Language Learning & Technology
ISSN 1094-3501

October 2017, Volume 21, Issue 3
pp. 195–216

ARTICLE

Effects of corpus-based instruction on phraseology in learner English

Katherine Ackerley, *University of Padova*

Abstract

This study analyses the effects of data-driven learning (DDL) on the phraseology used by 223 English students at an Italian university. The students studied the genre of opinion survey reports through paper-based and hands-on exploration of a reference corpus. They then wrote their own report and a learner corpus of these texts was compiled. A contrastive interlanguage analysis approach (Granger, 2002) was adopted to compare the phraseology of key items in the learner corpus with that found in the reference corpus. Comparison is also made with a learner corpus of reports produced by a previous cohort of students who had not used the reference corpus. Students who had done DDL tasks used a wider range of genre-appropriate phraseology and produced a lower number of stock phrases than those who had not. The study also finds evidence that students use more phrases encountered in paper-based concordancing tasks than in hands-on tasks. Unlike in previous DDL studies, observations of the learning of a specific text-type through DDL in the present study are based on the comparison with both a control learner corpus and an expert corpus. The study also considers the use of DDL with a large class size.

Keywords: *Data-driven Learning, Learner Corpora, Corpus Linguistics, Language Teaching Methodology*

Language(s) Learned in this Study: *English*

APA Citation: Ackerley, K. (2017). Effects of corpus-based instruction on phraseology in learner English. *Language Learning & Technology*, *21*(3), 195–216. Retrieved from http://llt.msu.edu/issues/october2017/ackerley.pdf

Introduction

The uses and benefits of corpora for language learning are multiple and widely reported in the literature (e.g., studies in Granger, Hung, & Petch-Tyson, 2002; Leńko-Szymańska & Boulton, 2015; O'Keeffe, McCarthy, & Carter, 2007). This paper is grounded in two areas of research within corpus linguistics: the role of learner corpora in second language acquisition research (Granger, 2002, 2015), and how data-driven materials can enhance language learning (Boulton, 2009a, 2012a). Granger and Meunier (2008) recommend that what is taught should make sense to learners and be useful and adaptable to their interests and level (p. 250). It is essential then, that attention be paid to how the results of corpus analysis may be made pedagogically relevant and effective. Perhaps the most extensively studied genre in learner corpus research is the argumentative essay,[1] but a call has been made for a wider range of genres to be dealt with (Gilquin & Paquot, 2008). The text type under investigation in this study is the public opinion survey report, which may be considered appropriate to university students' needs as it prepares them for writing in a formal register and reporting opinions objectively—both essential skills in academic writing. As the text type tends to be unfamiliar to students, much of its very specific phraseology is likely to be encountered only in corpus-based activities such as those focused on in this study. This study is based on three corpora of opinion survey reports: one written by professionals and the others by two different cohorts of first-year university students of English as a foreign language. One group of 223 students wrote reports following language awareness tasks based on data from expert writers (the expert corpus). These tasks shall henceforth be referred to as data-driven learning (DDL) tasks or activities. The other group of students (the control group) wrote reports but did not do any corpus-based activities. The study aims to investigate the extent to which

students' language is enriched following DDL activities. It looks, therefore, at the appropriateness and range of phraseology used by both sets of students, that is by those who were not exposed to corpus-based data and by those who did the DDL tasks. Though this exploratory study does not make a direct comparison between the use of paper-based and hands-on approach to DDL, it does attempt to investigate the use of both approaches with large classes of language learners.

Following a brief overview of genre-based language teaching, phraseology, and DDL, the paper presents the three corpora used in this study and the kinds of learning tasks developed. It goes on to identify some of the lexical choices and patterns that characterise the genre of opinion survey reports, investigating the differences between the language produced in both expert and learner texts. It then compares the two learner corpora to conclude whether or not DDL tasks may lead to a heightened awareness of the lexical and phraseological features of the genre.

Literature Review

Phraseology and a Genre-Based Approach to Teaching

The relevance of genre analysis in language learning is well established (e.g., Swales, 1990), and corpus-based approaches to identifying the typical linguistic features of a genre are by now common (Bondi, 2001; Flowerdew, 2001; Tribble, 2002). The analysis of a genre-based corpus can reveal recurring lexico-grammatical patterns, some of which can then be highlighted for students who are required to produce the particular text type, allowing them to adhere to the conventions of a discourse community. The combined analysis of an expert reference corpus and a learner corpus for the identification of features of a text type can be of significant pedagogical value for English for special purposes (ESP) and academic writing courses. Various studies (e.g., Flowerdew, 2001) support the comparison of such corpora for materials design and a number of studies have compared the lexico-grammatical features of text types in native-speaker and learner texts (Biber & Reppen, 1998; Gledhill, 1998; Meunier, 2002). Ackerley (2008) adopted a genre-based corpus analysis, comparing expert and learner corpora to inform the development of DDL materials in a university English course. However, fewer studies have taken these comparative studies further by creating DDL materials and then investigating the effects of these on learning.

In the context of this study, the term *phraseology* encompasses what have been referred to by various scholars as formulaic sequences (Schmitt, 2004), lexical bundles (Biber & Barbieri, 2006; Biber, Conrad, & Cortes, 2004), lexical chunks (Schmitt, 2000), and collocations—that is, the "occurrence of two or more words within a short space of each other in a text" (Sinclair, 1991, p. 170). Some such sequences may be learned together as single "big words" (Ellis, 1996, p. 111), while others may have *slots* or be composed of *collocational frameworks* (Renouf & Sinclair, 1991). According to Granger and Paquot (2008), phrases are made up of at least two words. For this article, a phrase refers to any "string of words whose status is not determined," such that a grammatical analysis of the words in a phrase is irrelevant (Sinclair, 2008, pp. 407–408), and a frequency-based view of collocation (Nesselhauf, 2005) is applied.

Several scholars have focused on phraseology in language learning, particularly how it concerns argumentative essay writing and academic writing (e.g., Allen, 2009; Paquot, 2008, 2013). For example, Allen (2009) notes that, in academic writing, learners are rarely able to use bundles competently and in a native-like way. Even if native-likeness is not a course objective, understanding writing conventions may well be. Hyland (2008) discusses how the use of lexical bundles may indicate "naturalness" in "competent participation in a given community" (p. 5), which might include a community of professional writers. On the other hand, he continues, a lack of such clusters may indicate "the lack of fluency of a novice or newcomer to that community" (p. 5). What is at stake for non-expert writers is revealing that they are not aware of the "specific norms, expectations, and conventions of a discourse community" (Bhatia, 2002, p. 37).

Inappropriate phraseology is one of many reasons why learner language may differ from the linguistic norms of a given genre. Stubbs (2002, p. 215) points out that language learners typically consider single

words as the traditional units of language. Students therefore tend to piece these units together, making direct translations from their first language (L1), but possibly failing to achieve the intended communicative purpose. A phrase-focused approach to teaching and learning may lead to more fluent, native-like, or expert production. Indeed, various aspects of phraseology may be considered in a description of native-like versus non-native-like production or novice versus expert writing.[2] O'Keeffe et al. (2007) argue that language chunks are of interest as they can be "register- (or genre-) specific" (p. 210). Moreover, Granger and Meunier (2008) stress how teachers should make students aware of the pervasiveness of phraseology—a field which, as Warren (2011) reports, is neglected in language teaching.

