TEACHING LINGUISTICS

Exploring the use of corpus tools for teaching language variation
to L2 Spanish majors

NAUSICA MARCOS MIGUEL

Denison University

Discussing language variation in introductory linguistics courses is a given in helping university students understand that speakers and communities use language differently. For Spanish majors/minors in the US, such content can be explored in linguistics courses taught in Spanish. However, little research has been done on pedagogy for teaching Hispanic linguistics in this context. Following an action-based design, this study contributes to research on linguistics pedagogy by exploring how students in four sections of an introductory Spanish linguistics course reacted not only to the topic of social and/or structural variation but also to the use of corpus tools in exploring this topic.*

Keywords: corpus tools, language awareness, linguistics pedagogy, materials research, Spanish L2, action research, language variation

1. INTRODUCTION. The teaching of variation is recommended in the Spanish second language (L2) classroom so that L2 learners recognize that a language encompasses more than one variety (e.g. del Valle 2014, Geeslin & Long 2014, Moreno Fernández 2007, Potowski 2003, 2005, Shin & Hudgens Henderson 2017). The study of variation may facilitate awareness of and appreciation for multiple varieties of the target language. This is a particularly important goal in the Spanish major, where native-English-speaking students and Spanish heritage speakers alike may have misconceptions about the value, quality, or correctness of nonstandard varieties of the target language (see Leeman & Serafini 2020). Moreover, the Modern Language Association’s report on foreign languages and higher education (MLA Ad Hoc Committee on Foreign Languages 2007), a core text for language departments, highlights the need to include linguistics content courses in modern language majors, as linguistics is considered to be fundamental knowledge for majors.

Linguistics courses that teach variation align with the five Cs of the World-Readiness Standards for Language Learning (The National Standards Collaborative Board 2015): communication, culture, comparisons, connections, and communities. These guidelines, promoted for K–16 language instruction by the American Council for the Teaching of Foreign Languages (ACTFL), aim to ‘guide learners to develop competence to communicate effectively and interact with cultural competence to participate in multilingual communities at home and around the world’ (p. 11).

The World-Readiness Standards emphasize that ‘successful communication’ is ‘knowing how (grammar), when (context) and why (purpose), to say what (vocabulary) to whom (audience)’ (The National Standards Collaborative Board 2015:11). In order to help learners communicate with different language communities, the standards aim at enhancing learners’ sociolinguistic knowledge. Teaching variation drives discussions about dialects, sociocultural varieties, historical varieties, and registers and is thus a good fit for teachers and language programs that incorporate the World-Readiness Standards into their instruction.

* I would like to thank my students for their participation in the study and for the feedback they provided. I am also thankful for the helpful comments of the referees and the associate editors, Dr. Hiramatsu and Dr. Temkin Martinez. All remaining errors are mine.

Printed with the permission of Nausica Marcos Miguel. © 2022.
At most US tertiary institutions, the study of variation could be incorporated in language majors either in the language courses taught during the first two years or in the content courses (usually culture, literature, and linguistics courses) taught during the later years. The latter approach was taken in this study, with participants being introduced to language variation in an upper-level linguistics-content course. Corpus-based activities in this course encouraged students to analyze language by inferring a rule or patterns of language use or by localizing the context of an expression. That is, students engaged in exploration of a language variety by locating examples in corpus data (e.g. Bennett 2010, Reppen 2010).

This empirical study examines learners’ perceptions of the effectiveness of corpus tools in helping them learn about variation within an introductory linguistics course taken during the third year or third-year equivalent of a Spanish major at a liberal arts college (LAC) in the US. The study contributes to our understanding of effective pedagogy in teaching language variation generally and Hispanic linguistics specifically, thereby helping to fill gaps in a research field that is still limited (see Knouse et al. 2015). It also explores the pedagogical possibilities of corpus linguistics methodologies, which are valuable tools for exploring variation in Spanish (e.g. Davies 2000, 2004a,b).

