsoft-Studie.
m Vorbild der Mobilfunkanbieter ihren Kunden die	**Computer**	kostenlos oder zu einem symbolischen Niedrigp
Der Vater kann auch in Herzhöfel am	**Computer**	arbeiten, und die Mutter ist nur vier Tage besc
Um Viertel nach fünf stand er auf, schaltete den	**Computer**	ein und begann, Seite für Seite von Alfred Dusl
rahren Bedürfnisse des Menschen im Umgang mit	**Computern**	und in der vernetzten Welt.

Figure 7.1. Excerpt from the DWDS corpus search results for *Computer* [1]

---

[1] Hyperlinks to all corpus search results, excerpts from which are represented in the figures, are listed as Figure References at the end of the chapter.

convincingly demonstrate that guided induction is superior to purely deductive or inductive teaching methods while focusing on lexico-grammar (see Cerezo et al., 2016, for an overview). Further, this method has been shown to be associated with learners' deep processing of and high cognitive engagement with the material (Leow, 2015), which accounts for Laufer's *Involvement*, the final "I" condition for successful L2 vocabulary acquisition. To summarize, Laufer's (2017) three conditions necessary for vocabulary learning are realized in DDL as follows:

- Input

  - Rich
  - Repeated
  - Concentrated
  - Enhanced
  - Authentic

- Instruction (guided induction)

  - Illustration
  - Interaction
  - Intervention
  - Induction

- Involvement

  - Cognitive engagement
  - Deep processing

Empirical DDL Research

Most DDL studies have employed variations of the guided induction method in their teaching interventions and explored the effectiveness of these variations in comparison with either one another or with non-DDL teaching methods, most frequently, deductive methods. To date, this research has accumulated a substantial body of evidence that DDL is a viable teaching approach. A detailed review of this literature is beyond the scope of this chapter, and the interested reader is referred to available narrative surveys (e.g., Boulton, 2017; Römer, 2011) and the first comprehensive meta-analyses (Boulton & Cobb, 2017; Lee, Warschauer, & Lee, 2018). Only a few relevant summative findings are discussed here. Boulton and Cobb's (2017) meta-analysis of research published until June 2014 shows that DDL is "a strong methodology for learning language per se, including lexicogrammar" (p. 381). More specifically, the meta-analysis demonstrates a large effect size of learners' improvement in their knowledge of single words and collocations following DDL interventions, and a medium effect size for DDL advantage over traditional materials and teaching methods. Boulton and Cobb also found that both hands-on DDL (learners' direct use of concordancers) and hands-off DDL (learners' work with teacher-printed concordances) are effective, with the hands-on

method being somewhat more efficient. Finally, the meta-analysis showed that DDL works not only for advanced but also for intermediate L2 learners. Lee et al. (2018) have confirmed and expanded Boulton and Cobb's (2017) findings in their meta-analysis that specifically targeted DDL vocabulary studies published through 2016. They found "a medium-sized effect on L2 vocabulary learning, with the greatest benefits for promoting in-depth knowledge to learners who have at least intermediate L2 proficiency" (p. 25) and also singled out a number of moderating variables such as corpus types and task types.

While these results are encouraging, the research syntheses also highlight some limitations of the field. The most drastic one relates to the target language: the overwhelming majority of DDL studies target English as a foreign language or an L2. Out of 64 studies that met Boulton and Cobb's (2017) criteria for their meta-analysis, only two addressed languages other than English (one Spanish, one mixed). All articles in the most recent journal special issue on DDL (Vyatkina & Boulton, 2017) also focus on English. While this trend reflects the worldwide spread and importance of English, the scarcity of documented DDL applications to other languages is regrettable because there are many excellent corpora in many languages that are being taught around the globe. This limitation is strongly interconnected with another one, namely, a limited understanding of lexico-grammar in DDL research. Since English is an analytic language without much inflection, DDL studies have primarily focused on the appropriate selection of content and function words but rarely on accuracy in inflectional morphology (Boers & Lindstromberg, 2012). Therefore, it is less clear how DDL fares with other languages. There are some notable exceptions, but overall, DDL studies and pedagogical suggestions for L2s other than English are few and far between (See, for example, Furniss, 2016, for Russian; Kennedy & Miceli, 2010, 2017, for Italian; Kerr, 2009 and Tyne, André, Benzitoun, Boulton, & Greub, 2014, for French; Mendikoetxea, 2014, for Spanish; and Neary-Sundquist, 2015 and Schaeffer-Lacroix, 2016, for German.)

The author of this chapter has contributed to filling this gap by conducting DDL studies using an open-access corpus to teach German to learners at different L2 proficiency levels at a U.S. university. In the first exploratory study, Vyatkina (2013) showed how this corpus was used with advanced learners (mostly graduate students) and outlined the progression of corpus-based activities from more teacher mediation to more learner autonomy. The next two studies focused on German verb–preposition collocations, a difficult lexico-grammatical construction for learners whose first language (L1) is English. Vyatkina (2016a) compared the effectiveness of teaching these collocations with hands-off DDL (printed concordances) and with a deductive method from the course textbook to low-intermediate learners. The study showed that the DDL method was better than traditional instruction for learning new collocations, but both methods were equally effective for improving the knowledge of previously learned collocations. Vyatkina (2016b) compared the effectiveness of hands-on and hands-off concordancing for teaching

the same verb–preposition collocations, but this time to high-intermediate learners. The results showed that both methods were equally effective. Some gains were also durable as evidenced on a delayed posttest. This study also found that learners improved regardless of their overall L2 proficiency, although there were some fine-grained differences depending on the test task and proficiency level. Furthermore, the study found that most learners very much liked DDL activities, and that even those who liked them less benefited from them. Finally, Vyatkina (2018) explored the feasibility of using corpus tools beyond concordancing for teaching different aspects of vocabulary knowledge to high-intermediate learners. This study found that learners successfully used a suite of DWDS tools (thesaurus, concordance, word profile, collocations) for researching German verb–noun collocations. As a result, they significantly improved their depth of knowledge of these collocations. It is important to note that these gains were equal to gains achieved from working with printed concordances in regard to word and collocation recall and surpassed paper-based gains in regard to morphological accuracy. Collectively, these studies have shown how an open-access German corpus was successfully used for teaching L2 vocabulary to learners at different L2 proficiency levels.

