Corpus-Based Approaches in Language Teaching: Outcomes, Observations, and Teacher Perspectives

Eric Friginal, Peter Dye and Matthew Nolen

Abstract

This paper is a synthesis of teacher perspectives, analyses of learner output and learner impressions, ideas for materials development, and summative researcher observations resulting from the implementation of a semester-long (around 8 weeks of instruction) corpus-based and data-driven English as a Second Language (ESL) instruction in two university-level and adult study abroad settings in the United States (U.S.). Case Study 1 investigates learner and instructor attitudes regarding the effectiveness of corpus-based instruction in developing academic writing skills specifically designed for a group of visiting Chinese scholars. Case Study 2 follows a mixed-methods, exploratory investigation into the use of a scaffolded student worksheet to guide learners with different proficiency levels in the use of corpora and corpus tools during a semester-long study abroad program at a non-profit, private institution based in the U.S. The worksheet was designed to regularly incorporate corpus-based lessons and data into the classroom instruction or into homework activities. Results reveal instructor enthusiasm for the potentially vast and promising benefits of incorporating corpus tools into their own language classroom instruction. With specific and clear learning goals and instruction on their use, corpus tools can be a valuable resource for appropriate students during and after completion of the course.

Keywords: Corpus linguistics, teacher perception, data-driven learning, technology in the classroom

Article info

Received: 03.04.2019

Revised: 23.08.2019

Accepted: 25.12.2019

Published online: 30.06.2020

Introduction

Corpus linguistics is primarily a methodological approach (more than a language model or a sub-field of linguistics) to the study of language structure, patterns, and use (Biber, Reppen, & Friginal, 2010; Lee & Swales, 2006). The use of corpora has become a popular methodology in the quantitative analysis of the linguistic characteristics of written and spoken academic discourse, resulting in the development of more authentic teaching materials, frequency-based dictionaries, and ESL textbooks, especially for university-level learners of English (Friginal, 2018; Friginal & Hardy, 2014; Tribble, 2015; Römer, 2011). *Corpora* (singular form corpus) are, in a sense, datasets of "systematically collected, naturally-occurring categories of texts" (Friginal & Hardy,

Eric Friginal, Georgia State University, College of Arts and Sciences, Department of Applied Linguistics, Atlanta, GA, USA, efriginal@gsu.edu, ORCID: 0000-0001-5956-3674
Peter Dye, Oglethorpe University, International Study Center, Atlanta, GA, USA, pdye@oglethorpe.edu
Matthew Nolen, Conexion Training, Panama City, Panama, mtnolen@gmail.com

2014, p. 20). That is, corpora are collections of written or spoken language-in-use, which are stored, analyzed, and utilized for a variety of purposes by researchers, teachers, and learners themselves. Direct applications of corpora and corpus tools in the classroom support various language teaching and language acquisition theories and concepts especially related to learner autonomy, use of *realia* and authentic texts, leaner-computer and learner-learner interactions, and explicit teaching of language features and patterns (Friginal, 2018).

This paper focuses on a synthesis of teacher perspectives, analyses of learner output and learner impressions, ideas for materials development, and summative researcher observations resulting from the implementation of a semester-long (or 8 to 10-week instruction) corpus-based and data-driven English as a Second Language (ESL) instruction in two university-level and adult study abroad settings in the United States (U.S.). As part of an ongoing pseudo-experimental, ethnographic, and longitudinal study of the "Corpus Linguistics for English Teachers (CLET) Project," this paper also highlights the role of teachers as classroom-based researchers and materials developers guided by the corpus approach to language teaching and learning. CLET was launched following the publication of Friginal's (2018) book manuscript, *Corpus Linguistics for English Teachers: New Tools, Online Resources, and Classroom Activities*, which describes corpus linguistics and its many relevant, creative, and engaging applications to language teaching and learning for teachers and practitioners in TESOL (Teachers of English to Speakers of Other Languages) and ESL/EFL (English as a Second/Foreign Language), and graduate students in applied linguistics. Friginal describes corpus approaches to the teaching of English vocabulary, grammar, and spoken-written academic discourse; identifies emerging tools, online resources, and classroom activities; and emphasizes the important contributions of the English teacher as a corpus-based materials developer and researcher to teachers and learners globally.

The corpus approach to researching the features and patterns of language benefits English language teachers as they facilitate the learning and acquisition of English. The number of teachers incorporating corpus-based materials in their classrooms has grown exponentially from the mid-1990s. However, as Flowerdew (2015), Friginal (2013), Geluso and Yamaguchi (2014), and Meunier and Reppen (2015) have noted, many teachers, even those who have received some training in corpus linguistics, are still not *regularly* using corpus-based activities in their classrooms for a variety of reasons, including a lack of confidence in the methodology, time constraints, difficulty learning and accessing tools, questions of relevance, and the challenges in orienting their students and re-designing their courses to incorporate corpus-based approaches (Friginal, 2018). Those who regularly use tools and materials developed from corpora are still not able to actively share their perspectives and observations with a broader audience worldwide. There are international conferences and various academic publications for corpora and teaching topics, but the number of these published studies and proceedings from international conferences is still extremely limited and mostly based on university-level language learners' writing classes in the U.S., United Kingdom, or parts of Europe. The intent here is, as Friginal (2018) emphasized, to continue to document and share relevant outcomes and teacher and learner experiences, including a wider-range of learners, English/language courses, and teaching goals.

A Corpus Described

The four main characteristics of a corpus are that it is authentic, relatively large, electronic, and conforms to specific criteria (Bowker & Pearson, 2002). There are corpora containing a variety of registers, also referred to as *text types* including academic English, spoken English, newspaper articles, novels and short stories, or legal cases. There is no specific rule regarding the size of a corpus but it should be large enough to promote a systematic analysis of relevant, target linguistic patterns. With the advent of personal computers as well as major innovations in internet technology, corpora have been freely shared and explored predominantly for language teaching and research purposes. One obvious benefit of this approach is that corpora allow for the observation and study of real-world language use, with relevant frequency distributions and access to actual occurrences of features, rather than relying only on limited intuition. Considering its potential, it is easy to envision the utility and benefit of corpus-based approaches in a variety of teaching contexts. In fact, Teubert (2005) noted that corpus linguistics is now held to be a "default resource" in linguistic research since it is a reflection of "real language data." For example, corpus tools have contributed greatly to studies of *phraseology* and *collocations* illustrating how such resulting datasets can inform language learning and teaching. Phraseology is not necessarily a new field, but corpus tools have enhanced the ability of learners to understand and visualize that "a word is not limited to the word itself but also the words around it" (Firth, 1957). Learners can, therefore, more readily comprehend that the "meaning" or utility of a word extends even beyond the "borders" of its neighboring words to include "chunks" or bundles commonly co-occurring in a language. Put plainly, Römer (2009) observes that "language is highly patterned" (p. 140), and often, these patterns are not rare.