Data-Driven Learning

DDL has been defined by Boulton (2012a) as "any use of language corpora by second or foreign language learners" (p. 263). The term, first coined by Johns (1990), refers to learning from information obtained from corpora, with the students acting as researchers to identify recurring patterns of language in concordances. It may also involve learners observing the frequency of items in a corpus, and differences between learner and native-speaker or expert-writer data. According to Boulton (2009b), DDL puts learners "at the centre of the process, taking an increased responsibility for their own learning rather than being taught rules in a more passive mode" (p. 2). Hunston (2002) posits that in this way students remember "what they have worked to find out" (p. 170). Such active learning is believed to result in more effective learning and is a tenet of autonomous language learning (Benson, 2001). Flowerdew (2015) discusses how it fits with language learning theories such as "the noticing hypothesis, constructivist learning, and Vygotskian sociocultural theories" (p. 16).

With DDL, students can either access corpus data indirectly (i.e., by examining concordances prepared by a teacher or materials developer) or directly (i.e., by using computer software to analyse corpora for themselves). Indirect and direct access are two approaches referred to by Boulton (2012b) as *hands-on* and *hands-off* use, respectively, and can be considered as extremes on a continuum where various levels of guidance can be provided. At one end, there are highly controlled conditions, with the teacher using corpora to identify language features to focus on in class and then providing carefully selected concordance lines on paper with questions guiding the learners towards predicted conclusions. At the other end, more experienced students can access corpora independently to suit their own needs, with serendipitous learning taking place as the desired result (Bernardini, 2000). An example of learners engaged in hands-on corpus work, acting "as language detectives or researchers investigating authentic examples of the target language on their own," is provided by Geluso and Yamaguchi (2014, p. 227). Their A2–B2 level students of English used the Corpus of Contemporary American English (Davies, 2008) to investigate formulaic sequences for use in a speaking project. Yoon (2011) provides an overview of the benefits of such direct access, or *learner concordancing*, on second language writing.

DDL, however, is not without its drawbacks. Boulton (2012b) highlights issues which may prevent the benefits of hands-on DDL, including "struggling with the interface and query syntax, conducting inappropriate searches, [and] misinterpreting data" as potential off-putting difficulties (pp. 153–154). What is more, as advocates of DDL admit, skills for using concordancing software and formulating appropriate queries "need time and effort to develop" (Leńko-Szymańska & Boulton, 2015, p. 4). A further issue may depend on class size. As Boulton (2012a) notes, the average number of student participants in DDL studies for ESP is 45, though this figure may be boosted by the study by Hafner and Candlin (2007) that includes 300 participants. Several studies that reveal the success of DDL are based on small classes (e.g., Vyatkina, 2016; Yoon, 2008).

Indirect access to corpus data, such as paper-based activities where the learners are provided with edited concordances, is also a valid form of DDL, the success of which has been noted in numerous studies (e.g., Boulton, 2010; Huang, 2014; Smart, 2014). An advantage of the hands-off approach is that learners can explore corpus data without the barriers posed by using technology (Boulton, 2010). With paper-based DDL, the above-reported problems of interface and knowing how to formulate queries can be avoided by the provision of worksheets with concordances that have been edited by the teacher for reasons of space

and comprehensible content.

The studies mentioned above make use of native-speaker or expert-writer corpora for DDL activities. However, as Seidlhofer (2002) notes, a learner corpus can be used to provide *learning-driven data*. This can highlight the language typically produced by learners, which can then be compared with an expert-writer model. Highlighting features of their own or their peers' language production can make some students' writing or speaking problems seem obvious and can give them impetus to avoid them in the future, while DDL materials based on a reference corpus can provide them with something concrete (i.e., an expert-writer model) to aim for.

Though Johns (1990) states that DDL gives learners direct access to data and is the "attempt to cut out the middleman as far as possible" (p. 18), this article reports on a study where the middleman (i.e., the teacher) maintains an important role in guiding learners in their use of corpora and their intended discovery learning. It investigates how, using DDL tasks, the language teacher can help students become more independent researchers and learners, developing their ability to recognise language patterns and note how words collocate so that they can then make their own informed choices about their language production.

Methodology

Context and Corpora

The study is based on three corpora: one expert and two learner corpora. The learner corpora both consist of texts produced by a large class of first-year students enrolled in the Linguistic and Cultural Mediation program at the University of Padova, Italy. The one-semester English language module, *An introduction to academic language skills*, focused on how lexico-grammatical features of different registers vary according to communicative purpose (Halliday, 1989). Prime objectives of the module were developing students' awareness not only of register variables, but also of the existence of disciplinary preferences and of how it is necessary for writers to follow the constraints of specialist genres (Ackerley, 2008). A large number of students enroll in the course each year (over 300), though not all attend lessons regularly. For the purpose of this study, only the texts produced by the 223 students who attended classroom and lab lessons regularly were selected for analysis. Although the students were expected to display B1+ level writing skills, according to the results of an in-house pre-course test their language competences ranged from pre-intermediate to upper-intermediate (A2 to B2 of the Common European Framework of Reference; Council of Europe, 2001).

The text type that received major focus in this module was that of public opinion survey reports. Though this was neither an ESP nor an academic writing course, the text type was selected because of certain similarities with academic writing (a future objective for the students), notably because of its formality and the objective reporting of findings. The students were expected to report on their classmates' opinions, as expressed in online class forums on topics selected by the students themselves. Dealing with topics that were well grounded in the students' personal experience allowed them to focus on the linguistic and structural features of the genre, rather than on any potentially demanding new academic content. Hyland (2002) reports that making reference to, building on, and reworking past utterances are necessary skills in academic writing and ones with which students often require assistance (pp. 129–130). A further important aspect of the task involved recognising the informal language produced in the forums and being able to synthesise it, re-elaborating it in more formal English, and using the phraseology that was suitable to the target text type.

Before beginning DDL tasks on the expert corpus, the students attended an introductory lesson on corpora, which aimed to raise their awareness of how words typically occur together as phrases, rather than existing as individual items that can be directly translated from the learner's L1. In this lesson, the students were introduced to the concept of corpora with a focus on how the keywords of a learner corpus of self-presentations compared with those from a corpus of native-speaker student self-presentations (Ackerley, 2015).[3] The aim was to introduce them to basic concepts in corpus linguistics, using a corpus of texts on a

familiar topic. This lesson was followed by two 90-minute lab lessons in a 90-seat computer lab, during which the students were trained to use AntConc (Anthony, 2011) to access the native-speaker self-presentation corpus. The lab lessons were attended by up to 90 students at a time (each lesson was repeated to allow full attendance). These were followed by three further 90-minute lab lessons in which they carried out a range of hands-off and hands-on tasks based on the opinion survey report corpus. The reason for including printed concordances in these three sessions, even though the students were already familiar with both AntConc and the technique of reading concordances on the computer, was to help the students deal with an unfamiliar text type, more demanding language, and a larger corpus than the one dealt with previously. The edited paper-based concordances, where lines were selected and sequenced, meant that the students were not overwhelmed by a high number of hits for their initial tasks on the particular corpus. After the hands-off tasks, the students worked on the corpus directly, guided by questions on worksheets that required them to search for words that would not produce an excessively high number of hits and where most answers were fairly immediate so as to keep both attention and motivation high.

As stated above, one expert and two learner corpora of opinion survey reports were used in this study. The corpora differ in terms of both word count and number of texts (see Table 1). The expert corpus was composed of 51 texts ranging from 685 to 5,661 words, while the two learner corpora had texts with an average of 222 and 204 words (the students had been instructed to produce reports of between 160 and 220 words).

Table 1. *Size of the Three Corpora*

	Expert Corpus	**Control Learner Corpus**	**DDL Learner Corpus**
Number of Words	58,000	53,350	45,400
Number of Texts	51	240	223

The 58,000-word expert corpus, considered here as an *exemplar corpus* as it served as a model for the students' language production (Tribble, 2002), was composed of 51 public opinion survey reports. These were retrieved from market research websites as well as from British and American news websites (for further details, see Ackerley, 2008).

