

Discovery Learning and Teaching with Electronic Corpora in an Advanced German Grammar Course

Nina Vyatkina

University of Kansas

Abstract

This study describes the design and implementation of a usage-based and corpus-based advanced German grammar course. Teaching materials for the course included DWDS, or *Digitales Wörterbuch der deutschen Sprache*: a large, representative, free and publicly available corpus of contemporary German texts. The article outlines specific theoretically informed principles for course design and presents a logical progression of corpus-based grammar-teaching activities as used in this course. It also reports participants' post-course reactions which were very positive. The article contains practical recommendations for educators interested in trying out corpus-based activities for teaching German grammar at different proficiency levels as well as to different participant groups. Furthermore, it promotes a more holistic perspective toward grammar as a meaning-making resource compatible with innovative approaches to grammar pedagogy.



Grammar instruction has always been a subject of interest and discussion in foreign language (FL) pedagogy and language teaching research. Recently, approaches that emphasize attention to context and explicit instruction, making clear meaning-form connections, have been on the rise. These approaches have been fleshed out in individual journal articles (e.g., Chavez, 2011; Zyzik, 2008), collections of empirical studies and thought pieces (Katz & Watzinger-Tharp, 2009), and monographs (Byrnes, Maxim, & Norris, 2010; Larsen-Freeman, 2003). Such approaches to grammar pedagogy are supported by prominent researchers who advocate developing multiple literacies as the core organizing principle of foreign language curricula, if those curricula are to lead to advanced proficiency levels (Byrnes, 2005; Kern, 2000; Swaffar & Arens, 2005). As Byrnes et al. (2010) explain, the “language-based literacy framework presents FL studies as a project involving an FL being learned by literate adults who wish to make meaning with their new language, an ability that requires them to know precisely how language forms enable making meaning in the first place and to learn language in that fashion” (p. 31).

The pedagogical approaches addressed above find conceptual grounding in usage-based theories, in which grammar is understood as an abstract representation of language in the mind of speakers which is, nevertheless, strongly tied to each speaker's experience with natural language (Bybee, 2006; Halliday, 1996; Tyler, 2010). One important corollary of this is inseparability of vocabulary and grammar. In other words, researchers advocating usage-based approaches argue that the use of language is always vocabulary specific and certain words tend to be used in certain constructions. In the applied sense, these principles mean that learning a

language goes far beyond the acquisition of underlying rules (cf. Chomsky, 1965) and involves substantial practical encounters of the learner with the target language. Although these encounters are notoriously limited in foreign language learning settings, various electronic resources have been coming to the rescue of the learner and teacher with recent technological advances (see Arnold & Ducate, 2011, for overview).

These resources include language corpora, or large electronic collections of natural texts organized according to certain principles and typically equipped with built-in search engines. Corpora can provide language learners with increased exposure to authentic language examples and invaluable information about the distribution and frequency of different linguistic features in different contexts. Furthermore, studies in corpus-based pedagogy have shown that engaging with corpora promotes development of learners' higher-level cognitive functions: inductive data-driven learning, or learning by discovery; problem-solving; and developing analytical skills and independent language learning strategies (e.g., Bernardini, 2002; Johns, 1986, 1997; Kennedy & Miceli, 2002; Mishan, 2004). Yet, despite these advantages, "corpus literacy" (Mukherjee, 2006; O'Sullivan, 2007) remains an underdeveloped aspect of contemporary multiple literacies in language education. The present article contributes to this line of research and pedagogy by reporting on the implementation of an advanced usage-based and corpus-based German grammar course and by making pedagogical suggestions.

Current Applications of Corpora in Language Pedagogy

The fact that corpora are treasure troves of language does not necessarily make them easy to use for instruction purposes. As Brown (2007) notes, "many of the widely accessible corpora were created as tools for linguistic research and not with pedagogical goals in mind" (p. 308). Therefore, corpora require considerable pedagogical mediation before they can be used in the classroom (see Widdowson, 2000, for discussion). Recently, a number of corpus-based dictionaries and student grammars have been published which crystallize and summarize relevant information for teachers and students without the need to search corpora directly (see Dodd et al., 2003, for German). For those who are interested in "getting their hands messy" with real corpus data, teacher guides are emerging (e.g., Bennett, 2010; Frankenberg-Garcia, 2012; McCarthy, 2004; McCarten, 2010). However, such resources are still few and far between, especially beyond the area of teaching English. Researchers describing the state of the art conclude that "considerably more corpora of English than other languages are publicly available and used in pedagogical contexts" (Römer, 2011, p. 207) and that "there still seems to be a long way to go before corpora can be understood and used by language teachers in general" (Frankenberg-Garcia, 2012, p. 460). Finally, the number of empirical studies exploring effectiveness of corpus-based teaching is small, albeit growing (see Kerr, 2009; Römer, 2011 for overviews).

Another important gap is a lack of suggestions for how to teach grammar in a narrower sense, i.e., morpho-syntactic features,¹ with corpora. An overwhelming majority of pedagogical suggestions involve searching corpora for specific lexical items, i.e., words, and exploring their use in context (see Dodd, 1997; Möllering, 2004, as examples for German). A notable exception is Zinggeler's (2006) *Grimmatik*: a pedagogical grammar with exercises based entirely on Grimm's fairy tales contained in the COSMAS corpus. Admittedly, searching for more

¹ Morpho-syntactic features are linguistic forms expressing certain functions such as verb inflections marking tense, person, number, etc.

abstract categories, such as “present perfect tense constructions,” requires morpho-syntactically annotated corpora and more sophisticated search rules. Despite these challenges, some researchers strongly argue in favor of corpora as “a wondrous resource for teaching grammar” (Nelson et al., 2002, p. 299, see also Greaves & Warren, 2007; Vannestål & Lindquist, 2007). Moreover, using corpora for grammar consultations may be even easier for learners than using reference grammars because it requires less sophisticated metalinguistic knowledge (Kaltenbock & Mehlmauer-Larcher, 2005).

This study addresses the abovementioned gaps by reporting on an innovative course that utilized corpora for integrated instruction in three interrelated dimensions of foreign language grammar: form, meaning, and use (Larsen-Freeman, 2003; Zyzik, 2008). It reports on the implementation of a course with learners who were advanced speakers of German but novices in dealing with electronic corpora. Although some examples of corpus-based activities in this study involve nuanced and subtle distinctions of grammar appropriate for higher proficiency levels, the suggested progression of activity types from ones with more teacher mediation to ones based on more learner autonomy is universal. In this way, the study aims to serve as a guide for educators interested in trying out corpus-based activities for teaching grammar at any proficiency level.

Course Description

Participants and Goals

The semester-long course was titled “Usage-based grammar of contemporary German” and was offered as an elective for both upper-level undergraduate students and graduate students at the department of Germanic languages and literatures at a large public university in the Midwest. Two undergraduate seniors (who graduated immediately upon finishing the course) and five graduate students enrolled in the course. The language of instruction was German. The number of grammar courses previously taken by the participants and their metalinguistic knowledge varied; however, none of them had had previous experience with learning grammar with corpora. The course met for three hours per week in a room equipped with a computer podium, projection screen, and both wired and wireless Internet access. At the beginning of the course it was decided that meetings in a computer lab with individual work stations were not necessary because all participants had personal laptop computers and agreed to bring them to class as requested by the instructor.

