

NINA VYATKINA

*University of Kansas, United States*

*(email: vyatkina@ku.edu)*

## **Abstract**

Research on Data-Driven Learning (DDL), or teaching and learning languages with the help of electronic corpora, has shown that it is both effective and efficient. Nevertheless, DDL is still far from common pedagogical practice, not least because the empirical research on it is still limited and narrowly focused. This study addresses some gaps in that research by exploring the effectiveness of DDL for teaching low-proficiency learners lexico-grammatical constructions (verb-preposition collocations) in German, a morphologically rich language. The study employed a pretest-posttest design with intact third- and fourth-semester classes for German as a foreign language at a US university. The same collocations were taught to each group during one class period, with one group at each course level taking a paper-based DDL lesson with concordance lines from a native-speaker corpus and the other one taking a traditional rule-based lesson with textbook exercises. These constructions were new to third-semester students, whereas fourth-semester students had been exposed to them in the previous semester. The results show that, whereas the DDL method and the traditional method were both effective and resulted in lexical and grammatical gains, DDL was more effective for teaching new collocations. The study thus argues in favor of using paper-based DDL in the classroom at lower proficiency levels and for languages other than English.

Keywords: data-driven learning, paper-based, guided induction, morphology, German

## **1 Introduction**

The concept of Data-Driven Learning (DDL) was introduced to the field of L2 (second language) learning by Johns (1990). DDL is associated with “using the tools and techniques of corpus linguistics for pedagogical purposes” (Gilquin & Granger, 2010: 359). Since Johns’ pioneering work, research exploring the effectiveness of DDL has been expanding and has by now convincingly shown that this method can be beneficial for various instructional foci and is applicable in many institutional contexts. Nevertheless, recent overviews show that “the field has yet to reach full maturity” (Boulton & Pérez-Paredes, 2014: 122) and that “the direct uses of corpora in language teaching are treated rather marginally in the literature in the field” (Leńko-Szymańska & Boulton, 2015: 3). In addition to the general scarcity of DDL studies in language teaching research, the foci of available studies have been narrow with regard to target language (English to the exclusion of other languages), L2 proficiency levels (intermediate to advanced), participating instructors (DDL researchers rather than regular teachers), and linguistic targets (primarily lexical and morphologically simple grammatical features). This study aims to expand the empirical DDL research body by addressing these limitations – targeting German as an L2, very low proficiency levels, interventions conducted by regular instructors, and complex lexico-grammatical items. Furthermore, this study investigates the feasibility of brief one-time DDL interventions within non-DDL curricula as well as DDL effectiveness for teaching new versus previously learned linguistic structures, another underexplored area.

## 2 Literature Review

### 2.1 Theoretical background

DDL is well compatible with usage-based Second Language Acquisition (SLA) theories that conceive of language as an open-ended dynamic system that emerges by way of probabilistic bottom-up abstraction rather than a fixed system that follows categorical top-down rules (Ellis, 2014). One of the most important tenets of usage-based approaches is inseparability of grammar and lexis, conceptualized by Ellis (2014: 399) as follows:

Language is intrinsically symbolic, constituted by a structured inventory of constructions as conventionalized form-meaning pairings used for communicative purposes. [...] Adult language knowledge consists of a continuum of linguistic constructions of different levels of complexity and abstraction. Constructions can comprise concrete and particular items (as in words and idioms), more abstract classes of items (as in word classes and abstract constructions), or complex combinations of concrete and abstract pieces of language (as mixed constructions). No rigid separation exists between lexis and grammar.

The second theoretical principle behind DDL is Schmidt's (1990) noticing hypothesis, which posits that some level of learner awareness of the target L2 construction is necessary for it to be learned. In order to attract learners' attention to L2 targets, specific instructional methods have been proposed, notably input enrichment and input enhancement. Input enrichment, or input flood, is increase of frequency of the target feature in the input (Trahey & White, 1993) and input enhancement refers to making the target feature more salient, for example with typographical means such as bolding, underlining, or color marking (Sharwood Smith, 1993). DDL is inherently conducive to both teaching techniques as corpora can supply a large number of attested language samples containing the target construction (input enrichment) and these samples can be retrieved from corpora with the help of concordance tools in form of stacked concordance lines with the target construction highlighted and centered (input enhancement). Although input enrichment and enhancement had been originally proposed as rather implicit instruction methods (i.e., those facilitating learner awareness at the level of subconscious noticing), empirical SLA research has shown that implicit instruction is often insufficient (especially for non-salient L2 targets) and that more explicit methods (i.e., those facilitating learner awareness at the level of conscious understanding) are necessary (Ellis, 2005; Sharwood Smith, 2013; Spada & Tomita, 2010). The specific method that has been adopted in most DDL interventions can be characterized as *explicit inductive* within the implicit–explicit / inductive–deductive taxonomy (DeKeyser, 2003: 314). In this method, also termed “discovery learning” (Bernardini, 2002), rich and enhanced input (e.g., concordance lines) serves as material for learners' noticing and analysis of language patterns. As a result, learners arrive at generalizations about the usage of the target constructions, which, in turn, lead to the complexification and expansion of their L2 knowledge (Flowerdew, 2015).

The third principle is that of learner autonomy, the development of which has been repeatedly pointed out as an important long-term benefit of DDL (e.g., Chambers & O'Sullivan, 2004). However, it has also been acknowledged that this development should be mediated through teacher and peer assistance. The adoption of the principle of the mediated nature of language

development from sociocultural theory (see Flowerdew, 2015) has led to the modification of the inductive discovery method to the so-called *guided* induction approach. This method, originally proposed by Herron and Tomasello (1992) in the area of general language teaching, has recently been adopted by many DDL teachers-researchers (Flowerdew, 2009, 2015; Frankenberg-Garcia, 2014; Huang, 2008; Yoon & Jo, 2014). Smart (2014: 187) describes guided induction as follows:

Guided induction is a particular evolution of the inductive approach where learners are first presented with language samples in the form of an interactive task that guides them to discover the language structure they contain; the teacher has an active role in facilitating these tasks, but typically does not present explicit grammar rules. Learners are then guided to produce the language structure in meaningful communicative tasks.

It must be noted that guidance implies not only teacher but also peer scaffolding (cf. “interactive task”), when learners collaboratively work on corpus tasks and assist each other in the inductive discovery process (Flowerdew, 2015; Kennedy & Miceli, 2001).

In sum, the abovementioned theoretical underpinnings inform the main pedagogical principles behind DDL, some of which stand in stark contrast to widely spread conventional teaching techniques as summarized by Flowerdew (2015: 15-16): (1) a lexico-grammatical approach as opposed to the separation of lexis and grammar into different curricular modules; (2) reliance on naturally occurring attested language as opposed to concocted textbook examples; and (3) guided inductive methods based on rich and enhanced input as opposed to rule-based deductive methods.

## ***2.2 Empirical DDL research***

Although DDL research is still a young area of inquiry, it has accumulated enough empirical studies for several research syntheses (Boulton & Pérez-Paredes, 2014; Chambers, 2007; Leńko-Szymańska & Boulton, 2015) and the first meta-analysis (Cobb & Boulton, 2015). Having integrated the results of quantitative studies published up to 2012, Cobb and Boulton showed large effect sizes for both learning gains with DDL methods and their superiority to conventional teaching methods. The overwhelming majority of L2 foci in these studies included English collocations (frequently co-occurring patterns of words) taught with DDL methods to learners in different countries and instructional settings. Most of the earlier studies explored the so-called “hard” (Gabrielatos, 2005) version of DDL, where learners searched corpora online on their own with minimal guidance from their teachers. However, this version was soon recognized as not equally feasible for all instructional settings. Online corpus searches by learners turned out to be challenging for many students and teachers alike due to high cognitive task demands, lack of technological expertise and support, or simply absence of computer-equipped classrooms (Farr, 2008; Tian, 2005). This hurdle has led to the emergence of numerous DDL modifications that can be positioned on a “cline” from hard to soft versions (Mukherjee, 2006), depending on the medium (computer or paper) and task difficulty (open-ended and controlled, addressing variable or fixed rules). The number of studies devoted to “softer” DDL versions remains small but it has been recently growing, as evidenced by the overview below.