The first learner corpus, referred to henceforth as the *control corpus* as the students did not have access to any corpus-informed learning materials on the target text type, was a 53,350-word collection of texts produced by 240 students during a previous academic year. For their end-of-module exam, they were required to produce a text that followed the linguistic conventions of a public opinion survey report. The students took the exam in a computer lab where they had access to online learner dictionaries. They received instruction on the kind of language to produce in their texts through exercises based on complete reports and parts of reports selected by the teacher (see Appendix A).

The second learner corpus, referred to henceforth as the *DDL corpus* as the students engaged in both hands-off and hands-on DDL tasks based on the expert corpus before producing their own reports, was smaller, at 45,400 words, and was composed of texts by 223 students. The students first compared word frequency lists extracted from the expert corpus and the control corpus (see Table 2). They considered which words commonly occurred in the expert corpus but not frequently in the learner corpus and, vice versa, which words tended to be over-represented in the control corpus. The teacher prepared worksheets for both hands-off and hands-on corpus exploration based on words selected from the frequency lists. As with the earlier cohort, the texts were written under exam conditions in a computer lab, but in addition to online learner dictionaries, students had access to AntConc and the expert corpus. The texts were written four weeks after the final DDL-based lesson.

Identification of Keywords

A stop list was used to eliminate function words from the frequency lists, while topic-specific words (such

as those related to the death penalty, abortion, immigration) were manually removed to avoid any bias towards topics in the phraseology. Because the aim of opinion survey reports is to report on and compare the views of a selected group of people, words that play roles in the representation of argumentative procedures (such as *favour*, *support*, *agree*) and the projection of ideas and meanings (such as *opinion* and *view*) were selected for analysis from these lists. The words with asterisks in Table 2 were chosen for DDL tasks, but only the words in Table 3 were focused on in this study.

The study initially considered the frequency of words in all three corpora. However, though this comparison of lists from expert and learner corpora could help understand whether the students were using "an appropriate variety of vocabulary in their written work" (Nation, 2001, p. 32), it did not allow us to see how the learners were actually using the language. Therefore, even if frequency of use was similar, as was the case with the word *majority*, this "[did] not necessarily imply any similarity in lexico-grammatical patterns" (Bondi, 2001, p. 144). Tribble (2002) argues how exploration of a concordance allows a more complete investigation of the patterns that contribute to the special identity of a text. To complete the study, then, concordances of the words focused on in class were first analysed to see how the control group's use of words compared with that of the expert writers and then to see how the DDL group's use of words compared with both the control group's use and with that of the expert writers in terms of frequency and phraseology.

Table 2. *Frequency of Top 20 Words in Three Corpora Normalised per 1,000 Words*

Rank	Expert Corpus		Control Corpus		DDL Corpus	
1	7.05	say	13.80	people	11.28	people
2	5.22	people	10.72	students	9.34	students
3	3.93	public	6.02	problem	8.66	university
4	3.59	support*	5.12	opinion	6.23	survey
5	3.14	issue*	5.10	think	5.97	majority
6	2.64	survey*	3.62	university	5.66	find
7	2.62	view*	3.52	survey	5.40	opinion
8	2.40	government	3.47	against	5.37	think
9	2.38	think	3.00	government	4.05	say
10	2.38	favo(u)r*	2.83	agree	3.15	different
11	2.03	opinion*	2.51	young	2.71	surveyed
12	1.94	respondent	2.44	different	2.53	prefer
13	1.91	poll*	2.38	say	2.47	young
14	1.86	believe	2.34	majority	2.44	like
15	1.84	majority*	2.08	hand	2.42	believe
16	1.69	research	2.06	fact	2.27	interview*
17	1.50	result	1.97	right	2.09	commission*
18	1.38	age	1.93	public	2.00	hand
19	1.34	percent	1.71	moreover	1.94	carry
20	1.29	compared	1.67	like	1.83	problem

Task Types

Four task types are relevant to this study, but, for reasons of space, only a general description of each will be given. The first concerns the observation of language in complete reports or parts of reports. Not being

corpus-based, in the present context, this task type is not considered to promote DDL. Both cohorts of students were given texts and asked to identify words or phrases that were of particular relevance to the genre studied (see example in Appendix A). They also focused on the structure of the text type.

The second type of task is based on the frequency lists of the expert and control corpus. The students observed notable differences in the lists between their peers' production and that of the professional writers. They were also asked to find alternative words in the expert corpus frequency list.

The third and fourth types of tasks were hands-off and hands-on concordance-based activities, respectively. The hands-off tasks consisted of carefully edited concordances (Appendix B). In the hands-on tasks, the students explored the expert corpus for themselves using AntConc (Appendix C). In both cases, the students were provided with worksheets designed to guide them through their queries and subsequent searches for noteworthy linguistic information within the results. In these tasks, the students were asked to consider both the collocation and colligation[4] of words selected from the frequency list.

Results

This section first presents observations on the range and frequency of vocabulary in the three corpora. Because of the considerable differences between the expert corpus and the two learner corpora (both in terms of text length and communicative purpose of the reports), the study did not aim to make statistical comparisons between the language produced by expert writers and learners. It did, however, look at the range of language used and tendencies, investigating how corpus-based focus on the lexis and phraseology produced by expert writers influenced students' writing. A comparison was then made between the phraseology in the expert corpus with that in the control corpus, focusing on those words that were of interest in the creation of the DDL materials (Table 3). To allow investigation of tendencies—that is, whether use of a word increases or decreases following DDL—the normalised frequency of the words analysed was given. A comparison was then made between the written production of the students who used the DDL materials, the texts produced by their peers (control corpus), and the texts of the expert writers.

Observations Based on Lexical Frequency Lists

One notable difference was the ranking and frequency of the verb *think* in the expert corpus compared with the control corpus: ninth and fifth place, respectively, with a rate of just 2.38 occurrences per 1,000 words (pkw) in the expert corpus and 5.10 pkw in the control corpus. The DDL group of students were encouraged to identify other words from the list that could be used as an alternative to *think*, with *view*, *opinion*, and *believe* being selected as possibilities. Though over-representation of the reporting verb *think* was pointed out to the DDL group of students and though they were made aware of alternatives, unexpectedly its use increased slightly in the DDL corpus (5.37 pkw).

Opinion was used frequently by both experts and learners, but considerably more often by the learners (5.12 pkw in the control corpus as opposed to 2.03 pkw in the expert corpus). Despite observation of over-representation in the control corpus frequency list and DDL exercises on the alternative word *view*, the frequency of *opinion* rose in the DDL corpus (up from 5.12 pkw to 5.40 pkw). One could argue, however, that this increase was to be expected, given the focus on the word's phraseology in the concordance-based activities.

The third word on the control corpus list was *problem* (6.02 pkw), a word frequently used by the students to introduce a topic but which did not appear in the top 20 words used by the professional writers (0.72 pkw). The students noted that experts favour the alternative *issue*, which is less overtly negative. Following observations in differences in the frequency lists and a concordance exercise on the word *issue*, use of the word *problem* dropped considerably in the reports written by the DDL group of learners (down to 1.83 pkw), yet their use of *issue* remained strikingly low (0.66 pkw) and failed to make an appearance in the students' top 20 words.[5]

It was noted that *majority* had a similar frequency and ranking in both the expert and control corpora. A

task on the collocation of *majority* (Appendix B) was devised for the second cohort of students and its use increased dramatically from 2.34 pkw in the control corpus to 5.97 pkw in the DDL corpus. Further attention to its collocates is given below.

