# Chapter 15
# Corpus-Informed Pedagogy in a Language Course:
## Design, Implementation, and Evaluation

**Nina Vyatkina**

*University of Kansas, USA*

## ABSTRACT

*Data-Driven Learning (DDL), or a corpus-based method of language teaching and learning, has been developing rapidly since the turn of the century and has been shown to be effective and efficient. Nevertheless, DDL is still not widely used in regular classrooms for a number of reasons. One of them is that few workable pedagogical frameworks have been suggested for integrating DDL into language courses and curricula. This chapter describes an exemplar of a practical application of such a pedagogical framework to a high-intermediate university-level German as a foreign language course with a significant DDL component. The Design-Based Research approach is used as the main methodological framework. The chapter concludes with a discussion of wider pedagogical implications.*

## INTRODUCTION

Corpora, or large electronic collections of texts, have been used in language pedagogy since their inception in the 1960s. Corpus-based language learning and teaching methods and the associated research strand dedicated to exploring their effectiveness have also been known as Data-Driven Learning, or DDL, since Johns (1990) adopted this term from computer science. The body of work on DDL has grown exponentially over the last three decades and has by now convincingly shown that this method can be effective for different teaching targets in different teaching contexts. Nevertheless, DDL is still not widely used in regular classrooms for a number of reasons. One of them is that few workable pedagogical frameworks have been suggested for integrating DDL into language courses and curricula. Although many DDL researchers report on studies conducted within the framework of their own courses, it is impossible to provide details of a course design in a typical research article due to length limitations. This chapter aims to fill this gap by providing a report on a practical application of such a pedagogical framework

to a language course with a significant DDL component. The chapter begins with a background section that presents a brief overview of empirical DDL research, including advantages and disadvantages of different types of corpora and DDL methods, thus explaining the rationale for selecting the DDL resources and the pedagogical approach for the course. Next, the study methodology is described. The Design-Based Research (DBR) approach was selected that has been broadly defined as "an emerging paradigm for the study of learning in context through the systematic design and study of instructional strategies and tools […, which] can help create and extend knowledge about developing, enacting, and sustaining innovative learning environments" (Baumgartner et al., 2003, p. 5). In the present study, this approach has been operationalized following ten steps suggested by Colpaert (2006) for Computer-Assisted Language Learning (CALL) courseware design. The bulk of the chapter presents a report on the design, implementation, and evaluation of a high-intermediate university-level German as a Foreign Language course with a substantial CALL (DDL) component that is mapped back onto these steps. In line with the DBR methodology, the chapter concludes with a reflection on how the affordances of the DDL technology were utilized for its meaningful integration into the focal language course, thus forming the basis for subsequent modifications to improve this specific instructional innovation.

The intended audience of this chapter includes language educators interested in implementing DDL in their courses as well as DDL researchers who plan to conduct classroom DDL research. Although the described course was designed and implemented in a specific language learning setting – teaching German as a foreign language at a US university – the suggested pedagogical framework and course design model are applicable to many different contexts with appropriate modifications.

## BACKGROUND

### Types of Corpora and Corpus Tools Used in DDL

Corpus-based language teaching, after a few first applications in the 1960s – 1980s, started developing rapidly after Johns (1990) had singled it out as a distinct strand of CALL: Data-Driven Learning (DDL). This development has been accompanied by the exponential growth of DDL empirical research (see Boulton, 2017, for a timeline) that by now has accumulated substantial evidence that corpus-based teaching is effective and, for certain learning targets, more efficient than non-DDL methods (see recent meta-analyses by Boulton & Cobb, 2017; Lee, Warschauer, & Lee, 2018).

Corpora and corpus tools (e.g., concordancers) used in DDL come in different configurations (see Godwin-Jones, 2017, for a recent list of available corpus resources). First, like any CALL technology, they can be dedicated or non-dedicated, i.e. custom-built for language teaching purposes or general-purpose (Colpaert, 2006). Second, either type of technology can be qualified as open-access or non-open-access, i.e. it can be freely and publicly accessed on the Internet or be proprietary and commercial. In practice, most open-access corpora and corpus tools are non-dedicated because they have been designed with research and not teaching purposes in mind. Therefore, although they are "open," they are not necessarily "educational." Dedicated corpora and tools, on the other hand, while being "educational," are rarely "open." The following section considers the effectiveness of DDL applications of different types of corpus resources.

## The Effectiveness of DDL With Different Types of Corpora and Tools

The two by far most popular open corpora in DDL are the Corpus of Contemporary American English, or COCA (Davies, 2008) and the British National Corpus, or BNC (BNC Consortium, 2007) – which is not surprising because the overwhelming majority of DDL interventions have been conducted for teaching English (Vyatkina & Boulton, 2017). Substantial parts of both the COCA and the BNC are freely accessible to the public and can be searched with a built-in concordancer and some other tools. Several studies report that learners of English were able to successfully generate concordance lines from these corpora and analyze them, which led to their outperforming learners in a non-DDL group on a posttest after a focused short-term intervention (Daskalovska, 2015; Kilimci, 2017) or improved accuracy and idiomaticity of collocations in their writing over a semester (Chang, 2014; Li, 2017). Learners in these studies were university students in different countries (China, Korea, Macedonia, Turkey) with a high-intermediate to advanced English proficiency, who had studied English for several years before. Even more impressively, Japanese learners of English with low-intermediate proficiency were able to successfully use COCA to improve their writing (Mueller & Jacobsen, 2016) and speaking (Geluso & Yamaguchi, 2014). In contrast, Quinn (2015) found COCA to be extremely difficult for her Japanese intermediate EFL learners, which has led her to switch to a dedicated, non-open-access corpus. These reservations about COCA are shared by Master's students in an EFL teacher education course in Iran: "Most of them appreciated the website for its features; however, they found it too complicated to be used in language classes because of its strict search rules" (Ebrahimi & Faghih, 2016, p. 126). Student teachers in that study had similar opinions about open access corpus tools such as Lextutor (Cobb, n. d.) and AntConc (Anthony, n. d.): they found them somewhat useful for teachers for correcting their students' writing (after a considerable time and effort invested into learning how to use them) but unsuitable for direct use by learners in a classroom. These reservations regarding non-dedicated corpus resources support Godwin-Jones's (2017) observation that "[s]tudents (and teachers) today are used to more attractive and intuitive (and mobile-friendly) interfaces than is normally the case for concordance software" (p. 21).

In light of the abovementioned limitations of non-dedicated software, it is not surprising that considerable efforts in CALL have been devoted to the development of dedicated resources. A number of recent studies attest to the fact that the use of dedicated user-friendly corpus software can lead to significant language learning gains and high learner satisfaction (e.g., Cotos, Link, & Huffman, 2017; Kennedi & Miceli, 2017; Mirzaei, Rahimi Domakani, & Rahimi, 2015). However, such software is rarely open-access. Furthermore, the lifespan of proprietary software can be, regrettably, limited. For example, Rezaee, Marefat, and Saeedakhtar (2015) used a commercial concordancer that they describe as "powerful and user-friendly" in their highly successful DDL course. However, the access to the tool was discontinued in 2016 due to compatibility issues with Windows. Open access corpora and tools, although also subject to this threat, usually are more sustainable as they have more reliable funding sources and large support teams behind them.

This brief overview has demonstrated that different corpus resources have different advantages and disadvantages, which makes it difficult for DDL educators to make the right choice in order to get "maximum return on minimum investment" (Boulton (2011, p. 70). Boulton proposes a few rules of thumb, stating that corpus software should be free and easy to use, and that techniques should be flexible and transferable. While open access corpora fulfill the first condition (they are free), they are not necessarily easy to use, which puts the onus on the teachers to equip their learners with suitable corpus exploration

techniques and methods. A closer look at DDL interventions that were successful with non-dedicated open resources provides insight into such teaching methods.

## DDL Methods Beneficial for Language Learning

The concept of guidance has been a key aspect of studies that have reported positive learning gains and learners' perceptions (e.g., Frankenberg-Garcia, 2014; Pérez-Paredes, Sánchez-Tornel, Alcaraz Calero, & Aguado Jiménez, 2011; Smart, 2014; Vyatkina, 2016a, 2016b). Teachers in these studies present their learners with a series of corpus tasks that progress from more scaffolded, mediated, and teacher-guided ones to non-mediated and autonomous ones. The level of this guidance varies considerably depending on specific learner populations and learning contexts. For example, Ebrahimi and Faghih (2016) point out that their Master's students in Iran needed much more CALL guidance than Boulton's (2011) Master's students in France due to the difference in their comfort level with technology. Learner motivation, learning styles, and other individual differences have been mentioned as further important factors (Flowerdew, 2015).