The effectiveness of paper-based DDL interventions, in which learners work with concordances selected and printed on worksheets by their teachers, has been investigated in a number of exploratory and experimental studies (see Boulton, 2010, for an overview). All studies that employed quantitative methods showed that paper-based DDL led to significant gains in

learners' knowledge of a variety of L2 English targets (e.g., phrasal verbs, connectors, and passive voice constructions), and that DDL was either as effective as or better than conventional rule-based methods. Furthermore, quantitative studies that compared computer-based and paper-based DDL found no difference in the effectiveness of these two methods (Vyatkina, in press; Boulton, 2012). Boulton (2008, 2009, 2010, 2012), in particular, has shown that "taking the computer out of the equation" (Boulton, 2010: 234) in paper-based DDL does not compromise its benefits for language instruction. Several studies have also shown that paper-based DDL works at both higher and lower proficiency levels (Koosha & Jafarpour, 2006; Tian, 2005; Yoon & Jo, 2014). Boulton (2008, 2009, 2010, 2012) has conducted systematic paper-based DDL research with L1 French university learners of English at relatively low levels of proficiency (levels A2-B1 of the Common European Framework of Reference, or CEFR; Council of Europe, 2001) despite up to 8 years of instruction. Boulton showed that DDL helped such learners to acquire certain linguistic targets that were often impermeable to traditional instruction methods. A number of studies focused on verb-preposition collocations, which are the target of the present study. These constructions are notoriously difficult for learners with any L1-L2 background because of form-meaning mapping mismatches between languages (Kennedy & Miceli, 2001; Nesselhauf, 2004). DDL studies that have targeted English verb-preposition collocations either alongside other foci (Boulton, 2010; Frankenberg-Garcia, 2014) or as a sole instructional focus (Koosha & Jafarpour, 2006) have shown positive DDL effects. It must be noted that most DDL interventions have been administered by the researchers themselves with two notable exceptions: Yoon (2008), who closely worked with an instructor who administered computer-based DDL to his students over a longer period; and Boulton (2010), who reports on a successful short paper-based DDL intervention administered by several teachers.

Several recent studies have fleshed out specific features of DDL conducive to language learning. Frankenberg-Garcia (2014) demonstrated that learners who worked with multiple concordance lines for English collocations achieved greater gains in L2 production than learners who worked with one concordance line, worked with definitions of the target structures, or were in a control group. Smart (2014) showed that for English passive constructions, the guided inductive DDL group outperformed two deductive groups, one which worked with corpus-informed materials and another with constructed examples from textbooks. The only study (conducted in the computer-based DDL context) that explicitly singled out the specific knowledge effect is Chan and Liou (2005). They found that learners with lower entry-level collocational knowledge made greater gains than learners with higher entry-level knowledge. However, the students in Chan and Liou's study had had some previous knowledge of the focal collocations, which is in line with virtually all DDL research that has focused on "'known' but error-prone items" (Boulton & Pérez-Paredes, 2014: 124) that have been claimed to be more amenable to DDL interventions than completely new items (Cobb, 1999; Nesselhauf, 2004). On the other hand, informal feedback from Boulton's (2008) participants suggested that they had had no previous knowledge of the target items (two phrasal verbs), which they successfully learned with paper-based DDL. Nevertheless, empirical comparisons of how DDL fares in teaching previously learned versus new items (a task especially relevant at early stages of instruction) are yet to be conducted.

The results of the research reviewed above thus suggest that the primary DDL benefits are richness of input and the use of a guided inductive approach, that both computer-based and paper-based DDL can be equally effective provided that they follow these principles, and that

paper-based DDL works even for lower proficiency learners. However, a number of research gaps still remain, some of which will be addressed in the present study.

### ***2.3 Remaining gaps and the goals of this study***

Despite the positive findings discussed above, Mauranen's (2004: 208) call, issued a decade ago, is still relevant and urgent: "to make a serious contribution to language teaching, corpora must be adopted by ordinary teachers and learners in ordinary classrooms". One way to bring this project to fruition is to promote paper-based DDL that is more feasible for "ordinary" settings. However, in addition to the general scarcity of studies that compare the effectiveness of paper-based DDL and non-DDL methods (see review above), there are a number of design limitations that have yet to be addressed. The first research gap concerns the target language focus. So far studies have been almost exclusively limited to English. This, in turn, has led to primary attention directed at lexical linguistic targets and analytical grammatical constructions (e.g., passive and subjunctive verb forms), whereas morphologically complex targets (e.g., nominal inflection), which are largely irrelevant to English, have flown under the radar. A case in point is verb-preposition collocations, which in English essentially belong to the lexical domain as they consist of a content word and a functional word. In contrast, German verb-preposition collocations represent a complex lexico-grammatical construction because either the verb or the preposition assigns grammatical case to the subcategorized noun phrase indexed by inflectional markers on that phrase. The German nominal inflectional paradigm represents a salient case of morphological complexity (Pallotti, 2015), and verb-preposition-case collocations have been shown to be difficult for L2 German learners at different proficiency levels (Baten, 2011; Vinagre & Muñoz, 2011). At the same time, explicit teaching has been shown to be necessary to attract learners' attention to inflectional morphology – an abstract, low-salience grammatical feature (DeKeyser, 2003; Ellis, 2005). Therefore, it is worth exploring whether DDL as an explicit inductive method may fare better than traditional explicit deductive methods to achieve this goal.

Furthermore, the empirical DDL research field is in urgent need of replication studies, as any classroom research is by nature small scale and quasi-experimental with intact classes functioning as participant groups. Partial replication studies can enhance the generalizability of classroom research findings (Chun, 2012; Porte, 2012), but such studies are still rare in DDL (see, however, Frankenberg-Garcia, 2014). This study aims to take this avenue by partially replicating two recent DDL studies. The first study is Boulton (2010) who compared the effectiveness of brief interventions for teaching a number of English collocations to L1 French EFL students at the A2-B1 proficiency levels. The study showed that the paper-based DDL method was marginally better than the non-DDL dictionary-based method, with more students improving their scores on more items, and that both methods were better than no instruction. The study thus argued in favor of applying paper-based DDL along with more traditional methods in teaching some lexico-grammatical targets to low-proficiency students. The second study is Vyatkina (in press), who compared the effects of computer-based and paper-based DDL for teaching German verb-preposition collocations to L1 English learners at the B1 proficiency level. The study showed that both methods were equally effective, with higher and longer lasting gains for controlled production tasks (gap filling) than for free production tasks (sentence writing). This study is similar to Boulton (2010) in that it compares the effects of a paper-based DDL and a non-DDL method administered by regular teachers as brief, narrowly focused interventions embedded within a non-DDL curriculum and without prior DDL training of the participants. It

also replicates Vyatkina (in press) in that it focuses on learners of German and targets German verb-preposition collocations as complex lexico-grammatical constructions. The novel features of the present are that its participants have very low L2 proficiency levels (CEFR A2 and below), that it assigns intact classes to different treatment conditions (unlike Boulton and Vyatkina who administered different treatments to the same cohort of learners), that it separates lexical and morpho-syntactic learning, and that it compares the DDL effects for teaching completely new versus previously learned collocations. The broader purpose of this study is, similar to Boulton's (2010: 541), to "counter a number of frequent objections to DDL and contribute to greater awareness of its potential".