While *agree* was overused by the learners in the control corpus (2.83 pkw as opposed to 0.86 pkw in the expert corpus and absent from the top 20 words), its frequency fell to 0.95 pkw in the DDL corpus. Alternatives to *agree* were sought in the expert frequency list and *support* and *favour*[6] were identified. To provide the students with alternatives for *agree*, hands-on DDL exercises were created based on the words *favor* and *support* (Appendix C), as these ranked high in the expert frequency list. The phraseology of these alternatives, along with those of other words dealt with in the concordance-based tasks (e.g., *opinion*, *view*, and *majority*), will be discussed in more detail below.

Table 3. *Normalised (pkw) Frequency of Words Selected for DDL Activities*

	Expert Corpus	Control Corpus	DDL Corpus
say	7.05	2.38	4.05
think	2.38	5.10	5.37
opinion	2.03	5.12	5.40
view	2.62	0.71	1.08
issue	3.14	0.94	0.66
problem	0.72	6.02	1.83
agree	0.86	2.83	0.95
support	3.59	1.21	0.95
favo(u)r	2.38	1.61	0.51
majority	1.84	2.34	5.97

Comparison of Phraseology: Expert and Control Corpora

As mentioned above, the fact that two groups of writers use a word does not mean that they use it in the same way. *Opinion* is a case in point. In the control corpus, the most frequently occurring cluster is *in his/their opinion* (27 times, 0.51 pkw), always used to project the opinion of a group of people. On the other hand, in the expert writer corpus the cluster *in their opinion* only occurs twice and with a different function—that is, as part of a phrase indicating difference of opinion:

- 5 of the 18 countries (i.e., Australia, United States, Canada, France, and Cameroon) appear divided in their opinion…
- People in Cameroon appear more split in their opinion compared to the other three countries…

Differences in the phraseology of *view* are also of note. Table 3 shows how the word was used 2.62 pkw in the expert corpus, but only 0.71 pkw in the control corpus (a total of 38 occurrences). Further analysis of the control group's use of the word showed that the cluster *point(s) of view* appeared 30 times, with 10 cases of *different points of view*. On the other hand, in the expert corpus there was only 1 instance of *point of view* in 152 occurrences of the word *view*.

What is interesting in the control corpus is that some learners displayed an expert-like use of *view* and *opinion*. There were 13 instances (0.24 pkw) of *share the (same) opinion* to express agreement between groups of respondents, and 4 instances (0.07 pkw) of *hold the (same) view/opinion*. Indeed, a look at the expert corpus revealed that *share* and *hold* both collocated with *view*/*opinion* 12 times (0.22 pkw). This relatively abundant use of expert-like collocations in the control group's writing can be traced back to an exercise done in class, where students were encouraged to identify phrases in a complete report to show that respondents agreed with an issue or with each other: *hold the same view* was identified in this single

text, while *share the same opinion* was added to a list of alternative expressions given to the students. This was an example, then, of students producing appropriate phraseology previously identified in a non-corpus-based task.

Another example of how the students were influenced by the language in this non-corpus-based task can be seen in their use of *majority*. In the exercise mentioned above, students in the control group added the phrase *overwhelming majority* to their list of expressions (meaning *many or most people*), and the completed list (Appendix A) was then sent to the whole class. This phrase occurred 29 times in the control corpus. However, it only occurred twice in the expert corpus for a normalised frequency of just 0.03 pkw in the expert corpus as opposed to 0.54 pkw in the control corpus. Though there was nothing wrong with the students all using the same phrase, the fact that only two other adjectives were used by the students to pre-modify *majority* (*vast*, occurring three times at 0.06 pkw, and *great*, occurring twice at 0.04 pkw) indicated a general lack of awareness of alternatives. Because of these observations, concordance-based tasks were developed to expose the students to a wider range of collocates and to broaden their knowledge of genre-appropriate phraseology (for an example of a hands-off DDL task on *majority*, see Appendix B).

Comparison of Phraseology: Expert, Control, and DDL Corpora

As stated above, after carrying out DDL activities based on the expert corpus, the second cohort of students wrote their own reports as part of their end-of-course exam. A comparison of aspects of phraseology identified in the three corpora and dealt with in the DDL tasks is presented below.

The control group of learners used *share * opinion* 18 times or 0.34 pkw (it is actually present only once in the expert corpus), and though its presence remains high in the DDL corpus (16 occurrences, 0.35 pkw), the DDL students also produced a range of alternatives. For example, an analysis of the clusters produced by AntConc reveals 32 instances (0.70 pkw) of *hold the opinion* and 22 instances (0.48 pkw) of *(to be) of the opinion*—both phrases identified by students in the paper-based DDL task (see Appendix B). This is a marked increase in use when compared to the control corpus (1 and 4 occurrences, or 0.02 pkw and 0.07 pkw, respectively). Also significant was the disappearance of the stock phrases *point(s) of view* (down to 6 occurrences, or 0.13 pkw, from 0.56 pkw in the control corpus) and *in * opinion* (just 8 occurrences, or 0.18 pkw, in the DDL corpus as opposed to 0.51 pkw in the control corpus).

The use of *express * opinion* was also noteworthy. In the expert corpus it was only used once and did not occur in the concordance-based tasks. However, the phrase occurred 13 times (0.24 pkw) in the control corpus—possibly because of positive L1 transfer (*esprimere un'opinione* translates directly to *express an opinion*). The occurrence of *express * opinion* dropped slightly, to 9 instances, or 0.19 pkw, in the DDL corpus.

A wide range of genre-appropriate phrases showing disagreement could be found in the DDL corpus (see Table 4). The phraseology of *opinion* to express disagreement in the control corpus, on the other hand, was far less varied: *different opinion* was found twice, and *dissenting opinion*, once.

Though *view* was used 19 times (0.33 pkw) as a verb in the expert corpus and its colligation was focused on in the hands-on exercise, it only occurred 4 times (0.09 pkw) as a verb in the DDL corpus.[7] When occurring as a noun, it was used in much the same way as *opinion*, with 15 instances (0.33 pkw) of *hold the (same) view* and 4 (0.09 pkw) of *share the view*. This was an increase in results compared to the control corpus, where *hold the (same) view* occurred 4 times (0.07 pkw) and *share the view* was not present.

Table 4. *Phrases Expressing Disagreement in the DDL Corpus*

Phrase	Frequency

(deeply) divided opinion	0.04	(2)
(dramatic/slight) differences in/of opinion	0.31	(14)
(severe) division in/of opinion	0.04	(2)
appear divided in their opinion	0.04	(2)
opinion is (evenly) divided	0.04	(2)
opinion is split	0.02	(1)

Note. Frequency is normalised (pkw); absolute frequency is in parentheses.

As for the collocation of adjectives with *majority*, Table 5 shows the variety and frequency of use in the three corpora. It can be seen that while just three different adjectives were found in the control corpus, 14 different pre-modifiers were found in the DDL corpus. In particular, there was an increase in the use of *great*, *large*, *overwhelming*, *solid*, and *vast*. *Vast* and *large* were the most frequent collocates of *majority* in the expert corpus and, indeed, the second- and third-most popular with the students who did the DDL exercise. However, *overwhelming*, one of the less frequent collocates in the expert corpus (2 occurrences, 0.03 pkw) underwent an increase from 29 occurrences (0.54 pkw) in the control corpus to 57 (1.30 pkw) in the DDL corpus—that is, it more than doubled in popularity.