DDL Research and Practice

One final limitation of DDL research that is directly relevant for this chapter is that a great majority of studies reporting on teaching interventions are conducted by the researchers themselves. Although a few studies did (successfully) involve regular teachers (e.g., Vyatkina, 2016a; Boulton, 2010), it is common for DDL researchers to "lament the fact that corpora have not become widespread in language education" (Boulton & Tyne, 2015, p. 305; see also Frankenberg-Garcia, 2012; Römer, 2011). The main reason for this resistance of language educators to DDL is that using corpora may require substantial corpus linguistic skills on the part of both teachers and learners. Even if teachers were willing to invest their time and effort into learning these skills (in order to share them later with their students), they would be hard-pressed to find appropriate manuals. Although a number of general teacher guides on using corpora and sample DDL exercises have been published (Bennett, 2010; Frankenberg-Garcia, 2012; Kerr, 2009; Thurstun & Candlin, 1997; see also http://sites.psu.edu/calpercorpusportal/corpus-tutorial), tutorials that accompany specific corpora are frequently fairly technical and may be found daunting by teachers (but see Shaw, 2011, for a notable exception). As Boulton and Tyne (2015) note, "it is surprising that there are not more materials to exploit the interactive potential of hands-on corpus use" (p. 308). Such guides could go a long way toward "bringing corpora to the masses" (Boulton, 2011, p. 69) if they capitalize on what is familiar and usual in corpus explorations as opposed to what is unfamiliar and unusual. As Boulton (2011) argues, corpus tools have much in common with other widespread electronic tools (e.g., dictionaries and Google search), and the web itself can be considered a type of corpus. As most language teachers

and learners habitually engage in using these tools and materials, tutorials drawing on such parallels could provide step-by-step guidance for using increasingly more sophisticated corpus tools. Finally, teachers and language program directors (LPDs) may shy away from DDL simply because they do not see how they can incorporate them in their already full and busy syllabi and curricula, so modular proposals for incorporating DDL activities into regular syllabi are needed.

This chapter presents such a first attempt as a roadmap for using an open-access corpus and a suite of associated electronic tools for teaching L2 vocabulary across the curriculum.

A Curricular Proposal for Teaching L2 Vocabulary with Corpus-based Resources

The *Digital Dictionary of the German Language*

The electronic resource used in this curricular proposal is more than just a corpus. It is a suite of lexicographic resources titled the *Digital Dictionary of the German Language* (*Digitales Wörterbuch der Deutschen Sprache*), henceforth, DWDS (http://dwds.de). It is an open-access resource, supported and regularly updated and expanded by the Berlin-Brandenburg Academy of Sciences in Germany. The DWDS front page reflects these updates by listing links to a word of the day (*Wort des Tages*) and the newest entries (*neueste Artikel*). The empirical basis of DWDS comprises a number of large-scale German corpora that include historical corpora that go as far back as the 17th century, several newspaper corpora, and a number of specialized corpora such as transcribed oral interviews, film subtitles, and blogs. Some of the corpora are annotated for various linguistic categories. Most notably, the core corpus (*Kernkorpus*) of the 20th- and 21st-century German is annotated for parts of speech (POS). It is also equipped with tools that allow the user to search for collocations of a particular word with nouns, adjectives, verbs, and other POS separately. Moreover, the core corpus is balanced by time and text type: each decade is represented by approximately 100,000 words, which are in turn equally divided between four text types: fiction, nonfiction (e.g., guides and manuals), science, and newspaper. While all these corpora can be searched in DWDS, the website also integrates other resources, including electronic dictionaries and visualization tools. The interface is intuitive and user-friendly. The entry page for each word contains its definition, etymology, pronunciation, relevant grammatical information, compounds containing this word, synonyms, collocates (words frequently occurring together with the search word), example sentences, and more. By clicking different links on the front page and setting search filters, the user can find statistical information on the word frequency in different genres, at different time points, in comparison with other words, and more. Due to the richness of DWDS information and presentation formats, it can suit the needs of both scholars (e.g., corpus linguists and lexicographers) and nonspecialist users (e.g., teachers

and learners of German). For example, while specialists may be interested in conducting searches that require a sophisticated search syntax (e.g., find all nouns with the ending *-er* separated by two words from the beginning of the sentence but excluding the word *Computer* in all science texts from the 1980s), nonspecialist users can elicit meaningful information from a few typical example sentences containing the search word or commonly used visualization tools, such as the time line or word cloud (illustrated in Figures 7.4 and 7.5, respectively). Over recent years, the DWDS interface has undergone several rounds of substantial revamping to make it more user-friendly. The DWDS website contains links to instructions for using all its resources as well as to publications about the resources and studies conducted with their help. While these references are extremely helpful to specialist users, many of them are too technical for nonspecialist users. Moreover, there currently is no user guide in English, which is a serious obstacle for German learners at lower levels of proficiency. The desire to make this rich vocabulary-learning resource and similar resources more accessible to language teachers and learners provided the main impetus for writing this chapter.

Pedagogical Principles

This proposal is based on the pedagogical principles discussed above that have been shown to benefit L2 vocabulary learning in both ISLA and DDL research. It consists of form-focused teaching modules all of which follow the guided induction method with teacher scaffolding adjusted to each respective curricular level. It is intuitive that tasks should progress from more scaffolded and controlled to less scaffolded and independent ones, also termed "soft" and "hard" DDL, respectively (Gabrielatos, 2005). However, as Boulton and Tyne (2015) note, "there is no single 'right' way to use corpora. It is important for each teacher to choose what is appropriate for him/herself given the learners' needs and available resources" (p. 309). Although activities are presented here according to the L2 proficiency level—from novice to advanced—learner familiarity with corpora should also figure into the equation. For example, if corpora are introduced to advanced L2 learners for the first time, the teacher will want to start with softer DDL versions. In contrast, even low-intermediate L2 learners can complete independent corpus searches after a certain amount of DDL practice. Respectively, all tasks presented here can be varied in regard to the presentation format (teacher-fronted presentation, partner work, individual work), task outcome (open-ended or closed-ended), medium (printed paper materials or online corpus searches), and other characteristics. The selection of vocabulary to be explored will depend on the course goals. Students can complete such worksheets with words preselected by the teacher in accordance with the topics covered in class or from textbook vocabulary lists as well as with self-selected words from class readings or independent readings. While working with textbook vocabulary, the tasks may involve both expanding

the depth of knowledge and verifying textbook information with real-life data. The following sections present sample DDL tasks as modules, each titled according to the DWDS tool used and the task to be completed.