Many studies have demonstrated that much attention needs to be paid to DDL task design. For example, Cho (2015) found that conceptual tasks such as comparing close synonyms in corpus data led to higher gains if implemented as collaborative group work, whereas individual consultation was better suited for procedural tasks such as using target items in translation. Bardovi-Harlig, Mossman, and Su (2017) showed that direct searches of the MICASE corpus were conducive to learners' improvement in using pragmatic routines but not speech acts, with the latter task requiring teacher-prepared corpus printouts. Other studies demonstrate that even low-proficiency learners (Wong & Lee, 2016) or young learners (Moon & Oh, 2017) can successfully search corpora and achieve significant learning gains in carefully guided DDL interventions focused on a narrow target structure. Kennedy and Miceli (2017) argue that DDL should not necessarily aim at converting learners into corpus researchers who can derive abstract patterns from empirical corpus examples. Instead, they propose a more modest goal of developing the "observe-and-borrow-chunks mentality" in learners and "encouraging their curiosity about language patterns" (p. 91). Finally, many researchers provide evidence that DDL may be especially beneficial in combination with other teaching methods such as concept-based instruction (Kilimci, 2017) and provision of written corrective feedback (Crosthwaite, 2017; Quinn, 2015), as well as with other CALL resources such as electronic dictionaries (Karras, 2016; Yoon, 2016).

The overall conclusion that can be drawn from the above literature review is that for corpus-based language teaching to be successful, it needs to be grounded in sound language pedagogy principles. Published research has identified teaching methods that can lead to success in DDL with open corpora. While empirical studies, understandably, contain only brief descriptions of implemented corpus activities, several teacher guides have been published to date that provide suggestions on how to conduct DDL activities and incorporate them into language syllabi and curricula (e.g., Bennett, 2010; Friginal, 2018; Poole, 2019; Reppen, 2010; Shaw, 2011; Thurstun & Candlin, 1997). What is still missing in the field is detailed reports on the implementation of DDL courses or courses with substantial DDL components integrated into a coherent pedagogical framework. This chapter aims to address this gap. While empirical results of several DDL teaching interventions conducted by the author have been published elsewhere (Vyatkina, 2013, 2016a, 2016b), this study presents a report on the planning of a DDL-enhanced language course, selection of corpus resources and teaching methods, course implementation, and lessons learned.

## THE STUDY

## Methodology

This chapter contributes to the theme of this volume by describing and discussing a practical exemplar of a DDL-enhanced course. Design-based research (DBR) has been adopted as the main research methodology, described by Pardo-Ballester and Rodríguez (2009, p. 87) as "an iterative cycle composed of multiple steps such as exploration of the design, enactment of interventions, evaluation and analysis of the outcomes, and redesign." In their introduction to the first collection of articles on DBR in CALL, Rodríguez and Pardo-Ballester (2013) argue that this methodology is eminently suited to CALL explorations due to the field's main focus on emerging technologies and, thus, on the implementation of complex instructional innovation and change. They also note that researchers frequently need to limit their focus in each particular study to certain segments of the complex, longitudinal, and iterative DBR process. In another contribution to that volume, Larsen-Freeman (2013, p. 25) adds that "DBR researchers adopt a retrodictive view, looking for the influence of prior activity on current activity," thus progressively refining instructional innovations in the next iterations of the design and research process. This chapter follows the abovementioned recommendations for DBR in CALL by presenting a detailed report on the design, implementation, and evaluation of a DDL-enhanced language course. The present report is limited to the first iteration of the course. The described course model that incorporated a technological innovation (DDL) is considered a product of the DBR process, while its retrodictive evaluation (Larsen-Freeman, 2013) serves as the basis for the process of course improvement in subsequent iterations, in accordance with the dual product-oriented and process-oriented nature of DBR (Hung, 2013).

It must be noted that, although the term DBR is relatively new to CALL, similar methodologies have been applied in earlier CALL research. One of the first applications was Colpaert's (2006) adaptation of the industrial design ADDIE (analysis, design, development, implementation, evaluation) model to the development of CALL courseware. Colpaert underscores that in educational settings, ADDIE must be "pedagogy-driven": "It starts from a detailed specification of what is needed for language-teaching and -learning purposes in a specific context, defines the most appropriate method, and finally attempts to describe the technological requirements to make it work" (p. 479). Colpaert operationalizes his model in 10 guiding steps:

1. Describe the Learning Environment
   a. Learner Characteristics
   b. Goals
   c. Selecting Appropriate Language Teaching Method
2. Describe the System Requirements for:
   a. Learners
   b. Teachers
   c. Pedagogy
   d. Technology
   e. Content
   f. Other Actors (e.g. Content Providers, Native Speakers, Parents, Training Managers, Software Providers, Policy Makers)

3.  Describe the Architecture of the Distributed Learning Environment
4.  Apply an Activity Framework
5.  Define Linguistic/Didactic Functionalities
6.  Define Personas and Their Goals
7.  Conceptualize
8.  Specify
9.  Perform Pre-Use Evaluation
10. Perform Post-Use Evaluation

In this model, most attention is dedicated to the Analysis, Design, and Development stages of AD-DIE (steps 1-9), while the Implementation stage is supposed to take place between steps 9 and 10, and the Evaluation stage is distributed between steps 9 (pre-evaluation) and 10 (post-evaluation). Several studies provide successful examples of applying Colpaert's model to the design of both CALL tools and technology-enhanced language courses (Heift, 2010; Schulze & Liebscher, 2010; Wood, 2011). Following suit, the above list serves as a roadmap and template for the report presented in this chapter.

Finally, this chapter follows the guidelines of the *Forum* rubric, recently introduced by one of the leading CALL journals, *Language Learning & Technology*. For this rubric, articles are solicited that "provide in-depth engagement with the intersection of technology and language teaching pedagogy" (Kessler, 2016, p. 192). The four principles are formulated for the *Forum* articles, which are very much in line with the DBR approach:

1.  Establish the empirical or conceptual framework for using the targeted technologies;
2.  Describe the instructional context, pedagogical objectives, and technological tools;
3.  Provide details of how the technology was integrated: which tools or approaches were used, how procedures were implemented, and how assessment or feedback was utilized;
4.  Comment on the challenges encountered and on the suggestions for additional integration. (Kessler, 2016, p. 192)

## Step 1: Describe the Learning Environment

Learner Characteristics

*General: University and Program*

The study was implemented in an upper-level German course of a German Studies program at a large public university in the Midwest of the United States. The majority of the students at this university come from the Midwestern region, speak American English as their L1, are ethnically Caucasian, and enter the university after graduating High School at the age of 18, although there also are students who are international and/or do not fit the above description. The level of comfort with technology is on average high, and most courses at this university include significant technology-enhanced components.

## Specific: The Course

The researcher taught this semester-long (16 weeks) course, titled Advanced German I. Eleven students enrolled in the course. Participant metadata were collected via a pre-course questionnaire. Their characteristics were representative of the general university and program student body profile described above. The gender distribution was equal (5 males, 6 females), and most students were German majors (6) and minors (4) as well as seniors (6) and juniors (4). The mean age was 21 years (ranging from 18 to 24). All students had English as their L1 and all of them spoke English as their family language, although five students indicated some German heritage. All students but one had visited German-speaking countries, with the length of their stay varying widely from several weeks to three years. The length of their college-level German study ranged between two and six semesters, with five students having also learned it in High School (2-4 years long). All but three students had also studied another L2 (Spanish, French, Latin, Russian, or Croatian) for a duration ranging from three weeks to eight years. The average L2 German proficiency of participants at the beginning of the course was intermediate at the B1 level CEFR (Council of Europe, 2001), as established via an online standardized test.

## Goals

### Learner Goals

On the pre-course questionnaire, all learners indicated a high level of intrinsic and/or instrumental motivation to study German. The answers to the question "Why are you studying German?" mentioned interest in the German language, history, and culture, German-speaking countries, and languages in general. Furthermore, all students expressed intent to use German after the course for the following purposes: travel and study, internship, or work in German-speaking countries, communication with family and friends, and teaching German.

### Pedagogical Goals

At the university level, broad pedagogical goals were formulated in accordance with the institutional strategic plan: to promote active student learning, take advantage of state-of-the-art pedagogical methods and information technology, and enhance students' higher-level cognitive and analytical skills. At the German Studies program level, the main goal of the course was to advance students' German language proficiency beyond the intermediate level and to prepare majors for the subsequent capstone course. The main pedagogical goal of the instructor was to address the abovementioned institutional, programmatic, and student goals. Furthermore, the instructor's overarching goal as a teacher was to instill in students a sense of responsibility, personal autonomy, and critical and creative thinking.