### **3 Design**

#### ***3.1 Research questions***

The study explores the following research questions:

1. Does learner lexical and morpho-syntactic knowledge improve following focused instruction (as demonstrated by written performance on a closed gap-filling task)?
2. Are the gains higher following the DDL or the non-DDL treatment?
3. Are there interactions between treatment, course level, and linguistic items (lexical or morpho-syntactic)?

#### ***3.2 Participants and institutional setting***

The study was administered at a large public North American university. Participants were recruited from the third- and fourth-semester German classes in a four-semester-long program that fulfills the foreign language requirement for certain majors. The classes met three times per week for 50 minutes. This is a multi-section program in which all sections of the same course follow a uniform syllabus and use the same textbook. The instructional approach combines a communicative approach with focus-on-form activities, and all courses devote an approximately equal amount of time to speaking, writing, reading, listening, vocabulary, grammar, and cultural learning. All syllabi also have a substantial learning-with-technology component including an electronic workbook and biweekly computer lab meetings, mostly devoted to searching German websites for cultural information, but without a DDL component. All seven experimental classes in this study were taught by graduate student instructors under the researcher's supervision. Some instructors had basic knowledge of corpora and DDL from their graduate coursework (see Vyatkina, 2013) but none of them used DDL in teaching the focal classes.

Altogether, 88 students participated in the two iterations of the experiment. All of them had American English as their L1. The average age was 21 (18-35 range). Gender was not considered a variable in this study, however it can be mentioned that although the proportion of females and males in each class differed, the overall distribution was balanced (43 females and 45 males). The L2 German proficiency of the participants was fairly low. Students who enroll in the first semester of this program have no or almost no knowledge of German. Some of the students progress through all four semesters in the program while others join it at later time points via a placement test. To obtain a more general measure of the participants' L2 proficiency, a standardized German proficiency test was administered at the end of the experimental semester.

All fourth-semester classes took an official diagnostic test which was administered by the onDaF Institute in Bochum, Germany ([www.ondaf.de](http://www.ondaf.de)) and proctored at the researcher's institution. Participants take this online cloze test over 40 minutes and are then automatically placed within CEFR bands (Eckes & Grotjahn, 2006). The results showed that approximately two thirds of all fourth-semester students reached the A2 level by the end of the semester, one third did not reach it, and only two participants reached the B1 level. Since the onDaF test only measures proficiency at or above the A2 level, it was not considered meaningful to test third-semester students. Overall, although proficiency is not considered a variable in this study, it is safe to state that both third- and fourth-semester students in the focal instructional program were generally at or below the A2 CEFR level, i.e. had roughly low-intermediate L2 proficiency.

Importantly, third- and fourth-semester participants differed not only in their course level but also in their entry-level knowledge of the target items. The third-semester students were explicitly taught them for the first time, whereas the fourth-semester students were reviewing material taught in the previous semester. Although previous incidental exposure to the target items by the third-semester students cannot be excluded, the bulk of the engagement of this student population with their L2 is limited to the classroom. As far as the fourth-semester groups are concerned, the overwhelming majority of the students progressed to their fourth-semester course immediately after the third-semester course, which means that they were first exposed to the target items at the same time and in the same fashion during the previous semester. As the intervention was administered during a spring semester, only a relatively short (one-month-long) winter break separated the semesters, which does not allow much time for knowledge attrition or additional out-of-class exposure to the L2 (e.g., traveling to German-speaking countries, which none of the participants undertook). Therefore, most of the residual knowledge of the target items in the fourth-semester groups can be confidently attributed to the instructional effects from the previous semester. The only exception is three participants who enrolled in the course via a placement test, and who therefore did not follow the same instructional sequence as other participants. However, they showed an entry-level knowledge of the target items similar to other participants and thus were included in the study.

### **3.3 Target items**

The intervention focused on 11 verb-preposition collocations (Appendix 1) from the regular textbook used in the first three semesters in this program (Di Donato, Clyde & Vansant, 2012: 359). In ten of these collocations, prepositions are not congruent in German and English: For example, *to wait for* is equivalent to *warten auf*, although the prototypical translation of the preposition *for* is *für*. Furthermore, the German noun phrase or the pronoun following the preposition carries an obligatory gender, case, and number marker (e.g., *Ich warte auf **meinen** Bruder / **ihn*** [masculine, accusative, singular] – *I am waiting for **my** brother / **him***). However, if the pronoun refers to an inanimate object, German uses pronominal adverb-preposition contractions, the so-called *da*-compounds, instead (e.g., *Ich warte **darauf*** – *I am waiting **for that***), in which no inflectional markers are present. Furthermore, German pronominal adverbs (e.g., *davon*, *dabei*) are extremely frequent, not genre-restricted, and can combine with most prepositions, unlike their very infrequent and genre- and item-restricted English counterparts (e.g., *thereof*, *thereby*). German *da*-compounds are considered an important part of active vocabulary for learners and are typically taught in conjunction with prepositional verbs along with prepositional phrases. Needless to say that, being an important instructional focus due to the

sheer frequency of verb-preposition collocations in language usage, such lexically, grammatically, and semantically complex constructions present considerable difficulties to learners, although typical syllabi and textbooks allocate only one to two lessons to introducing them.

### ***3.4 Procedures***

There were two iterations of the experiment. In the first iteration, two sections of the third-semester course and three sections of the fourth-semester course participated. The experimental sections were assigned to four conditions based on treatment (D = DDL; T = textbook) and course level (L = low, i.e. third semester; M = mid, i.e. fourth semester). Only the data of the students who agreed to participate and took both the pretest and the posttest were included in the study. The DL (n=15) and TL (n=13) sections were taught by the same instructor. Two fourth-semester sections taught by a different instructor were assigned the TM condition due to the low number of students (combined n=16). Finally, the third fourth-semester section, taught by yet another instructor, was assigned the DM condition (n=13). It must be noted that the DM section followed a syllabus different from the TM sections as it constituted a different course (introduction to German for the professions). However, since the intervention was administered on the tenth day of the semester in fourth-semester classes, no noticeable difference in instruction had yet taken place. Also, since the regular instructor of the DM section was out of town on the day of the experiment, the researcher administered the intervention. Since she was known to the students as program coordinator, and the regular instructor was still new to the students, the intervention did not cause any considerable disruption to the regular instructional flow. After the first iteration of the experiment, the researcher decided to replicate the study with another cohort of third-semester students with slight design changes (see explanations under *Results* below). Two sections taught by an instructor who did not participate in the first iteration of the study were assigned to the DL2 (n=16) and TL2 (n=15) conditions.

The intervention in the third-semester classes was administered at the time designated for the target items in the course syllabus (after about one third of the semester was over), and in the fourth-semester classes as part of the start-of-the-semester review. Prior to the intervention, the researcher met with all participating instructors individually for about 30 minutes, discussed the interventions, gave them packets with detailed lesson plans, test sheets, and worksheets, and answered their questions. In the T-groups, the procedure followed the regular syllabus (barring the pretest and the posttest), so the instructors did not have to change anything in their teaching. Regarding the D-groups, instructors showed much interest in trying out the new teaching method. The ten-minute-long pretest was administered at the end of the class immediately preceding the intervention class. The homework assignment in all groups was unrelated to the target items. The intervention, administered during the next class, lasted for 40 minutes and the class concluded with a ten-minute-long posttest. Both tests were announced to the students as ungraded quizzes. Additionally, all participants filled out a brief electronic personal and language learning background questionnaire during one of their regular computer lab meetings.