DDL Tasks for Novice L2 Learners and/or Novice Corpus Users

Although there may be many ways of introducing corpora, this chapter presents an approach in which learners receive "a gentle introduction [to] corpus use rather than being dropped in at the deep end" (Boulton & Tyne, 2015, p. 308). To introduce DWDS to the students, the teacher can project its main page on a big screen and explain that this is a new generation electronic dictionary that provides many different types of information about words. The key to success is referring to operations and tools that are familiar to students from their everyday technology use, such as Internet and electronic dictionary searches, word clouds, time lines, and so forth. The teacher also needs to explain that DWDS is a monolingual resource; that is, it does not provide L1–L2 translations. For those, learners should use bilingual dictionaries (e.g., http://dict.leo.org/englisch-deutsch/).

Basic search: Form and meaning.

Many foreign language programs start introducing target language vocabulary with cognates: words that have similar meaning and form in both the L1 and L2 (either as borrowings or due to the shared historic origin, as is often the case for German and English, two Germanic languages). Starting on the very first day of a beginning German class, the teacher can enter the cognate word *Internet* in the DWDS search line and then draw the students' attention to the very top line of the resulting page (Figure 7.2) that will list *Internet*, *das*. The teacher can explain that German has many English borrowings, particularly in the fields related to electronic technology. The teacher may want to point out that the spelling of these borrowings in German is often similar to English, but all nouns in German have to be capitalized and assigned grammatical gender (therefore, one needs to learn German nouns together with their articles—*der*, *die*, *das*). The teacher may also click on the speaker icon to listen to the pronunciation of the word. In addition to cognates, good candidates for corpus demonstrations even at beginning stages of German proficiency are so-called pseudoborrowings, or false anglicisms (Furiassi & Gottlieb, 2015), that is, words that have an English form but a different meaning or no meaning in English (e.g., *Handy*—"cell phone," *Twen*—"person between 20 and 30 years of age," *zappen*—"to flip through channels").

Time line: Word frequency over time.

Next, the teacher can point to the word frequency (*Worthäufigkeit*) tool represented as a time line graph in the upper-right corner of the screen (Figure 7.3). Some students may be familiar with this visualization technique from the Google

Internet, das

Grammatik	Substantiv (Neutrum) · Genitiv Singular: **Internet/Internets** · wird nur im Singular verwendet
Aussprache	◄))
Worttrennung	In-ter-net
Herkunft	Englisch
Wortbildung	mit ›Internet‹ als Erstglied: ↗Internet-Adresse ... 86 weitere

[Bedeutung]   [Thesaurus]   [Typische Verbindungen]

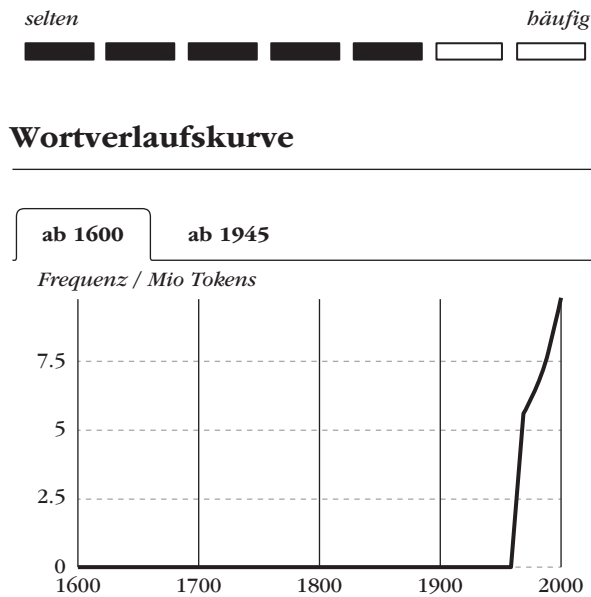Figure 7.2. Top part of the DWDS vocabulary entry page for *Internet*

selten   *häufig*

■ ■ ■ ■ ■ □ □

Wortverlaufskurve

ab 1600	**ab 1945**

Figure 7.3. Word frequency time line graph for *Internet*

Books Ngram Viewer tool (https://books.google.com/ngrams) or any time line graph. It should be easy to interpret this DWDS view that illustrates that German did not have the word *Internet* until the mid-20th century but that its frequency has been growing exponentially. At this point, the teacher may want to mention that all examples and visualizations in DWDS draw from millions of words from texts from historical and contemporary German in different genres. Differences between curated corpora such as DWDS and the web are also worth noting. Frequency data retrieved with this DWDS tool is stable and controlled, and the user can easily access source texts from which this numerical data has been harvested (including examples in context, bibliographic citations for each text, and

sometimes the full text if freely available on the Internet). In contrast, frequency data retrieved from the web with the Google Books Ngram Viewer is unstable (it may change any day if new texts are added), largely undifferentiated by text type, and not traceable back to the original texts.

Word cloud: Collocations.

Finally, the teacher can scroll down to the word cloud view that shows typical word combinations (*Typische Verbindungen*), or collocates, of the search word (Figure 7.4). Many students will be familiar with the word cloud view as well, so volunteers can be called on to explain that the words that most frequently appear together with *Internet* are represented in larger letters. While looking at the collocates of *Internet*, the students will recognize further English borrowings such as *Computer* and *Homepage* and other cognates such as *Telefon*, *Informationen*, and *Adresse*. Students may also notice that the word cloud contains both the forms *Surfen* and *surfen*. The teacher can explain that the first one stands for the noun ("surfing")—because it is capitalized—and the second one for the verb ("to surf"). In other words, this corpus tool differentiates between different POS.

Activities similar to those described above are a good starting point for introducing corpus-based activities to learners at any L2 proficiency level if they are not familiar with corpora or with this particular corpus. For intermediate and advanced learners, corpus-based activities can then be expanded to include more independent work and analysis as well as progressively more sophisticated DWDS tools. Sample activities are described in the following sections.

Adresse Computer E-Mail Fernsehen

Handel Handy Homepage

Informationen Infos Medien Mobilfunk

Nutzung Surfen Telefon Zugang

abgerufen abgewickelt anbietet bestellen

buchen ermöglicht heruntergeladen informieren

mobile nutzen surfen verbreitet

vertreiben veröffentlicht übertragen

Figure 7.4. Collocates of *Internet*

DDL Tasks for Intermediate L2 Learners and/or Intermediate Corpus Users

Concordances: Lexico-grammatical analysis.