## Selecting Appropriate Language Teaching Method

The Literacy, or Multiliteracies (ML), pedagogical approach (Cazden et al., 1996; Kern, 2000; Paesani, Allen, & Dupuy, 2015) aligns itself well with multiple goals described above, especially because this approach has been shown to be conducive to the development of advanced foreign language capacities (Byrnes, 2005). The ML framework has been developed to fill gaps in the Communicative Language Teaching (CLT) approach that is widespread in the US. Most importantly, the ML approach pushes learners to develop their language abilities beyond fluent oral communication about everyday personal

matters toward interpretation and production of texts in multiple genres (including secondary, public discourses) and multiple modes (including various technologies) through collaboration, problem solving, and critical reflection (Schicker, 2018). This approach capitalizes on explicit methods of language teaching that include analysis of linguistic resources employed for purposeful textual meaning making. Furthermore, the ML approach is compatible with Guided Induction (GI) – a teaching method (Herron & Tomasello, 1992) frequently employed in DDL, in which learners induce abstract linguistic patterns while analyzing samples of actual language use through partner work and under teacher guidance.

## Step 2: Describe the System Requirements

This step is the most elaborate one in Colpaert's (2006) model as it requires separate specifications for various participants and aspects of the process (learner, teacher, pedagogy, technology, content, and other actors) at four levels (general, local, differential, and targeted requirements), as shown in the so-called GLDT grid (see Table 1).

*Table 1. GLDT Grid*

|  | **General** | **Local** | **Differential** | **Targeted** |
|---|---|---|---|---|
| Learner |  |  |  |  |
| Teacher |  |  |  |  |
| Pedagogy |  |  |  |  |
| Technology |  |  |  |  |
| Content |  |  |  |  |
| Other Actors |  |  |  |  |

Data Source: Colpaert (2006, p. 484)

*In another study, Colpaert (2004, pp. 138-139) has formulated guiding questions for filling out each cell of the above table. These questions are used as subheadings for the following section of this report for clarity.*

## Learners

### G: What Are Generally Accepted Findings and Principles for Learners?

Adult learners in Foreign Language (FL) environments typically have limited naturalistic exposure to the target language but they can acquire it in instructed settings (Ellis & Shintani, 2014; Loewen, 2015). Motivation and individual differences play a big role in FL acquisition, and the method of instruction can enhance or hinder learners' success.

### *L: What Are Common Characteristics of the Learners in This Particular Design Space?*

The learners are young adult university students learning German as a foreign language whose first and dominant language is American English. German Studies is their primary or secondary subject of study. They have had limited experience learning foreign languages and none of them started learning them as young children. They have little exposure to German outside of class but they are motivated to learn the language and plan to use it in the future for personal or professional purposes. Despite years of learning German in instructed settings, students come to their junior and senior year of German Studies still lacking idiomatic vocabulary and advanced grammar appropriate for understanding authentic German texts in public, literary, and academic genres as well as expressing themselves in these genres. They have a high comfort level with technology.

### *D: Which Distinctions Must Be Made Within This Design Space, or Which Elements Are Subject to Change?*

The L2 proficiency level of the students coming to the course can broadly be defined as intermediate (on average, B1 CEFR) but it varies widely from very low to high intermediate. At the beginning of the course, 2 students were at the A2 CEFR level, 4 at the B1, and 3 at the B2 level. Although all of them were interested in German, the level of motivation to perform well in this particular course may have varied. Other inter-individual differences (e.g., language aptitude, learning styles and preferences) were also expected to be large as in any classroom – although they were not measured in this study. Some students had special needs, for which the university provides accommodations (e.g., flexible deadlines and attendance schedules, special exam arrangements, etc.).

### *T: Which Characteristics Are Amenable to Improvement (e.g., Vocabulary, Topics, Skills)?*

All students were expected to broaden and deepen their vocabulary and grammar knowledge that would enable them to understand and produce German texts in public, literary, and academic genres. The course was also aimed at preparing them to write a capstone research paper (a short-term goal) and to provide them with tools and strategies for life-long language learning beyond their university studies (a long-term goal). Due to differences in the L2 proficiency and individual differences (see above), the ideal course design should have allowed all students to improve independently from their starting point and the pace of their progress.

## Teachers

### *G: What Findings and Principles Are Generally Accepted for Teachers?*

Typically, advanced language courses for majors and minors at US universities are taught by professorial staff. Professors are highly independent and flexible in syllabus design and material selection. The positive aspects of this are that most professors are motivated to teach effectively and to improve the design of repeatedly taught courses. The negative aspects are that not all professors keep abreast of innovative teaching methods (especially if their research is outside Applied Linguistics) and that they do not have enough time for teaching innovations due to a high research and service workload. Furthermore, this independence frequently leads to a lack of coordination between courses in a program.

*L: What Are Common Characteristics of Teachers in the Design Space?*

The case course was taught by a teacher-researcher whose main field of inquiry is Instructed Second Language Acquisition, who regularly follows new developments in language pedagogy, and engages in a research-application cycle while testing the effectiveness of new teaching methods, specifically CALL methods. The instructor was thus highly motivated to design this new technology-enhanced course and had the freedom to do so as a member of the professorial staff.

*D: Which Distinctions Must Be Made Within This Design Space, or Which Elements Are Subject to Change?*

The syllabus was to be prepared in advance of the course, including materials, activities, assignments, and assessments, but the teacher was ready to make changes and adjustments in accordance with the class dynamics and individual student differences.

*T: Which Characteristics Are Amenable to Improvement?*

While planning the course, the teacher/researcher was going to convert her theoretical knowledge of ISLA and CALL principles (as a researcher) into practical knowledge (as a teacher) by implementing them under specific local conditions, to monitor this implementation, analyze the results, and make improvements for future course iterations as well as to contribute to future theory development.

## Pedagogy

*G: What Are Generally Accepted Findings and Principles for Learning and Teaching?*

It is an established fact that rich exposure to the target language is a necessary condition for successful ISLA (Ellis & Shintani, 2014; Loewen, 2015). However, specific teaching methods can be more or less effective depending on the teaching and learning goals, learner characteristics, and local conditions.

*L: Which Learning/Teaching Method Is Currently Being Used Within the Design Space?*

The instructor followed an approach that aimed at a balanced development of different language skills (reading, writing, speaking, and listening) and knowledge areas (vocabulary, grammar), incorporated a significant culture component, and included both explicit (e.g., grammar analysis) and implicit (e.g., reading authentic texts) teaching. A considerable portion of class time was devoted to collaborative work in partner groups, while most individual reading, writing, and grammar practice was allocated to homework.

*D: Which Distinctions Must Be Made Within This Design Space, or Which Elements Are Subject to Change?*

In designing this new course, the teacher/researcher aimed at implementing a more principled ML and GI pedagogy approach while continuing to focus on multiple language teaching targets (listed above).

### T: Which Aspects Are Amenable to Improvement?

The students needed more exposure to authentic target language. More effective methods for developing advanced vocabulary and grammar, writing skills, as well as critical thinking, reflection, and analysis activities needed to be incorporated. Students needed to learn strategies and tools that they could use for life-long language learning beyond the classroom. Corpus-based CALL (see next section) was thought to be the necessary technology that could help improve these pedagogical aspects.

## Technology

### G: What Does a Strength-Weaknesses-Opportunities-Threats (SWOT) Analysis of Technology in Learning Show? What Does a SWOT Evaluation Determine About Existing Comparable Courseware?

A SWOT analysis of corpus-based CALL (DDL) was presented in the literature review at the beginning of this chapter. It can be summarized as follows: *Strengths*: DDL can lead to significant learning gains and high student satisfaction. With sufficient teacher mediation at the beginning, corpora can gradually become a valuable resource for autonomous language learning. Open access corpora are preferable to proprietary corpus software. *Weaknesses*: Open corpora are not (exclusively) dedicated to educational purposes and may be difficult to use for language learners and teachers. *Opportunities*: Pedagogical mediation, guidance, and training can alleviate difficulties connected to (initial) corpus use. Corpora can be especially useful in combination with other CALL resources and other teaching methods. Corpora can provide flexibility for learners to study at their own pace and time. *Threats:* If not tailored to local conditions and goals and not based on sound pedagogy, corpus use may be detrimental to learner success and lead to teacher and learner frustration.

A SWOT analysis of comparable CALL – e.g., dedicated corpora and corpus tools – shows that they may have some advantages (being more tailored to teacher and learner needs, having a friendly user interface). However, if the ultimate goal is to prepare learners to language use outside the classroom and life-long learning, the ecological validity of open corpora outweighs their disadvantages.

### L: What Infrastructure and Equipment Are Available in the Design Space? What Courseware Has Been Used Before?

The university is well equipped with educational technology, fast internet connection, and tech support. All classrooms are equipped with computer stations and large projection screens, and instructors can also reserve computer labs with computer stations for each student or check out laptops or tablets for regular classrooms. Almost all students feel comfortable using technology and possess multiple electronic devices. Most university courses utilize an electronic course management system (*Blackboard*) that is mandatory for all students as well as multiple technology-based assignments and assessments. All of the above applies to the courses in the focal program. Examples of courseware that are used include electronic writing assignment submission, online quizzes, audio and video clips from German media, electronic dictionaries, and more. Furthermore, the teacher/researcher had successfully piloted using the target German corpus in a graduate course (Vyatkina, 2013), also attended by two undergraduate students, which increased her confidence in its suitability for this planned undergraduate course.