Before we proceed further, a definition of linguistic variation is needed. This definition is necessarily broad since it can refer not only to different language varieties but also to varied patterns of language use within a variety. Variation refers both to the choices made by a speaker in different social situations and to the choices made by different communities of speakers. It is often difficult or impossible to separate structural from social variation, as language choices tend to be influenced by social factors. Therefore, the term ‘variation’ necessarily deals with lexical, morphosyntactic, and phonological phenomena across different contextual, social, and geographical levels. In the classroom, comparisons of structures that express impersonal values like construyes ‘you build (generic)’, construyen ‘they build’, uno construye ‘one builds’, and se construyó and fue construido ‘was built’ can be productive learning exercises (see Lunn & DeCesaris 2007). Pronominal forms are also very useful structures to analyze (see Potowski & Shin 2018).

This study on student perceptions of the use of corpus tools to learn about language variation draws on scholarship that addresses the following topics: (i) pedagogical practices in higher education and in language majors, (ii) implementing the study of language variation in modern languages programs, and (iii) corpus-based approaches to teaching L2s and linguistics. Literatures relevant to each of these topics are discussed next.

2. Literature review.

2.1. Pedagogical practices in higher education and language majors. In order to situate linguistics instruction at the university level, it is necessary to contextualize the teaching of the discipline within general pedagogical theories, making it relevant for linguists in different departments. Moreover, this study also needs to be situated within the field of L2 content and language instruction, which is particularly relevant for linguists working in language departments.

Because of the positive effect of student engagement in learning (Freeman et al. 2014), active learning methodologies have been heavily promoted across disciplines in the last decades (e.g. McCarthy & Anderson 2000 for history and political sciences, Michael 2006 for physiology, and Prince 2004 for engineering). These pedagogical practices situate the learner’s activity as fundamental in developing, testing, and refin-
ing the mental models needed to understand a topic (Michael 2006). Although a variety of activities can be found under the active learning umbrella, these activities share an important element: namely, they are carried out in the classroom in collaboration with others (Prince 2004). In this student-centered approach, instructors design activities that promote learners’ discovery, allowing ‘students [to do] (at their level, of course) what biologists, engineers, philosophers, political scientists, and sociologists do’ (Weimer 2013:77). Learning takes place while students engage in these activities.

For L2 learning, communicative approaches to language teaching also promote the idea that having students complete meaningful tasks in interaction with others will contribute to their language learning (Richards & Rodgers 2001). That is, active learning pedagogies are also beneficial for language learning. However, research on tasks tends to focus not on whether learners acquire the content discussed during the completion of the task, but rather on whether their language develops because of those interactions.

For all faculty working in a language department, there is always a struggle when teaching in the learners’ nondominant language: learning a new topic in an L2 requires that the student understand the L2 (e.g. Grabe & Stoller 1997, Lyster 2007). For example, reading new content in the L2 can be less conducive to acquisition of content than reading in the L1, according to the cognitive-load hypothesis (see Roussel et al. 2017). This hypothesis suggests that both second language learning and content learning use the same cognitive resources, with detrimental effects for learners. This implies that when the subject matter has not been studied in the L1, which tends to be the case with linguistics for Spanish majors, the learning process will be more challenging.

Macaro’s (2018) comprehensive review of content courses taught in English L2 in non-English-speaking countries indicates that very little research has been undertaken in higher education to analyze the cost-benefit ratio of learning content through the L2. He concludes that current research does not allow for generalizations, as pedagogical practices vary significantly in courses that must address both content learning and L2 proficiency.

Because of the duality of learning goals in the language major, both proficiency level and prior content knowledge play important roles in ensuring student success. Klee and Tedick’s (1997) study on a postsecondary Spanish program in the US suggests that students need at least an intermediate-high proficiency level in order to follow content classes in the L2. Although this study had a small sample size, it drew data from a variety of sources, including diaries, interviews, focus groups, and pre and post language tests. The researchers recommend including texts in the English L1 and offering additional language courses or language-learning support to reduce learners’ frustration and facilitate the learning of both content and language. These recommendations have been echoed by other researchers (e.g. Dupuy 2000, Grabe & Stoller 1997, Tedick & Lyster 2020).

In brief, teaching linguistics in a language major presents specific challenges that instructors teaching through the students’ dominant language do not face. Perspectives that recognize the importance of social interaction for language and content learning suggest that by making use of the multiple affordances of the classroom environment (physical and virtual), both content and language can be successfully learned (e.g. Thoms 2014). Active learning theories are therefore relevant both for linguists teaching in the students’ L1 and for those teaching in the students’ nondominant languages.