Concordancers and concordances are the most frequently used corpus tools and output format in DDL applications. The introduction to concordancing activities should, like with any new tools and activities, begin with a teacher's demonstration on the big screen. To show students specific usage examples of the search word, the teacher can scroll down below the time line and display the list of corpora with associated search word frequencies (Figure 7.5). If, for instance, *chatten* ("to chat") is the search word, students can see at a glance that there are hundreds of examples of this verb in corpora of newspapers, blogs, and film subtitles. By clicking on a hyperlink associated with a corpus, the teacher can display these examples in the form of concordances (Figure 7.6a). The first view displays complete sentences containing the search word, but an extended cotext view or a Key Words In Context (KWIC) view can be selected as well (Figures 6b and 6c). The KWIC view lends itself to form-focused tasks, such as the one described in this section. By seeing different verb forms—which are bolded and centered—*chatten*, *chattet*, *chattete*, *gechattet*, students become aware of how a verb borrowed from English takes on German morphology. Students can be asked to identify specific verb forms (past tense, third-person singular, etc.). The teacher should underscore that at this stage of the analysis, the students should be concerned not with translating each example sentence but rather with identifying general patterns of usage. In contrast, the sentence view or the extended context view lends itself to activities that combine the focus on both meaning and form (described later in this chapter). The activity described here can be conducted in a teacher-fronted format with concordances displayed on a big screen. Alternatively, the teacher may print preselected concordances and assign them to partner groups or as individual homework. The latter format will be especially productive if the aim is to cover several different words as they can be divided between groups or individual students and the results can be later shared with the whole class.

The same procedure can be followed for analyses of lexico-grammatical collocations rather than individual words. A good candidate is verb–preposition collocations (e.g., *warten auf*—"to wait for"), a notorious area of difficulty for English learners of German (see Ecke, 2015, p. 99; Ecke & Hall, 2000, p. 31). In the first iteration of this activity, students work with teacher-printed concordances that best illustrate the usage of the focal collocations. Later (as with any activity presented in this chapter), the assignment can be changed to student online searches for the target verb–preposition collocations in the corpus. Detailed guidelines and sample worksheets for these activities are available in Vyatkina (2015a, 2015b), an open-access online publication.

Korpustreffer

Referenzkorpora

  | DWDS-Kernkorpus (1900–1999) (5)
  | DWDS-Kernkorpus 21 (2000–2010) (30)
  | Deutsches Textarchiv (1473–1927) (8)

Zeitungskorpora

  | Berliner Zeitung (1994–2005) (135)
  | Tagesspiegel (1996–2005) (197)
  | Die ZEIT (1946–2016) (310)

Spezialkorpora

  | Referenz- und Zeitungskorpora (aggregiert) (685)
  | Blogs (396)
  | Polytechnisches Journal (2)
  | Filmuntertitel (68)

Figure 7.5. Frequencies of *chatten* in different DWDS subcorpora

Die Zeit, 24.10.2016 (online)
Der Münchner Amokschütze hat im Vorfeld seiner Bluttat im Internet vermutlich mit sich selbst **gechattet**.

Die Zeit, 13.10.2016, Nr. 43
Im Netz wird gesurft, gespielt, **gechattet** – und leider oft auch gehasst.

Die Zeit, 27.09.2016 (online)
Am Ende liegt es im Ermessen der Nutzer, ob sie lieber abhörsicher(er) **chatten** und dafür auf Komfort verzichten.

Die Zeit, 14.08.2016 (online)
Wir wissen nicht, wer miteinander **chattet**", sagte Maaßen.

Figure 7.6a. Concordances with *chatten* (sentence view)

Suite of tools: Get to know a word.

With the help of DWDS, students can significantly deepen their knowledge of both new and previously learned words, since each entry page contains a plethora of lexicographic information. Students can be assigned a word and asked to complete a worksheet in which they list its definition, relevant grammatical information, compounds containing this word, synonyms, collocates, and example sentences with

Die Zeit, 13.10.2016, Nr. 43

Im Netz wird gesurft, gespielt, **gechattet** – und leider oft auch gehasst.
Doch Vorsicht: Wer fiese Kommentare und Fotos postet, kann sich strafbar machen und Besuch von der Polizei bekommen. von Johanna Schoener

Die Zeit, 27.09.2016 (online)
Die künstliche Intelligenz, mit der Facebook und Google seine Dienste in Zukunft für Nutzer verbessern möchten, steht also gewissermaßen der sicheren Kommunikation im Weg.
Am Ende liegt es im Ermessen der Nutzer, ob sie lieber abhörsicher(er) **chatten** und dafür auf Komfort verzichten.
Oder ob Chatbots und der Versand von Gifs für sie unerlässlich sind.

Figure 7.6b. Concordances with *chatten* (extended view)

Wir wissen nicht, wer miteinander	**chattet**	", sagte Maaßen.
...ht zufolge mit möglichen IS-Kontakten in Saudi-Arabien	**gechattet**	.
... Person, die etwas tun könnte, sitzt in New Jersey und	**chattet**	mit ihrer Facebook-Gemeinde!"
	Gechattet	hatte sie demnach mit einem der Verdächtig
...chten beantworte, swipen sie nach links oder rechts und	**chatten**	mit ihrenmatches . Für Sex, für Romantik, m
Mit Luther per SMS	**chatten**	oder Buchstaben fühlen - in der Grimmwelt H
...Sie Ihr Snapchat-Videos, folgen Sie Ihr auf Instagram,	**chatten**	Sie bei WhatsApp mit ihr.