Table 5. *Normalised (pkw) Pre-Modification of Majority*

	Expert Corpus		**Control Corpus**		**DDL Corpus**	
no modifier	0.90	(54)	1.69	(90)	1.40	(64)
broad	0.03	(2)	0.00	(0)	0.02	(1)
clear	0.07	(4)	0.00	(0)	0.20	(9)
great	0.03	(2)	0.04	(2)	0.60	(25)
large	0.20	(10)	0.00	(0)	0.70	(30)
narrow	0.03	(2)	0.00	(0)	0.10	(4)
two-to-one	0.03	(2)	0.00	(0)	0.02	(1)
overwhelming	0.03	(2)	0.54	(29)	1.30	(57)
slight	0.02	(1)	0.00	(0)	0.20	(7)
slim	0.03	(2)	0.00	(0)	0.10	(4)
small	0.05	(3)	0.00	(0)	0.04	(2)
solid	0.05	(3)	0.00	(0)	0.30	(12)
substantial	0.02	(1)	0.00	(0)	0.10	(5)
vast	0.20	(10)	0.06	(3)	0.90	(39)
wide	0.02	(1)	0.00	(0)	0.00	(0)
widespread	0.00	(0)	0.00	(0)	0.02	(1)

Note. Frequency is normalised (pkw); absolute frequency is in parentheses.

Further hands-on tasks were devised requiring students to investigate the phraseology of *support* and *favour*. The participants were first asked to find three pre-modifying adverbs for the verb *favour*. The students then observed how the gerund is used after the verb *favour*. There were no occurrences of any of these pre-modifying adverbs, and just one example of *favour* + *gerund* in the DDL corpus. Students were also expected to identify the expression *in favour of*, and 13 instances (0.29 pkw) were found in the DDL corpus. Interestingly enough, it occurred 70 times (1.31 pkw) in the control corpus—indicating, on the one hand, that the students were already familiar with this lexical bundle and, on the other, that the students who did the corpus-based activities had possibly acquired alternative options which, for reasons of space,

cannot be dealt with here.

In their hands-on investigation of *support*, the students were expected to identify the verb *express* as a collocate. In the expert corpus, this collocate occurred 16 times (0.03 pkw), but not at all in the control corpus. There was just 1 instance of *express * support* in the DDL corpus and, despite a question focusing on adjectives that collocate with *support*, there were only 2 occurrences of *support* used as a noun in the DDL corpus and no instances of pre-modifying adjectives. Despite focus on *support* as a noun, the students used it more frequently as a verb, leaving little evidence of any effects of hands-on DDL tasks on their written production.

Table 6. *Summary of Observed Effects of DDL on Students' Written Production*

Word	**Task Type**	**Results**
opinion	paper-based	Reduction of stock phrases that were not appropriate to genre: *in * opinion* down to 0.18 pkw in the DDL corpus from 0.51 pkw in the control corpus Increase in frequency and range of genre-appropriate phraseology: *hold the opinion* up to 0.70 pkw from 0.02 pkw; *(to be) of the opinion* up to 0.48 pkw from 0.07 pkw Increase in frequency of single word, despite low frequency in the expert corpus: up to 5.40 pkw from 5.12 pkw (2.03 pkw in the expert corpus)
view	hands-on	Reduction of stock phrases that were not appropriate to genre: *point(s) of view* down to 0.13 pkw from 0.56 pkw Slight increase in frequency and range of genre-appropriate phraseology: *hold the (same) view* up to 0.33 pkw from 0.07 pkw; *share the view* up to 0.09 pkw from 0.0 pkw (*hold the view* also occurred in the single-text task) No increase in use of *view* as verb
majority	paper-based	Considerable increase in frequency of genre-appropriate phraseology: 75.5% of instances of *majority* had genre-appropriate pre-modifiers, up from 27.4% in the control corpus Considerable increase in range of genre-appropriate phraseology: *majority* has 14 different pre-modifiers, up from three in the control corpus Over-representation of the phrase *overwhelming majority*: *overwhelming majority* occurred 1.30 pkw in the DDL corpus, and 0.03 pkw in the expert corpus
favour	hands-on	Decrease in frequency of word: down to 0.51 pkw from 1.61 pkw No increase in frequency of genre-appropriate phraseology No increase in range of genre-appropriate phraseology
support	hands-on	Decrease in frequency of word: down to 0.95 pkw from 1.21 pkw No increase in frequency of genre-appropriate phraseology No increase in range of genre-appropriate phraseology

Table 6 summarises the results for each word examined and illustrates the kinds of exercises used for each one. The comments in the results column are based on observations of students' phraseology following the concordance-based exercises. It would appear that the most noteworthy positive changes were for *opinion* and *majority*. There were only slight changes in the genre-appropriate phraseology of *view*, and searches for *favour* and *support* in the DDL corpus produced disappointing results. It would appear that the phrases dealt with in the hands-off exercises were those that the students chose to focus on in their exam.

Discussion

The results of the present study seem to indicate that the DDL group of students learnt to make more genre-appropriate use of some of the items in their concordance-based tasks, notably the words *opinion* and *majority*. That is, they displayed a wider range of suitable collocations and a higher usage of typical phrases used to project opinions and present preferences. However, not all items had the same levels of success. Although the aim of this study was not to make a direct comparison between hands-off and hands-on approaches to DDL, it would appear that paper-based concordance tasks led to a higher use of items studied than the hands-on tasks. It would also appear that factors influencing students' use of phraseology included a phrase's occurrence in a language-awareness exercise based on a single text (i.e., a non-corpus-based exercise), as was the case with the high frequency of *hold the (same) view*. Both groups of students observed this phrase in a report studied in class and it is likely that this—in combination with reinforcement found in the concordance-based task—led to its high frequency in the DDL corpus. As for the items encountered in the hands-on tasks (see Table 6), students seem to have paid little attention to the phraseology of *support* and *favour*, so there was less evidence that hands-on corpus use led to the adoption of phrases by students in their own writing. This could have a number of explanations, including the fact that their presence on the computer screen was fleeting. Though students may notice a pattern and be intrigued by what they observe, if they do not save their results or take detailed notes, then these phrases and any contextual information that should also be learnt may be lost. Concordances on a worksheet "provide something tangible" (Boulton 2010, p. 560)—that is, they may be underlined, looked at again, added to with a pen, and used for revision. As there is evidence in other studies (see Boulton & Cobb, 2017) that hands-on DDL is more effective, this study indicates that attention needs to be paid to how the learners store their discoveries when engaged in hands-on DDL so that they can be accessed again. Explicit instruction about note-taking may prove beneficial to students working with a concordance (for an example of how this may be promoted, see Geluso & Yamaguchi, 2014, p. 231).

A further issue highlighted by Boulton (2009a) is that learners may have difficulty dealing with the authentic language and truncated lines in a concordance. Problems could also be posed by the number of lines and the amount of language students have to deal with in a directly-accessed concordance. That is, it could be that students struggled to find the answers within the time limits of the lesson. Students need training in managing the time they spend dealing with lengthy concordances and should be encouraged to work independently on tasks at home (see also Kennedy & Miceli, 2010). Student training is of fundamental importance in the corpus-based coursework of Kennedy and Miceli (2010) and is seen as an apprenticeship, with the development of skills being actively supported in subsequent courses. This would be desirable in a context where students are at the beginning of their university language studies and where they would benefit from the reinforcement and development of the skills acquired in their first year.