In the classroom, once it is collected by an ESL/EFL teacher, for example, a corpus can be analyzed by students using computer-based tools such as a concordancing program. A *concordancer* is the most basic form of a corpus tool which allows the corpus, with all its texts and total number of words to be searched based on word frequencies, collocations and word units (e.g., lexical bundles or n-grams), and any number of usage-related topics. There are, additionally, numerous applications that will provide learners the opportunity to compare various corpora (i.e., various registers or text types) to progress into more technical analyses, requiring additional, more advanced computer programs. Recently, online computer programs and corpus databases with built-in concordancers have been shared online. (Some are proprietary or may require a license, but there are several which are accessible for free.) Briefly discussed in this paper are popular corpus databases such as the "Corpus of Contemporary American English" (COCA) developed by Mark Davies of Brigham Young University (See other databases from Davies here: https://www.english-corpora.org/.) and the concordancer AntConc (See Figure 1.), created and freely-shared by Laurence Anthony from Waseda University. Many other accessible and newly-developed packages and databases are also available, with user manuals and tutorials. Search online, for example, for the following: Sketch Engine, LancsBox, Coh-Metrix, TextSTAT, MAT Tagger, "The International Corpus Network of Asian Learners of English," or "MICASE/MICUSP".

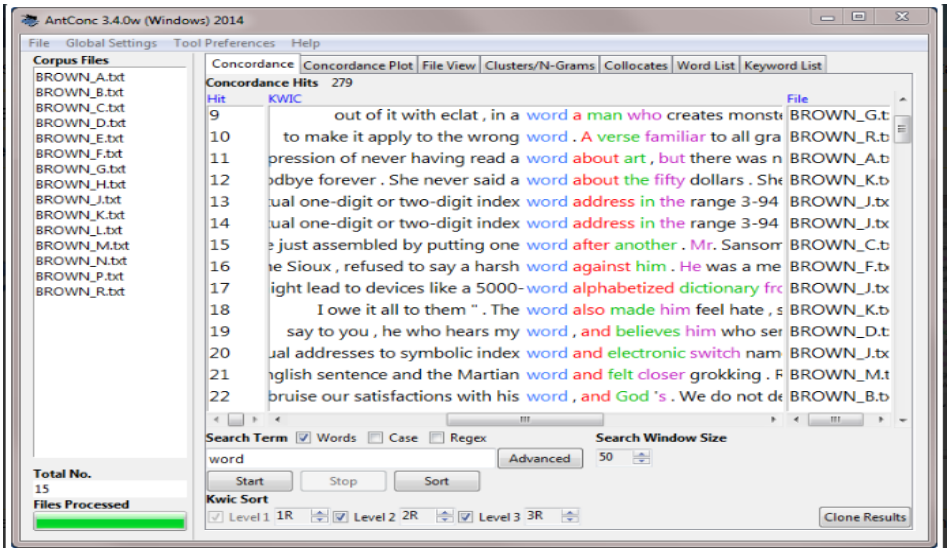

Figure 1. Screenshot of the AntConc concordancer (Anthony, 2018) available for free download from http://www.laurenceanthony.net/software/antconc/

Corpora, Instructional Technology and Data-Driven Learning

In the broader field of English teaching across learners and contexts, corpora and corpus tools have been incorporated into three primary instructional approaches: (1) educational or instructional technology-based learning, (2) computer-assisted learning, and (3) data-driven instruction. These three strategies, especially the first two, share common characteristics: both are machine-specific (i.e., computers) and they also align well with and support other instructional approaches such as learner-centered instruction or autonomous learning. Specifically, Instructional Technology emphasizes the role of tools and their integration into the learning process; Computer-Assisted Language Learning (CALL) focuses on learning languages with the aid of computers with a particular emphasis on software design and evaluation, and data-driven learning (DDL) focuses on learners' direct discovery and use of linguistic information/data in the language classroom and beyond. These three have been the most common instructional approaches in which corpora and corpus tools have been situated in various studies over the past two decades (Friginal, 2018).

Smart (2014) observed that, "DDL allows learners to inductively discover language structures and patterns through interacting with concordancing software or with concordance-based instructional materials" (p.184). In this sense, DDL presents learners with actual concordance lines of written authentic language that centers literally on a particular word or phrase. Friginal and Hardy (2014) noted that concordancers "provide the user with the organized contexts of items that are searched. Often, one might be interested in exploring the words before and after a given word" (p. 39). At the

same time, concordancers provide the immediate, yet limited, context surrounding a target word or phrase. Context is placed on the word or phrase of interest and not on the meaning of the sentence or paragraph as a whole (Kaltenböck & Mehlmauer-Larcher, 2005). This may seem confusing and limiting at first, but concordance lines typically yield enough context to inform learners and teachers of the various uses of words or phrases in relation to collocations and other multi-word patterns. The process, then, clearly provides a focus on form and meaning in short, multiple contexts, showing various usages simultaneously and without the distraction of longer stretches of discourse (Boulton, 2009). Therefore, while the entire discourse may not be comprehensively attended to, the patterns found between words are highlighted, allowing the learners to discover nuanced meanings of the real-word language they encounter.

Researching Corpus-Based Approaches in the Classroom

Corpora have been put to practical use, especially in the writing classroom, as described in a number of studies since the early 2000's. Many of these studies highlight the classroom experiences of non-native speakers (NNS) of English. A great deal of linguistic variation exists across academic disciplines, and this can be particularly challenging for NNSs working to improve their writing within a specific field. Lee and Swales (2006) designed an experimental course entitled "Exploring Your Own Discourse World" to help doctoral students compare their own writing to that of more established writers in their fields through the use of corpora. With the use of a concordancer, the students were able to examine the use of linguistic elements like common verbs and their conjugations, definite article usage, and collocates used in their disciplines. In a similar study, Cortes (2011) also had students compile and compare their own writing to published research articles in their disciplines. Selected texts from the field of applied linguistics were also used to analyze and reveal organizational patterns of research articles.

These types of NNS courses are particularly useful for identifying and examining patterns in specific disciplines that may differ from more general linguistic patterns found across other academic fields. By comparing their own writing to those of experts, students can identify, refine, and adapt their linguistic choices enabling them to enhance their overall written presentation of ideas and research processes. The benefit of acquiring this skill is that students can continue to use the approach more independently and universally, well after a course or workshop has finished. For example, Gilquin, Granger and Paquot (2007) examined the effectiveness of using NNS learner corpora in conjunction with native corpora in an English for Academic Purposes (EAP) context. They found the approach to be useful in expanding NNSs' linguistic repertoire and in avoiding falling into common writing traps that typical NNSs face. Similar methods have been put into practice for improving the discipline-specific writing skills for native-speakers of English as well. Friginal (2013) used corpus tools to develop research report writing skills for a group of college-level students in a professional forestry program. A concordancing program was used to analyze specific linguistic patterns including linking adverbials, reporting verbs, verb tenses, and passive

sentence structures. The results of the study showed improvements in the students' report writing abilities after the corpus instruction.

Overall, research shows a great deal of enthusiasm from teachers regarding corpus use, and there are some data, although still limited, showing that university-level learners also tend to respond positively to these types of corpus-based courses and approaches. The participants of Lee and Swales' (2006) course described above found many useful applications of the corpus approach after gaining familiarity with using the tools and their own databases. They found corpora to be empowering as a reference tool as well as convenient because they can be accessed at any time inside or outside of the classroom. Yoon and Hirvela (2004) also confirmed students' positive attitudes regarding the use of a corpus tools. They do note that the use of corpora worked more effectively with advanced learners and that the teacher has a responsibility to adequately explain the merits of using a corpus as a supporting mechanism to improve writing. They conclude by cautioning against placing too much prominence on corpus-based instruction. It is one tool among many, but when implemented properly, it can provide students with a practical way to improve their writing ability and, consequently, create a positive impression of the approach. In the same vein, Flowerdew (2005) examines some of the criticisms of corpus-based studies including the potential excessive focus on the bottom-up approach to writing. The utilization of corpus methods in the classroom can be regarded as relying too heavily on the micro-level elements of texts to the exclusion of the examination of their broader contexts. Flowerdew suggests, however, that top-down, genre-based approaches to writing instruction, when done correctly, could still be compatible with and complement bottom-up, corpus-based analysis and *vice versa*, with proper planning and preparation.