The test instruments were paper worksheets with 14 sentences from the DWDS corpus: a large, freely and publicly available corpus of contemporary German ([www.dwds.de](http://www.dwds.de)). Ten sentences contained the target verbs and four sentences contained control items (see Boulton, 2010) – verb-preposition collocations not taught in the intervention. Only ten target constructions were tested because two out of the 11 collocations contained the same verb with different



prepositions that indicated different meanings. To remove this additional difficulty, only one collocation of these two was tested. The verb was followed by a preposition and a noun phrase with a definite article in half of the sentences and by a *da*-compound in the other half. Only singular nouns were used and the gender of each noun was indicated in parentheses so that participants had to only think about case while working on the grammatical items. The target prepositions, articles, and *da*-compounds were blanked out and learners had to fill the gaps. A model with a non-target prepositional verb introduced each part of the test. The verbs that were used with prepositional phrases in the pretest were used with *da*-compounds in the posttest and vice versa. During scoring, one point was given for each correctly supplied lexical item (preposition or *da*-compound), and one point for each correctly supplied grammatical item (article) after correct prepositions.

The difference between the 40-minute-long treatments was the following. The TL and TM groups followed the sequence of activities in the textbook (Di Donato et al., 2012: 359-362). This lesson followed the typical *deductive* ‘triple P’ (*Presentation-Practice-Production*) model (e.g., DeKeyser, 1998). The target verb-preposition-case collocations were first presented as a list with English translations. A practice drill followed, and then *da*-compounds were introduced in contrast with preposition-pronoun collocations with a few examples. Next, students practiced using *da*-compounds in individual, pair-work, and whole-class drills. The sequence concluded with an oral pair-work exercise in which students exchanged questions and answers about everyday topics using the target constructions. In contrast, the DL and DM groups worked with DDL worksheets containing 5 to 7 concordance lines for each collocation (Appendix 2; see also Vyatkina, 2015) copied and pasted from the DWDS corpus. The intervention followed the inductive ‘triple I’ (*Illustration-Interaction-Induction*) model proposed by Carter and McCarthy (1995) and extended by Flowerdew (2009) to the *guided induction* model with an additional ‘*Intervention*’ step between Interaction and Induction. The instructor in each D-class briefly introduced the idea of a corpus as a rich repository of language usage examples and distributed the worksheets. Then, she instructed the students to find patterns in the concordance lines and modeled that with one focal verb. The students then worked in pairs discussing the verb-preposition-case patterns and wrote individual examples for each verb. Next, the instructor discussed the results with the whole class, making sure that everybody arrived at the right pattern. A similar procedure was followed with concordance lines for verb-preposition-pronoun collocations and *da*-compounds. The lesson concluded with a Q&A oral pair-work exchange similar to that in the T-groups. Therefore, the lessons in all groups were designed as in Boulton’s (2010: 547) study: “The intention was thus that all students should come away with essentially the same final information; the main differences lay in the way it had been reached and the materials from which it was derived.”

## **4 Results**

### ***4.1 Iteration 1***

First, the pretest-posttest scores for the control items that were not part of the intervention were analyzed. 88% of all participants scored zero points on these items (out of the possible 6) on both tests and the maximum score was 1 for the lexical knowledge and 1 for the grammatical knowledge. Furthermore, only four participants showed a gain of 1 or 2 points. These very low frequencies show that the test effect was minimal in this study. Therefore, the scores for the

control items were not included in the statistical analysis and the following report only refers to the experimental items.

Participants could earn up to 10 points for the lexical knowledge and up to 5 points for the grammatical knowledge on each test. The descriptive statistics for the pretest and posttest scores are presented in Table 1 and illustrated in Fig. 1 and Fig. 2. The overall scores on the pretest were very low, with even the mid-level students getting, on average, only about a third of both the lexical and grammatical items right.

Table 1 *Iteration 1: Pretest and posttest scores by treatment, level, and outcome (lexical and grammatical)*

treat- ment	level	n	pretest				posttest			
			lexical (max. 10)		grammatical (max. 5)		lexical (max. 10)		grammatical (max. 5)	
			mean	SD	mean	SD	mean	SD	mean	SD
D	L	15	1.07	1.03	0.80	0.77	3.80	2.04	1.33	1.23
T	L	13	0.15	0.38	0.08	0.28	1.23	1.54	0.69	1.03
D	M	13	3.85	3.21	1.15	1.46	5.38	3.15	2.54	1.61
T	M	16	3.50	1.55	1.50	0.89	6.13	2.87	2.19	1.52

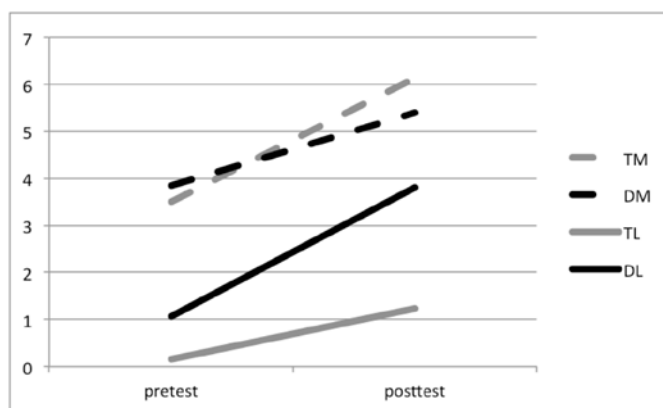


Fig. 1. Iteration 1: Pretest and posttest lexical scores by treatment and level

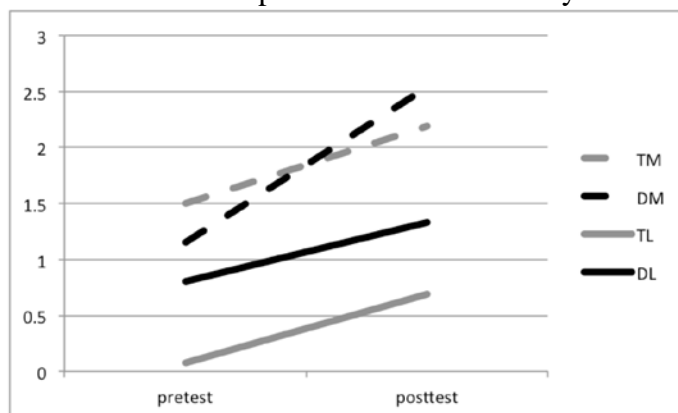


Fig. 2. Iteration 1: Pretest and posttest grammatical scores by treatment and level

The data was analyzed with multilevel modeling methods (Cunnings, 2012). Raw test scores were not normally distributed, so zero-inflated Poisson regression models with associated  $z$ -tests were used. Gain scores were more normally distributed, so a multilevel linear regression with associated  $t$ -tests was used. Although all variables were fed into the models together, the results will be presented separately for lexical and grammatical gains for clarity.

*Lexical knowledge.* Regarding the pretest scores, several differences across groups were revealed. Mid-level students scored significantly (ca. 6 times) higher than low-level students on the lexical outcome (95% CI [3.55, 10.34],  $z = 6.61$ ,  $p < .0001$ ). Whereas this result was expected, a difference between low-level groups was unexpected: the DL group scored higher than the TL group by 0.9 points (95% CI [0.39, 1.51],  $z = -2.54$ ,  $p = .01$ ). On the other hand, the mid-level groups were not different from one another ( $z = -0.29$ ,  $p = .77$ ). Second, all groups improved on the posttest, and the overall posttest scores were (on average) ca. 2 times higher than pretest scores (95% CI [1.57, 2.42],  $z = 6.14$ ,  $p < .0001$ ). Next, lexical gains were compared across treatments controlling for course level. It turned out that among low-level students, the D-method resulted in significantly greater lexical gains than the T-method by 1.7 points on average (95% CI [0.28, 3.03],  $t(104) = -2.39$ ,  $p = .02$ ). Among mid-level students, the D-method resulted in somewhat lower lexical gains than the T-method (by 1 point on average), but this difference was not significant (95% CI = [-0.27, 2.44],  $t(104) = 1.59$ ,  $p = .11$ ).