The first objective of the course was a review of the grammar of contemporary standard German in order to further increase the level of linguistic proficiency and awareness of the participants who were already proficient speakers of the language. The second objective was to provide these proficient speakers with tools for autonomous life-long learning.² Furthermore, the course syllabus listed a number of specific learning outcomes formulated as abilities participants were expected to acquire by the end of the course (see Appendix A).

² Another, somewhat subsidiary, objective was to encourage participants to look at corpus tools from the pedagogical perspective as future teachers of German. However, this discussion is beyond the scope of this article.

Course Components and Materials

All course components (class discussions and presentations, problems and other homework, quizzes, the midterm and final examination) included material and exercises from the main course book and electronic corpora as well as additional materials (such as journal articles and newspaper columns). The thematic progression of the course followed the sequence of grammar topics in the main course book: Rug and Tomaszewski's (2009) pedagogical grammar *Grammatik mit Sinn und Verstand*. This book adopts a functional approach by presenting advanced German grammar structures in examples with context and exercises interspersed with humorous explanations and illustrations. Importantly, the book systematically points out alternative forms which can be used in German to express similar meanings, some of which, however, may be more or less felicitous in specific contexts.

The Corpus

Most corpora used in the course were components of DWDS, or *Digitales Wörterbuch der deutschen Sprache*. DWDS is an electronic corpus of German of the 20th and 21st centuries, developed, published, and maintained by the *Berlin-Brandenburgische Akademie der Wissenschaften* (<http://www.dwds.de>). I chose DWDS because this is a large, representative, free and publicly available corpus of German supported by a respectable linguistic institution and with built-in corpus search and analysis tools. Moreover, the corpus is annotated for Parts-of-Speech (POS) and lemmata. In other words, it allows searches not only for specific word forms but also for word classes (e.g., past participles) and all inflected forms of one base form, or lemma (e.g., forms *gehe*, *gehst*, *ging*, *gegangen*... of the lemma *gehen*).³

DWDS contains several subcorpora, although we mostly worked with the core corpus (*Kernkorpus*): the German corpus of the 20th century, containing 100 million words. I focused on the core corpus because it is balanced chronologically and by text type (genre). More specifically, it contains approximately 10 million words per decade of the 20th century with approximately 25% of the texts in each of four genres: journalism (*Zeitung*), literary texts (*Belletristik*), scientific literature (*Wissenschaft*), and other non-fiction (*Gebrauchsliteratur*). For some tasks, I used additional DWDS corpora such as *Wendekorpus* (the corpus of transcribed oral interviews with people from East and West Berlin in the 1990s) and *ZEIT-Korpus* (the corpus of articles published in the paper and online editions of *Die Zeit* newspaper from 1946 to 2009).⁴ Additionally, a corpus of computer-mediated communication (separate from DWDS) was used that will not be discussed here⁵.

³ See <http://www.dwds.de/ressourcen/korpora/> and Klein and Geyken (2010) for a comprehensive description of DWDS. For an alternative German corpus with similar robust characteristics, see COSMAS (<http://www.ids-mannheim.de/cosmas2/>).

⁴ It is important to note that DWDS is a live project, and the corpus is being continuously updated and expanded (for a description see <http://www.dwds.de/>). The information given here reflects the state at the time of the course (Spring 2011). DWDS also includes subcorpora of Austrian and Swiss German which were not explored in the course.

⁵ Dortmund Chat-Korpus: <http://www.chatkorpus.tu-dortmund.de/>

The Pedagogical Sequence

This study, unlike many corpus-based teaching publications which utilize more or less randomly selected examples of instructional activities, presents a logical progression of corpus-based grammar-teaching activities as used in this course. They were carefully sequenced from activities that involved much teacher mediation to ones assuming more learner autonomy (Breyer, 2009; Vannestål & Lindquist, 2007). Another related principle was a progression from indirect to direct corpus applications. In indirect applications, the teacher prepares corpus-based handouts and worksheets for student work, whereas in direct applications, students peruse corpora under teacher's guidance or independently while completing teacher-designed or self-designed tasks. The sequence of activities is listed below and described in detail in the following sections.

1. Introduction to DWDS and corpus searches.
2. Teacher-designed indirect exercises.
3. Teacher-designed and teacher-led collaborative direct corpus searches.
4. Teacher-designed independent direct corpus searches.
5. Student-designed independent direct corpus searches.

Introduction to DWDS and Corpus Searches

At the beginning of the course, the students were given a brief introduction to corpora with a summary handout (containing the information about the corpus given above). They were then asked to familiarize themselves with the DWDS information pages describing the project and the corpora. I also asked the students to register as users because a free registration provides access to more corpus texts. Furthermore, I explained and demonstrated basic principles of corpus queries and search output formats, as described below.

If one enters a word into the search line of DWDS, a number of “panels” with a range of information about the search word appears (*DWDS-Wörterbuch*, *DWDS-Wortprofil*, and others). Initially, I only demonstrated how to work with the panel *DWDS-Kernkorpus* by enlarging it to the full screen mode. This panel presents the search output in form of ‘concordances’: stacked lines of text with the search word appearing in the middle and highlighted as well as with a left and right context of 5–10 words (see Fig. 1). In the left corner of each line, one can see the year and genre of the source text. The user can specify some output settings (*Darstellung*) such as number of lines appearing on the screen, length of context (a line or a paragraph), ascending or descending order according to date, etc. Furthermore, the user can change the filter settings for the source text (*Suchfilter*) so that only texts from a certain period or genre (or even a specific author) are searched.

Next, being a POS-annotated corpus, DWDS allows searches for specific word classes. The POS tags (assigned by corpus developers to each word) are invisible to the user. However, if she enters, for example, the symbols \$p=VVPP into the search line, all past participles of full (lexical) verbs from the focal corpus will appear one under another in concordance lines. At the beginning of the course, I gave participants a short list of basic search string examples for lemmas, word forms, POS, as well as combinations thereof (Appendix B). Furthermore, I showed them where to find a comprehensive list of search strings and examples (*Hilfe anzeigen* and *STTS-Tagset*).

Teacher-Designed Indirect Exercises

For this type of exercise, the teacher finds clear-cut corpus examples of a particular linguistic feature, prints them on worksheets, and adds stepwise guidelines for inductive awareness-raising activities. It is advisable to start with brief corpus excerpts that can be expanded later. Modality was our first grammar topic in the course, and one of the focal constructions was *sein + zu + Infinitiv*. The textbook briefly states that this construction can express both possibility and obligation and is synonymous to passive voice constructions with modal verbs. Based on my teaching experience, distinguishing between the two modal meanings (also called epistemic and deontic modality, respectively) of this construction is difficult for even advanced learners of German. To draw learners' attention to the use of this construction in context, I prepared the following concordance activity.