2.2. IMPLEMENTING THE STUDY OF LANGUAGE VARIATION IN MODERN LANGUAGES PROGRAMS. Undoubtedly, introducing linguistics in modern languages programs is not without challenges. Students probably have not been exposed to linguistics as a field of study in primary and secondary education, or they might have been exposed from a pre-
scriptive perspective but not from an exploratory perspective (e.g. Denham 2015). Heritage language learners may have had even less exposure than L2 learners to linguistics terminology and explicit grammar instruction (e.g. Shin & Hudgens Henderson 2017). This lack of exposure may make linguistics topics harder to understand, if students have never reflected on language as an object of study (see Correa 2014, Shin & Hudgens Henderson 2017).

Compounding the challenges of introducing an unfamiliar topic in students’ L2 is the question of when to introduce the topic of language variation. While upper-level content courses may seem like the most natural place to incorporate a linguistics topic, doing so in first- and second-semester courses could be the best way to introduce this topic to a large number of students. These courses may be students’ only exposure to a language other than English, since they may be taking a language simply to fulfill a college-wide language requirement.1

First- and second-semester courses can easily include the following sociolinguistics topics: forms of address, turn-taking, politeness norms, idioms, and different registers and dialects (Council of Europe 2002). Personal pronouns in the Spanish-speaking world are also a tremendous source of variation. For example, following the World-Readiness Standards, Shenk (2014) proposes a series of activities for working on recognition of vos forms at the intermediate level, given that this topic is minimally addressed in textbooks. Activities such as this, which help students learn about variation across language varieties, can easily be included in both lower- and upper-level courses within the Spanish major. Yet these activities can also help students learn about variation within a single variety. Two other easily accessible examples of such activities are the Open Educational Resources (OER) ‘Spanish in Texas’ project (Bullock & Toribio n.d.) and the sociolinguistics activities designed by Holguín Mendoza et al. (2017) to target heritage language learners. A combination of activities to study variation across varieties of language and variation within a single variety of language should inform Spanish L2 instruction.

Textbooks may also offer a way to help students explore both of these aspects of variation. Although a neglect of language variation in Spanish L2 textbooks might make learners believe that the standard variety is the only type of Spanish that exists (see Burns 2018, Fernández Martín 2014), textbooks designed for use in Hispanic linguistics courses can correct this misconception. In advanced grammar courses, introducing sociolinguistics can make learners more sociolinguistically informed while improving their grammatical knowledge (Shin & Hudgens Henderson 2017). Audiovisual OERs designed to complement, and even substitute for, textbooks focus to a great extent on language variation. Examples include Acceso (Open Language Resource Center 2019), Voices of the Hispanic world (Morgan n.d.), and Catálogo de voces Hispánicas (Instituto Cervantes n.d.).

Regarding classroom practices, various descriptions of how to teach sociolinguistics for Spanish learners can be found (e.g. Checa-García 2015, Leeman & Serafini 2016). Furthermore, Knouse et al. (2015) proposed several strategies to teach linguistics using a student-centered approach, such as collaborative learning activities, online discussion boards, reflective journals, and mini-research projects. Student research and service-learning projects can have a positive impact on learners in linguistics courses (see Knouse 2018).

In brief, although there seems to be a lack of research on pedagogical practices for introducing social and structural variation (see Knouse et al. 2015), scholars agree that

---

1 In 2009–2010, 50.7% of US schools had a college-wide language requirement (Lusin 2012).
these topics can and should be discussed at any level. In linguistics courses, learners’ previous knowledge of pedagogical grammar, that is, the explicit, simplified rules that are taught in the L2 classroom, can be expanded into a descriptive grammar, a description of the way speakers use language structures that is gleaned from different varieties of the target language. Learners can then see more clearly the commonalities across and within varieties and become more aware of what a language system is, how it works, and how social and contextual factors influence it. Furthermore, learners can reflect on the role of prescriptive grammar, or the explicit rules about how learners should use the language in written and formal spoken registers. In Spanish L2 textbooks, prescriptive grammar tends to be identified with the Real Academia de la Lengua Española (RAE), the Royal Academy of the Spanish Language. Reflections on pedagogical, prescriptive, and descriptive grammars can help learners develop critical thinking beyond the discipline, a skill that is fundamental in higher education.