*Grammatical knowledge.* Pretest comparison yielded results very similar to the lexical knowledge results. Mid-level students scored significantly (3 times) higher than low-level students (95% CI [1.57, 5.91],  $z = 3.28$ ,  $p = .001$ ) and the DL group had significantly (0.7 points) higher pretest scores than the TL group (95% CI [0.31, 1.15],  $z = -2.34$ ,  $p = .02$ ), whereas the mid-level groups were not different from one another ( $z = 0.80$ ,  $p = .43$ ). Another parallel result was that the overall posttest scores were 1.9 times higher than pretest scores (95% CI [1.34, 2.63],  $z = 3.63$ ,  $p = .0002$ ). However, in contrast to the lexical gains, there was no difference in grammatical gains across treatments. Although the D-groups had higher gains than the T-groups, this difference was not significant either between mid-level groups (95% CI [-0.66, 2.05],  $t(104) = -1.20$ ,  $p = .23$ ) or low-level groups (95% CI [-1.46, 1.29],  $t(104) = 0.12$ ,  $p = .91$ ).

In summary, all groups improved both types of knowledge following instruction. There is no evidence to support the effect of teaching method on grammatical items or lexical items in the mid-level classes, but the DDL method led to significantly more improvement of lexical scores in the low-level classes. Since the participant number was low and because the low-level groups turned out to be different already on the pretest, it was decided to replicate the study with another cohort of low-level students.

## 4.2 Iteration 2

First, only low-level classes were included, so the course level variable was eliminated. Second, since control items in the first iteration of the study had already shown test effects to be negligible, only experimental items were included in the testing materials. Third, the data for several participants were eliminated prior to statistical analysis for the following reasons. First, the data for three participants who scored 5 or more points (out of a possible total of 15) on the pretest were eliminated since the study focused on participants with low entry-level knowledge of the target items, and because none of the low-level participants in the first iteration had scored more than 4 points on the pretest. Furthermore, the data for four participants were eliminated since they simply copied and pasted the preposition and the article from the model into all gaps

in the posttest. On the other hand, upon consultation with a statistician, it was decided to keep the data for the participants who did not participate either in the pretest or in the posttest because the statistical methods used in this study allow for missing data points (Cunnings, 2012). This resulted in the DL2 (n=16) and the TL2 (n=15) design. The descriptive statistics are presented in Table 2 and illustrated in Fig. 3 and Fig. 4.

Table 2 *Iteration 2: Pretest and posttest scores by treatment and outcome (lexical and grammatical)*

treat- ment	level	n	pretest				posttest			
			lexical (max. 10)		grammatical (max. 5)		lexical (max. 10)		grammatical (max. 5)	
			mean	SD	mean	SD	mean	SD	mean	SD
D	L	16	0.62	0.77	0.23	0.60	2.79	1.48	1.64	1.28
T	L	15	0.64	0.93	0.29	0.47	1.21	1.12	0.71	0.99

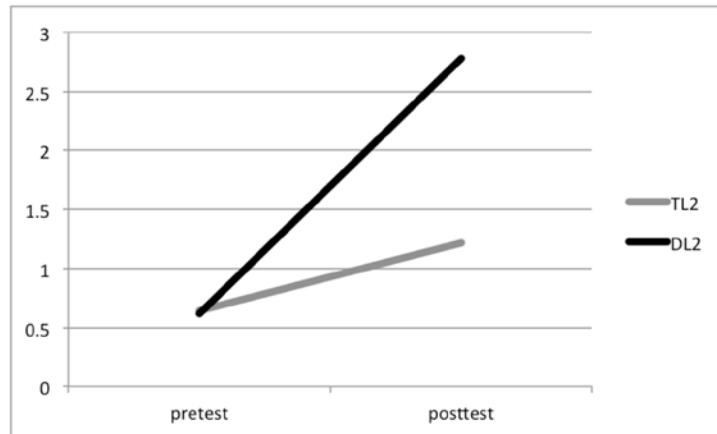


Fig. 3. Iteration 2: Pretest and posttest lexical scores by treatment

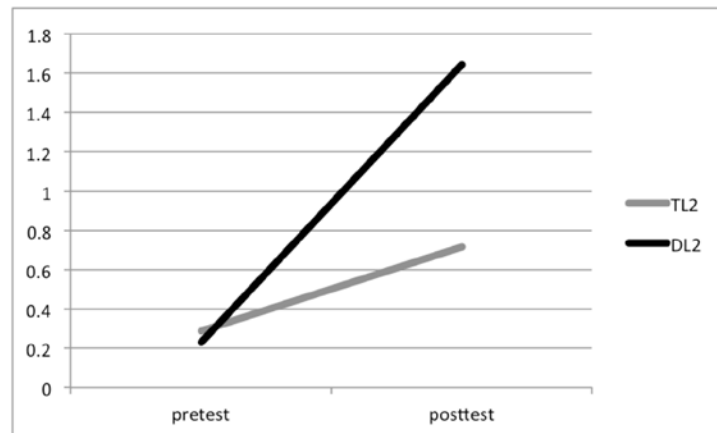


Fig. 4. Iteration 2: Pretest and posttest grammatical scores by treatment

The multilevel modeling results showed that in this cohort, the groups were much more uniform on the pretest: although the DL2 group scored somewhat lower than the TL2 group on both the lexical and grammatical items, no significant difference was discovered for either (95%

CI [-0.57, 0.64],  $z = 0.09$ ,  $p = .93$ , and 95% CI [-0.45, 0.37],  $z = 0.28$ ,  $p = .78$ , respectively). Next, there was an overall improvement on the posttest in comparison with the pretest. Regarding the lexical outcome, overall posttest scores were (on average) 3.2 times higher than pretest scores (95% CI [1.87, 5.63],  $z = 4.13$ ,  $p < .0001$ ). Regarding the grammatical outcome, overall posttest scores were (on average) 4.6 times higher than pretest scores (95% CI [2.16, 11.45],  $z = 3.66$ ,  $p = .0002$ ). Finally, the DL2 group improved significantly more than the TL2 group on both the lexical outcome, by 1.7 points (95% CI [0.82, 2.65],  $t(42) = 3.81$ ,  $p = .0004$ ), and the grammatical outcome, by 1.2 points (95% CI [0.26, 2.09],  $t(42) = 2.58$ ,  $p = .01$ ). This result stands in contrast to the result from study iteration 1, where the DL group had a significantly higher gain than the TL group only on the lexical outcome. A possible explanation of this difference is that the DL1 group scored higher on the pretest than DL2 and thus cannot be compared directly to the 'true' beginner DL2 group.

In summary, the results of study iteration 2 confirm the result for the low-level groups from iteration 1 by showing a significant DDL treatment effect for lexical learning, and extend it to grammatical learning. The result from the latter iteration is also more reliable because there were no significant differences between the groups on the pretest. Moreover, although the DDL group scored slightly lower on the pretest on both outcomes, it scored significantly higher than the non-DDL group on the posttest on both outcomes. Finally, the distribution of gains was much more even for the DDL group: all 11 participants who took both the pretest and the posttest improved, with ten participants showing lexical gains and nine participants showing grammatical gains. In contrast, only five participants in the non-DDL group out of 13 who took both the pretest and the posttest improved, all of them showing both lexical and grammatical gains, whereas five other participants showed no gains and two participants scored lower on the posttest.