First, I established the search string that retrieves the target grammatical construction from DWDS. Components of this string have the following functions:

| | |
|---------------------|---|
| sein with \$p=VAFIN | -> any form of the lemma <i>sein</i> but restricted only to auxiliary verb occurrences (to block <i>sein</i> occurrences as a possessive pronoun) |
| #3 | -> no more than 3 words should occur between <i>sein</i> and <i>zu</i> ⁶ |
| zu | -> any form of the lemma <i>zu</i> (which is an uninflected particle, therefore does not change) |
| #0 | -> no words should occur between <i>zu</i> and the infinitive |
| \$p=VVINF | -> an infinitive of any full (lexical) verb |

The whole string also needs to be put into quotation marks.

Next, I selected a few representative concordance lines from a recent time period that included the focal construction used to express both epistemic and deontic modal meanings. Then, I created a worksheet with concordance lines (Fig. 1) and associated questions (Excerpt Set 1).

Excerpt Set 1. Worksheet Questions to Concordance Lines with *sein + zu + Infinitiv*

1. Lesen Sie die Sätze mit *sein + zu + Infinitiv* in Zeilen 20, 21, 27, 28. Geht es hier um Notwendigkeit oder (Un-)Möglichkeit? Formulieren Sie die Sätze anders, mit anderen Grundverben, die eine modale Bedeutung haben können.
2. In welchem Sprachstil können diese Konkordanz-Beispiele gebraucht werden? Schriftsprache, Umgangssprache, im neutralen Stil?
3. Wie sind die Sätze mit *sein + zu + Infinitiv* in Zeilen 30, 31 anders? Geht es hier um Notwendigkeit oder (Un-)Möglichkeit? Formulieren Sie die Sätze anders, mit anderen Grundverben, die eine modale Bedeutung haben können.
4. In welchem Sprachstil können diese Konkordanz-Beispiele gebraucht werden? Schriftsprache, Umgangssprache, im neutralen Stil?
5. Können Sie bestimmen, ob es in Zeilen 18, 19 um Notwendigkeit oder (Un-) Möglichkeit geht?
6. Wie ist die Bedeutung der Sätze in Zeilen 21, 22, 23, 24, 25 ähnlich?

⁶ Corpus linguists evaluate corpus searches in terms of *precision* and *recall*. The former measure addresses the extent to which a search retrieves ONLY relevant examples, and the latter the extent to which it retrieves ALL relevant examples. Limiting search string length enhances the precision but reduces the recall. For example, if one limits the search to elements co-occurring in a span of 1-5 words, elements separated by 10 words will not be retrieved. However, most irrelevant examples (such as elements belonging to different clauses) will be blocked. Users are advised to find a strategic 'golden middle' according to their purposes.

| | | | |
|------------|--|------|---|
| 18 1999 ZE | ...von Sozialleistungen zur Arbeit motiviert werden. Da | ist | tatsächlich viel zu verändern . Tony Blair hat außerd... |
| 19 1999 ZE | ... schafft man kein Vertrauen bei der Bevölkerung. Was | ist | zu tun ? Kompromisse müssen gefunden werden. |
| 20 1999 ZE | ...n Medizin im Frankfurter Krankenhaus Nordwest. Dabei | ist | der Säurerückfluß gut zu therapieren . Protonenpumpen... |
| 21 1999 ZE | ...h in der Wüste; der Dissens zwischen Metz und Ratzinger | war | nicht zu schlichten . Beide stritten darüber, an wel... |
| 22 1998 ZE | ...edite zu verkaufen, für unter ihrer Würde. Die Kluft | ist | kaum zu überbrücken . Traditionelle Bankiers legen im... |
| 23 1998 ZE | ...mit geringem Metallanteil anzusprechen. Vom Auto aus | sei | das kaum zu schaffen . Vermintes Gelände ist selten s... |
| 24 1998 ZE | ...em fusselfigen Hemd herumzubürsten und zu klopfen. Da | ist | nichts zu machen . Sascha findet irgendwo noch ein an... |
| 25 1998 ZE | ..., Lehmhütten, vor denen Brennholz liegt. Kein Mensch | ist | zu sehen . Der Kopilot gibt mir ein Zeichen, und ich ... |
| 26 1998 ZE | ...er und präsender war, weil es "den" Feind gab? Daran | wäre | doch ernstlich zu zweifeln - und so nun wiederum Eppler... |
| 27 1998 ZE | ...garter Landesarbeitsamtschef. Auch aus der Industrie | ist | Selbstkritik zu hören . Es stimmt schon, daß manche u... |
| 28 1998 ZE | ...nd Vevey bis zum 10. Januar 1999. Die Botschaft? Sie | ist | nur schwer zu formulieren . Nachdenken über Gentechni... |
| 29 1998 ZE | ...erhalb der Staatengemeinschaft auseinander. Was also | ist | von ihr bestenfalls zu erwarten ? Die Außenministerin... |
| 30 1998 ZE | ...Mullah verkündet sie. Verfügung: Alle Mädchenschulen | sind | zu schließen . Verfügung: Öffentliche Bäder sind zu s... |
| 31 1998 ZE | ...ulen sind zu schließen. Verfügung: Öffentliche Bäder | sind | zu schließen . Verbot: Kein Körperteil einer Frau dar... |

Figure 1. Selected Concordance Lines with *sein* + *zu* + *Infinitiv*

In class, I distributed the worksheets to the students and asked them to work individually. After they were finished, we discussed the results together as a group. We arrived at the conclusion that the first group of examples conveys the modal meaning of possibility and the second group the meaning of necessity (obligation), which could be decided based on the minimal context presented by concordance lines. Next, we agreed that whereas the first group of examples is not tied to a specific speech register, the second group represents the bureaucratic style of orders and regulations (explicitly expressed by the word *Verfügung* occurring in lines 30 and 31). Furthermore, we concluded that nuances of modal meaning cannot always be easily disambiguated. For instance, line 19 can be reformulated either as *Was kann getan werden?* or as *Was muss getan werden?* Finally, the learners could see that the focal construction is also frequently (lines 21–25) used to express negative epistemic modality, i.e., impossibility (in conjunction with the negation *nicht* or *kein*) or near impossibility (in conjunction with the adverb *kaum*).

This activity allowed the students to analyze how a construction is used, find alternative expressions, as well as arrive at generalizations. This inductive approach applied to authentic language examples has led to a much better understanding of this construction than the textbook presentation.

Teacher-Designed and Teacher-Led Collaborative Direct Corpus Searches

With these activities, I wanted to show students what types of information can be gained from corpora about frequency and distribution of linguistic features. I demonstrated these activities on the big screen while students followed all steps on their computers. I will illustrate this approach by an activity with a focus on impersonal expressions. The course book first contrasts active and passive expressions and then presents alternatives for the latter by giving examples of different impersonal constructions such as the passive voice (*werden* + *Partizip II*), constructions with the indefinite pronoun *man* and adjectives with the suffix *-bar*, as well as constructions with *sich lassen* + *Infinitiv* and *sein* + *zu* + *Infinitiv*. So I selected these constructions for corpus demonstrations.

In preparation for class, I established search strings for all focal constructions, outlined all activity steps for myself, and prepared worksheets with a grid for the students to enter the informa-

tion found. In class, I opened the corpus panel *DWDS-Kernkorpus* and the statistics panel (*Wortverlauf*) and projected them onto the big screen. I entered each target construction in turn into the core corpus search line, retrieved concordance lines with this construction highlighted, and asked the students to note the total number of hits (*Treffer*) found in the corpus. Then we looked at the timeline (*Wortverlauf*) panel which has the shape of a stacked columns chart (Figs. 2–3). The height and colors of each column visualize the frequency (number of hits) of the target construction by genre and decade. We discussed what we saw in each graph and the students entered the following information into their worksheets: In what genres is the construction used more often? How does the use of the construction change over time? Is the trend ascending or descending? Are there any peaks or dips? Having repeated all steps for each of the target constructions, we discussed what we found in *plenum*.