2.3. Corpus-based approaches to teaching. Corpus linguistics has contributed to the development of L2 materials that reflect naturally occurring language (see Flowerdew 2012, O’Keeffe et al. 2007, Römer 2010). Furthermore, the use of corpora serves to raise learners’ language awareness, which is an important factor for language learning (see Svalberg 2007, 2012). One of the main methodologies encouraging the use of corpus tools in this context has been data-driven learning (DDL; Boulton 2011, Johns 1991). Similar approaches can also be part of linguistics classes (e.g. Contreras Izquierdo 2020, Davies 2000, 2004a,b).

Corpus-based approaches to L2 teaching. Following a DDL approach, a learner should be able to deduce patterns of use from the concordances of a corpus, that is, lists of examples from the corpus showing where a searched term appears in context. As Boulton and Cobb (2017:350) put it, ‘DDL reflects current linguistic theory’, showing language as ‘dynamic, complex, probabilistic and interactive’. Moreover, DDL offers possibilities to include focus-on-form instruction in the classroom, focusing on language patterns instead of utilizing rules.

DDL is based on the metaphor of the learner as a researcher (Johns 1991), which relates language learning with the work that linguists do. Thus, DDL clearly connects with active learning practices as DDL highlights the role of the student in their own learning (see Vyatkina 2013, Weimer 2013). Through their hands-on research, learners are exposed to tools used by linguists, which can facilitate learning about the discipline while learning language (see Boulton & Cobb 2017).

Work with DDL started in the 1990s with Johns 1991 and is still blooming, but it is not a mainstream L2 teaching practice. DDL implementation is not without challenges, since it requires basic knowledge of the field of corpus linguistics (see Godwin-Jones 2017, Meunier 2019). Despite this challenge, Boulton and Cobb’s (2017) metastudy and Boulton’s (2017) research timeline of the use of corpora in language teaching and learning clearly show that corpora help in learning grammar and serve as reference tools.

Overall, research on learners’ reactions has revealed some challenges when learners work with corpus tools because corpora and their interfaces might not be user-friendly (Boulton & Cobb 2017, Godwin-Jones 2017, Pérez-Paredes et al. 2011). To alleviate this problem, it has been proposed that instructors develop their own materials and pre-select concordances for their students. However, this approach may reduce the effectiveness of working with corpus tools. Although Boulton and Cobb (2017) found good results in studies utilizing teacher-made materials, the most positive results were found in those studies that required learners to use the corpus tools by themselves.
The challenges learners face in working with corpus tools may be compounded by teachers’ own need for training in their use (e.g. Callies 2016, Frankenberg-Garcia 2012, Leńko-Szymańska 2017, Mukherjee 2006). There are a number of resources directed toward training L2 teachers. For example, the CALPER Corpus Portal (Center for Advanced Language Proficiency Education and Research n.d.) introduces teachers to corpus linguistics and offers them tools and reports on how other teachers have incorporated this methodology, and also models sample activities. Bennett (2010) provides sample lessons on using concordances to search for linking words in academic speaking and to look at differences in definite and indefinite articles. Reppen (2010) also presents numerous pedagogical ideas. For example, students can create their own frequency lists from assigned readings and analyze which words are more specific to a genre such as a newspaper. Lamy and Mortensen (2012) also offer some model activities for French, German, and English L2 that target lexical, grammatical, cultural, and register elements. These activities can be easily adapted for a Spanish L2 language course or for Hispanic linguistics courses.

In the case of Spanish L2, two recent articles relevant to teaching an advanced Spanish grammar class utilizing corpus tools are Paz 2013 and Benavides 2015. Paz 2013 uses examples from four semesters of individual learners’ assignments that show students’ abilities to infer rules from concordances, as well as a selection of learner comments on the discussion board of their institution’s learning management system (LMS). In addition to using the internet as a corpus, learners in all four iterations of this study used the Corpus Diacrónico del Español (CORDE), Corpus de Referencia del Español Actual (CREA), Corpus del Español, Base de Datos Sintácticos del Español Actual, and Archivo de Textos Hispánicos de la Universidad de Santiago. The article is largely a reflection on the structure and effectiveness of the course for other instructors who are interested in using corpus linguistics methodologies. For this reason, it includes no information on the number of students taking the courses, the number of discussion boards analyzed, the number of assignments analyzed, and/or methods used for analyzing artifacts.