## ***5. Discussion, implications, and conclusion***

This study explored the effectiveness of brief DDL interventions for teaching collocations to students at low L2 proficiency levels. On the theoretical level, the results lend support to the benefit of usage-based approaches (Ellis, 2014) and the guided induction DDL approach (Flowerdew, 2009, 2015; Smart, 2014) in teaching lexico-grammar. The results demonstrate that such interventions are not less effective and, on some parameters, more effective than a traditional deductive teaching method. More specifically, this guided induction method worked better than a deductive method during initial exposure of low-level learners to a new lexico-grammatical construction. Arguably, the combination of guided induction and input enrichment and enhancement by the teacher has led to enhanced learners' perception and understanding (Sharwood Smith, 2013) and, therefore, to a higher level of awareness (Schmidt, 1990), as evidenced in their improved performance.

This study also confirms the finding from a previous study (Vyatkina, in press) that DDL can be extrapolated beyond English as an L2 and is effective for teaching lexico-grammatical collocations in inflectional languages. The novel finding of this study is that it shows that paper-based DDL works with students at very low proficiency levels (at and below CEFR A2) rarely considered in research before and that, moreover, it is *especially* effective with students at the lowest proficiency level. This result is in line with Yoon and Hirvela's (2004) study of learner DDL perceptions, who found intermediate learners to be more receptive to DDL than advanced learners. More specifically, this study showed that DDL was significantly more effective for teaching new collocations (thus contradicting Cobb, 1999, and Nesselhauf, 2004), whereas both

the DDL and the non-DDL method were equally effective for improving the knowledge of previously learned items. Since the pretest scores of mid-level students were higher than the posttest scores of low-level students, we can infer that all students gradually improve their knowledge of the target collocations even with traditional instruction methods but this learning is extremely slow, as shown by still very low pretest scores of mid-level students. However, the low-level DDL groups showed a much higher initial learning which may have given them a better jumpstart. This study has only explored the short-term development of explicit knowledge (measured by a controlled production test), therefore no claims can be made about long-term development and acquisition of implicit knowledge (DeKeyser, 2003; Doughty, 2003). However, if later supported by recycling of the target items (with either a DDL or non-DDL method), this can potentially lead to better long-term learning.<sup>1</sup> Especially in regard to difficult German inflectional morphology, this short explicit instructional intervention played the role of “enhancing later implicit acquisition by increasing chances of noticing” (DeKeyser, 2003: 332).

On the practical level, this study supports an argument in favor of integrating brief paper-based DDL interventions into non-DDL syllabi (thus corroborating Boulton, 2008, 2009, 2010). Similar to Boulton (2010), this study also shows that regular teachers were able to successfully implement these lessons following brief oral and written instruction. Furthermore, all instructors in this study commented positively on their experience with this new teaching method. This finding shows that, if more ready-made DDL materials were available, it could be more widely implemented in mainstream language teaching. However, as Boulton (2010: 560) notes, “DDL materials are extremely time-consuming to prepare” and “published materials are virtually nonexistent”, which is especially true for languages other than English. Therefore, we hope that positive results from more DDL studies like this one will inspire publishers to produce more DDL materials and more researchers to share them with open access (see, e.g., Vyatkina, 2015).

This study has a number of limitations that should be addressed in future research. The first limitation, typical of all classroom research, is that the results need to be interpreted with caution due to the low number of participants. More partial replications with other participants, proficiency levels, and target items are needed to increase the generalizability of the results. Second, there is a dire need for studies of long-term DDL effects (Boulton & Pérez-Paredes, 2014; Leńko-Szymańska & Boulton, 2015), although these are notoriously difficult to design. In particular, it is worth exploring the feasibility of a wider integration of various paper-based and computer-based DDL tasks into L2 syllabi across the curriculum and the effects of “more substantial training or repeated use of such materials” (Boulton, 2010: 559). Third, more studies that single out specific DDL effects (e.g., guided induction vs. deduction, input richness, input enhancement) are needed. Finally, future studies should investigate the DDL effectiveness for various grammatical targets, especially those in inflected languages. One such direction to expand the present study is an item-based analysis comparing the DDL effects for collocations of different levels of grammatical complexity in morphologically rich languages (e.g., verbs and prepositions governing fixed vs. variable cases).

To conclude, this study shows that guided induction learning based on the analysis of authentic language use patterns is conducive to learning not only at advanced but also at incipient proficiency levels. Hopefully, these findings will inspire greater adoption of DDL by language teachers and more studies on DDL by language researchers.

---

<sup>1</sup> Although a delayed posttest was not formally part of this study, an indication of this trend can be seen in the fact that the DL2 group retained more target *da*-compounds (scoring on average 4.33 points out of 5) than the TL2 group (3.31 points) on a chapter test administered two weeks after the intervention.

## Acknowledgements

This study was supported in part by the Department of Germanic Languages and Literatures of the University of Kansas. I would also like to acknowledge Terrence D. Jorgensen Jr. for his invaluable help with the statistical analysis.

## References

- Baten, K. (2011) Processability theory and German case acquisition. *Language Learning*, **61**(2): 455–505.
- Bernardini, S. (2002) Exploring new directions for discovery learning. In: Kettemann, B. and Marko, G. (eds.), *Teaching and learning by doing corpus analysis*. Amsterdam: Rodopi, 165–182.
- Boulton, A. (2008) Looking for empirical evidence of data-driven learning at lower levels. In: Lewandowska-Tomaszczyk, B. (ed.), *Corpus linguistics, computer tools, and applications: State of the art*. Frankfurt: Peter Lang, 581–598.
- Boulton, A. (2009) Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, **21**(1): 37–51.
- Boulton, A. (2010) Data-driven learning: Taking the computer out of the equation. *Language Learning*, **60**(3): 534–572.
- Boulton, A. (2012) Hands-on / hands-off: Alternative approaches to data-driven learning. In: Thomas, J. and Boulton, A. (eds.), *Input, process and product: Developments in teaching and language corpora*. Brno: Masaryk University Press, 152–168.
- Boulton, A. and Pérez-Paredes, P. (2014) Editorial: Researching new uses of corpora for language teaching and learning. *ReCALL*, **26**(2): 121–127.
- Carter, R. and McCarthy, M. (1995) Grammar and the spoken language. *Applied Linguistics*, **16**(2): 141–158.
- Chambers, A. (2007) Popularising corpus consultation by language learners and teachers. In: Hidalgo, E., Quereda, L. and Santana, J. (eds.), *Corpora in the foreign language classroom*. Amsterdam: Rodopi, 3–16.
- Chambers, A. and O’Sullivan, Í. (2004) Corpus consultation and advanced learners’ writing skills in French. *ReCALL*, **16**(1): 158–172.
- Chan, P.-T. and Liou, H.-C. (2005) Effects of web-based concordancing instruction on EFL students’ learning of verb–noun collocations. *Computer Assisted Language Learning*, **18**(3): 231–251.
- Chun, D. (2012) Review article: Replication studies in CALL research. *CALICO Journal*, **29**(4): 591–600.
- Cobb, T. (1999) Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning*, **12**(4): 345–360.
- Cobb, T. and Boulton, A. (2015) Classroom applications of corpus analysis. In: Biber, D. and Reppen, R. (eds.), *The Cambridge handbook of English corpus linguistics*. Cambridge: Cambridge University Press, 478–497.
- Council of Europe. (2001) *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg: Language Policy Unit. [http://www.coe.int/t/dg4/linguistic/source/framework\\_en.pdf](http://www.coe.int/t/dg4/linguistic/source/framework_en.pdf)