The Focus of This Paper

The CLET Project aims to document qualitative and quantitative, and especially measurable learning (i.e., learning gains) attributable to the use of corpora in the classroom. Clearly, there is still a lot of research to be done here. Longitudinal studies are critical to drawing relevant connections between language learning and corpus-based approaches, but up to this point, data, especially from true experimental studies, are still very limited, given the clear challenges. One way to continue to contribute data is to qualitatively document and describe teacher and learner perspectives and outputs in corpus-based classrooms. The two case studies presented here follow this methodology, with two instructor-participants who received training in corpus linguistics and technology and language teaching and learning from their coursework in a master's program in Applied Linguistics and English as a Second Language at the U.S.-based urban university.

Case Study 1 investigates learner and instructor attitudes regarding the effectiveness of corpus-based instruction in developing academic writing skills specifically designed for a group of visiting Chinese scholars. The study took place over an eight-week English instruction class from a Faculty Mentoring Program in a university setting. Both the COCA (Davies, 2018) and AntConc (Anthony, 2018) were introduced to and used by the participants. The participants also compiled their own corpora from academic articles in their disciplines to search for field-specific linguistic

patterns. The primary goal of Case Study 1 is to determine whether or not corpus tools can contribute positively to the development of professional/academic writing, editing, and research skills of the participating Chinese scholars across a range of disciplines. Data analyzed were collected from the instructor's reflections and journal of classroom observations, student email responses, recorded interviews with learners, classroom worksheets involving the use of corpus tools, as well as a pre- and post-course surveys.

Case Study 2 follows a mixed-methods, exploratory investigation into the use of a scaffolded student worksheet in order to guide learners with different level of proficiencies to use corpora and corpus tools during a semester-long study abroad program at a non-profit, private institution based in the U.S. The worksheet was designed as an "Explorer's Daily Journal," regularly incorporating and recording learners' responses to corpus-based lessons, data, and various online and computer-based tools in the classroom or for work outside the class as homework activities. In Case Study 1, AntConc, COCA, MICASE, and others and teacher-developed materials are utilized by learners. As the tools were being presented and learned during various phases of the course, student and teacher attitudes and perceptions were documented through a teacher's journal and interviews with the learners. The specific goal of Case Study 2 is to examine how learners and the teacher perceive the process of using the worksheet as integral part of learning English, together with an exploratory look at general learning or performance improvements and challenges.

Case Study 1: Corpus-based Instruction to Develop Professional and Academic Writing Skills

Scenario

This university, located in the Southeastern U.S., has been hosting visiting scholars under a "Faculty Mentoring Program" in an effort to promote professional development and strengthen the university's global presence and relationships. Each visiting scholar (all from China, in this study) is paired with a faculty mentor from the same academic field or discipline to work with during his/her stay in the U.S. As a part of the program, there is also an Intensive English Instruction (IEI) component lasting eight weeks, for four hours per week. The IEI involves English training for written and oral comprehension and production. This part of the program focuses on topics such as writing professional emails, leading academic discussions, and oral research presentations. Other topics include specific elements of academic English such as pronunciation issues, field-related writing conventions, and cultural aspects of communication. The IEI course was the setting within which corpora and corpus tools were introduced and incorporated as part of the coursework. The instructor is a native speaker of English, with extensive teaching experiences across various levels of learners. He received training in corpus linguistics in the U.S., as part of his master's degree in applied linguistics and English as a second language. The participants in Case Study 1 were nine Chinese professors from a diverse group of academic fields including art, biology, computer science, English, early childhood education, history, and physics. The majority of the visiting scholars are competent in comprehending academic reading (especially related to their field) and several of them have previously published research

articles in English. Despite this, academic writing and its specific conventions are an aspect that most members of the group expressed interest in developing or improving. They also expressed a desire to publish future academic papers in English in competitive international publications, and improve their ability to do so accurately and effectively. The participants also noted that they would like to receive support in editing research articles and conference presentation abstracts, selecting the appropriate expressions in academic writing, translating ideas from Chinese to English, and using English grammar correctly in both speech and writing. These overall comments from participants as part of an initial needs assessment provided the motivation to incorporate corpus-based methods in the IEI component and demonstrate how corpora may be useful in helping the instructor to achieve participants' English goals.

Outline of Corpus Instruction

The participants utilized corpus tools and corpus-based materials throughout the eight-week IEI program. Specifically, the corpus tools introduced were the COCA (Davies, 2018) and the AntConc (Anthony, 2018) concordancing program.

Corpus of Contemporary American English. Because the course included a variety of skills (listening, speaking, reading, etc.), the corpus instruction was planned specifically around the days of writing instruction. In order to set the stage for the introduction to corpus tools, the second week of class involved analyzing patterns and features of different English written registers. Handouts were distributed that included samples of text/SMS messages, formal and informal emails, an online sports article, and one page from a research article. The participants were instructed to identify any unique elements or features of each written register and compare their observations with each other. They were guided to examine a range of aspects such as abbreviations, tone, verb tenses, and intended audience. The primary aim was to have the class analyze various texts and identify particular features by hand so that they would better understand the merits of using a corpus tool for a more systematic way of analyzing linguistic patterns and characteristics introduced in the following class session.

The corpus instruction was held in a computer lab and the class began with a summary of the reading about corpora and a brief explanation of how and why corpus tools can be useful to language learners. Next, a handout instructed the scholars to discuss in pairs any aspect of English writing that they felt was unique to their disciplines. The handout also contained examples from some of the participants' own writing that contained potentially inappropriate linguistic elements. They were guided to discuss what could be edited or changed and how to check to determine if their corrections were appropriate. Following this discussion, the class was instructed to access COCA on their workstations. The participants were guided to perform simple searches demonstrated by the instructor on the projector. After everyone had practiced conducting a few searches, they were directed back to the handout to follow instructions in order to complete a scavenger hunt with the academic corpus. With a partner, participants searched through various words and phrases to determine which were most frequent and how they were used in context. The activities revolved around tasks like recording the number of search results (e.g., "*For example*: 48, 376"), writing down

examples of the terms used in context, and documenting any punctuation patterns used with different search terms. The class concluded with a discussion of general and specific findings and anything "new" that was learned or discovered by individual participants.

AntConc and Learner-Compiled Corpora. Over the next few weeks, other necessary oral and written skills (e.g., pronunciation, writing emails, and leaving voicemail messages) were discussed, with COCA regularly referenced during class as a resource when questions came up regarding usage and frequency of various vocabulary or phrasal terms. During the fourth week of class, AntConc was presented as another useful tool for searching through teacher or learner-collected corpora. Participants were introduced to the concept of a stand-alone concordancer, and after becoming familiar with AntConc (by means of a tutorial led by the instructor), the class performed various searches and activities from written texts provided to them by the instructor. Working with a partner, they recorded wordlists and collocations, multi-word units, and distributions of frequent phrasal units (e.g., searches to determine which prepositions and verbs were used in differing contexts: *move in* vs. *move on*). The instructor showed how to analyze frequency "hits" to help identify differences in meaning across contexts. A follow-up activity allowed participants to conduct their own searches to identify words and patterns they were curious about, check their intuitions against actual data, and further familiarize themselves with using the program.