Vyatkina's more structured study (2016) of the effects of paper-based and hands-on DDL on the learning of collocations finds that both hands-on and hands-off approaches are equally effective. However, among the differences in the study are the kinds of tasks used to test students' knowledge and class size. As with many DDL studies, Vyatkina's is based on short-answer activities (gap-filling and sentence-writing), designed to force the production of what should have been learnt. Though the writing task in the present study was structured, the choice of what language to produce was left open to the students. They were not obliged to use any of the phrases dealt with in the DDL activities. Other studies may test what students have learnt in more controlled conditions with short-answer items designed to elicit specific vocabulary or phrases. In a future study, greater control over the language produced by students in their texts may be obtained by obliging them to use some of the words encountered in their DDL tasks (see Huang, 2014).

A factor influencing the students' apparent preference for language dealt with in the hands-off tasks may be the class size. As stated above, Boulton (2012a) found that the average number of students in DDL studies for ESP is 45, with some studies on hands-on DDL focusing on much smaller classes (e.g., 11 students in Vyatkina, 2016; 14 in Yoon, 2008). The teacher-researcher in the present study was dealing with groups of up to 90 per lab session, with large university classes being a common situation in both

Italian and some other European universities. Though the students had been trained to use AntConc and collaborative work was encouraged, it was difficult to ensure that all students were managing to find the intended answers and that all were paying full attention during class feedback time. It is possible that the success of hands-on DDL may be facilitated by smaller class numbers, but this is an area for further investigation.

A word should also be said about the items that were selected for concordance-based analysis. The first item that the students encountered in these DDL tasks was *opinion*, a word that was already significantly more present in the control corpus than in the expert corpus. Its use was higher in the DDL corpus, even though students were encouraged to explore alternatives. It is likely that students are keen to use a word in their written production because they have studied it and feel confident about its phraseology. Conversely, students may also be keen to use completely new words within a simple phrase structure (e.g., adjective + noun) such as *overwhelming majority*.

The issues discussed here indicate that further research is necessary. A more careful research design would allow more precise conclusions about whether exercises based on single texts, paper-based concordances, or direct access to corpora are more effective for learning with large classes. The evidence would suggest that much of the students' preparation for their exam was based on the language in the paper-based concordances and even, to a lesser extent, on non-corpus-based tasks. Another issue to be considered is that the texts in the learner corpora were produced for an exam. It is highly likely, therefore, that the students had learnt key phrases from their worksheets in order to perform well and it is not clear whether this approach to studying phraseology has lasting effects. Huang (2014), though concluding that hands-off DDL can provide an "effective approach to helping learners obtain and retain lexico-grammatical patterns" (p. 175), does concede that a two-week delay between the concordance task and the post-test "is not sufficient to detect the development of learners' writing ability" (p. 177). Indeed, Callies (2015) also notes that there is still a scarcity of longitudinal studies in learner corpora. This observation is confirmed in the 2017 meta-analysis of 64 DDL studies by Boulton and Cobb, which finds that very few studies reported on the results of delayed post-tests, which would be essential to understand the long-term effects of DDL on students' output.

A further issue to address is that this study makes generalisations about the apparent beneficial effects of DDL in a group of students rather than looking at the dispersion of phrases across the group as used by individual students. The student texts are too short to produce relevant results and such a study would work better on longer texts. Though it cannot be claimed here that each student has broadened their vocabulary and knowledge of genre-specific phraseology, as a group, benefits can be seen from their exposure to a far wider range of expressions than could be provided by other types of exercises. In the light of this, there is encouraging positive evidence that phrases from the corpus-based activities are being reproduced.

Conclusions

This study has shown how DDL materials have fostered a heightened awareness of phraseology, with evidence of learners putting their new-found knowledge of sequences of words into practice. The comparison of two similar cohorts of students—one of which (the control group) did not have access to corpus-based exercises and the other which had both indirect and direct access to a corpus—revealed that DDL did indeed appear to lead to beneficial effects on students' written production, in that their phraseology more closely reflected what is expected in the genre studied. Students also showed knowledge of a wider range of vocabulary and suitable collocates than those in the control group.

The more extensive use of phraseology concerning words covered in the paper-based DDL exercises suggests that students possibly preferred a hands-off approach and that this may be more effective for their learning. Phrases identified by students in a task based on a single text, rather than in corpus-based activities, occurred frequently, indicating that such activities were also useful. Such tasks, however, may do little to broaden students' range of vocabulary and phraseology. Indeed, following the DDL tasks, a wider range of

vocabulary and suitable collocates is evident.

This study highlights how the students had a heightened awareness of the lexis and phraseology of the genre and appeared to learn to use phrases that were not produced by the control group. However, the language that students can be exposed to through hands-off tasks and tasks based on single texts is limited. What is more, the meta-analysis of 64 DDL studies by Boulton and Cobb (2017) reveals that hands-on tasks appear to lead to more beneficial effects than hands-off tasks, which indicates that there is potential for a more successful application of a hands-on approach in the context of a study such as this. The present study has highlighted areas that require more attention when applying a hands-on DDL approach, such as how to store and retrieve this information and how to deal with time constraints. Further approaches, particularly to promote the use of DDL with large classes, need to be sought to enhance the effectiveness of DDL, since students can be fully empowered to make discoveries and learn more for themselves only as independent users.

Notes

1. The comprehensive Learner Corpus Bibliography hosted by the Centre for English Corpus Linguistics at the Université Catholique de Louvain currently contains 30 entries that refer to argumentative essays in their titles alone.
2. Though studies based on contrastive interlanguage analysis tend to compare learner production with native speaker production (for discussion of the comparative fallacy, see Granger, 2015), I prefer to speak of expert and non-expert production in the context of this study, where the aim is for students to follow the norms expected of professional writers of a text type, rather than to appear native-like.
3. The self-presentations in these corpora are short messages written by students to introduce themselves to fellow students in an online forum.
4. Colligation has been defined by Sinclair (2004) as "the co-occurrence of a member of a grammatical class—say a word class—with a word or phrase" (p. 142).
5. Single-word substitutes for *problem* or alternatives for *issue* were not found in the DDL corpus. In the control corpus, the word *problem* was used to introduce a topic. One hypothesis, which would require further research, is that more effective use of genre-appropriate phraseology enabled the students to introduce a topic without a head noun such as *issue* or *problem*.
6. Students were specifically instructed to search for *favor* when using AntConc (see Appendix C) to facilitate the identification of significant phrases (for which there were no occurrences if *favour* was searched for). Both spelling varieties were investigated in the two learner corpora, though reference is made to the British spelling.
7. The students were asked to identify both whether the verb *to view* was used more frequently in the passive or active voice and what function word occurred to the right of *view* (see Appendix C).

References

Ackerley, K. (2008). Using comparable expert-writer and learner corpora for developing report-writing skills. In C. Taylor Torsello, K. Ackerley, & E. Castello (Eds.), *Corpora for university language teachers* (pp. 259–273). Bern, Switzerland: Peter Lang.

Ackerley, K. (2015). Short-term effects of students' exploration of corpora: A longitudinal study of pre- and post-modification of noun phrases in learner English. In E. Castello, K. Ackerley, & F. Coccetta (Eds.), *Studies in learner corpus linguistics: Research and applications for foreign language teaching and assessment* (pp. 199–218). Bern, Switzerland: Peter Lang.

Allen, D. (2009). Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education*, *1*, 105–107. Retrieved from: http://park.itc.u-tokyo.ac.jp/eigo/KJEE/001/105-127.pdf

Anthony, L. (2011). *AntConc (Version 3.2.4w)* [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Benson, P. (2001). *Teaching and researching autonomy in language learning*. Harlow, UK: Longman.

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) Frankfurt am Main, Germany: Peter Lang.

Bhatia, V. (2002). A generic view of academic discourse. In J. Flowerdew (Ed.), *Academic discourse*, (pp. 21–39). Harlow, UK: Pearson.