Figure 7.6c. Concordances with *chatten* (KWIC view)

this word. As a final task, the students can write their own example sentences with the word. These worksheets can be compiled in a vocabulary log to be submitted to the teacher iteratively over a semester, or students can be asked to enter this information in a class Wiki for everybody to see. As opposed to all activities described above, this activity can lead to divergent results as learners will search the corpus directly and independently and have a choice of which hyperlinks to click and which examples to select. The key to this activity's success is careful scaffolding. First, the students should be warned that DWDS provides a large amount of information about each word, and covering all of this information would be beyond the limits of this particular assignment. Next, they should be given clear directions and a model—a completed worksheet prepared in advance by the teacher. Ideally, the class would meet in a computer lab (or in a regular room with each student having a laptop, tablet, or smartphone with Internet access) where the teacher could demonstrate the completion of the assignment for the model word on the big screen with students mirroring all steps on their individual computers. In this way, they will learn how to be selective and not become overwhelmed by the richness of information in DWDS. This guided induction approach, in which students receive careful teacher guidance and then work with corpora independently and inductively, has been shown to lead to higher levels of learner engagement with and attention to the L2 material. This results in higher learning gains in comparison with deductive methods in which the teacher provides all information to the students (see Section "ISLA and DDL Principles Relevant to L2 Vocabulary Learning").

DDL Tasks for More Advanced L2 Learners and Corpus Users

More advanced students, after they have become familiar with DWDS by completing some of the activities described above, can be assigned more complex DDL tasks that are open ended and independent as well as involve more sophisticated DWDS tools. In what follows, the target L2 proficiency level will be referred to as "advanced," although this designation is relative depending on the definition of the term. For example, Vyatkina (2016b) has shown that students who have reached the German proficiency at or above the Common European Framework of Reference (Council of Europe, 2001) B1 level (or intermediate high on the ACTFL scale) can successfully complete such tasks if given sufficient initial guidance by the teacher.

Word profile: Lexico-grammatical collocations.

Research results show that although advanced learners have good receptive knowledge of collocations (i.e., they understand their meaning), they have much poorer productive knowledge (Schmitt, 2010). For example, students of German can easily translate the verb–noun collocation *Hausaufgaben machen* as "to do homework" because they are being exposed to the noun *Hausaufgaben* early on and repeatedly. In contrast, in their own German production, calque errors (i.e., using a cognate in contexts that require a noncognate) are frequent because the prototypical translation of "to do" is *tun* and not *machen* (Nesselhauf, 2003; Rott, 2016). Moreover, many learners are not aware that the collocation *Hausaufgaben erledigen* ("to complete homework") is much more frequent than *Hausaufgaben machen* and that the noun is predominantly used in German in its plural form. Form-focused DDL activities can help raise learners' awareness of the verb component in such collocations and improve their productive knowledge (Vyatkina, 2018). As usual, the activity should first be demonstrated by the teacher, then completed by learners under teacher's supervision, and only then assigned as partner work or an independent task.

For this task, the learners will work with the DWDS word profile (*Wortprofil*) tool. By iteratively clicking certain hyperlinks, the user is taken to the word cloud view (Figure 7.7a), then to the view that arranges collocations in ranked lists according to their POS and syntactic role (Figure 7.7b), and then to concordances with usage examples for each collocation (Figure 7.7c). This last step is important because it shows that members of a verb–noun collocation (bolded in this view), though directly connected syntactically, can be separated in a sentence by a considerable number of words due to the German word order rules (Rott, 2016). The instructions direct the learners' attention to the most frequent verbs that go together with the search noun and typical usage patterns, including word order and inflectional morphology. As a result, learners remember more verb collocations of the target nouns and use them more accurately in subsequent production tasks (Vyatkina, 2018).

Erledigen   Erledigung

Klassenarbeiten   Klausuren   Mittagessen

Referate   abfragen   abgearbeitet   abschreiben

aufbekommen   aufhatte   aufkriegen

beaufsichtigen   betreuen   brütet   erledigen

erledigte   finanzpolitischen   gemacht

gemachten   haushaltspolitischen   helfen   nachgeholt

nichtgemachte   strukturpolitischen

unerledigten   vernachlässigt

Figure 7.7a. Collocates of *Hausaufgabe* (word cloud)

hat Adjektivattribut	logDice	Freq.
1. unerledigt	9.9	50
2. erledigt	9.2	26
3. nichtgemacht	8.2	7
4. strukturpolitisch	7.6	9
5. gemacht	6.1	25
6. finanzpolitisch	5.9	16
7.	5.3	5

ist Akk./Dativ-Objekt von	logDice	Freq.
1. erledigen	10.4	666
2. machen	6.5	3096
3. vernachlässigen	6.2	22
4. aufhaben	6.1	12
5. aufbekommen	5.7	7
6. abschreiben	5.6	13
7. nachholen	5.6	11

hat Genitivattribut	logDice	Freq.
1. Kind	2.6	14
2. Politik	1.9	7

in Koordination mit	logDice	Freq.
1. Klassenarbeit	11.5	16
2. Referat	9.3	19
3. Note	7.5	8
4. Unterricht	7.2	12

Figure 7b. Collocates of *Hausaufgabe* by POS/syntactic role (word profile)

14:	Die Welt, 17.09.2004
	Ganztagsschulen sollten eingeführt, **Hausaufgaben** zusammen mit den Lehrern **erledigt** werden.
15:	Bild, 21.08.2004
	Doch für unseren Traum vom Pokal-Endspiel müssen wir solche **Hausaufgaben erledigen**.
16:	Der Tagesspiegel, 08.07.2004
	Vor allem aber begutachten sie, wie Leth seine **Hausaufgaben erledigt** hat.
17:	Berliner Zeitung, 21.06.2004
	Wichtig dabei ist, dass man vorher seine **Hausaufgaben erledigt** hat.
18:	Die Welt, 14.05.2004
	Bevor Mehdorn mit einem Börsengang liebäugelt, sollte er endlich seine **Hausaufgaben erledigen**.
19:	Der Tagesspiegel, 09.05.2004
	Und dann hat Ihr Sohn auch noch seine **Hausaufgaben** nicht **erledigt,** obwohl Sie ihn fünf Mal ermahnt haben.

Figure 7c. Concordances for the collocation *Hausaufgaben erledigen*

Word profile: Semantic prosody of near-synonyms.

The word profile tool also supports comparisons of the usage patterns of near-synonyms, or words with very similar meanings (Storjohann, 2009). In particular, students can explore the "semantic prosody" (Louw, 2000) of such words, or their tendency to appear with similar or different collocates. For example, if one explores the word profile of the adjective *gewaltig* ("powerful, enormous"), one can enter its near-synonym *heftig* ("forcible, fierce") in the field *Gemeinsamkeiten mit* ("common-alities with") on top of the page (Figure 7.8a). The resulting view displays lists of common collocates for *gewaltig* and *heftig*. One can see that both adjectives frequently appear together with nouns from the semantic fields "natural disasters; explosions."