*D: Which Distinctions Have to Be Made Within the Design Space or What Is Subject to Change (e.g., Operating Systems, Network Types, Processor Type, Software Versions)?*

No distinctions were to be made because the corpus software worked well with local infrastructure conditions.

*T: Which Aspects Are Amenable to Improvement (e.g., Less Network Traffic, Faster Execution)?*

No infrastructure changes were needed.

## Content

*G: What Content Is Available Worldwide?*

In most foreign language classes, the main content is provided through textbooks with a variety of supplementary materials. With the recent exponential growth of CALL resources, the proportion of such resources in comparison with traditional paper-based resources has been growing, too. There are also various German corpora available.

*L: What Kind of Content Is Being Used in This Design Space (e.g., Textbook, Syllabus)?*

All courses in the focal program use syllabi, textbooks, and a variety of supplementary paper-based and CALL materials as described above. The Digital Dictionary of the German Language (the German acronym DWDS) is one of the largest and best developed and supported open corpora. DWDS was selected for the course based on the instructor's experience with this corpus and promising results of its pilot use in a graduate course (Vyatkina, 2013). DWDS has other important pedagogical advantages such as a user-friendly interface, a wide variety of represented text types (genres), and the availability of multiple integrated and cross-linked corpus tools and associated electronic resources. In fact, it is more than a corpus, being rather a suite of lexicographic electronic resources, including thesauri, timeline graphs, and statistical analysis and visualization tools (see Vyatkina, 2018, for details).

*D: Which Distinctions Must Be Made Within This Design Space, or Which Elements Are Subject to Change?*

In accordance with the main course goal – development of learners' multiple literacies – the instructor made a decision not to use a traditional textbook. Instead, the required course materials included the syllabus, a reference grammar, a German novel, the DWDS corpus accompanied by printed and electronic assignment guidelines. Additional materials were also used in class such as audio and video clips from German media.

*T: Which Aspects Are Amenable to Improvement?*

The corpus and the novel were intended to provide students with increased exposure to authentic target language. Furthermore, because the whole text of the novel is integrated in DWDS, it helped contextualize some corpus assignments: the students could read a passage in the book (see Appendix 1) and then focus on specific patterns of language use while working with concordance lines generated from

the text of the novel using the corpus interface. To connect the corpus with another course text, a reference grammar was provided and some assignments asked the students to compare grammar rules as formulated in the book and as evidenced in actual usage examples from the corpus. Corpus assignments aimed at the development of vocabulary and grammar knowledge as well as of writing and speaking skills. Beyond purely linguistic goals, corpus assignments targeted the development of critical thinking, reflection, and analysis. Furthermore, the students were explicitly advised to continue using the corpus for consultation and further development of their advanced German capacities after finishing the course and upon graduation.

## Other Actors (e.g., Content Providers, Native Speakers, Parents, Training Managers, Software Providers, Policy Makers)

### G: What Is Their Generally Accepted Overall Role During Implementation, Use, And Evaluation?

University administrators can support or hinder new course design efforts. They are, in turn, influenced by the general population's demand for language education services as well as policy maker decisions. Software providers are also important actors in the process of designing CALL-enhanced courses.

### L: What Is Their Presence and Role in the Design Space?

The institutional administrators were supportive of designing new courses that incorporated technology (see step 1). Moreover, the design of the case course was supported by an internal institutional course redesign grant for introducing Digital Humanities skills, methods, and tools to undergraduate students.

### D: Which Types of Actor Can Be Distinguished and Which Actor Characteristics Are Subject to Change?

The position of university administrators and policy makers is subject to change due to various factors. The DWDS corpus has been developed and is being supported and regularly updated by a large team of researchers at the Berlin-Brandenburg Academy of Sciences in Germany. They regularly improve the corpus interface and upgrade its content.

### T: What Is Amenable to Improvement (e.g., Teacher Guidance, Parent Control, Mediated Communication with Native Speakers)?

Foreign language advocacy can influence the position of university administrators and policy makers to the benefit of language programs. The instructor's department has been actively engaged in such advocacy efforts. The DWDS support team is very responsive to feedback. They promptly respond to customers' suggestions for improvement and provide helpful explanations when asked clarification questions.

## Step 3: Describe the Architecture of the Distributed Learning Environment

Colpaert (2006) describes this step as follows: "Online technologies allow the design of 'distributed' learning environments. The term 'distributed' not only refers to the physical aspect of location (inside or outside the classroom), but also to the distribution of our focus four components: (a) the learner, (b)

the teacher, (c) the colearner [sic], and (d) the content. […] These eight possibilities amount to numerous flexible combinations […]. Every combination leads to a new setting which can be appropriate for a specific moment in the learning process or for a specific task" (p. 485). The following combinations were utilized in the case course:

**Configuration 1** (Figure 1): All four components were connected in-classroom as in any face-to-face lesson. In this setting, the learners worked with their co-learners with corpus printouts under the teacher's guidance. This configuration was used at the beginning of the course to familiarize the students with the corpus output format as well as later on when the task warranted offline work.
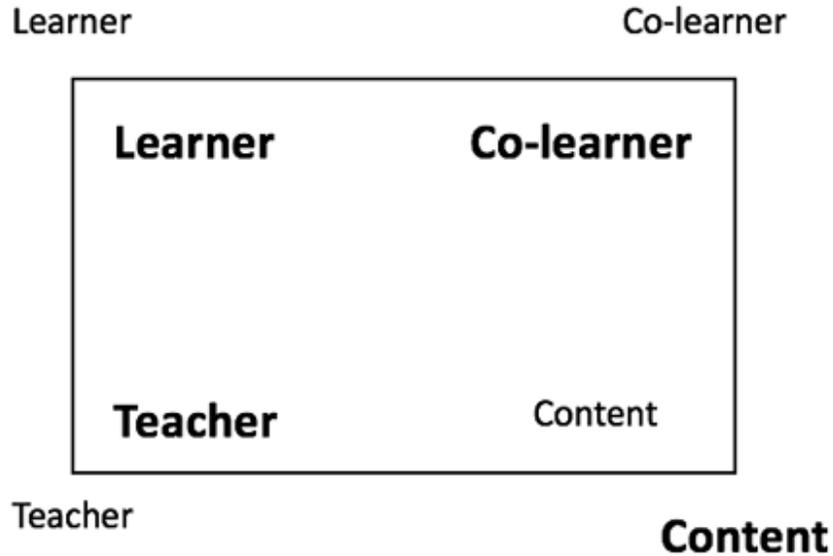
*Figure 1. Architecture of the distributed learning environment 1: Traditional classroom; corpus printouts*
*Modified from Colpaert (2006, p. 485)*
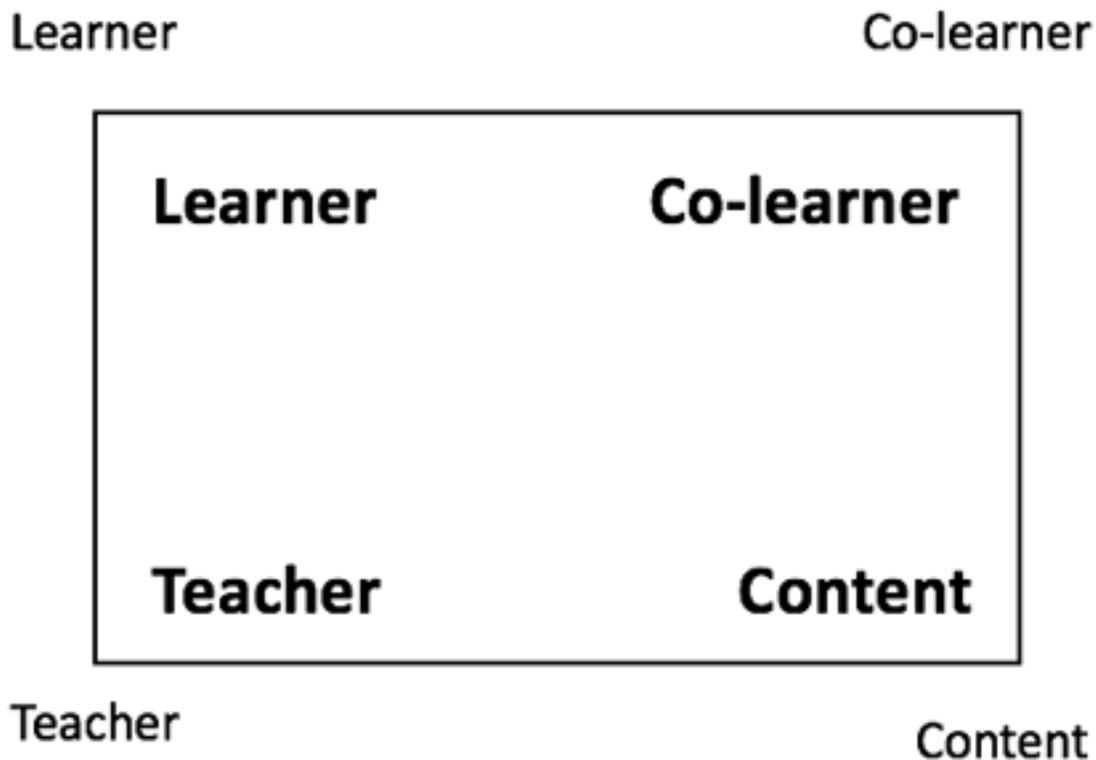


**Configuration 2** (Figure 2): The in-classroom learner connected to in-classroom co-learners and in-classroom teacher as well as to remote content. This configuration was used when the teacher demoed direct corpus searches to learners on a big screen as well as when learners completed direct corpus searches in the computer lab in partner work and under teacher's guidance.