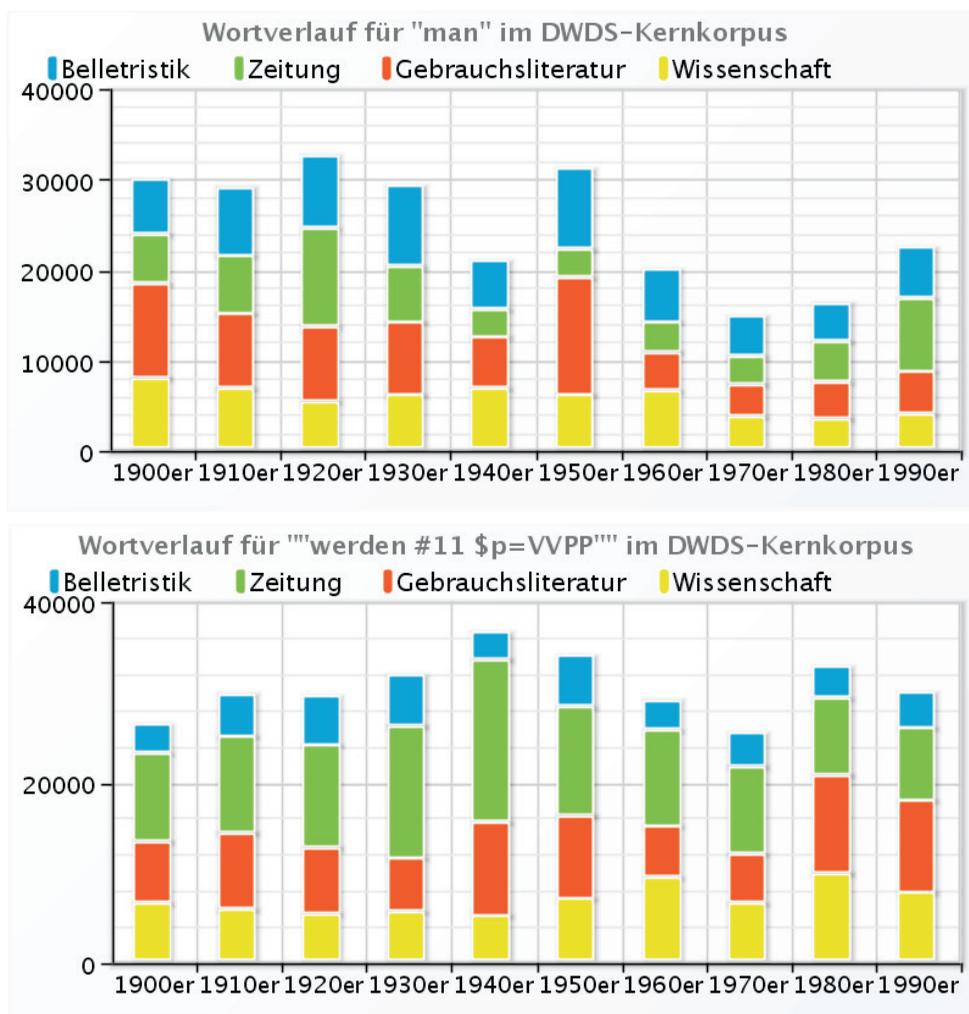


Figure 2. Frequency Graphs across Genre and Decade for *man* (above) and the Passive (below)

The following results emerged from the observation and discussion. First, the *man* constructions (Fig. 2, above) and the passive voice constructions (Fig. 2, below) turned out to be the most frequent impersonal constructions in the core corpus. With respect to genres, all imper-

sonal constructions are relatively more frequent in journalism and infrequent in literary texts, although *sich + zu + Infinitiv* and *man* constructions are used with somewhat higher frequency in fiction. Some genre-related information gained from the corpus confirmed our expectations as well as the information from the course book. For instance, *man* appears more frequently in non-fiction (e.g., in instructions and recipes) and adjectives with the suffix *-bar* in scientific texts. The timeline, however, afforded discoveries that were not necessarily anticipated either by the teacher or the students. We found that *man* was used more frequently in the first half of the 20th century than in its second half. For passive constructions, the frequency was increasing from decade to decade in the first half of the century (or showed a rising dynamic) but started falling in the second half. However, there appears to be a trend toward a rising dynamic again toward the end of the century (in the 1980–90s).

Our next, most striking, discovery was for the passive voice construction with the sentence-initial impersonal pronoun *Es*. The course book simply states that *Es* is frequently used in passive constructions. However, the core corpus timeline shows a discernable peak for *Es + Passiv* in the 1940s and 1950s (Fig. 3), more pronounced than for other passive constructions.⁷ A similar, although smaller, surge can be observed for the 1910s and 1920s. When one maps this observation onto the history of the 20th century, one can clearly see a prevalence of this impersonal construction in the war and post-war years. Additionally, whereas it is more frequently used in journalism across the timeline, the 1940s peak is also associated with a much more frequent use in non-fiction. Extracting longer contextualized examples from that period (Excerpt Set 2) helps shed more light onto this latter finding: *Es + Passiv* constructions are typical not only for journalism reporting about war events (a) but also for soldier letters and diaries (b) as well as for the language of the Nuremberg proceedings (c) which constitute the bulk of the non-fiction texts for that period.

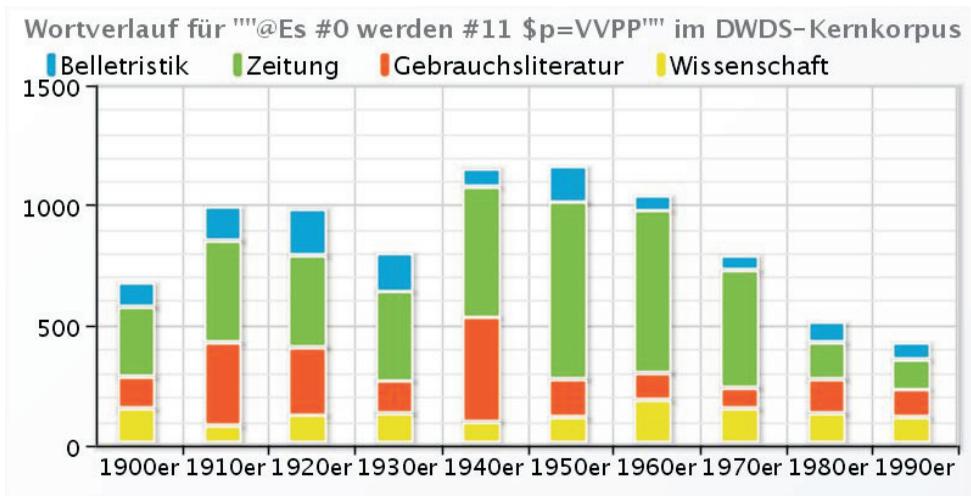


Figure 3. Frequency Graph across Genre and Decade for *Es + Passiv*

⁷ Note that the data in Fig. 3 (*Es + Passiv*) represent a subset of the data in the right half of Fig. 2 (all passive constructions).