Then, each participant was asked to collect a mini-corpus of research articles in English from his/her discipline, with the ultimate goal of building a discipline-specific corpus intended as a reference corpus for future activities and comparisons. They were told to collect up to 15 articles published in top, peer-reviewed international journals. This activity culminated with the main assignment which was to conduct their own searches using their self-compiled corpora, guided by questions developed by the instructor. The following section (Figure 2) shows a sample handout for this activity distributed to the participants as they utilize AntConc to explore their learner-collected corpus.

Analyzing patterns in writing: Verbs
This is an open research activity. Everyone may identify different patterns. The goal is to determine which types of verbs are more appropriate in specific situations to help our English writing.

1. News
Read through the news sample and underline all of the main verbs in each sentence. What grammatical patterns are most common (For example: do/did/have done/had done/is doing/was doing/was done/etc.)? Are there any verbs that are used more than once?

2. Fiction
Read through the short story sample and underline all of the main verbs in each sentence. What grammatical patterns are most common (For example: do/did/have done/had done/is doing/was doing/was done/etc.)? Are there any verbs that are used more than once?

3. Academic
Read through the academic article sample (applied linguistics) and underline all of the main verbs in each sentence. What grammatical patterns are most common (For example: do/did/have done/had done/is doing/was doing/was done/etc.)? Are there any verbs that are used more than once?

4. Your Discipline
Analyze the verb patterns of your discipline using AntConc.

1. Open AntConc (from your USB or download from http://www.laurenceanthony.net/software/antconc/)
2. Select **File** and click on **Open File(s)...**
3. Select all of the files from the corpus of your field (Biology, Computer Science, etc.)
4. Go to the **Word List** tab
5. Click the **Start** button to see the most frequent words in your corpus
6. You can use the **Concordance** tab to search for specific verbs/words/phrases
7. Answer the following guide questions:
   - What are some examples of common verbs? How are these verbs used in context?
   - What grammatical patterns are the most frequent?
   - What patterns jumped out to you during your search? Did you find anything interesting or surprising?
   - How can you use this information to improve your writing?

Figure 2. AntConc handout developed by the instructor

Table 1 shows a mini-synthesis of three participants' responses to the AntConc guide questions, illustrating some similarities and differences in how they completed the activity.

Table 1. Analyzing patterns in writing from participants' disciplines with AntConc (samples from three participants)

Participant	**What are some examples of common verbs?** [Participants were told they could look up any types of words if preferred.]	**What grammatical patterns are most common?**	**Did you find anything interesting or surprising?**	**How can you use this information to improve your writing?**
1	is are study learn	passive voice the subjects are usually something or objects	TMM MEM WOW (what are these?)	-to check the patterns if I am not sure about them
2	(41) development (51) study (60) teach (77) growth (78) effect	past pattern	Verbs come out on 51 rank	-first find the verb, then check it
3	1. Defined parameter (*d* is the thickness of a tilted? Form?) 2. Passive voice (the maximum value is improved) 3. Is followed by adjective (but the reason is not clear so far)	passive voice present present perfect	Both A and C increase slightly with **increasing** oxygen content The seebeck coefficient decreases due to the increased carriers concentration With the increasing steps [more "increasing" examples and other -ing adjectives]	gerund The doping element Heavy doping Lower doping increas* [further notes]

Outcomes and Observations

Instructor's Impressions. Throughout the course, the instructor kept a journal to record his impressions about corpus instruction and participants' learning outcomes. Overall, the instructor's notes contain positive impressions in terms of direct benefits and applications, especially for participants who have shown major interest in the approach. The instructor noted a generally mixed reception to some aspects of corpus instruction at the start of the course, but he sought to adapt the approach throughout the course "to better meet the needs and desires of the class as a whole." He noted that "more time committed specifically to using corpora would have been helpful." Overall, he argued that the IEI program was implemented effectively and was completed with generally encouraging results. The participants appeared to have responded positively to discovering linguistic patterns and distributions in academic written texts that they would not have normally found by means of a manual analysis or traditional instruction. Regarding the participants' level of commitment and expected learning curve, the instructor during the earlier stages of instruction, expressed doubts about how focused the participants would be in learning to use corpus tools since they were already working with such a large number of new skills and tasks within a new context to consider.

Because of the range of disciplines and high-level academic backgrounds represented by the scholars in the program, the instructor had high expectations that corpus tools would be directly applicable and immediately accessible. However, several challenges presented themselves throughout the course as described in this entry from the instructor's journal:

> We had a bit of a rocky start with the introduction to COCA. One participant was new and hadn't had enough time to acclimate to [the university] to digest the information. Also, there is quite a mix of levels in the group and some of the instructions went over a few people's heads. I need to slow down my instruction and explanation of procedures. I also failed to plan on the limited number of queries that COCA allows and should have allotted more time for them all to register before going through my demonstrations and examples. We'll try again, and hopefully at least some of the group will be able to use COCA as a resource. (Instructor's Journal, Week 3)

When introducing COCA, careful planning was found to be critical to make it meaningful and effective. One recurring issue from the instructor's perspective was that the scholars had to learn to use corpus tools while learning a great deal of other skills simultaneously, including the ins and outs of campus life in a new country. About halfway through the course another journal entry from the instructor describes this observation:

> One note that I will need to come back to is the context of these meetings. I feel some of the members are overwhelmed by the number of new things to learn (culture, technology, communication with mentors, ID cards/numbers, etc.) and trying to learn and apply corpus tools might be too much for such a short course that covers so much. (Instructor's Journal, Week 5)

As the course progressed, the instructor attempted to continually weave in the themes of genre or register analysis, linguistic frequency, and appropriate usage in context to better set up the implementation of corpus tools. The introduction to AntConc was framed around examining a sample Spoken American Conversation corpus and later the learner-collected corpus. The group as a whole showed a great deal of interest in the spoken aspects of English and how to know when they were using terms appropriately or not. The goal was to use this learned skill to eventually transition into analyzing discipline-specific writing from corpora they personally collected. The instructor's observations from the initial introduction to AntConc are described below:

> The activities went over pretty well in the [computer lab] today. The challenge is and always will be properly selling the merits of corpus tools. I think I did a good job of explaining the importance of finding linguistic patterns and how challenging it can sometimes be. The preposition analysis was a good place to start because the one word (move in vs. move on) makes a big difference in meaning and usage. I'm also glad we started with the spoken corpus because it served as a nice transition from our recent discussions of spoken English and will (hopefully) transition nicely into the analysis of the written corpora that we'll be looking at next. I emailed all of the participants to start compiling their own collection of articles from their fields, and several have already emailed me a nice sampling of pdfs. I think the ability to collect, keep, and search through their own corpus will be a practical tool for their future writing and research. (Instructor's Journal, Week 4)

The scholars eventually began compiling their articles based on their own fields and research. Even as non-writing skills were taught, the idea of using context to determine usage and meaning was emphasized. The more these ideas were repeated in the class, the more they appeared to resonate with the participants, as can be seen in the following entry:

> We used a Ted Talk today to go over listening strategies and learning vocabulary through context. I further stressed the importance of using the context to help discern meaning and linked the idea back to the use of corpora. At the end of class I reviewed how I would like them all to continue sending me academic articles from their fields so that we can analyze them through AntConc. I showed them how to search for PDFs through the library page and attach them to an email. Everyone seemed to get it, and I think everyone is starting to see the benefits of this type of analysis as we continue to do it in different ways. (Instructor's Journal, Week 5)

After each class member had compiled their own corpus, they were instructed to begin conducting their own searches. Several participants noted some advantageous applications of the corpus analysis, while one did not seem to find it practical, as described below.