- Cunnings, I. (2012) An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, **28**(3): 369–382.
- DeKeyser, R. M. (1998) Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In: Doughty, C. J. and Williams, J. (eds.), *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press, 42–63.
- DeKeyser, R. M. (2003) Implicit and explicit learning. In: Doughty, C. J. and Long, M. H. (eds.), *The handbook of second language acquisition*. Oxford: Blackwell, 313–348.
- Di Donato, R., Clyde, M. and Vansant, J. (2012) *Deutsch, na klar!: An introductory German course* (6th edition). Boston: McGraw-Hill.
- Doughty, C. J. (2003) Instructed SLA: Constraints, compensation, and enhancement. In: Doughty, C. J. and Long, M. H. (eds.), *The handbook of second language acquisition*. Oxford: Blackwell, 256–310.
- Eckes, T. and Grotjahn, R. (2006) A closer look at the construct validity of C-tests. *Language Testing*, **23**(3): 290–325.
- Ellis, N. (2005) At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, **27**(2): 305–352.
- Ellis, N. (2014) Cognitive AND social language usage. *Studies in Second Language Acquisition*, **36**(3): 397–402.
- Farr, F. (2008) Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language Awareness*, **17**(1): 25–43.
- Flowerdew, L. (2009) Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics*, **14**(3): 393–417.
- Flowerdew, L. (2015) Data-driven learning and language learning theories: Whither the twain shall meet. In A. Leńko-Szymańska and A. Boulton (eds.), *Multiple affordances of language corpora for data-driven learning*. Amsterdam: John Benjamins, 15–36.
- Frankenberg-Garcia, A. (2014) The use of corpus examples for language comprehension and production. *ReCALL*, **26**(2): 128–146.
- Gabrielatos, C. (2005) Corpora and language teaching: Just a fling, or wedding bells? *TESL-EJ*, **8**(4): 1–37.
- Gilquin, G. and Granger, S. (2010) How can data-driven learning be used in language teaching? In: O’Keeffe, A. and McCarthy, M. (eds), *The Routledge handbook of corpus linguistics*. London: Routledge, 359–370.
- Herron, C. and Tomasello, M. (1992) Acquiring grammatical structures by guided induction. *The French Review*, **65**(5): 708–718.
- Huang, L.-S. (2008) Using guided, corpus-aided discovery to generate active learning. *English Teaching Forum*, **46**(4): 20–27.
- Johns, T. (1990) From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, **10**: 14–34.
- Kennedy, C. and Miceli, T. (2001) An evaluation of intermediate students’ approaches to corpus investigation. *Language Learning and Technology*, **5**(3): 77–90.
- Koosha, M. and Jafarpour, A. (2006) Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL Journal Quarterly*, **8**(4): 192–209.
- Leńko-Szymańska, A. and Boulton, A. (2015) Introduction: Data-driven learning in language pedagogy. In: Leńko-Szymańska, A. and Boulton, A. (eds.), *Multiple affordances of language*



- corpora for data-driven learning*. Amsterdam: John Benjamins, 1–14.
- Mauranen, A. (2004) Speech corpora in the classroom. In: Aston, G., Bernardini, S. and Stewart, D. (eds.), *Corpora and language learners*. Amsterdam: John Benjamins, 195–211.
- Mukherjee, J. (2006) Corpus linguistics and language pedagogy: The state of the art – and beyond. In: Braun, S., Kohn, K. and Mukherjee, J. (eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods*. Frankfurt am Main: Peter Lang, 5–24.
- Nesselhauf, N. (2004) Learner corpora and their potential for language teaching. In: Sinclair, J. M. (ed.), *How to use corpora in language teaching*. Amsterdam: John Benjamins, 125–152.
- Pallotti, G. (2015) A simple view of linguistic complexity. *Second Language Research*, **31**(1): 117–134.
- Porte, G. (ed.) (2012) *Replication in applied linguistics: A practical guide*. Cambridge: Cambridge University Press.
- Schmidt, R. (1990) The role of consciousness in second language learning. *Applied Linguistics*, **11**(2): 129–158.
- Sharwood Smith, M. (1993) Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, **15**(2): 165–179.
- Sharwood Smith, M. (2013) Possibilities and limitations of enhancing language input: A MOGUL perspective. In Benati, A., Laval, C. & Arche, M. J. (eds.), *The grammar dimension in instructed second language learning*. London: Bloomsbury, 36–57.
- Smart, J. (2014) The role of guided induction in paper-based data-driven learning. *ReCALL*, **26**(2): 184–201.
- Spada, N. and Tomita, Y. (2010) Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, **60**(2): 263–308.
- Tian, S. (2005) The impact of learning tasks and learner proficiency on the effectiveness of data-driven learning. *Journal of Pan-Pacific Association of Applied Linguistics*, **9**(2): 263–275.
- Trahey, M. and White, L. (1993) Positive evidence and preemption in the second language classroom. *Studies in Second Language Acquisition*, **15**(2): 181–204.
- Vinagre, M. and Muñoz, B. (2011) Computer-mediated corrective feedback and language accuracy in telecollaborative exchanges. *Language Learning and Technology*, **15**(1): 72–103.
- 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. (2015). *Corpus Activity: German Verb-Preposition Collocations*. Center for Applied Second Language Studies (CASLS), University of Oregon.  
<http://caslsintercom.uoregon.edu/content/18927>
- Vyatkina, N. (in press). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology*, **20**(3).
- Yoon, H. (2008) More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning and Technology*, **12**(2): 31–48.
- Yoon, H. and Hirvela, A. (2004) ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, **13**(4): 257–283.
- Yoon, H. and Jo, J. (2014) Direct and indirect access to corpora: An exploratory case study comparing students' error correction and learning strategy use in L2 writing. *Language Learning and Technology*, **18**(1): 96–117.

## Appendix 1: Verb-preposition-case collocations used in the study

Angst haben vor (+ dat.)	<i>to be afraid of</i>
sich ärgern über (+acc.)	<i>to be annoyed about</i>
sich beschäftigen mit (+ dat.)	<i>to occupy oneself with</i>
sich bewerben um (+acc.)	<i>to apply for</i>
denken an (+acc.)	<i>to think of</i>
sich freuen auf (+acc.)	<i>to look forward to</i>
sich freuen über (+acc.)	<i>to be happy about</i>
sich interessieren für (+acc.)	<i>to be interested in</i>
verzichten auf (+acc.)	<i>to do without</i>
sich vorbereiten auf (+acc.)	<i>to prepare for</i>
warten auf (+acc.)	<i>to wait for</i>

## Appendix 2: DDL teaching materials

Many German verbs require the use of fixed prepositions; these verb-preposition combinations are usually different from their English equivalents. For example: *to wait for* -> warten auf.

I. Look at the following examples of such verb-preposition combinations. They come from a German **corpus** (a large electronic collection of German texts) – *Das Digitale Wörterbuch der deutschen Sprache*, or DWDS ([www.dwds.de](http://www.dwds.de)). The examples are organized in **concordances** – stacked lines of examples from different texts with the search words in the middle.

In each line, underline the noun phrase immediately following the preposition: i.e., the noun with the article or adjective that modifies it. Then, write out the verb + preposition + case combination. **NB:** you do **not** need to translate the sentences. Finally, write an example sentence.