Excerpt Set 2. Examples for *Es* + *Passiv* from the 1940s

- a. ***Es wurden*** Polizei- und Militärverstärkungen **angefordert**. (o.A., Britisches Blutbad in Indien, in: Völkischer Beobachter (Berliner Ausgabe) 20.03.1940, S. 2, S. 2)
- b. 5.9. 44 — *Das Lazarett bleibt nicht lange. Es wird nichts ausgepackt. Die Stadt liegt unter Beschuß.* (Tagebucheintrag Ernst G. vom 01.09.1944, Feldpost-Archive mkb-fp-0270, S. 139)
- c. ***Es wurde*** schon über alle Affidavits zu einem früheren Zeitpunkt **gesprochen**, den ich während der Pause werde feststellen lassen. (o.A., Zweihundertneunter Tag, Donnerstag, 22. August 1946, in: Der Prozess gegen die Hauptkriegsverbrecher vor dem Internationalen Militärgerichtshof Nürnberg, Nürnberg: Internationaler Militärgerichtshof 1947, S. 27768).

Thus, corpus visualization tools helped us see that not all impersonal constructions are “created equal.” *Es* + *Passiv* has a double impersonal effect by using both an indefinite pronoun and agentless passive. The DWDS timeline clearly shows that this construction has surged during war times but has been gradually falling out of favor (more so than other impersonal constructions, cf. Fig. 2) in all genres since the post-war years.

Teacher-Designed Independent Direct Corpus Searches

After completing a number of teacher-led exercises, the students were ready to conduct independent corpus searches in response to teacher-defined tasks. A task that will be demonstrated here was focused on German pronominal adverbs, or *da*-compounds (*daran*, *darauf*, *dagegen*...). *Da*-compounds function as an important cohesive and reference device in German but are essentially absent from contemporary English (barring the genre-restricted usage of *thereof*, *thereby*...). This linguistic feature was chosen because research has shown that American learners of German often misuse and/or misunderstand appropriate contextual conditions for using *da*-compounds (Belz, 2005; Belz & Vyatkina, 2008). The quoted research also suggested some corpus-based activities for awareness-raising which I applied in this course to the DWDS-corpus. I selected one *da*-compound⁸ for each course participant and distributed the following instructions for a homework assignment (Excerpt Set 3).⁹

Excerpt Set 3. Worksheet Questions to *da*-compounds

1. Tragen Sie Ihr *da*-Kompositum in die Suchbox des DWDS-Kernkorpus ein und starten Sie die Suchmaschine. Notieren Sie die *Trefferanzahl*.
2. Klicken Sie auf *Wortverlauf*. Welche Tendenzen merken Sie in Bezug auf Textsorten und Zeitverlauf?
3. Klicken Sie auf *Kollokationen*. In der Kolonne *w2* sehen Sie die Liste von 25 häufigsten “Nachbarn” Ihres Suchwortes. Notieren Sie aus dieser Liste 5–7 Vollverben (keine Modalverben!); 2–4 Nomen; 2–4 andere Wortarten (z.B. Adjektive, Adverbien).
4. Klicken Sie auf *Suche* neben dem häufigsten Nachbarwort, wählen Sie 1 Beispiel mit dem Wort und klicken Sie auf das rot markierte Wort. Kopieren Sie das ganze Beispiel im Kontext und fügen Sie es in ein Word-Dokument ein.
5. Wiederholen Sie Schritt 4 für noch 9 Nachbarn Ihres Suchwortes, die Sie in (3) aufgelistet haben.

⁸ I performed a preliminary corpus search and assigned the most frequent *da*-compounds. However, I eliminated compounds that are frequently used in functions other than pronominal adverbs such as *damit* which also appears as a subordinating conjunction. Differentiating homonymic grammatical functions was the focus of another activity not reported here.

⁹ The corpus interface was substantially revamped when our course was in progress. The *Kollokationen* feature can be accessed via the old interface (<http://retro.dwds.de/>). A new tool presenting similar information which is being continuously improved is *Wortprofil*.

6. Jetzt haben Sie 10 Beispiele mit Ihrem *da*-Kompositum und seinen häufigsten Nachbarn. Drucken Sie diese Seite aus.
7. Unterstreichen Sie das Wort/ die Wortgruppe, auf die Ihr *da*-Kompositum verweist. Steht dieses Wort / diese Wortgruppe vor oder nach dem *da*-Kompositum?

During the next classroom session, participants presented their results and we discussed them as a group. The first batch of findings was related to *da*-compound collocates, or their frequent contextual neighbors. Since *da*-compounds are functional words consisting in part of prepositions that are governed by specific verbs, nouns, and adjectives, it was important to raise learners' awareness as to what such typical collocations are. This information served as a pointer for the learners to what collocations they may want to prioritize for learning (e.g., *darauf* most frequently goes together with the verbs *hinweisen* and *achten* and the adverbs *bald* and *gleich*).

Next, and perhaps most importantly in this task, learners extracted paragraph-long examples and explored *da*-compound use in this extended context. In textbooks, *da*-compounds are typically presented as 'linguistic pointers' (or reference devices) referring to a preceding noun (Excerpt 4a: *davon* refers back to *Kaffee*). After our activity, the learners became aware that actual *da*-compound functions are much more varied. In particular, they found many examples in which the compounds pointed forward and/or to elements longer than single words (Excerpt 4b: *davon* refers forward to the *daß*-clause).

Excerpt Set 4. Examples for *da*-compounds

- a. *Als ich anfang, wußte ich nichts über Kaffee. Inzwischen verstehe ich mehr **davon** als mancher, der in der Branche arbeitet.* (Hars, Wolfgang, Nichts ist unmöglich! Lexikon der Werbesprüche, Frankfurt a.M.: Eichborn 1999, S. 407)
- b. *Der Unterzeichnete ist **davon** überzeugt, daß die Staatspolizeistelle von einer Überstellung des Verurteilten an die Staatsanwaltschaft abgesehen und die Hinrichtung selbst angeordnet haben würde.* (Hannover, Heinrich, Die Republik vor Gericht 1975 - 1995, Berlin: Aufbau-Verl. 1999, S. 293).

Therefore, this activity gave the learners an opportunity not only to obtain general distribution and frequency information about the focal feature but also to observe and make generalizations about its use in context.

Student-Designed Independent Direct Corpus Searches

After several rounds of teacher-designed activities like ones described above, the students were ready to try to formulate their own corpus queries. Our grammar topic of the week was sentence connectors. I decided that this topic was well suited for the first fully independent activity because searches for a separate POS or word are easier to formulate for novice corpus users than searches for syntactic constructions. We completed an introductory *in plenum* activity devoted to the differentiation between contextual uses of the conjunction *nachdem*, the combination of the preposition and the article *nach dem*, and the adverb *danach*. In this activity, students filled gaps in printed corpus concordance lines with appropriate connectors. Then, as homework I asked the students to come up with a corpus search question about conjunctions or prepositions, perform the search, and prepare to present the results during the next class. I also gave them some pointers as to what questions lend themselves to direct corpus queries (Appendix C).