*Figure 2. Architecture of the distributed learning environment 2: Computer lab; direct corpus searches*
*Modified from Colpaert (2006, p. 485)*



**Configuration 3** (Figure 3): The out-of-classroom learner connected to remote content while completing direct corpus search homework assignments.

*Figure 3. Architecture of the distributed learning environment 3: At home; direct corpus searches*
*Modified from Colpaert (2006, p. 485)*
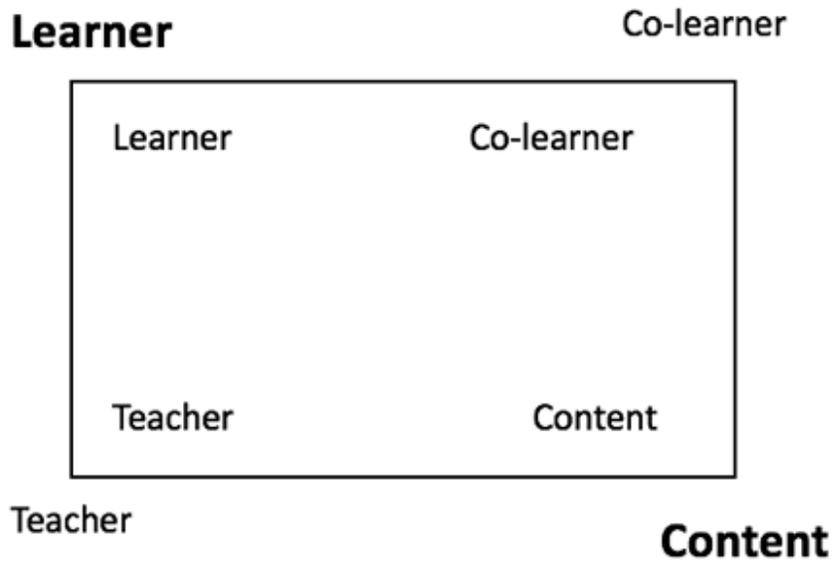
## Step 4: Apply an Activity Framework

According to Colpaert (2006), the activity framework includes three components - information, interaction, and communication – each of which can be mediated (include a human or technological didactic intervention) or non-mediated (include incidental learning or activities in which the student practices in real situations).

In the target course, information was non-mediated because corpus texts are authentic materials (i.e., they are produced in real-life situations and not for educational purposes). Interaction with content was also non-mediated because the target corpus was not designed for language learning purposes and because the software did not provide learners with feedback or help (in contrast to intelligent CALL tutors, see Heift, 2010). Communication, in contrast, was mediated because learners worked with this authentic content following the teacher's guidelines. Furthermore, each learner engaged in mediated oral communication with both the teacher and co-learners, receiving scaffolding while completing in-classroom corpus tasks. It is worth noting that due to this mixture of mediated and non-mediated aspects of corpus work, both explicit and implicit teaching and learning was expected to occur. Explicit learning occurs when learners achieve learning gains in accordance with the outcomes explicitly aimed at in pedagogical tasks, whereas incidental learning occurs during direct corpus searches when learners make unintended serendipitous discoveries (cf. Bernardini, 2000).

## Step 5: Define Linguistic/Didactic Functionalities

Colpaert (2006) explains that dedicated intelligent CALL software may play a range of roles (such as monitor or tutor), whereas non-dedicated systems "never offer appropriate guidance, feedback, tracking, or reporting functionalities" (p. 488). Therefore, corpus software, being non-dedicated, remains only a tool, whereas didactic functionalities and mediation need to be provided by the teacher or materials designer.

## Step 6: Define Personas and Their Goals

This step provides a "link between system requirements (analysis) and the concept of the solution (design)" (Colpaert, 2006, p. 489). Personas are "hypothetical users" that the courseware designer has to have in mind as an average course participant. This hypothetical user should represent the intersection between differential learner characteristics (see step 2), and the goals of this hypothetical user should embody a compromise between various learner, teacher, and institutional goals (see step 2). In this course, the hypothetical user was a learner with average characteristics described in 2.1, who was motivated to improve their L2 proficiency with the goal to use German for personal and occupational purposes. Colpaert also suggests using standardized benchmarks and frameworks to specify practical pedagogical goals in this step. In this study, the goal was set to increase learners' average L2 proficiency by half a CEFR level (from B1.1 to B1.2) by the end of the course. In particular, learners were expected to improve their vocabulary, grammar knowledge, and writing ability. Furthermore, they were expected to develop their corpus literacy (as a particular instance of digital literacy, see Mukherjee, 2006) and to learn how to use corpus resources for autonomous language learning.

## Step 7: Conceptualize

This step "consists of two concurrent and iterative activities: concept development and the application of usefulness criteria" (Colpaert, 2006, p. 490). The elaboration of scenarios for how the personas will use the courseware to achieve their goals (see step 6) is central to the concept development. Scenarios then need to be translated into system tasks with an emphasis on the integration of the tools. In the present case, conceptualization can be operationally defined as the course syllabus and the progression of pedagogical activities. These activities were scheduled to progress from more mediated to less mediated (see step 4) and from predominantly in-classroom to varied in-and-out-of-classroom configurations (see step 3). Six computer lab class meetings were scheduled throughout the semester at two-week intervals with 60 minutes lab time dedicated to corpus-based activities.

Next, the developed concept has to be checked against four usefulness criteria: usability, usage, user satisfaction, and didactic efficiency (Colpaert, 2006, p. 491). The usability of the selected suite of corpus resources (DWDS) was evaluated as high due to its free and open nature, i.e. its availability to all targeted users. Colpaert's usage criterion also checked out because the possible usage modes corresponded to the intended use: DWDS is easily accessible on and off campus and on any electronic devices (computer, tablet, mobile phone). Moreover, DWDS was expected to satisfy the intended user because the software is reliable (few connectivity issues, high speed connection), the interface is user-friendly, and the content is appropriate (see step 2). Due to these features, the users could continue using the software long-term, beyond the course time boundaries. Finally, the didactic efficiency was expected to be high based on a good learner, teacher, and teaching method fit (see step 2).

## Step 8: Specify

In this study, the concept outlined in the syllabus (see step 7) was specified in lesson plans and worksheets, each including detailed guidelines for students for what corpus tools to use and what tasks to complete. The tasks were designed according to three main learning environment configurations (see step 3): as in-class partner work, in-class individual work, and out-of-class individual work. Some assignments included indirect DDL activities with corpus printouts, while others included direct corpus searches in a computer lab. While in class, students were asked to work on corpus tasks individually but also to consult with their partners and to ask the teacher any questions that might arise. They were allowed to finish each corpus task at home if they ran out of time in class. They then brought the completed worksheets to the next class and shared the results with co-learners. The learning targets included word profiles (grammatical information, collocations, changes in word frequency over time, Appendix 1), grammatical and stylistic error correction and prevention (Appendix 2), idiomatic expressions (Appendix 3), and more. The capstone DDL assignment of the course was using the corpus for consultation during learners' work on their out-of-class final paper: an analysis of the course novel. Students were instructed to submit a 1-2 page-long reflective log about how they used DWDS (in English) in addition to this final paper (in German). Sample worksheets are included in Appendices, and more assignment descriptions can be found elsewhere (see Vyatkina, 2018 and 2020, for examples and references).

## Step 9: Perform Pre-Use Evaluation

In this step, teachers evaluate the blueprint developed by the researcher as outlined in the previous steps (Colpaert, 2006, p. 493). The teacher in this study was also the course developer and the researcher. Therefore, this step was modified as pre-use data collection to obtain a baseline for future post-use evaluation (see step 10 below). Colpaert relies on Hubbard's (2003) study for specifying the evaluation steps of his model. Hubbard had surveyed 120 specialists from around the world asking them to list most pressing issues in CALL and then classified their responses into four categories that emerged in the data:

1)   Effectiveness Issues: the relation between technology use and learning gains
2)   Learner-Centered Issues: learner attitudes toward technology
3)   Design-Centered Issues: issues of CALL course, software, or activity design
4)   Research-Centered Issues: aspects and directions of CALL research.