If, instead, the operator is changed from "commonalities" to "differences" (*Unterschiede zu*), one can see that while *gewaltig* appears much more frequently with nouns such as *Summe*, *Ausmaß* (finances), *heftig* goes together with *Kontroverse*, *Auseinaderstzung*, *Debatte* (polemics) (Figure 7.8b). All these collocations can be explored in context by clicking on each collocate and perusing the resulting concordance view. Students can be asked to note and discuss their findings with their partners and then share the results with the class. The teacher can select pairs of near-synonyms for this task from relevant curricular topics and texts as well as with the help of DWDS itself, as front pages for each word entry (*Wortinformation*) contain lists of synonyms.

Time line: Word usage history.

To quote DWDS developers, their time line tool supports explorations of "word careers" (*Wortkarrieren*) over more than 400 years (https://www.dwds.de/d/neues). The time trajectory of the word usage frequency allows for tracking a word's emergence, waxing and waning, and even disappearance from the German language. The basic time line view (Figure 7.2) automatically appears in the upper-right corner of the front page for each word entry and presents the word history at a glance
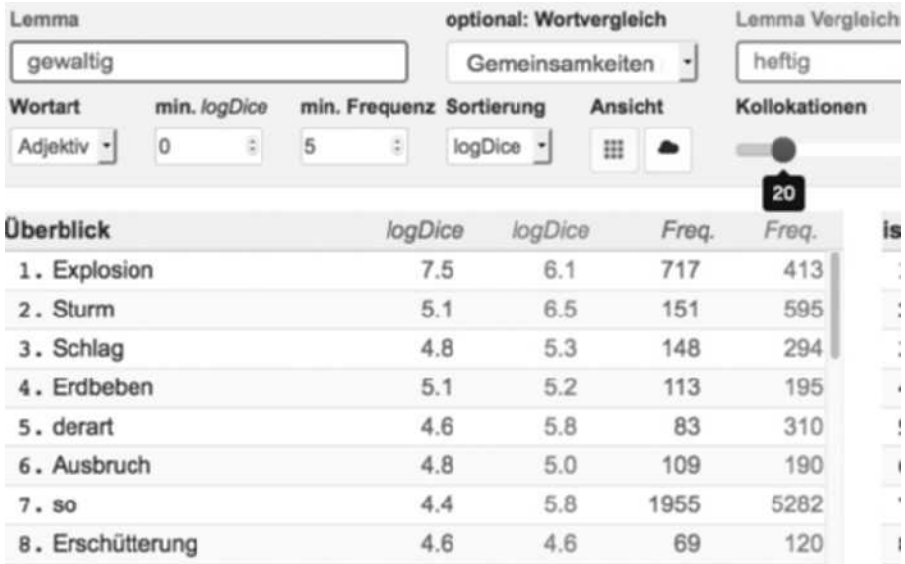
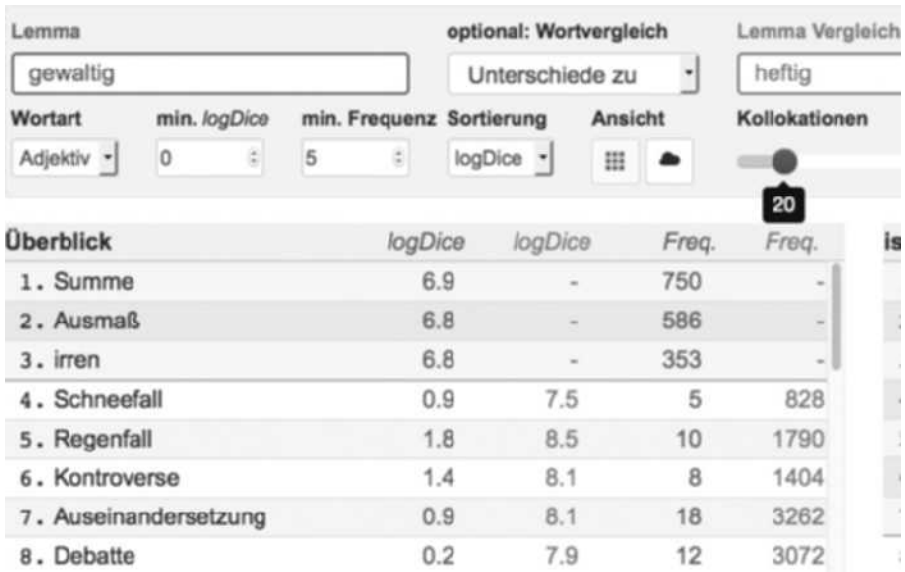Figure 7.8a. Common collocates of *gewaltig* and *heftig* (word profile)

Figure 7.8b. Different collocates of *gewaltig* and *heftig* (word profile), marked by different colors (not visible here)

while also marking whether it is rare or frequent (*selten* or *häufig*). This line graph is easily interpretable for users at any proficiency level. More advanced learners can explore further functionalities of this tool. By clicking on the graph, the user gets to

a full-screen view in which one can set various filters. For example, while researching the word *Herausforderung* (challenge), one can select a view containing separate time lines for the relative word frequency (per million words) in different text types (Figure 7.9). One can see that the word first emerged around 1765 in newspaper texts. Its frequency was slowly growing over the next centuries but it really took off after the 1970s. A divergence between text types is also noticeable, with the newspaper frequencies growing rapidly and frequencies in other text types lagging behind. Further, by moving the cursor over the curves, one can see frequencies of the word at each specific point in time. Figure 7.9 shows that in the present decade's newspaper texts, *Herausforderung* has so far been used with an average frequency of 36.59 times per million words, which amounts to 11,978 occurrences in total. If the user clicks the box with this information, a link to corpus examples from the respective decade and in the respective text genre appears. By following this link, the user gets the view of concordances and can explore the use of *Herausforderung* in context. A comparison of the 17th–19th-century examples with more contemporary examples will lead the user to conclude that while, earlier, this word was used in reference to fights and wars (as in "challenge to a duel"), its contemporary usage is almost exclusively restricted to the meaning of "challenge" in the sense of "an issue to be overcome" which had originally been absent in German but was relatively recently borrowed from English (see Kramsch, 1993, pp. 31–32). The teacher can assign such a historical analysis of different words to different students, who will then share the results of their research with the class.