Most students chose to compare the distribution and frequency of connectors that are listed in dictionaries and textbooks as synonyms or words with a similar meaning. For example, one

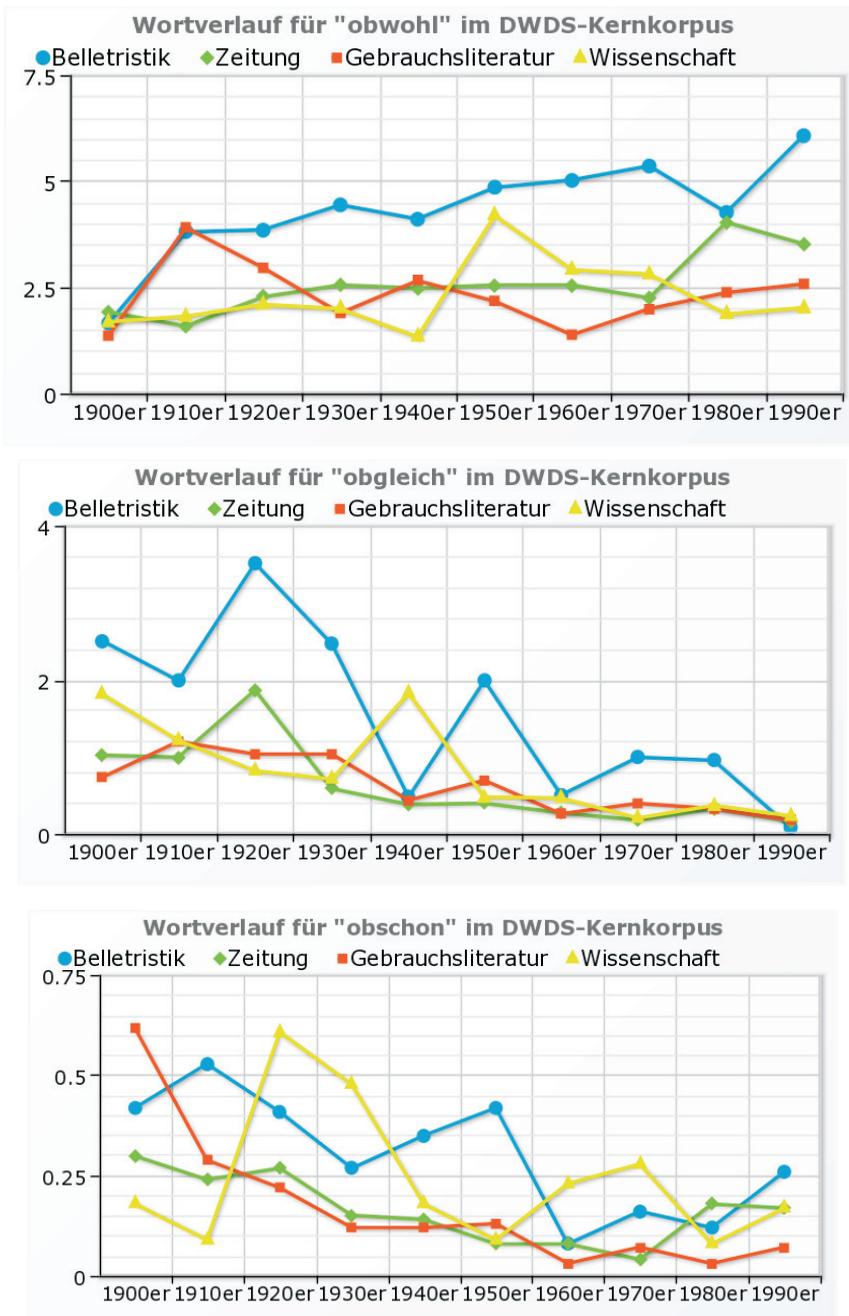


Figure 4. Frequency Graphs across Genre and Decade for *obwohl* (top), *obgleich* (center), and *obschon* (bottom)

student was interested in temporal changes in the usage of the synonymous subordinating conjunctions *obwohl*, *obgleich*, and *obschon*. He worked with the DWDS panel *Zeitverlauf* and retrieved the graphs pictured in Fig. 4 for each of the search words.

The student showed the charts to the class on the big screen and commented on the results. He demonstrated that *obwohl* not only appears much more frequently (8180 total hits in

comparison with 2555 for *obgleich* and 653 for *obschon*) but also exhibits a clear increase in frequency (rising dynamic) over the course of the 20th century. The *obwohl* curve rises especially sharply in literary texts, followed by journalism. In contrast, *obgleich* and *obschon* show a steep falling dynamic. However, whereas *obgleich* virtually disappears in all genres by the 1990s, *obschon* (although lower in total frequencies) shows signs of revival in the same decade. In my opinion, this search was very successful. First, it confirmed the general intuition of proficient speakers of German that *obwohl* is more frequent than its two synonyms. Most importantly though, it demonstrated how drastic this contrast is and visualized two opposite diachronic patterns for *obwohl*, which rises from virtually zero at the beginning of the century to the status of a common word at the end of the century, and its synonyms. This finding showed the students the importance of contextual information for word and construction usage. If one randomly chooses a word from a dictionary list of synonyms, one may sound obsolete even if literate and formally correct. The student also illustrated his results with extended corpus excerpts containing the focal words.

As the course progressed, the students learned to formulate more sophisticated corpus queries for syntactic constructions. All autonomous activities were assigned for home and were then presented and discussed in class so that students could get feedback from both the instructor and the other participants.

Summary and Discussion

This study reported on the design and implementation of a sequence of corpus-based activities in an advanced German grammar course. The main organizing principle was a progression from more teacher-guided to more learner-centered activities. Independent direct corpus searches progressed from ones that were easier to formulate (such as searches for separate words and POS) to those requiring a more sophisticated search syntax (such as lexico-grammatical constructions). As learners gradually became more comfortable and experienced in working with the corpora, the teacher could alternate and mix different types of activities depending on specific needs arising in the course of instruction.

These activities helped the teacher and the participants achieve various instructional goals. First, students compared textbook materials and ‘real-life’ language use under the teacher’s guidance, which helped to validate, fine-tune, or refute the textbook information (Gavioli and Aston 2001; Tognini-Bonelli 2001; Sinclair 2004; O’Keeffe, McCarthy, & Carter, 2007). For example, the course book links the use of certain prepositions to written speech styles. A direct corpus search activity, in which each student was assigned a pair of prepositions and conjunctions expressing a similar meaning (e.g., *trotz* / *obwohl*; *infolge* / *so dass*; *im Fall(e)* / *falls*) confirmed the information from the textbook. Comparing frequencies of these linguistic connectors in written and spoken parts of DWDS, students could see that these prepositions are indeed either absent or very infrequent in spoken texts but frequent in written texts. In the discussion phase of this activity, the teacher made connections between its results and genre characteristics. Students could see that certain prepositions appear in noun phrases characterizing the nominal style of formal written genres, whereas conjunctions are typical of the verbal style of neutral written or informal spoken genres (see Byrnes et al., 2010).