In this study, the last category was not singled out since all evaluation issues were considered research-centered (in line with the DBR method). Only a brief summary of the data is presented here (see Vyatkina, 2016b, for a detailed description of the quantitative study and results).

### Effectiveness Issues

These were explored in a quantitative study of student learning outcomes in their overall L2 German proficiency (measured with a standardized online test) as well as their knowledge of one specific lexico-grammatical target – verb-preposition collocations (measured with a written production test). The pre-use data showed that the overall L2 proficiency of the learners was, on average, at the B1.1 CEFR level. Regarding the specific target, the learners knew only about 35% of the target verb-preposition collocations despite explicit teaching of these forms in previous years as well as immediate prior exposure in the text of the course novel.

### Learner-Centered Issues

These were explored through questionnaires. The pre-use data showed that learners had a very high intrinsic as well as instrumental motivation for learning German; no prior familiarity with corpora; and fairly high expectations for using corpora for language learning.

### Design-Centered Issues

These were operationally defined as the pedagogical goals described in step 1. The overarching goal was for the students to improve their German proficiency and their corpus literacy, as well as their critical and creative thinking and learner autonomy.

## Step 10: Perform Post-Use Evaluation

Post-use evaluation should be performed to ascertain whether the goals set for using the courseware were met. This step includes evaluation of the improvement on all issues evaluated at the pre-use stage (step 9). Only a brief summary of the findings is presented here (see Vyatkina, 2016b, for detailed results).

### Effectiveness Issues

Post-use results showed that students improved their overall German proficiency by half a CEFR level, namely from B1.1 to B1.2 (with 3 students reaching the B2 level and only one remaining at the A2 level). The results also showed significant gains in the accuracy of controlled productive use of the target verb-preposition collocations immediately after DDL instruction and also retention of some of these gains a month later. More specifically, learners knew about 70% of the target forms on the immediate posttest and about 45% of the forms on the delayed post-test. Furthermore, all learners improved: those with lower and higher overall proficiency as well as those with lower and higher DDL receptivity.

### Learner-Centered Issues

These were explored through analysis of learner questionnaire and reflective log responses. The question-naire responses showed that students had a very positive experience with corpus activities. On average, they rated their satisfaction at about 4 on a 5-point-scale, showing that they found DWDS activities interesting, useful, and easy to use for language learning purposes. Moreover, these post-course ratings were on average higher than those in the pre-course questionnaire that measured students' expectations. Prose comments from reflective logs matched these numerical ratings, as can be seen from the following excerpts:

- . . . dwds.de was great because it filled in for me everything that a dictionary can't.
- . . . a corpus such as DWDS is extremely useful. Its particular insight into usage trends and stylistic norms allow an insight into the writing of native speakers that dictionaries and grammar guides simply cannot.
- It would certainly be to my benefit to spend further effort in figuring out the ways in which the German language defines itself. I believe that this would clear up quite a few usage issues as well, especially where one-to-one translations of words or phrases simply do not exist. Further, I expect to continue using corpora such as DWDS, in conjunction with dictionaries and grammar refer-ences, as the tools with which I self-correct my writing.

In summary, these positive comments showed the students' appreciation of the DWDS suite of re-sources which they call "extremely useful," "enormously useful," and even "the most important thing" they learned about in the course. Further important comments were those showing that students came to understand the difference between dictionaries and grammars (that provide prescriptive rules) and corpus-based resources (that provide insight into actual usage patterns). Finally, most of them expressed the intent of using DWDS in the future. Encouragingly, at least five participants have started implement-

ing this plan: after the next course (with no explicit focus on corpora) they took with the researcher, they reported having used DWDS independently for consultation throughout the course.

The noted difficulties included some uncertainty about how to use the corpus independently (one student) as well as a difficulty resulting from insufficient L2 proficiency (one student). However, even the least proficient student indicated that he very much liked using DWDS. This fact shows that learning about DWDS was still useful as this student may return to this resource in the future if his L2 proficiency develops.

## Design-Centered Issues

These were explored through mapping the course outcomes back onto the pedagogical goals described in step 1. It can be concluded that these goals were successfully realized: students improved their German proficiency and their corpus literacy. Their ability to use open access corpus resources for exploring the nuances of the German language also contributed to the development of their critical and creative thinking and personal autonomy. This evaluation led to the researcher's decision to continue using DWDS in future iterations of the course. After this first iteration of the course, the course was taught three more times with slight modifications implemented based on each iterative post-use evaluation. For example, even more guidance was incorporated into instructions and scaffolding to assist students with lower proficiency and some specific linguistic foci have changed.

## SOLUTIONS AND RECOMMENDATIONS

This report on the design, implementation, and evaluation of a DDL-enhanced course contributes to the small but growing body of DBR studies in CALL. It explored how a technological and pedagogical innovation was implemented to suit the specific institutional and instructional context of the study. In order to "provide in-depth engagement with the intersection of technology and language teaching pedagogy" (Kessler, 2016, p. 192), a conceptual framework for using the targeted technologies was established; the instructional context, pedagogical objectives, and technological tools were described; and details of how technology was integrated were provided. Following the DBR approach, this study examined "multiple variables in order to develop a qualitative account that links different instructional conditions with different effects on learning, all the while acknowledging the uniqueness of individual learners and of contexts" (Larsen-Freeman, 2013, p. 23). By mapping the course on a set of carefully defined design specifications (Colpaert, 2006), it has shown how DDL components can be successfully integrated into a language course. Overall, the course was successful as the pedagogical goals set for the course were realized. These positive outcomes support findings from previous research that DDL with open corpus resources leads to learning gains and is positively received by learners (e.g., Bardovi-Harlig et al., 2017; Boulton, 2011; Chang, 2014; Daskalovska, 2015; Kilimci, 2017; Li, 2017). Furthermore, the learners developed their corpus literacy (Mukherjee, 2006) and some of them continued using the corpus autonomously after the course.

Methodologically, the chapter also shows that a pedagogy-driven design model as specified by Colpaert (2006) is a viable approach to designing and implementing CALL courses, thus corroborating Schulze and Liebscher's (2010) study. The present project's success can be attributed to the fact that the course

was carefully conceptualized, adequately tailored to the target learner population and learning context, and implemented using appropriate teaching methods, technology (open access corpus resources), and content. As Colpaert (2006) notes, filling the ADDIE template with specific design details "is a challenging, thought provoking, and revealing activity" (p. 484). It is, however, also rewarding: if technology is used in a judicious and theoretically grounded way, it may benefit language learners by helping them achieve various learning goals from specific aspects of L2 knowledge to developing their overall L2 proficiency to general analytical skills and learning autonomy.

From the DBR perspective, the main outcome of the study is both "a methodological framework and a mediational artifact" (Hung, 2013, p. 77) that combines product- and process-oriented aspects. This successful model of a DDL-enhanced course (product) was used by the author herself and can be used by language educators as a basis for developing new courses and improving existing courses with DDL components (process). More broadly, it is hoped that the study will serve yet another important DBR goal – teacher professional development, since only a deep understanding of pedagogy, technology, and their interaction by teachers can lead to a successful implementation of instructional innovations (Rodríguez & Pardo-Ballester, 2013). More DBR studies that focus on DDL coupled with corpus user guides and corpus activities collections designed for and accessible to language teachers can go a long way toward achieving this goal.

## FUTURE RESEARCH DIRECTIONS

This chapter contributes to the main theme of this edited volume by exemplifying how DDL, a relatively new technology, can be implemented in a language class following a pedagogy-driven approach. It presented a report on one specific course design experience. As such, it proposes a potential model for future replications, although its success cannot be generalized to all possible language teaching settings. This study is conceived of as the first stage in a longitudinal iterative DBR process that can progress through a systematic altering and refining of contextual design aspects (Larsen-Freeman, 2013, p. 27). Future studies should report on other cases of DDL implementation in various contexts, which would collectively lead to generating a robust knowledge base. Furthermore, although the DDL research body is constantly growing (see Boulton & Cobb, 2017; Vyatkina & Boulton, 2017), more empirical studies reporting on DDL processes, perspectives, and outcomes are needed, including DBR studies. Publication of open access materials for DDL as well as studies reporting on the effectiveness of these materials are also needed. Such a project devoted to the development and testing of teaching resources based on the DWDS corpus is currently under way, with new teaching modules being regularly released with free and open access (Vyatkina, 2020).