Figure 7.9. Extended view of the time line graph for *Herausforderung*. Curves for different genres are marked by different colors (not visible here)

Beyond a historical-linguistic interest, such activities have practical implications for learners' choice of words in their own L2 use. For example, dictionaries list *obwohl* and *obgleich* as alternatives for translating the English adverb "although." However, the DWDS time line tool shows that whereas the frequencies of the former have been continuously rising since 1800 throughout the 20th century, the picture for the latter is exactly the opposite (see Vyatkina, 2013). This information provides a clear indication that *obwohl* is the preferred alternative today, which may lead the learner to opt in favor of using *obwohl* instead of *obgleich*.

Suite of tools: Long-term student projects.

Finally, advanced students may be assigned group projects or individual projects that can last from several weeks to an academic term. For example, students can keep vocabulary logs in which they record various types of information about newly learned words or expanded information about previously learned words. This activity can be further modified if applied to course texts that are incorporated in the target corpus. In such cases, the teacher can design DDL activities in which learners focus on specific words they encountered while reading, taking a top-down, text-to-corpus approach (Charles, 2007; Johns, Hsingchin, & Lixun, 2008; Vyatkina, 2016b). Such an approach would address Widdowson's (2000) notorious criticism of corpus examples being taken out of a larger context and thus lacking relevance to learner interests. Further, if deemed necessary for the course goals, advanced students may also be taught how to use more advanced search filters (e.g., https://www.dwds.de/d/suche). Another idea for a project is using corpora as a reference resource while completing writing assignments. Students can be asked to work on the appropriateness of their vocabulary use by consulting corpora. They can be instructed to document their searches and attach brief descriptions of these searches and corpus snapshots to their rough or revised writing assignment drafts. Although a detailed outline of such projects is beyond the scope of this chapter, a number of DDL studies attest to their benefits (see Yoon, 2016, for a recent study and overview). Finally, a group project can be designed in which students create multimedia tutorials for teaching languages with corpora and which would culminate in end-of-semester presentations and/or posting the tutorials online (see https://scholarblogs.emory.edu/germangrammar/ for outcomes of a similar project). Such projects would prepare students for life-long learning with corpora and using them as a reference resource beyond the classroom.

Conclusion

Empirical research has by now convincingly demonstrated that DDL with electronic corpora is effective for many areas of ISLA including vocabulary. All necessary conditions for successful L2 vocabulary acquisition are present in DDL; corpora and associated visualization tools provide learners with rich, dense, and

authentic input, and the preferred DDL method—guided induction—ensures high levels of learner involvement with and deep processing of this input. However, the paucity of accessible DDL application models for language teachers, especially for languages other than English, has so far hindered a wide spread of DDL in teaching practice. LPDs may be especially hesitant to implement DDL in their programs thinking that it may require a complete revamping of the curriculum.

This chapter intended to counter this misconception and contributed to building a bridge between DDL research and teaching practice. It presented sample modules using an open-access corpus and a suite of associated electronic tools for teaching German vocabulary across the curriculum. Preparation of a more comprehensive teacher guide is currently underway. This proposal presents a case study situated in a specific local context: teaching German as a foreign language to North American university students with a specific electronic resource. All activities described here (or their variations) have been used in a German Studies program in German courses at different proficiency levels and the results were very positive (Vyatkina, 2013, 2016a, 2016b, 2018). However, the principles and activities described here can be adapted to various target languages for which electronic corpora are available.[2] The model presented here conceives of DDL as a "corpus apprenticeship" (Kennedy & Miceli, 2010) that is introduced into a curriculum in a gradual way without radically changing the existing teaching approach. It is hoped that this proposal will help language teachers and program directors become familiar with open-access corpus resources and tools and start using their rich potential in DDL applications to the benefit of their learners.

Figure References (hyperlinks to DWDS search results retrieved May 26, 2018)
Figure 7.1: https://www.dwds.de/r?q=Computer&corpus=korpus21&-date-start=2000&date-end=2010&genre=Belletristik&genre=Wissen-schaft&genre=Gebrauchsliteratur&genre=Zeitung&format=kwic&sort=ran-dom&limit=50
Figure 7.2–7.4: https://www.dwds.de/wb/Internet
Figure 7.5: https://www.dwds.de/wb/chatten
Figure 7.6a: https://www.dwds.de/r?corpus=zeit;q=chatten
Figure 7.6b: https://www.dwds.de/r?q=chatten&corpus=zeit&date-start=1946&date-end=2016&format=max&sort=date_desc&limit=50&p=1
Figure 7.6c: https://www.dwds.de/r?q=chatten&corpus=zeit&date-start=1946&date-end=2016&format=kwic&sort=date_desc&limit=50&p=1
Figure 7.7a: https://www.dwds.de/wb/Hausaufgabe
Figure 7.7b–7.7c: https://www.dwds.de/wp/Hausaufgabe

---

[2] For example, American English (COCA, https://corpus.byu.edu/coca/), Portuguese (http://www.corpus-doportugues.org/), Russian (http://www.ruscorpora.ru/en/), Spanish (http://www.corpusdelespanol.org/).

Figure 7.8a: https://www.dwds.de/wp?q=gewaltig&comp-method=inter-section&comp=heftig&pos=0&minstat=0&minfreq=5&by=logDice&lim-it=20&view=table

Figure 7.8b: https://www.dwds.de/wp?q=gewaltig&compmethod=inter-section&comp=heftig&limit=20&minstat=0&minfreq=5&by=log-Dice&view=table

Figure 7.9: https://www.dwds.de/r/plot?view=1&norm=date%2B-class&smooth=spline&genres=1&grand=1&slice=10&prune=0&win-dow=3&wbase=0&logavg=0&logscale=0&xrange=1600%3A2016&q1=Herausforderung

References

Barcroft, J. (2003). Effects of questions about word meaning during L2 Spanish lexical learning. *The Modern Language Journal*, *87*(4), 546–561.

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

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, *32*, 83–110.

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, *60*(3), 534–572.

Boulton, A. (2011). Bringing corpora to the masses: Free and easy tools for interdisciplinary language studies. In N. Kübler (Ed.), *Corpora, language, teaching, and resources: From theory to practice* (pp. 69–96). Bern, Switzerland: Peter Lang.

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.