> The AntConc analysis of their corpus appeared to provide some useful data. They each identified some common words and were able to find some patterns with the passive voice. A few seemed particularly optimistic about how they could use corpus tools in the future to improve

> their own writing, while one noted that she did not find it interesting or useful. (Instructor's Journal, Week 6)

Case Study 2: Corpus Tools and Digital Journaling Beyond the Classroom

Scenario

Case Study 2 was conducted in a *study abroad* setting, also based in the Southeastern U.S., within a private, third-party academic organization. "Study abroad" is a general term that serves as an umbrella term for a variety of different academic and professional programs. Following Engle and Engle's (2003) operationalization of study abroad programs, the participants in the current study are enrolled in a "Level 5: Cross-Cultural Immersion Program." Here, learners stay between an academic semester to a year, are instructed in the target-language in all "curricular and extracurricular activities," are required to participate regularly in a cultural integration program, and receive ongoing orientation and mentorship (p.11). The learners (all adults) who attend this program are typically professionals and students hoping to eventually take and pass the Test of English as a Foreign Language (TOEFL). Most learners have attended universities in their home countries and are familiar with more traditional means of studying English. The result is that learners have very limited exposure to corpora and the tools needed to explore them. For this study, there were 10 participants enrolled in the program who all attended the course facilitated with the use of corpus tools. These participants all in one heterogeneous group were categorized as novice-level learners (5), intermediate-level learners (3), and advanced-level learners (2). Demographically, 7 out of 10 learners were women, 6 out of 10 were Brazilians, and all learners were 28 years old or older. As in Case Study 1, the instructor for Case Study 2 had received a master's degree in applied linguistics and ESL in the U.S. and various training in corpus linguistics and technology and language learning and teaching.

Instruction for about an hour was dedicated every week to the study and use of corpora, framed within a DDL approach. Overall, a total of 10 hours of instruction on corpora/DDL, together with a series of homework assignments based on the "Explorer's Digital Journal" worksheet (discussed below) was provided. Once the worksheet was presented, learners were encouraged to explore words or phrases that they found in their other classes or in "the real world," i.e., other forms of language use and instruction outside the classroom. Learners understood that the aim of including DDL in their lessons was to foster a habit of using data and technology-based tools as an additional resource while learning English in the U.S.

Outline of Corpus Instruction

The Explorer's Digital Journal. *The Explorer's Digital Journal* (EDJ) is a worksheet that was designed by the Case Study 2 instructor to help learners navigate some basic processes for using corpora *outside* the classroom (see Figure 3 for a sample EDJ Handout developed by the instructor). The EDJ was scaffolded into three parts or "levels" to allow for all learners, including the novice-level, to participate in the activities. Each stage introduced a new element to using corpus tools such as

concordancers and various online corpora (e.g., "Michigan Corpus of Academic Spoken English" or MICASE). Level 1 is simply "pattern-hunting" or identifying possible phraseological items. The participants were not expected, at this level, to even fully understand the words they were looking for—they must simply find *patterns*. They were, however, encouraged to write original sentences for examples, which required them to try to apply their understanding of linguistic patterns in the context of a sentence. The novice-learners were also encouraged to talk to native speakers of English or their instructors about the patterns they found, but they were not expected to induce the exact meanings of the words and the collocations they found on their own.

Level 2 required participants to investigate words or phrases and their possible collocations in more detail. Here, they tried to recognize some of the lexico-grammatical patterns, word semantics and prosody, and possibly even pragmatic features of collocations. The novice-level class was not expected to complete the Level 2 EDJ. The second level of the EDJ followed the same initial steps as the Level 1, but participants were additionally asked to look up collocations and make open-ended observations about the collocations found. Finally, Level 3 was designed specifically with advanced-level learners in mind. This version of the EDJ replicated the first steps that Level 2 introduced and built upon it three new aspects: First, participants compared English collocations with the literal translation in their first language (L1). This was done to isolate each collocation in the L1 so that they could make direct comparisons and hopefully discover syntactic, semantic, or pragmatic differences for the "same" phrase. This activity was used to show learners that some phrases created from their L1 may not carry the same semantic prosody or meaning as it does in English. Then, they were asked to take the same three collocations they found and interpreted and search for their occurrences in a different corpus. This activity served to help learners to understand the impact that register has, not only within different corpora but also within the language as a whole. Finally, the participants were asked to plan and attempt to use one of the collocations that they had explored in their writing or speaking. In this activity, learners considered how they could appropriately use a collocation within the correct context.

But how do you complete an entry?

Don't panic! The following handout provides step-by-step instructions for your Journal Entries to explore more and more about English. Please understand that an example was provided on the handout AND another example will be completed during this class.

STEP 1 - Select the Word or Phrase
Start with a word or short phrase that you are interested in exploring! For example, let's look at the often-confusing preposition "by".

Hint: Where can you find words? Look in your corrected homework for words or phrases that you have trouble with. Phrases should not be greater than 3 words.

1. What WORD or PHRASE would you like to explore?
By
(Example for Class)

STEP 2 - Identify the Part of Speech
Let's start with the basics. Before analyzing the word or phrase, try to figure out the part of speech. This will help you understand the context(s) where you may find the word.

Hint: What do you know about the word or phrase that you are exploring? Can you tell me the part of speech? (Noun, verb, adjective, etc.) There may be more than 1, so underline the part of speech that you are most interested in. If you are having trouble finding the Part of Speech, look it up in a dictionary.

2. What is the PART OF SPEECH of a word?
Preposition & Adverb
(Example for Class)

STEP 3 - Select a Corpus & Identify the Register(s)
Now we have to choose where we get our information. It is very important that we choose from the right "body of information" for our interests. For the example of "by" we will look at its use in a Spoken, Academic context through the Michigan Corpus of Academic Spoken English (MICASE).

Hint: What context do we want to study? Are we more interested in the written or spoken contexts? Do we want formal contexts like Academic Journals or informal narrative contexts like works of fiction?

3. What CORPUS will you use?
Michigan Corpus of Academic Spoken English (MICASE)
(Example for Class

4. What is the REGISTER(S) of the WORD or PHRASE you are interested in?
Spoken English in classrooms
(Example for Class)

Figure 3. An excerpt from an EDJ handout

Corpus Tools. The EDJ was designed for several available online corpora (e.g., COCA, MICASE) to be explored systematically in a fashion that would benefit the learners. Before any corpus tools were introduced, a basic "crash course" considering the concepts behind corpora and DDL was conducted during the first week of instruction. Like the EDJ worksheets, different corpora and corpus tools were introduced during the eight-week course. All learners were explicitly taught corpus tools in class to train them on how to access and utilize corpora on their own.

Outcomes and Observations

Various data sources were compiled from Case Study 1: Teacher's notes and observations collected as the teacher demoed and taught corpus tools over eight weeks of instruction; audio-recorded interviews, administered at the first week, third week, and after the class concluded in the ninth week; copies or access to the EDJs that learners completed over the eight weeks; a simple pre-test to see which materials students accessed before learning about corpus, followed by a quick performance-based quiz and survey; and a post-test to see if there were any observable differences. Pre- and post-test data are not presented in this paper.