Moreover, corpora turned out to be “great sources of serendipitous findings” (Stewart, Bernardini, & Aston, 2004, p. 13) when searches revealed information not necessarily anticipated based on intuition or information from reference sources. A particular tool that helps lead to such findings is the new DWDS panel *Wortprofil* recently added by the corpus developers. This panel succinctly and visually presents close lexical and grammatical associations of a word.

For example, the class discovered that the noun *Tendenz* (which has a neutral semantic value on its own) is more frequently combined with adjectives with negative connotations such as *antidemokratisch*, *apologetisch*, *destruktiv*. Thus, using corpus tools, one can gain insights into semantic prosody, or contextual meaning (Louw, 2000), of words and constructions which are hard to obtain by other means. In this way, the learners gain access to “alternative organizers” other than traditional inflection paradigms that can help “to establish order in the face of the seeming untidiness of language” (Chavez, 2011, p. 96).

Furthermore, corpora are a good source of advice in problematic cases when traditional reference materials are of little help. Spelling rules are a good example of such an application. For example, German compound verbs can be spelled either as two separate words or as one word (*kennen lernen* or *kennenlernen*, *stehen bleiben* or *stehenbleiben*). The course book mentioned the history of this rule and concluded that, after the recent spelling reform, both variants are considered acceptable, and native speakers follow their own preference. However, it is harder for non-native users of German to make such a choice. After polling the core corpus for these spelling variants, course participants found out that the one-word variant is more frequent in general and shows a rising dynamic in recent years, whereas the two-word spelling is less frequent in general and is declining in popularity. Therefore, students may decide to adopt the more frequent variant for their own use.

Because this was an innovative course presenting a pedagogical experiment, a post-course written survey (based on Heather & Helt, 2012) was administered to learn about student reactions. In general, the reactions were very positive. Among the benefits of corpus-based activities, the students listed gaining insights into frequency of grammar forms and nuances of meaning; checking their intuitions; seeing actual examples of usage; and learning about contextual combinations and phraseological expressions. As one participant put it, corpora help “collect historical and linguistically analytical data.” Among challenges and limitations, the students mentioned formulating their own grammatical queries and finding questions appropriate for corpus searches. They also expressed a wish to learn more about spoken corpora because we mostly focused on written text types. It is encouraging that all participants expressed a strong intent to engage with corpora in the future as reference materials. All those envisioning careers as teachers also wrote that they planned to use corpora in teaching.

Conclusion and Pedagogical Implications

The preceding section illustrates that the main goals of the course were successfully achieved (see Appendix A). The most important result is that all participants have reached some degree of corpus literacy as formulated by Mukherjee (2006). All of them acquired the basic aspects of this literacy: know how to “read” and analyze concordances and make sense of basic frequency information. Some participants achieved a higher degree of corpus literacy—knowing what can and cannot be done with corpora—while others felt that this remained an area of difficulty. Especially encouraging were post-course survey responses confirming that all participants felt confident in using corpora in the future for their own learning or in teaching. Therefore, the course has stimulated autonomous research skills by advanced learners (Bennett, 2010; Mauranen, 2004). Their level of self-confidence was higher than reported by participants in similar courses (cf. Heather & Helt, 2012) which also supports a positive evaluation of the course outcomes.

These results present evidence that corpus-based grammar courses have great potential in foreign language education, provided that instructors follow a principled approach in their design and implementation. First, such courses require a great time investment and careful

preparation. The instructor should follow a logical progression in the course by strategically combining various degrees of teacher mediation and learner autonomy as outlined in this article. Much attention should be paid to guidelines, handouts, and worksheets. Even after switching to more independent activities, modeling by the teacher should be continued on a regular basis. One more piece of advice is in order: if possible, combining undergraduate and graduate courses should be avoided because of the differences in both cognitive skills and language proficiency levels. Offering separate seminars for graduate student instructors would allow for a more systematic incorporation of activities involving the development of pedagogical activities and materials. Finally, it should be noted that designing a corpus-based grammar course for participants at lower language proficiency levels is possible and has an equally great instructional potential but would constitute a separate project.

In general, working with corpora promotes a more holistic perspective toward grammar compatible with cutting-edge literacy-based and usage-based pedagogical approaches described in the introduction to this article. In particular, it can help both language educators and students to make a mental transition from conceptualizing grammar as a dry and strict collection of rules toward understanding language as a meaning-making resource. As Thorne argues in his interview with Antoniadou, “[w]hat we typically refer to as grammar is better understood as observable, recurrent and malleable patterns of language use. An entailment to this view is that different genres and different kinds of communicative engagement are structured or patterned in different ways” (Antoniadou, 2011, p. 101). If teachers can convey this understanding to students, this will contribute to “resolving the tensions between learners’ varied understandings of grammar, language professionals’ objectives, and the demands of the language itself, as it is embedded in cultural practices” (Chavez, 2011, p. 84).

To conclude, this article has sought to make a case for a wider application of corpora in foreign language instruction, especially beyond the area of teaching English. There are a number of excellent freely available corpus resources for German. Teachers of German are encouraged to improve their own corpus literacy and to help their students to do so. Although working with corpora does require learning a new skill, it is a worthwhile endeavor to embrace for language educators. This article should serve as another contribution to the pool of resources which both provide teachers with examples of carefully sequenced corpus-based activities and guide them toward becoming independent corpus users and teaching material developers.

Acknowledgment

I would like to acknowledge the students at the University of Kansas who participated in this course.

Works Cited

- Antoniadou, V. (2011). New media, new communicative genres and inclusive technology-mediated L2 pedagogy: A conversation with Steve Thorne. *Bellaterra Journal of Teaching & Learning Language & Literature*, 4 (4), 99-109.
- Arnold, N., & Ducate, L. (Eds.). (2011). *Present and future promises of CALL: From theory and research to new directions in language teaching*. San Marcos, TX: CALICO.
- Belz, J. A. (2005). Corpus-driven characterizations of pronominal *da*-compound use by learners and native speakers of German. *Die Unterrichtspraxis / Teaching German*, 38 (1), 43-59.
- Belz, J. A., & Vyatkina, N. (2008). The pedagogical mediation of a developmental learner corpus for classroom-based language instruction. *Language Learning and Technology*, 12 (3), 33-52.
- Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Ann Arbor, MI: University of Michigan Press.

- Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 165–182). Amsterdam, Netherlands: Rodopi.
- Breyer, Y. (2009). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning*, 22 (2), 153–172.
- Brown, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL*, 19 (3), 307–328.
- Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82 (4), 711–733.
- Byrnes, H. (2005). Literacy as a framework for advanced language acquisition. *ADFL Bulletin*, 37 (1), 11–15.
- Byrnes, H., Maxim, H., & Norris, J. M. (2010). Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment. *The Modern Language Journal*, 94 (s1, Monograph).
- Chavez, M. (2011). German grammar in the students' words: The essentialization of German grammar by American college-level learners. *Die Unterrichtspraxis / Teaching German* 44 (2), 83–97.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Dodd, B., Eckhard-Black, C., Klapper, J., & Whittle, R. (2003). *Modern German grammar: A practical guide (second edition)*. London, England: Routledge.
- Dodd, B. (1997). Exploring a corpus of written German for advanced language learners. In A. Wichmann, S. Figelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 131–145). London, England: Longman.
- Frankenberg-Garcia, A. (2012). Raising teachers' awareness of corpora. *Language Teaching*, 45 (4), 475–489.
- Gavioli, L., & Aston, G. (2001). Enriching reality: Language corpora in language pedagogy. *ELT Journal*, 55 (3), 238–246.
- Greaves, C., & Warren, M. (2007). Concorgramming: A computer driven approach to learning the phraseology of English. *ReCALL*, 19 (3), 287–306.
- Halliday, M. A. K. (1996). On grammar and grammatics. In R. Hasan, C. Cloran, & D. G. Butt (Eds.), *Functional descriptions: Theory in practice* (pp. 1–38). Amsterdam, Netherlands: John Benjamins.
- Heather, J., & Helt, M. (2012). Evaluating corpus literacy training for pre-service language teachers: Six case studies. *Journal of Technology and Teacher Education*, 20 (4), 415–440.
- Johns, T. (1986). Micro-concord: A language learner's research tool. *System*, 14 (2), 151–162.
- Johns, T. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Figelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100–115). London, England: Longman.
- Kaltenbock, G., & Mehlmauer-Larcher, B. (2005). Computer corpora and the language classroom: On the potential and limitations of computer corpora in language teaching. *ReCALL*, 17 (1), 65–84.
- Katz, S. L., & Watzinger-Tharp, J. (Eds.). (2009). *Conceptions of L2 grammar: Theoretical approaches and their application in the L2 classroom*. Boston, MA: Heinle Cengage Learning.
- Kennedy, C., & Miceli, T. (2002). The CWIC project: Developing and using a corpus for intermediate Italian students. In B. Kettemann, & Marko, G. (Eds.), *Teaching and learning by doing corpus analysis* (pp. 183–192). Amsterdam, Netherlands: Rodopi.
- Kern, R. (2000). *Literacy and language teaching*. Oxford, England: Oxford University Press.
- 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.
- Klein, W., & Geyken, A. (2010). Das Digitale Wörterbuch der Deutschen Sprache (DWDS). In U. Heid, S. Schierholz, W. Schweickard, H. E. Wiegand, R. H. Gouws, & W. Wolski (Eds.), *Lexikographica* (79–93). Berlin: De Gruyter. <http://www.dwds.de/veroeffentlichungen/>
- Larsen-Freeman, D. (2003). *Teaching language: From grammar to grammaring*. Boston, MA: Heinle.
- 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, England: University of Birmingham.
- Mauranen, A. (2004). Speech corpora in the classroom. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and language learners* (pp. 195–211). Amsterdam, Netherlands: John Benjamins.
- McCarten, J. (2010). Corpus-informed course book design. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 413–427). London, England: Routledge Taylor & Francis Group.
- McCarthy, M. (2004). *Using a corpus in language teaching*. CALPER Professional Development Document (CPDD-0410). University Park, PA: The Pennsylvania State University, Center for Advanced Language Proficiency Education and Research. <http://calper.la.psu.edu/publications.php>
- Mishan, F. (2004). Authenticating corpora for language learning: A problem and its solution. *ELT Journal*, 58 (3), 219–227.
- Möllering, M. (2004). *The acquisition of German modal particles*. Bern: Lang.
- Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art – and beyond. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods* (pp. 5–24). Frankfurt am Main: Peter Lang.

- Nelson, G., Wallis, S., & Aarts, B. (2002). *Exploring natural language. Working with the British component of the International corpus of English*. Amsterdam, Netherlands: John Benjamins.
- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge, England: Cambridge University Press.
- O’Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL*, 19 (3), 269–286.
- Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205–225.
- Rug, W., & Tomaszewski, A. (2009). *Grammatik mit Sinn und Verstand: Übungsgrammatik Mittel- und Oberstufe*. Stuttgart, Germany: Klett.
- Sinclair, J. (Ed.). (2004). *How to use corpora in language teaching*. Amsterdam, Netherlands: John Benjamins.
- Stewart, D., Bernardini, S., & Aston, G. (2004). Introduction: Ten years of TaLC. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and language learners* (pp. 1–18). Amsterdam, Netherlands: John Benjamins.
- Swaffar, J. K., & Arens, K. (2005). *Remapping the foreign language curriculum: An approach through multiple literacies*. New York, NY: Modern Language Association of America.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam, Netherlands: John Benjamins.
- Tyler, A. (2010). Usage-based approaches to language and their applications to second language learning. *Annual Review of Applied Linguistics*, 30, 270–291.
- Vannestål, M. E., & Lindquist, H. (2007). Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course. *ReCALL*, 19 (3), 329–350.
- Widdowson, H. G. (2000). On the limitations of linguistics applied. *Applied Linguistics*, 21 (1), 3–25.
- Zinggeler, M. (2006). GRIMMATIK: German grammar through the magic of the Brothers Grimm fairy tales and the Online Grimm Corpus. In E. Fitzpatrick (Ed.), *Corpus linguistics beyond the word: Corpus research from phrase to discourse* (pp. 167–184). Amsterdam, Netherlands: Rodopi.
- Zyzik, E. (2008). A novel format for teaching Spanish grammar: Lessons from the lecture hall. *Foreign Language Annals*, 41 (3), 434–453.

Appendix A. Course Goals

On completion of this course, students will be able to:

- understand grammar as “the cognitive organization of one’s experience with language” (Bybee, 2006) and not only an abstract system of rules
- discover patterns of use and establish meanings of lexico-grammatical means of expression in different contexts
- understand language variability and choose appropriate linguistic patterns from the options available to speakers of German
- compare grammar as presented in course books and as used in authentic texts
- independently work with authentic German texts to analyze grammar usage and to create teaching materials
- formulate and solve linguistic problems
- develop, explore, and reflect on using strategies of learning a language at an advanced level

Appendix B. Hauptregeln der DWDS-Korpussuche (Syntax)

| | |
|---|-------------------------------|
| 1. Einfache Wortsuche | man |
| 2. Lemma | werden |
| 3. Wortform | @wird |
| 4. Teilwort | *bar; be* |
| 5. Wortart | \$p=VVPP |
| 6. Ein Wort als eine bestimmte Wortart | *bar with \$p=ADJD |
| 7. 2 Suchwörter in einem Satz | werden && \$p=VVPP |
| 8. 2 Suchwörter in einem bestimmten Abstand | “werden #5 \$p=VVPP” |
| 9. Wort ausschließen (X aber nicht Y) | werden && \$p=VVPP && !gebaut |
| 10. Interpunktion | \, |

Appendix C. Fragen für Korpusrecherchen

1. Regeln aus Grammatikbüchern überprüfen. Kommas, Rechtschreibung, Kasus, Wortfolge...
2. Wie oft wird ein Wort / eine Wortart / eine grammatische Konstruktion gebraucht?
3. In welchen Textsorten wird dieses Wort öfter gebraucht?
4. Welche Konstruktion von 2 Konstruktionen mit ähnlichen Bedeutungen wird öfter gebraucht?
5. Wie verändert sich der Gebrauch dieses Wortes mit der Zeit?
6. In welchem Kontext wird das Wort gebraucht? Welche Wörter sind häufige Nachbarn dieses Suchworts?