Boulton, A., & Tyne, H. (2015). Corpus-based study of language and teacher education. In M. Bigelow & J. Ennser-Kananen (Eds.), *The Routledge handbook of educational linguistics* (pp. 301–312). New York: Routledge.

Carter, R., & McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics*, *16*(2), 141–158.

Cerezo, L., Caras, A., & Leow, R. P. (2016). The effectiveness of guided induction versus deductive instruction on the development of complex Spanish *gustar* structures. *Studies in Second Language Acquisition*, *38*, 265–291.

Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes*, *6*(4), 289–302.

Council of Europe. (2001) *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg: Language Policy Unit. Retrieved from http://www.coe.int/en/web/common-european-framework-reference-languages

Ecke, P. (2015). Was (oft lustige) Fehler und Wortfindungsprobleme über Wortschatzlern- und Verarbeitungsprozesse enthüllen. In M. Löschmann & M. Löschmann (Eds.), *Humor im Fremdsprachenunterricht* (pp. 96–111). Frankfurt am Main: Peter Lang.

Ecke, P., & Hall, C. J. (2000). Lexikalische Fehler in Deutsch als Drittsprache: Translexikalischer Einfluss auf drei Ebenen der mentalen Repräsentation. *Deutsch als Fremdsprache*, *37*(1), 30–36.

Ellis, N. C. (2014). Cognitive AND social language usage. *Studies in Second Language Acquisition*, *36*(3), 397–402.

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics*, *14*(3), 393–417.

Flowerdew, L. (2015). Data-driven learning and language learning theories: Whither the twain shall meet. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 15–36). Amsterdam, Netherlands: John Benjamins.

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

Furiassi, C., & Gottlieb, H. (2015). *Pseudo-English: Studies on false anglicisms in Europe*. Berlin, Germany: De Gruyter.

Furniss, E. A. (2016). Teaching the pragmatics of Russian conversation using a corpus-referred website. *Language Learning & Technology*, *20*(2), 38–60.

Gabrielatos, C. (2005). Corpora and language teaching: Just a fling, or wedding bells? *TESL-EJ*, *8*(4), 1–37.

Herron, C., & Tomasello, M. (1992). Acquiring grammatical structures by guided induction. *The French Review*, *65*(5), 708–718.

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., Hsingchin, L., & Lixun, W. (2008). Integrating corpus-based CALL programs in teaching English through children's literature. *Computer Assisted Language Learning*, *21*(5), 483–506.

Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate Italian students 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.

Kim, Y. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, *61*(1), 100–140.

Kramsch, C. (1993). *Context and culture in language teaching*. Oxford: Oxford University Press.

Laufer, B. (2017). The Three "I"s of second language vocabulary learning: Input, instruction, involvement. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning*, (Vol. 3, pp. 343–354). London: Routledge.

Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*, 1–26.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*, 307–322.

Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. *Language Teaching Research*, *15*, 391–411.

Lee, H., Warschauer, M., & Lee, J. H. (2018). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, 1–34. doi:10.1093/applin/amy012

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. New York: Routledge.

Louw, B. (2000). Contextual prosodic theory: Bringing semantic prosodies to life. In C. Heffer, H. Sauntson, & G. Fox (Eds.), *Words in context: A tribute to John Sinclair on his retirement* (pp. 48–94). Birmingham, UK: University of Birmingham.

Mendikoetxea, A. (2014). Corpus-based research in second language Spanish. In K. L. Geeslin (Ed.), *The handbook of Spanish second language acquisition* (pp. 11–29). Malden, MA: Wiley-Blackwell.

Nation, P. (1990). *Teaching and learning vocabulary*. Rowley, MA: Newbury House.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.

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

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, *24*, 223–242.

Peters, E. (2012). Learning German formulaic sequences: The effect of two attention-drawing techniques. *Language Learning Journal*, *40*, 65–79.

Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research*, *18*, 75–94.

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

Rott, S. (2016, March). *Explicit learning of multi-word chunks: Insights into the development of phraseological competence in L2 German*. Paper presented at the Pennsylvania State University, State College, PA.

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.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge, UK: Cambridge University Press.

Schmitt, N. (2010). *Researching vocabulary*: *A vocabulary research manual*. Basingstoke, UK: *Palgrave* Macmillan.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, *15*(2), 165–179.

Shaw, E. M. (2011). Teaching vocabulary through data-driven learning. Provo, UT: Brigham Young University. Retrieved from http://corpus.byu.edu/coca/files/Teach- ing_Vocabulary_Through_DDL.pdf

Sinclair, J. M. (1987). The nature of the evidence. In J. M. Sinclair (Ed.), *Looking up: An Account of the COBUILD Project in Lexical Computing* (pp. 150–159). London: Harper Collins.

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, *63*(1), 121–159.

Storjohann, P. (2009). Plesionymy: A case of synonymy or contrast? *Journal of Pragmatics, 41*, 2140–2158.

Thurstun, J., & Candlin, C. N. (1997). *Exploring academic English: A workbook for student essay writing*. Sydney, Australia: National Centre for English Language Teaching and Research.

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. (2013). Discovery learning and teaching with electronic corpora in an advanced German grammar course. *Die Unterrichtspraxis/Teaching German*, *46*(1), 44–61.

Vyatkina, N. (2015a). *Corpus-driven instruction*. Center for Applied Second Language Studies (CASLS), University of Oregon. Retrieved May 26, 2018, from http://caslsintercom.uoregon.edu/content/18926

Vyatkina, N. (2015b). *Corpus activity: German verb-preposition collocations*. Center for Applied Second Language Studies (CASLS), University of Oregon. Retrieved May 26, 2018, from http://caslsintercom.uoregon.edu/content/18927

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, March). *Data-driven learning for instructed second language acquisition: Deepening the L2 vocabulary knowledge*. Paper presented at the American Association of Applied Linguistics conference, Chicago, IL.

Vyatkina, N., & Boulton, A. (Eds.). (2017). Corpora in language learning and teaching [Special issue]. *Language Learning & Technology*, *21*(3).

Webb, S., & Kagimoto, E. (2011). Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics*, *32*, 259–276.

Widdowson, H. G. (2000). On the limitations of linguistics applied. *Applied Linguistics*, *21*(1), 3–25.

Yoon, C. (2016). Concordancers and dictionaries as problem-solving tools for ESL academic writing. *Language Learning & Technology*, *20*(1), 209–229.