Example: to wait for

13 1999 GE	...se), und er ist in der Rolle des Schlägers (des batters) und <u>wartet auf den Ball</u> . Da wirft ihm der Werfer (pitcher) den Ball n...
14 1999 ZE	...Publikum fietscht die Zähne, es hat Bock auf bösen Humor und <u>wartet auf Bedienung</u> . Man weiß das in den Sendern und sucht na...
15 1999 ZE	...m? Die vielen Antworten, die unsere Sammlungen enthalten, <u>warten auf die richtigen Fragen</u> . Die besondere und wahrhaft in...
16 1999 BE	...er wird in eine dunkelblaue Kontrolleursuniform gesteckt und <u>wartet auf seine Rente</u> . Wenn er einen Waggon entert, ruft er, ...
17 1999 BE	...er Stadtteil Berlins. Ich stehe an der Bushaltestelle und <u>warte auf den 180er</u> nach Lichterfelde Ost, 20 Minuten lang, denn...
18 1998 BE	... von unten einen grellen Pfiff. Mein Vater, sagte sie, er <u>wartet auf seine Zigaretten</u> . Und hielt die R 6-Packung in die ...
19 1998 BE	...en, Alfred hatte schon einen Hof in Höhenreute gepachtet und <u>wartete auf sie</u> . Johann hatte auf der Continental schon das Zeu...

Verb-preposition-case combination: warten + auf + acc. (accusative)

Write an example sentence about yourself: Ich warte auf die Frühlingsferien.

### Tasks:

#### 1. to be interested in

21 1989 ZE	...spielsweise den Flüchtling Mehmet Naim Sakir,* aber niemand <u>interessiert sich für</u> uns. Keiner kennt uns, keiner gibt uns Arbei...
22 1985 ZE	...aus Meinersdorf, Michael (15) aus Magdeburg und viele andere <u>interessieren sich für</u> den Beruf = Sänger = bzw. = Sängerin = und möch...
23 1984 WI	...als Differenz von Anzeichen und Ausdruck. Wir <u>interessieren uns für</u> den Unterschied dieser begrifflichen Disposition...
24 1984 ZE	...n Teilen Deutschlands hat gemeinsame Ideen und Ziele. Sie <u>interessieren sich für</u> das Leben und die Kultur ihrer Generation im an...
25 1981 WI	...dingungen der Rationalität der Wissenschaft, aufzuklären; er <u>interessiert sich für</u> das, was Max Weber die » Kulturbedeutung « der ...
26 1981 WI	...tündigung, sie bezieht sich auf kommunikatives Handeln: Mead <u>interessiert sich für</u> sprachliche und sprachförmige Symbole nur insow...

Verb-preposition-case combination: \_\_\_\_\_ + \_\_\_\_\_ +

Write your example sentence:

Fig. 5. Corpus exercise I (tasks 2-10 not included)

II. Analyze the concordances with the prepositional verb *hoffen* (to hope) below. What is the difference between concordance sets 1 and 2?

1)

1 2006	Ob das Revolutionspaar dieses kleine Wunder erreichen kann, bleibt fraglich. Schon während der orangefarbenen Revolution haben viele Menschen in Kiew Straßen gesagt: Wir <b>hoffen auf sie</b> , denn wir haben keine besseren.
2 2005	Da sah ich dein Gesicht, Elisabeth, an der Wand, die wohl durchsichtig war. Ich <b>hoffte auf dich</b> , inbrünstig, aber plötzlich warst du weg, und ehe ich etwas anderes spüren konnte, geschah mir eine große Erleichterung, so sanft, dass ich glaubte, jemand würde mich vergiften. Und warum ich trank, war, weil es jemand zu sein schien: ein Haltepunkt, eine winzige Vergewisserung meines Ichs.
3 1999	Er wollte zu den Bayern, er verdient eine Menge Geld, und er ist jung, aber alt genug, um zu wissen, was für ein Spiel Fußball ist. Er wird es schwer haben, aber wir <b>hoffen auf ihn</b> . Wir hoffen, mehr noch als ohnehin, dass die Bayern in der 80. Minute mit zwei zu null vorne liegen.
4 1984	Hubert Schwarz indessen wurde beim Sprung von einer Windbö erfasst, die seinen Flug bremste, seine Ski zu Boden drückte, und mußte bei 72 Metern notlanden - 20 Meter weniger als zuvor, fern jeder Medaillenchance. Die Außenseiter <b>hoffen auf ihn</b> , die Favoriten fürchten ihn: den Zufall. Schnee zum Beispiel, der alle Bedingungen ändert, in Sarajevo auch den Zeitplan.

2)

1 2008	Aber das Unwahrscheinliche kann eintreten. Nicht nur eine Nation <b>hofft darauf</b> . Die Brasilianer, die Italiener, die Holländer, die Deutschen und sogar die Schotten "rechnen sich Chancen aus" obwohl eben gerade "Rechnen" allein glücklicherweise nicht weiterhilft.
2 2007	Ob ihm diese Haltung auch beim neuen Job hilft? Nicht nur das Weiße Haus und die Weltbank-Mitarbeiter <b>hoffen darauf</b> . Auch Europas Regierungen wünschen sich, dass der Mann wieder Ruhe in den Laden bringe schon aus Eigennutz.
3 2006	Ich kann Ihnen keine Hoffnung machen. - Ich <b>hoffe darauf</b> . Ich habe keinen Job.
4 2003	Glauben Sie an die Fähigkeit der Amerikaner, die Demokratie im Land zu installieren? Viele Iraker im Exil <b>hoffen darauf</b> . Elias Khoury: Der Krieg wird länger dauern, als es jetzt scheint, und seine zerstörerische Kraft erst nach und nach zeigen.
5 2002	Ob man das Speichermedium jemals wird kaufen können, ist bei IBM noch nicht entschieden. Binnig <b>hofft darauf</b> . Ein kommerzielles Produkt sei eine Art von Experiment: "Ich träume mir irgendwas zusammen.

The word *darauf* from concordance set 2 is called a **pronominal adverb**, or simply a **da-compound**. Formulate a rule: when is the combination "preposition + personal pronoun" used, and when is it replaced by a *da-compound*?

III. Now look at other *da*-compounds below.

2 2009	...t uns nie. Und niemand auf der Welt <b>interessiert sich dafür</b> . Millionen armer Palästinenser werden nie aus ihrem Elen...
7 2005	...twickeln wir eine neue Sendung, und ich <b>beschäftige mich damit</b> . Aber eines Tages werde ich mich bestimmt mal hinsetzen ...
5 2009	...sogar in Grün. Die kühnsten Fremdenverkehrsleute <b>denken daran</b> , die Saison zu verlängern, die bis jetzt mit dem August end...
2 2009	...Madam Wang Qi, eine resolute Frau Anfang 40, <b>ärgert sich darüber</b> . Vom Staatsanwalt erwartet sie, dass er seiner Rolle gem...
5 1965	...nächsten Olympischen Spiele. Nach München <b>bewarb sich darum</b> auch die spanische Hauptstadt Madrid.
1 2008	...t - das habe ich noch nie gemacht -, aber ich <b>habe davor Angst</b> und ich halte es auch für möglich. Heinlein: Ist SARS übe...
10 1906 BE	...Nächste Woche kommt er wieder. Ich <b>freue mich darauf</b> . Ich kann ja nicht sagen, daß ich mit d...
9 2008	...as tut die Welt? Sie sieht die Zeichen und <b>freut sich darüber</b> . Und verändert sich viel schneller, als es den Architekt...
5 2009	...et für immer auf die Unabhängigkeit. Aber er <b>verzichtet darauf</b> nicht. Wir reden mit ihm, um ihm die Gelegenheit zu geben...
2 2009	... kündigte sich an, beide Länder <b>bereiteten sich darauf vor</b> . Im Juli hielten sie Manöver ab, Georgien mit 1000 Solda...

What prepositions combine with *dar-?*

What prepositions combine with *da-?*

What is the rule?

Are there English equivalents of *da*-compounds?

IV. Work in partner groups. Take turns and ask each other questions based on your example sentences from I.

Example:

A. *Ich warte auf die Frühlingsferien. Und du?*

B. *Ich warte auch darauf. Or: Ich warte auf den Sommer.*

Fig. 6. Corpus exercises II-IV