Biber, D., & Barbieri F. (2006). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, *26*, 263–286. doi: 10.1016/j.esp.2006.08.003

Biber, D., & Reppen, R. (1998). Comparing native and learner perspectives on English grammar: A study of complement clauses. In S. Granger (Ed.), *Learner English on computer* (pp. 145–158). London, UK: Longman.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at… Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*, 371–405. doi: 10.1093/applin/25.3.371

Bondi, M. (2001). Small corpora and language variation: Reflexivity across genres. In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies and ELT* (pp. 135–174) Amsterdam, Netherlands: John Benjamins.

Boulton, A. (2009a). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, *35*(1), 1–27.

Boulton, A. (2009b). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, *21*(1), 37–51. doi: 10.1017/S0958344009000068

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, *60*(3), 534–572. doi: 10.1111/j.1467-9922.2010.00566.x

Boulton, A. (2012a). Corpus consultation for ESP. A review of empirical research. In A. Boulton, S. Carter-Thomas, & E. Rowley-Jolivet (Eds.), *Corpus-informed research and learning in ESP: Issues and applications* (pp. 261–291). Amsterdam, Netherlands: John Benjamins.

Boulton, A. (2012b). Hands-on / hands-off: Alternative approaches to data-driven learning. In J. Thomas & A. Boulton (Eds.), *Input, process, and product: Developments in teaching and language corpora.* (pp. 152–168). Brno, Czech Republic: Masaryk University Press.

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

Callies, M. (2015). *Learner corpus methodology*. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 9–34). Cambridge, UK: Cambridge University Press.

Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, and assessment.* Cambridge, UK: Cambridge University Press.

Davies, M. (2008) *The Corpus of Contemporary American English (COCA): 520 million words, 1990-present*. Retrieved from http://corpus.byu.edu/coca/

Ellis, N. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18,* 91–126.

Flowerdew, J. (2001). Concordancing as a tool in course design. In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies and ELT* (pp. 71–92). Amsterdam, Netherlands: John Benjamins.

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

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

Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction, 1*(1), 41–61. doi: 10.1075/etc.1.1.05gil

Gledhill, C. (1998). Learning a genre as opposed to learning French. What can corpus linguistics tell us? In W. Geertz & L. Calvi (Eds.), *CALL, culture, and the language curriculum* (pp. 124–137). Berlin, Germany: Springer.

Granger, S. (2002). A bird's-eye view of learner corpora research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition, and foreign language teaching* (pp. 3–33). Amsterdam, Netherlands: John Benjamins.

Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research, 1*(1), 7–24. doi: 10.1075/ijlcr.1.1.01gra

Granger, S., & Meunier F. (2008). Phraseology in language learning and teaching. Where to from here? In S. Granger & F. Meunier (Eds.), *Phraseology in foreign language learning and teaching* (pp. 247–252). Amsterdam, Netherlands: John Benjamins.

Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 27–49). Amsterdam, Netherlands: John Benjamins.

Granger, S., Hung, J., & Petch-Tyson, S. (Eds.) (2002). *Computer learner corpora, second language acquisition, and foreign language teaching*. Amsterdam, Netherlands: John Benjamins.

Hafner, C., & Candlin, C. (2007). Corpus tools as an affordance to learning in professional legal education. *Journal of English for Academic Purposes, 6*(4), 303–318. doi: 10.1016/j.jeap.2007.09.005

Halliday, M. A. K. (1989). *Spoken and written language*. Oxford, UK: Oxford University Press.

Huang, Z. (2014). *The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. ReCALL, 26*(2), 163–183. doi: 10.1017/S0958344014000020

Hunston, S. (2002). *Corpora in applied linguistics.* Cambridge, UK: Cambridge University Press.

Hyland, K. (2002). Activity and evaluation: Reporting practices in academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp. 115–130). Harlow, UK: Pearson.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27,* 4–21. doi: 10.1016/j.esp.2007.06.001

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., & Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning & Technology, 14*(1), 28–44. Retrieved from: http://llt.msu.edu/vol14num1/kennedymiceli.pdf

Leńko-Szymańska, A., & Boulton, A. (2015). Introduction: Data-driven learning in language pedagogy. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 1–14). Amsterdam, Netherlands: John Benjamins.

Meunier, F. (2002). The role of learner and native corpora in grammar teaching. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition, and foreign language teaching* (pp. 119–142). Amsterdam, Netherlands: John Benjamins.

Nation, P. (2001). Using small corpora to investigate learner needs: Two vocabulary research tools. In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies and ELT* (pp. 31–46). Amsterdam, Netherlands: John Benjamins.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam, Netherlands: John Benjamins.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *Corpora in the classroom: Language use and language teaching*. Cambridge, UK: Cambridge University Press.

Paquot, M. (2008). Exemplification in learner writing: A cross-linguistic perspective. In S. Granger & F. Meunier (Eds.), *Phraseology in foreign language learning and teaching* (pp. 101–119). Amsterdam, Netherlands: John Benjamins.

Paquot, M. (2013). Lexical bundles and L1 transfer effects. *International Journal of Corpus Linguistics*, *18*(3), 391–417. doi: 10.1075/ijcl.18.3.06paq

Renouf, A., & Sinclair, J. (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics* (pp. 128–143). London, UK: Longman.

Schmitt, N. (2000). Key concepts in ELT: Lexical chunks. *ELT Journal*, *54*(4), 400–401. doi: 10.1093/elt/54.4.400

Schmitt, N. (Ed.) (2004). *Formulaic sequences*. Amsterdam, Netherlands: John Benjamins.

Seidlhofer, B. (2002). Pedagogy and local learner corpora: Working with learning-driven data. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition, and foreign language teaching* (pp. 213–234). Amsterdam, Netherlands: John Benjamins.

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

Sinclair, J. (2004). *Trust the text: Language, corpus, and discourse*. London, UK: Routledge.

Sinclair, J. (2008). Envoi. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 407–410). Amsterdam, Netherlands: John Benjamins.

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

Stubbs, M. (2002). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics*, *7*(2), 215–244. doi: 10.1075/ijcl.7.2.04stu

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge, UK: Cambridge University Press.

Tribble, C. (2002). Corpora and corpus analysis: New windows on academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp. 131–149). Harlow, UK: Pearson.

Vyatkina, N. (2016). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology*, *20*(3), 159–179. Retrieved from http://llt.msu.edu/issues/october2016/vyatkina.pdf

Warren, M. (2011). Using corpora in the learning and teaching of phraseological variation. In A. Frankenberg-Garcia, G. Aston, & L. Flowerdew (Eds.), *New trends in corpora and language learning* (pp. 153–166). London, UK: Continuum.

Yoon, C. (2011). Concordancing in L2 writing class: An overview of research and issues. *Journal of English for Academic Purposes*, *10*, 130–139.

Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology*, *12*(2), 31–48. Retrieved from http://llt.msu.edu/vol12num2/yoon.pdf

Appendix A. Task based on Single Text

Both cohorts of students were asked to read a single opinion survey report and complete the second column with expressions related to those provided in column one. The third column has been completed with example responses provided by both the teacher and students during a lesson with the control group. The responses in the third column were then made available in a file shared with the control group of students.

Expressions for Report Writing: Vocabulary Task

It is important to use a variety of structures and vocabulary in your formal writing. Study the text on the previous page, and identify some key expressions used by the writer in the report. Add other expressions you know to the third column.