The teacher's first and primary challenge in the corpus instruction focused on the unique course dynamic of having multiple learner levels in the class. These levels provided noticeably different reactions to corpus tools and DDL as technical concepts to introduce in the classroom. For example, novice-level students were unable to understand the material in English, despite the efforts of the instructor (including attempts to deliver instructions in the students' first language). It took advanced-level students only about five to ten minutes to understand and use the "KWIC" ("Key Word in Context") feature in COCA, but it took approximately two class periods for novice-learners to use this feature independently.

In summary, whereas the intermediate and advanced-levels progressed with corpus tools and DDL through the EDJ relatively well, the novice-level learners needed much more time and practice. This observation is clearly not surprising, and as noted above, the teacher was able to explain the tools and fundamentals of DDL in the learners' L1. Corpus-based instruction with heterogeneous groups (and especially with novice learners) was very difficult to successfully conduct across the board, and may have to be reconsidered in this type of study abroad setting. The teacher noted and documented his various approaches for simplifying lessons on the use of corpus tools such as extended group work, a need for more detailed handouts and simplified instructions, and the use of teacher-mediated paper concordance lines (e.g., as Smart, 2014 advocated).

Despite this major challenge, feedback from all learners during the eight weeks of instruction was largely favorable to DDL and the EDJ worksheet. The teacher noted that, "at numerous points, learners form the novice and intermediate levels approached me before and after class to discuss how to use corpus tools for other classes." As far as lesson planning with the EDJ, all three levels started from a similar starting point, but soon established to their own pace and required different foci and mediation from the teacher. Advanced-level learners had their blogs working and were already experimenting on their own with COCA by the second week. The teacher demonstrated

how to input phrases into COCA to get more specific searches on longer n-grams (3-grams and 4-grams) and assigned homework.

The completed EDJs were analyzed in various ways, e.g., based on the quality of their content, detailed responses to questions, types of words/collocations explored, and so on. Table 2 shows the words and phrases that learners investigated in each level. Only two of the words included in the list were taken directly from advice given by the teacher as possibilities to explore in COCA: *take* and *ubiquitous*. The teacher noted that "recommending words for students in completing their EDJ prevents them from freely exploring and discovering patterns. It may seem like they are trying to appease me." This list provided a glimpse of the differences by examining what the learners investigated, and revealed more complex or technical vocabulary as the levels went up. Novice-level learners tended to explore common English verbs, while the intermediate and advanced levels were prepared to explore more complicated words or phrases like *make out* and *flabbergasted*.

Table 2. Summary of words explored in the EDJ per learner level

Novice-Level		Intermediate-Level	Advanced-Level
Take	Found	Get down	Pious
Need	To	Make out	Take it in
Been	Do	Bear up	Give forth
Put	Come	Lead	Kept in
To go	Keep	Bring	Set on
Use		Though	Grieved by
Began		Also	Subject to
Call		Up	Sit on the fence
Find		Figured	Flabbergasted
Remain		Give	Tangy
Attract		Had had	Apparel
Walk		Possessed	To shake off
All		Fierce	Casket
Due		Nodding	To sort out
Once lived		Draw	To bestow
To ask		Ubiquitous	Hedging
Wise		Go	Amid
Forward		Try	To tackle
Have		Done	Dreadful
Hide		Wondering	Enough
			To undertake

There were 66 words and phrases that were investigated by all 10 students over the eight weeks. This list excluded repetitions by students in the same level. For

example, 4 students out of 5 in the Novice-level investigated the word *take*. Interestingly, 75% of all EDJ explorations in all levels were tied to verbs either independently (e.g., *take*), in phrasal verbs (e.g., *shake off*), or as a part of a phraseological item (e.g., *sit on the fence*). This finding may suggest that learners struggled to develop an in-depth interest in the proper use of verbs at this point. One of the novice-level students commented that she was interested in the use of phrasal verbs, which she discovered by talking about her EDJs with other students and teachers during the course. Of the 25% of EDJs that were not tied to verbs, 14% were related to adverbs or prepositions, 8% were tied to adjectives, and 3% were tied to nouns.

Clearly, there was an observable pattern between learner levels and the number of phrases investigated in the EDJs. The novice-level had only one inquiry of a phrase (*once lived*), whereas the intermediate-level investigated four phrases and the advanced-level investigated nine phrases. The more proficient learners, as expected, explored longer phrases using the EDJs. This result, coupled with the increase in the worksheet's complexity, may actually support the potential continuing benefits of the EDJ approach as learners progress in their learning of English and acquisition of higher-level grammatical skills. In general, all three levels of learners from various interviews acknowledged that the EDJ was an excellent tool to use in tandem with class discussions and lessons, as well as opportunities to practice speaking and listening in English. All maintained an optimistic, if not eager, disposition to learn about corpus tools and DDL and to complete their EDJs outside of the classroom. The teacher observed that, "their attitude towards the worksheet may, in part, be attributed to that fact that lessons on the matter were spread to once a week, which reduced the sense of feeling overwhelmed by allowing time to process the information and homework tasks."

The EDJs provided a good resource for determining the participants' abilities and range of discoveries within DDL. As homework assignments, it was relatively easy for the teacher to determine who understood the material and who needed additional guidance. The value of properly scaffolding the EDJ is demonstrated by the fact that novice learners were able to complete the first level of the EDJ and understand the prominence of phraseological items, albeit, in some cases, with some struggle and additional trials or attempts. On the other hand, the intermediate and advanced levels were not only finding patterns, but were also already reaching conclusions as far as possible collocations, the appropriate contexts for the phraseological items, and shifts in meanings were concerned. The selection of a word or phrase and the depth to which participants examined these respective words or phrases were also very productive aspects to examine in the worksheet. There were encouraging signs that the DDL approach was encouraging advanced learners to conduct their own research with limited support from the teacher (see the Appendix for a sample worksheet from an advanced-level participant).

Finally, participants commented positively on their experience completing the EDJs and they also acknowledged the purpose of DDL and the EDJ as useful tools to help them explore the characteristics and nuances of English that they do not typically observe in traditional classroom settings. This sentiment recurred throughout the three interviews conducted during the course. Among the neutral or expected negative comments from participants, the novice-level learners commented that the activity was "too advanced" for them. Another student at the intermediate-level commented that he

believed "corpus tools could never replace dialogue and conversation." Another criticism at the intermediate and advanced levels was that the interface for corpus tools used (AntConc, COCA, and other online corpora) were not very user-friendly for students or those who are non-native speakers of English. They were able to understand the tools, especially with support from the teacher, by the end of the eight weeks, but they thought that a more simplified or translated version directly developed for learners of English would be more appropriate in this setting. Ultimately, overall outcomes and observations in Case Study 2 illustrate that the teacher and his learners saw the potential for corpus tools and DDL in the classroom and beyond, but they also saw various limitations and areas for potential improvement.

Conclusion and Recommendations

The instructors in Case Studies 1 and 2 both expressed that the possibilities of incorporating corpus tools in their own language classrooms are vast and very promising. For the right students with specific and clear learning goals, corpus tools can provide them with a practical resource to use relatively well and productively during and after the course has finished. An additional focus, then, on strategies and skills acquisition in corpus-based classrooms is needed. Classroom teachers should assess the needs of the students in a particular class in order to determine whether corpus-based instruction is appropriate and to what extent it will have to be incorporated in various activities and modified to accommodate different skill levels. Sufficient training for students to be familiar and at ease with data and tools is a must, and teachers need to prepare well-developed handouts and instructions accessible outside of the classroom. Students need to have a buy-in to the approach; this was achieved in general in the two case studies presented in this paper, but it took the two instructors' serious commitment, creativity, and dedication to stay the course. Keeping these factors in mind, there are a great deal of potential positive learning outcomes for students when exposed to corpora as a tool for improving writing/speaking skills and effective acquisition of English vocabulary and grammatical patterns.