## CONCLUSION

This chapter has described an exemplar of a practical application of a pedagogy-based model to the design, implementation, and evaluation of a high-intermediate university-level German as a Foreign Language course with a significant DDL component. It is hoped that this example will inspire teachers to use DDL in their own classes at the level and to the extent warranted by their own teaching and learning context: from brief single DDL interventions to full-fledged DDL courses and from highly scaffolded activities

for beginning and younger students to more complex and autonomous activities for advanced and older students. Furthermore, the ideal development would be engaging these teachers in conducting their own classroom-based studies for the DBR cycle to continue and the collective knowledge base to grow.

## REFERENCES

Anthony, L. (n. d.). *AntConc* [Computer software]. Tokyo, Japan: Waseda University. Retrieved from https://www.laurenceanthony.net/software

Bardovi-Harlig, K., Mossman, S., & Su, Y. (2017). The effect of corpus-based instruction on pragmatic routines. *Language Learning & Technology*, *21*(3), 76–103.

Baumgartner, E., Bell, P., Brophy, S., Hoadley, C., His, S., Joseph, D., ... Tabak, I. (2003). Design-based research: An emerging paradigm for educational inquiry by the design-based research collective. *Educational Researcher*, *32*(1), 5–8. doi:10.3102/0013189X032001005

Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Ann Arbor, MI: University of Michigan Press. doi:10.3998/mpub.371534

Bernardini, S. (2000). Systematising serendipity: Proposals for concordancing large corpora with language learners. In L. Burnard, & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 225–234). New York: Peter Lang.

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. doi:10.1017/S0261444817000167

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, *67*(2), 348–393. doi:10.1111/lang.12224

Byrnes, H. (2005). Literacy as a framework for advanced language acquisition. *ADFL Bulletin*, *37*(1), 11–15. doi:10.1632/adfl.37.1.11

Cazden, C. B., Cope, B., Fairclough, N., Gee, J. P., Kalantzis, M., Kress, G. R., ... Nakata, N.The New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, *66*(1), 60–93. doi:10.17763/haer.66.1.17370n67v22j160u

Chang, J. Y. (2014). The use of general and specialized corpora as reference sources for academic English writing: A case study. *ReCALL*, *26*(2), 243–259. doi:10.1017/S0958344014000056

Cho, H. (2015). Task dependency effects of collaboration in learners' corpus consultation: An exploratory case study. *ReCALL*, *28*(1), 44–61. doi:10.1017/S0958344015000130

Cobb, T. (n. d.). *Compleat lexical tutor* [Computer software]. Retrieved from http://www.lextutor.ca

Colpaert, J. (2004). *Design of online interactive language courseware: Conceptualization, specification, and prototyping: Research into the impact of linguistic-didactic functionality on software architecture* (Unpublished doctoral dissertation). University of Antwerp, Belgium.

Colpaert, J. (2006). Pedagogy-driven design for online language teaching and learning. *CALICO Journal*, *23*(3), 477–497. doi:10.1558/cj.v23i3.477-497

Consortium, B. N. C. (2007). *The British National Corpus, version 3* (X. M. L. Bnc, Ed.). Oxford, UK: Bodleian Libraries, University of Oxford; Retrieved from http://www.natcorp.ox.ac.uk/

Cotos, E., Link, S., & Huffman, S. (2017). Effects of DDL technology on genre learning. *Language Learning & Technology*, *21*(3), 104–130.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg, France: Language Policy Unit.

Crosthwaite, P. (2017). Retesting the limits of data-driven learning: Feedback and error correction. *Computer Assisted Language Learning*, *30*(6), 447–473. doi:10.1080/09588221.2017.1312462

Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, *28*(2), 130–144. doi:10.1080/09588221.2013.803982

Davies, M. (2008). *The corpus of contemporary American English (COCA): 560 million words, 1990-present*. Retrieved from https://corpus.byu.edu/coca/

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

Ebrahimi, A., & Faghih, E. (2016). Integrating corpus linguistics into online language teacher education programs. *ReCALL*, *29*(1), 120–135. doi:10.1017/S0958344016000070

Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. New York: Routledge.

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, The Netherlands: John Benjamins. doi:10.1075cl.69.02flo

Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. *ReCALL*, *26*(2), 128–146. doi:10.1017/S0958344014000093

Friginal, E. (2018). *Corpus linguistics for English teachers*. New York: Routledge. doi:10.4324/9781315649054

Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL*, *26*(2), 225–242. doi:10.1017/S0958344014000044

Godwin-Jones, R. (2017). Data-informed language learning. *Language Learning & Technology*, *21*(3), 9–27.

Heift, T. (2010). Developing an intelligent language tutor. *CALICO Journal*, *27*(3), 443–459. doi:10.11139/cj.27.3.443-459

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

Hubbard, P. (2003). A survey of unanswered questions in CALL. *Computer Assisted Language Learning*, *16*(2-3), 141–154. doi:10.1076/call.16.2.141.15882

Hung, H.-T. (2013). Capitalizing on the dual goals of design-based research in computer-assisted language learning contexts. In J. C. Rodríguez, & C. Pardo-Ballester (Eds.), *Design-based research in CALL* (pp. 67–80). San Marcos, TX: Computer Assisted Language Instruction Consortium.

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

Karras, J. N. (2016). The effects of data-driven learning upon vocabulary acquisition for secondary international school students in Vietnam. *ReCALL*, *28*(2), 166–186. doi:10.1017/S0958344015000154

Kennedy, C., & Miceli, T. (2017). Cultivating effective corpus use by language learners. *Computer Assisted Language Learning*, *30*(1-2), 91–114. doi:10.1080/09588221.2016.1264427

Kern, R. (2000). *Literacy and language teaching*. Oxford, UK: Oxford University Press.

Kessler, G. (2016). Call for papers for language teaching & technology forum. *Language Learning & Technology*, *20*(2), 192.

Kilimci, A. (2017). Integrating cognitive linguistics insights into data-driven learning: Teaching vertical prepositions. *Journal of Language and Linguistic Studies*, *13*(2), 681–719.

Larsen-Freeman, D. (2013). A promising combination: complexity theory, design-based research, and CALL. In J. C. Rodríguez, & C. Pardo-Ballester (Eds.), *Design-based research in CALL* (pp. 23–29). San Marcos, TX: Computer Assisted Language Instruction Consortium.

Lee, H., Warschauer, M., & Lee, J. H. (2018). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, *40*(5), 721–753. doi:10.1093/applin/amy012

Li, S. (2017). Using corpora to develop learners' collocational competence. *Language Learning & Technology*, *21*(3), 153–171.

Loewen, S. (2015). *Introduction to instructed second language acquisition*. New York: Routledge.

Mirzaei, A., Rahimi Domakani, M., & Rahimi, S. (2015). Computerized lexis-based instruction in EFL classrooms: Using multi-purpose LexisBOARD to teach L2 vocabulary. *ReCALL*, *28*(1), 22–43. doi:10.1017/S0958344015000129

Moon, S., & Oh, S.-Y. (2017). Unlearning overgenerated *be* through data-driven learning in the secondary EFL classroom. *ReCALL*, *30*(1), 48–67. doi:10.1017/S0958344017000246

Mueller, C. M., & Jacobsen, N. D. (2016). A comparison of the effectiveness of EFL students' use of dictionaries and an online corpus for the enhancement of revision skills. *ReCALL*, *28*(1), 3–21. doi:10.1017/S0958344015000142

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, Germany: Peter Lang.

Paesani, K., Allen, H., & Dupuy, B. (2015). *A multiliteracies framework for collegiate foreign language teaching*. Upper Saddle River, NJ: Pearson.

Pardo-Ballester, C., & Rodríguez, J. C. (2009). Using design-based research to guide the development of online instructional materials. In C. A. Chapelle, H. G. Jun, & I. Katz (Eds.), *Developing and evaluating language learning materials* (pp. 86–102). Ames, IA: Iowa State University.

Pérez-Paredes, P., Sánchez-Tornel, M., Alcaraz Calero, J., & Aguado Jiménez, P. (2011). Tracking learners' actual uses of corpora: Guided vs non-guided corpus consultation. *Computer Assisted Language Learning*, *24*(3), 233–253. doi:10.1080/09588221.2010.539978

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

Quinn, C. (2015). Training L2 writers to reference corpora as a self-correction tool. *ELT Journal*, *69*(2), 165–177. doi:10.1093/elt/ccu062

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

Rezaee, A. A., Marefat, H., & Saeedakhtar, A. (2015). Symmetrical and asymmetrical scaffolding of L2 collocations in the context of concordancing. *Computer Assisted Language Learning*, *28*(6), 532–549. doi:10.1080/09588221.2014.889712

Rodríguez, J. C., & Pardo-Ballester, C. (2013). Introduction. In J. C. Rodríguez, & C. Pardo-Ballester (Eds.), *Design-based research in CALL* (pp. 1–7). San Marcos, TX: Computer Assisted Language Instruction Consortium.

Schicker, J. (2018). Teaching writing and grammar through critical thinking in a genre-based upper-level undergraduate German class. *Die Unterrichtspraxis / Teaching German, 51*(1), 63-75.