	Expressions from the text	**Other expressions you know**
Quantity		
Many people/ most people	almost two thirds around nine out of ten most people a high number of participants	a high number of several a great deal of an overwhelming majority of people
Not many people	only a minority only one person in 20 only half that number a mere 5%	a few not many
Opinions		
Saying that people agree with the issue/each other	there is little or no difference… on attitudes to… **half the public would like** the Government to go even further over 47% **hold the same view about** cannabis a minority **favour** these options a majority **agree that**	people agree with people share the same opinion
Saying people disagree with the issue/each other	**the survey finds widely differing attitudes to** soft drugs a high number of participants… **dispute** the version of the "gateway"	*people are against* people reject people condemn people disapprove of

Stating what people think [note also the use of the passive voice]	*people think* most people **consider** **one argument often advanced** participants **say** heroin and cocaine **are commonly regarded as** cannabis **is seen as**	as far as people are concerned people believe as people see it in their opinion as far as X is concerned
Structuring the text		
Giving reasons	there seem to be two main reasons the real reason for... is that... one argument often advanced for continuing…	I think so because...[inf]
Enumerating points	the first is secondly finally	first of all in addition to this following this afterwards subsequently moreover furthermore
Adding contrasting ideas	on the other hand however	on the contrary in contrast nevertheless

Appendix B. Hands-Off DDL Exercises Based on Opinion and Majority

Opinion

The concordance below shows *opinion* in the Report Corpus, sorted 1L, 2L, 3L. What patterns do you notice?

```
ortant than the rights of early embryos. However, opinion is evenly divided on
ionally, the South has seen the biggest change in opinion on this issue. In
9 percent. Indonesia shows the greatest change in opinion, moving from 33
there are nonetheless considerable differences in opinion between the countries
an important factor in determining differences in opinion, with 52% of people
n new Member States reveals slight differences in opinion, particularly when it
 this opinion poll reflect the general balance of opinion we have witnessed
 t attitudes are in line with both the balance of opinion and intensity of
ly unchanged from 1999. Meanwhile, the balance of opinion among Catholics has
 by three-to-one (72% to 23%). But the balance of opinion has shifted in favor
utional law while 36% are opposed. The balance of opinion among other
younger than 15. The survey shows a difference of opinion among Welsh speakers
take this view. There is a dramatic difference of opinion over gay adoption
lder gap emerges, and only a slight difference of opinion is seen across age
```

```
ons are largely stable, so too are differences of opinion on the issue across
o the nation, but this masks a severe division of opinion within the party -
   mania. 1.6 Religion or beliefs European public opinion is divided when it he
 the basis of ethnic origin. Here as well, public opinion differs between the f
ion is widespread are far more likely to hold the opinion that being a woman is
eople across the 18 countries surveyed are of the opinion that nuclear power is
LUSION A large proportion of Europeans are of the opinion that discrimination
ocent people executed did not at all affect their opinion on the death penalty.
r opposing the death penalty did not affect their opinion and less likely to
easons to oppose the death penalty affected their opinion. Respondents living
penalty. Over 80 percent said this affected their opinion a lot or some.
a, France, and Cameroon – appear divided in their opinion, not significantly
us. People in Cameroon appear more split in their opinion compared to the other
eme, less than 4 out of 10 respondents share this opinion in Malta (32%) and
ppose overturning Roe v. Wade. But many with this opinion favor stricter limits
```

1. Find 2 expressions that mean people "think/believe".
2. Find an expression that means people agree with others.
3. Find 5 expressions that mean people disagree with each other.
4. In these 5 expressions (Question 3), find 3 examples of the passive voice.
5. Find 3 adjectives that collocate with difference/s of/in opinion.

Majority

A concordance can be useful to examine the collocations of a particular word. As a result concordances can help learners expand their vocabulary. The extracts below come from a concordance of majority from a corpus of reports on public opinion surveys.

```
003. For the first time, a majority (53%) favors permitting gays and
ort. But nearly as large a majority (54%) supports allowing homosexual
 over the past 11 years. A majority (55%) say they are at least "fairly
 progress to hard drugs. A majority, 58%, agree that cannabis users are
Only among seculars does a majority (63%) express support for gay
ociety On average, a broad majority of European Union citizens believe
g a lesser extent. A broad majority of European citizens believe that
ple. For instance, a clear majority (56%) continues to oppose allowing
blic this is true; a clear majority, 56%, think that people in other large
largely unchanged. A clear majority (56%) says it is more important to
h Catholics (61%). A clear majority of the public (68%) continues to
n, and of these, the great majority, 83%, admitted to drug use on the
r survey is that the great majority of British Muslims want to be loyal,
ate, for instance, a large majority (58%) supports allowing gays and
y. way. • In 1997, a large majority (59%) supported the execution of
who a moral issue. A large majority (60%) of those who believe that
and as MPs (72 %). A large majority also think that disabled people (74%)
```

```
tant role to play. A large majority of European Union citizens are willing
          In 1997, a large majority (59%) supported the execution of
(by 66%-50%). And a narrow majority of seculars (51%) feel it would not be
-Assisted Suicide A narrow majority of Americans (51%) favor making it
es of events, a two-to-one majority of adults (59% to 28%) and an even
l courts, the overwhelming majority of Americans (74%) indicate that
ed, while the overwhelming majority of liberals (71%) disagrees.
e these concerns, a slight majority (52%) feels that there are much bigger
 hot-button issues. A slim majority (52%) opposes allowing gays and
   Similarly, only a small majority (54%) of gay marriage opponents favor
ligious component. A small majority of conservatives (52%) says
ed in their views. A solid majority long have felt that Roe v. Wade should
he greater danger, a solid majority of conservative Republicans (57%) cite
   And while a substantial majority (57%) agrees that there are basic
ey had expected to be. The majority of students said they were so cash
(from 42% to 49%). But the majority of the public still rejects the idea
tion found that, while the majority of Americans support capital 0
35; FSU sample, M_38). The majority of the participants in the two samples
es available, a two thirds majority (66%) says that entertainment TV shows
                  The vast majority - 91% - of 12-year-olds now own a 38
help and support, the vast majority of abortions are not performed due to
ots don't smoke. "The vast majority of Scots who do smoke want to give
 be ended (though the vast majority think only non-violent means should be
```

Look at the percentages in the concordance lines. Find left collocates used to describe a:

- 51-54% majority
- 55-57% majority
- 57-60% majority
- majority of over 70%
- majority of over 80%

Appendix C. Hands-On DDL Exercises Based on View, Support, and Favor

View

Look up *view*. Sort 1L, 2L, 3L.

1. Is *view* used more as a noun or a verb?
2. How many times is it used as a verb?
3. What function words (prepositions, pronouns, auxiliary verbs, articles, or conjunctions) typically collocate to the right of *view* when it is used as a verb?

Look up *viewed*. Are there more examples of the verb in the active or passive voice?

When *view* is used as a noun, which adjectives does it most typically collocate with?

4. Which verb/s does it collocate with? _____

Support

Wordlists from comparable corpora of reports show Italian learners overuse agree and underuse possible alternatives.

Use the concordancer to find *support* in the report corpus.

1. Find a verb commonly found to the left of *support* [in Antconc: Level 1 = 1L > SORT].
2. Three adjectives that pre-modify support.

Favor

Use the concordancer to find *favor* in the report corpus [in Antconc: Level 1 = 1L > SORT].

1. Find adverbs that pre-modify the verb *favor*.
2. Find the most common clusters (groups of words). [in Antconc: click on cluster at top of screen; Cluster size: min. 3, max. 3 > SORT]

About the Author

Katherine Ackerley is Assistant Professor of English Language at the Department of Linguistic and Literary Studies and Deputy Director of the Language Centre at the University of Padova. Her research interests include applied corpus linguistics (particularly learner corpora), DDL, and English-medium instruction.

E-mail: katherine.ackerley@unipd.it