Data-driven learning is not easy to teach in the language classroom. The lower the proficiency of learners, the greater the need—as well as the difficulty—for them to invest in learning about corpus tools and, consequently, the greater the challenge for the instructor. What the data, especially in Case Study 2, has shown, however, is that learners benefitted to some degree or another simply from exposure to corpora and related tools. It appeared that lower-level learners could grasp the clear message that learning to explore linguistic data will help them take charge of their acquisition of English, especially once they get to the next stage of the learning process. The amount of time, energy, and preparation maybe daunting for novice-level learners and their teachers in similar settings; nonetheless, with promising learner feedback and completed worksheets (i.e., EDJs), it may be well worth the effort.

The scholars' AntConc output in Case Study 1 and the completed EDJs in Case Study 2 all showed that learners in various levels were gaining an understanding of how English is interconnected and patterned across registers. Interestingly, the gains for the intermediate- and advanced-levels (Case Study 2) were reflected in how these learners were pushed to dig deeper and come to conclusions about patterns and features they

discovered themselves. Those who invested sufficient time and dedicated additional work in comparing and contrasting various distributions on their own produced outstanding worksheets. Scholars in Case Study 1 who were receptive to the research-based focus of AntConc and COCA commented positively on the new information they observed, also, on their own—information that they never would have received in traditional classroom instructional approaches. It was also encouraging to note that, one of the novice-level learners in Case Study 2 asked the instructor on the last day of the course, "Now that I can see […] pattern. What I do?" This question led to an informative conversation about using corpus tools as a means to begin investigating and exploring language. The instructor guided the student to personally answer the question *Now what?* or *So what?* that evolved into the student articulating the value of word collocations, shifts in meaning, and observable conventions or grammar rules in English (e.g., *move on* vs. *move in* and situations in which these phrases are used). The instructor noted that, "At this point – at the end of the day, it was the student who determines the effect and direct application of the EDJ." In summary, if students superficially participate in the EDJ or AntConc activities, these tools may lose their primary purpose. If students, however, are invested and willing to explore--their exploration perceptively monitored and facilitated, when needed by the instructor—they will discover a variety of different useful aspects of English, leading to the acquisition of greater relevant, autonomous skill in language learning and usage.

In Case Study 2, the use of the learners' L1 was critical to engage students in the processes necessary to learn how to use the tools. Arguably, it would be considered "common (pedagogical) sense" to not overwhelm international students in the U.S., with something as technical as DDL instruction. An aspect to improve the EDJs, however, is finding a means to tie in the learners' own findings with some practical application that extends beyond writing examples. This may be done if the teacher is able to take the EDJs and similar worksheets vis-à-vis lessons that are based on the exploratory habits and learning styles of students. In this case, teachers will be expected to become much more involved with what's going on and track how learners are directly interacting with data and applying patterns they are discovering to various language learning situations.

Recommendations

By the conclusion of the two case studies presented, the majority of the participants had positive impressions towards using corpora and corpus tools. However, there clearly were participants who did not find this approach beneficial to their own learning goals. This observation seems to summarize corpus-based instruction well: **It is not for everyone**. For scholars or students that had clear and immediate academic English-related goals or for the students already confident in the use of various types of software for research, understanding and applying the use of corpus tools came naturally and was *exciting*, *creative*, and *fun*. For others who did not have pressing English writing/speaking goals and preferred to spend their time in the U.S. practicing spoken English skills, the advantages of corpus use were a harder sell. This is important for instructors to keep in mind when planning to implement corpus instruction in the classroom across a range of learners. Bearing this in mind, there are a number of factors that instructors should consider before adding a corpus component to their curricula.

Ample Time. To effectively introduce students to corpus tools requires sufficient time for explanation, demonstration, and practice. For these two case studies, several hours of class time were committed exclusively for this purpose. Obviously, depending on the instructor's goals, different elements will require varying amounts of time. The most basic aspects of COCA could be introduced within a single class, but if the students are expected to compile their own corpora to be analyzed with a program like AntConc, more time will need to be allotted. The scholars in Case Study 1 were given two weeks to collect their research articles and then two full classes were spent practicing various analyses. Only after multiple opportunities to practice using AntConc did the scholars begin to realize its potential for their own writing and research.

Proper Explanation of Merits. Although, as noted, some students will intuitively discover the benefits and applications of this approach as they progress further into their learning, one of the most challenging initial responsibilities of the instructor to get student buy-in in this context is to properly and convincingly explain to students why these tools can be helpful. There is a time-commitment on their part to learning the process, so they need to understand why their time is being spent towards a potentially practical application. In these case studies, the participants practiced examining linguistic features of various writing samples by hand so that when they gained access to the concordancer, the gains would be more apparent and more easily achieved with these tools. The instructors had to spend a significant amount of time explaining why factors like frequency and authenticity of texts were valuable in learning English or specifically in writing or editing their own papers. Without the appropriate explanation of the tools and materials, it would be easy for students to feel resentful, bored, or overwhelmed.

Appropriate English Level. The students need to have a sufficiently strong foundation of English before setting out to analyze millions of words of text for specific linguistic features. Otherwise, they may not know what features to search for or how to interpret the results. For this reason, it is recommended that these types of courses be designed for at least intermediate students, but preferably more advanced learners. Several of the lower level students in the two case studies required extra time and attention from the instructor (and more advanced classmates) before they were able to comprehend the tools, materials, and objectives.

Relevant Student Goals and Access to Tools and Materials. Lastly, the students' desired outcomes in English should relate specifically to the instruction and learning contexts. The two classes here were particularly well suited for corpus instruction and both instructors were trained and prepared to address all types of course/learner needs. They also had easy access to computer labs, the internet, and related materials. Research articles were easily accessible and these immediately lent themselves to corpus analysis. A speaking or listening course could incorporate corpora, but may not be as suitable as a vocabulary/grammar and writing course. It would also be helpful if the students are at least minimally computer savvy. In Case Study 1, scholars with technology and computer science backgrounds were able to more quickly understand and use the concordancing program than some of the other students from different fields or those who were less proficient technically.

Authors' Note

The following ethical considerations were strictly observed in conducting the study. First, the two case study instructors volunteered to participate (as focal participants) and are both co-authors of the paper and related projects. Both are trained to conduct research following necessary ethical standards. They have provided permissions to use primary observational data, classroom materials, and journal entries for analyses. Second, actual learner data were not directly utilized in this study (as the focus was instructor experiences and perspectives); however, learners were provided with sufficient information pertaining to the goal of the study and the benefits they would reap from their participation. A larger study related to this paper will include university-approved review and duly-completed participant consent forms. All participants, study locations, and all other data sources have been kept confidential and will not be used for any other purposes.

References

Anthony, L. (2018). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Accessed July 9. http://www.laurenceanthony.net/

Biber, D., Reppen, R., & Friginal, E. (2010). Research in corpus linguistics. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics (2nd Ed.)* pp. 548-570. Oxford: Oxford University Press.

Boulton, A. (2009). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL, 21*(1), 37-54.