Schulze, M., & Liebscher, G. (2010). Going in cycles: Courseware and material development for written communication. *CALICO Journal*, *27*(3), 554–563. doi:10.11139/cj.27.3.554-563

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

Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL*, *26*(2), 184–201. doi:10.1017/S0958344014000081

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.

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. (2016a). Data-driven learning for beginners: The case of German verb-preposition colloca-tions. *ReCALL*, *28*(2), 207–226. doi:10.1017/S0958344015000269

Vyatkina, N. (2016b). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology*, *20*(3), 159–179.

Vyatkina, N. (2018). Language corpora for L2 vocabulary learning: Data-driven learning across the cur-riculum. In P. Ecke, & S. Rott (Eds.), *Understanding vocabulary learning and teaching: Implications for language program development* (pp. 121–145). Boston, MA: Cengage Learning.

Vyatkina, N. (2020). *Incorporating corpora: Using corpora to teach German to English-speaking learners [Online instructional materials]. Lawrence, KS: KU Open Language Resource Center.* https://corpora.ku.edu

Vyatkina, N., & Boulton, A. (2017). Corpora in language learning and teaching. *Language Learning & Technology*, *21*(3), 1–8.

Wong, T., & Lee, J. S. (2016). Corpus-based learning of Cantonese for Mandarin speakers. *ReCALL*, *28*(2), 187–206. doi:10.1017/S0958344015000257

Wood, P. (2011). Computer assisted reading in German as a foreign language, developing and testing an NLP-based application. *CALICO Journal*, *28*(3), 662–676. doi:10.11139/cj.28.3.662-676

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

## ADDITIONAL READING

Baumgartner, E., Bell, P., Brophy, S., Hoadley, C., His, S., Joseph, D., ... Tabak, I. (2003). Design-based research: An emerging paradigm for educational inquiry by the design-based research collective. *Edu-cational Researcher*, *32*(1), 5–8. doi:10.3102/0013189X032001005

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, *67*(2), 348–393. doi:10.1111/lang.12224

Colpaert, J. (2006). Pedagogy-driven design for online language teaching and learning. *CALICO Journal*, *23*(3), 477–497. doi:10.1558/cj.v23i3.477-497

Godwin-Jones, R. (2017). Data-informed language learning. *Language Learning & Technology*, *21*(3), 9–27.

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.

Paesani, K., Allen, H., & Dupuy, B. (2015). *A multiliteracies framework for collegiate foreign language teaching*. Upper Saddle River, NJ: Pearson.

Rodríguez, J. C., & Pardo-Ballester, C. (2013). Introduction. In J. C. Rodríguez & C. Pardo-Ballester (Eds.), *Design-based research in CALL* (pp. 1–7). San Marcos, TX: Computer Assisted Language Instruction Consortium.

Vyatkina, N. (2018). Language corpora for L2 vocabulary learning: Data-driven learning across the curriculum. In P. Ecke & S. Rott (Eds.), *Understanding vocabulary learning and teaching: Implications for language program development* (pp. 121–145). Boston, MA: Cengage Learning.

## KEY TERMS AND DEFINITIONS

**ADDIE:** An industrial design model that includes the steps of analysis, design, development, implementation, and evaluation.

**Computer-Assisted Language Learning (CALL):** The method of teaching and learning languages with the help of electronic tools and media.

**Corpus:** A large collection of texts selected in a systematic way and stored as an electronic database.

**Corpus literacy:** A particular aspect of multiliteracies; ability to use corpus resources for autonomous language learning.

**Data-Driven Learning (DDL):** The method of teaching and learning languages with the help of corpora.

**Design-Based Research (DBR):** A research method used to study innovative learning environments that involves iterative steps of the exploration of the design, enactment of interventions, evaluation and analysis of the outcomes, and redesign.

**Multiliteracies:** A pedagogical approach that capitalizes on collaboration, problem solving, and critical reflection and aims at developing learners' cultural, linguistic, communicative, and technological proficiency.

## APPENDIX 1

### Sample DDL Worksheet: Word Profile

(instructions translated into English)

1. Choose a new noun (unfamiliar to you) from the first chapter of the course novel. Enter this word in the search line on the frontpage of the DWDS corpus (www.dwds.de) and click on the search icon.
2. Write down the grammatical information about your search word (grammatical gender - *der, die, das*, the genitive form, and the plural form):
3. Write down the word definition in German:
4. Scroll down to the word profile (*Wortprofil*) tool (presented as a word cloud). Write down some collocations of your search word, i.e., its most frequent neighbors. These are words that appear in the LARGEST font.
   a. 3-4 nouns:
   b. 3-4 adjectives:
   c. 3-4 verbs:
   d. 3-4 other parts of speech:
5. In the word profile tool, click, in turn, on three words from those you noted above and write down one example sentence with each of them:
   a.
   b.
   c.
6. Bring this worksheet to the next class.

## APPENDIX 2

### Sample DDL Worksheet: Error Correction and Prevention

(instructions translated into English)

   **Grammar and spelling.** Analyze the **concordances** from the ZEIT corpus (Figures 4, 5, and 6)[1] and decide how you can correct the following erroneous sentences. Rewrite the sentences and explain the rule.

## Model

*Figure 4. Concordance set 1*
Source: DWDS corpus

| Am | **Anfang stand** die Anfrage, ob sie als Gastdozentin Vorlesungen in interkultureller Ko... |
| Am | **Ende läuft** die Debatte auf die Frage hinaus, wie wichtig Arbeitnehmern Freizeit und ... |
| Im | **März kehrte** er nicht mehr von einem Heimaturlaub zur Truppe zurück. |
| Im | **Nachwuchs versuchen** wir, die Antizipationsfähigkeit zu schulen. |
| Im | **Motel schlafe** ich den Schlaf der Gezechten. |

**Erroneous sentence:** *Am Morgen, ich gehe zur Uni.*
**Correction:** *Am Morgen gehe ich zur Uni.*
**Explanation:** In German, the verb always takes the second position in the main clause and there is no comma after the sentence constituent that occupies the first position.

*Figure 5. Concordance set 2*
Source: DWDS corpus

| Er | **ist Zuhörer**, Therapeut, Trostspender. |
| Er | **ist Familie**." |
| Er | **ist Ausstellungsmacher**, der Exponate präsentiert. |
| Er | **ist Experte** für das weithin Unsichtbare. |
| Er | **war Alkoholiker**, er war Nazi, hatte dieses knallharte Tattoo: "wollte lieben lernte has... |

*Figure 6. Concordance set 3*
Source: DWDS corpus

| Sie | **ist Opernregisseurin** geworden. |
| Sie | **ist Wirtschaftswissenschaftlerin**, das kühle Vokabular des Kapitalismus geht ihr lei... |
| Sie | **ist Protestantin**. |
| Sie | **ist Geschäftsführerin** des Economic Security Project aus San Francisco, einer Initiati... |
| Sie | **ist Teil** des Walk of Modern Art, eines Parcours der Gegenwartskunst im öffentlichen ... |

**Task**

**Erroneous sentence:** *Er will ein Lehrer sein.*
**Correction:**
**Explanation:**

**APPENDIX 3**

**Sample DDL Worksheet: Idiomatic Expressions**

(instructions translated into English)

Look at the concordances from the DWDS corpus (Figure 7)[1]. What expression is repeated in each example? Is it repeated verbatim or are there variations? Write down the results of your analysis below.

*Figure 7. Concordance set 4*
*Source: DWDS corpus*

| | | |
|---|---|---|
| Man sprach aus der Machtposition, bis sie manchem | **zu** | **Kopf stieg**. |
| Das kann einem ja nur | **zu** | **Kopf steigen**. |
| Netanjahu, dem die Triumphe seiner Schlauheit | **zu** | **Kopf gestiegen** sein mögen, wird die Brüchigkeit seine… |
| Hier hat einer bauen lassen, dem die Macht | **zu** | **Kopf gestiegen** war. |
| …ädchen der Nation wird die Zicke, der der schnelle Ruhm | **zu** | **Kopf gestiegen** ist. |
| …herweise ist diese Glanztat dem Herrn Liu Wenzhe etwas | **zu** | **Kopf gestiegen**, in jedem Fall veröffentlichte er das Bu… |
| …litikern im Westen seien Euphorie und Triumphalismus | **zu** | **Kopfe gestiegen**. |
| … damit zusammen, dass ihm Erfolg und Wertschätzung | **zu** | **Kopf gestiegen** seien, wie ein Autor der Süddeutschen… |
| …e irgendwann nicht wiedererkennen, weil ihm der Erfolg | **zu** | **Kopf gestiegen** ist. |
| …g gibt den Pinken recht, auch wenn er ihnen hie und da | **zu** | **Kopf gestiegen** scheint. |

**The expression:**
**The variations:**
**The meaning:**