Bowker, L., & Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora*. New York: Routledge.

Cortes, V. (2011). Genre analysis in the academic writing class: With or without corpora?. *Quaderns de Filologia-Estudis Lingüístics, 16*, 65-80.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 520 million words, 1990-present*. Available online at https://corpus.byu.edu/coca/.

Engle, L., & Engle, J. (2003). Study abroad levels: Toward a classification of program types. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *9*(1), 1-20.

Firth, J. (1957). *Papers in linguistics*. Oxford, UK: Oxford University Press.

Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes, 24*(3), 321-332.

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

Friginal, E. (2013). Developing research report writing skills using corpora. *English for Specific Purposes, 32*(4), 208-220.

Friginal, E. (2018). *Corpus linguistics for English teachers: New tools, online resources, and classroom activities.* New York: Routledge.

Friginal, E., & Hardy, J.A. (2014). *Corpus-based sociolinguistics: A guide for students*. New York: Routledge.

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

Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes, 6*(4), 319-335.

Kaltenböck, G., & Mehlmauer-Larcher, B. (2005). Computer corpora and the language classroom: On the potential and limitations of computer corpora in language teaching. *ReCALL, 17*(1), 65-84.

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes, 25*(1), 56–75.

Meunier, F., & Reppen, R. (2015). Corpus vs. non-corpus informed pedagogical materials: Grammar as the focus. In D. Biber & R. Reppen (Eds.) *The Cambridge handbook of English corpus linguistics* (pp. 498 – 514). Cambridge: Cambridge University Press.

Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics, 7*, 140-162.

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

Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL, 26*(02), 184–201.

Teubert, W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics, 10*(1), 1-13.

Tribble, C. (2015). Teaching and language corpora: Perspectives from a personal journey. In Leńko-Szymańska, A. & Boulton, A. (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 45-67). Philadelphia, PA: John Benjamins North America.

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing, 13*, 257–283.

Dil Eğitiminde Derlem Temelli Yaklaşımlar: Çıktılar, Gözlemler ve Öğretmen Görüşleri

Öz

Bu makale, Amerika Birleşik Devletleri'nde iki üniversite ve bir yurtdışı yetişkin eğitimi programında bir sömestr süresince (yaklaşık 8 haftalık) verilen derlem temelli ve veriye dayalı İkinci Dil Olarak İngilizce eğitiminden elde edilen öğretmen kanaatleri, öğrenci çıktıları ve öğrenci izlenimlerinin analizleri ile materyal geliştirmeye dair fikirler ve araştırmacı gözlemlerinin sentezini sunmaktadır. 1. Vaka Çalışması, bir grup misafir Çinli akademisyene özel tasarlanmış akademik yazma becerilerini geliştirme dersi bağlamında derlem temelli eğitimin etkililiğine yönelik öğrenci ve öğretmen tutumlarını incelemiştir. 2. Vaka Çalışması ise, ABD'de bulunan kâr amacı gütmeyen özel bir eğitim kurumunda bir sömestr süresince verilen yurtdışı eğitim programı sırasında, derlemleri ve derlem araçlarını farklı düzeylerde kullanma becerilerine sahip öğrencilere yardımcı olmak üzere tasarlanan yönlendirici çalışma şablonu kullanımını karma yöntemlerle araştırmıştır. Bu şablon, derlem temelli dersler ile verilerin düzenli olarak sınıfta işlenen derslere veya ev ödevlerine dahil edilmesini sağlayacak biçimde tasarlanmıştır. Sonuçlar, derlem araçlarının dil sınıflarında sunabilecekleri faydalar hakkında öğretmenlerin heyecanlı ve büyük beklentileri olduğunu göstermektedir. Öğrenme hedefleri ve eğitimlerinin belirgin ve açık olması durumunda, derlem araçları uygun öğrenciler için ders sırasında ve sonrasında değerli bir kaynak olabilir.

Anahtar Kelimeler: Derlem dilbilim, öğretmen algıları, veri güdümlü öğrenme, sınıfiçi teknoloji

Appendix

Sample completed EDJ of *to undertake* from an advanced-level participant

Level 3 of the Explorer's Journal Entry

1.**DICTIONARY** – What **WORD** or **PHRASE** would you like to explore?

To Undertake

2. **DICTIONARY** – What is the **PART OF SPEECH** of a word?

Verb

3. **DICTIONARY** – What are the possible definitions of the **WORD** or **PHRASE** in the dictionary or Google?

To take a task for yourself; to make a promise; to guarantee something

4. **CORPUS** – Which **CORPUS** will you use?

AntConc – MICASE
Corpus – NOW

5. **CORPUS** – What is the **REGISTER(S) / CONTEXTS** of the **WORD** or **PHRASE** you are interested in? If possible, how many **HITS** or **TOKENS** of the **WORD** or **PHRASE** were in the corpus?

Register(s)	**Hits or Tokens**
NOW	
COCA	200+

6. **CORPUS** – List any **PATTERNS / BLOCKS / "FRIENDS OF THE WORD" / COLLOCATIONS** of the **WORD** or **PHRASE** that are common in the corpus. (5 minimum)

[RIGHT]		[LEFT]
Be/is/was/will/would		Any
Being	[BOTH]	Of
Had		[objective pronouns]
Have		By
To		For
		In
		The/this
		No

7. **CORPUS** – Choose 3 of the above (number 6) phrases that you find interesting. Using the same corpus, look up the three phases. *List any interesting findings* – does the meaning change? Is there a longer phrase? Does it only occur in a specific context?

Phrase 1 – w* undertak* [200]	**Phrase 2 – undertak* in [200]**	**Phrase 3 – undertak* for [200]**
1. x was undertaken by	**1.**undertaken in collaboration / conjunction	**1.** undertakings for
2. y will undertake a/an	**2.**undertaken in order to	**2.**undertaken for
3. written undertaking [as a noun]	**3.**undertaken in [dates or places]	**3.**

8.**CORPUS**–What are the meanings of the 3 original phrases from number 7?

1. The expression is used to say that someone is taking charge of/ conducting/doing something, or the opposite when in passive voice.	**2. it is used to give further explanations or purpose (dates/places or in order to), but keeps the same meaning as before.**	**2. When used as a verb, it means that something was put under someone authority for a certain purpose. When used as a noun, it refers to the actions itself that were performed by someone.**
9.**CORPUS**—What are the **LITERAL TRANSLATIONS** in your first language for the 3 original phrases from *step 7*? How are they similar? How are they different?		
1.tomar (para si) assumir(indo) algo / conduzir	**2.**medidas tomadas em conjunto / tomadas para…	**3.**medidas tomadas / garantias para (quando substantivo)
10.**CORPUS** — Open a *different* corpus using AntConc or an online concordancer. Return to the phrases you chose for *step 7* and look them up*. Are there any difference between the corpora? Why?*		
Phrase 1 — w* undertak* [0]	**Phrase 2 — undertak* in [0]**	**Phrase 3 — undertak* for [0]**
11. Write a sentence as an **EXAMPLE** with the **WORD** or **PHRASE** you explored.		
After the investigations, many actions will be undertaken by the Congress.		
12. Write a second **EXAMPLE** with a different collocation		
Mitigating procedures were undertaken in June by the company.		
13. What is your plan to try and use one of the phrases from *step 7* in your own language production (writing or conversation)?		
It seems like that this verb is basically used in the passive voice, which emphasizes the taken actions rather than who is taking the action. In other words, it's a good verb to emphasize that something is being done.		