Benavides 2015 examined six intermediate to advanced L2 learners and three native speakers in a similar setting who completed a semester’s worth of deductive and inductive homework assignments, seeking either to describe language-use rules based on the concordances in the Corpus del Español (Davies 2002) or to confirm whether textbook rules were consistent with the concordances. At the end of the semester, the students completed a questionnaire, which included both closed questions using a Likert scale and open-ended questions, that gauged their reactions to learning with the corpus. In addition, their homework assignments were collected. Student comments on the questionnaire and the assignments were analyzed using four themes: ‘perceived impact on students’ learning’, ‘perceived helpfulness of the corpus as a learning tool’, ‘incidental learning’, and ‘challenges’. The questionnaire revealed mixed views on using the corpus. In some instances, students found using the corpus challenging and thought it did not contribute to their learning, while other students found the corpus helped them ‘to “prove” the rules of grammar’ offered in the textbook (Benavides 2015:227), and some were interested in using the tools beyond the class. With their detailed comments in the homework assignments, learners showed that they were aware of the way context prompted the use of one form or another. This type of awareness indicates that students recognized language variation. Based on data collected both from the students’ assignments and from their responses to the open-ended items on the questionnaire, Benavides concludes that the use of the corpus did promote student learning.
Both Paz (2013) and Benavides (2015) reported that learners encountered difficulties with the technology, but both authors drew positive conclusions about teaching and learning with corpus tools. These small-scale case studies explore the use of corpus tools in a language course, however, not in a linguistics course. But there is clear overlap between topics and reflections on using corpus approaches in language courses and in linguistics courses, especially when the linguistics courses are not taught in the learners’ dominant language. Because of this overlap, studies analyzing corpus tools in language courses are relevant to linguistics courses.

**Corpus-based approaches to teaching linguistics.** Studies of DDL rarely focus on linguistics courses. However, Contreras Izquierdo (2020) utilized the bases of DDL and active learning theories to design two courses (undergraduate and postgraduate, respectively) for Spanish L1 speakers in a Spanish university to explore language variation using corpora and other online resources. Contreras Izquierdo shows some examples of student work and tallies the resources learners used to explore language varieties. Using the average score of the final paper in the undergraduate class (four iterations, 121 students) as a proxy of learning, as well as the averages of end-of-semester student evaluations on teaching effectiveness, the author concludes that this active approach to teaching variation was successful. However, he points out that students preferred other online resources to the corpora when completing the assignments. As the data collected does not include the students’ perspectives on the use of the online tools, the study does not give their insights into their learning process. But this study is an illustrative case of different ways to approach teaching variation using several online resources beyond corpora, and it reminds us of the importance of specific training in the use of corpus tools if we want students to take advantage of them, even if the class is in their dominant language.

In these L1 linguistics courses, corpora were clearly used as a reference tool. Similarly, in L2 linguistics courses, much like in translation courses (e.g. Frankenberg-García 2005, 2015), the use of corpora as a reference tool, rather than an L2 learning tool, is highlighted. But the focus on the reference function does not preclude in any way the language-learning function.

For example, Davies 2000, 2004a,b are good examples of the use of corpora in Hispanic linguistics courses. Davies 2000 discusses the role of the *Corpus del Español* for historical linguistics courses and syntax courses. More specifically, in the context of an online Spanish historical linguistics course, Davies 2004a notes that the *Corpus del Español* can be used to help students emulate the work of historical linguists: that is, to ‘1) investigate the relationship between different stages of the language, and 2) accurately model several different types of linguistic change in Spanish’ (2004a:210). Both morphological and syntactic shifts were explored in the corpus. Davies 2004b reports on a different online course with similar learning goals, showing that advanced students used the corpus to study synchronic syntactic variation. Thus in both courses, the *Corpus del Español* and other corpora complemented the textbook by allowing students to search for morphological and syntactic patterns and, consequently, to ‘test the textbook rules’ (Davies 2004a:208). In this way, the corpora could serve not just as reference tools for these advanced learners but also as tools for enhancing their knowledge of and proficiency in the L2. Davies’s articles present the instructor’s perspective in depth, including his pedagogical goals and the language structures and kinds of activities that he chose to include in the courses. These articles do not, however, offer insight into learners’ views of the effectiveness of corpus tools for language learning.

Literature on pedagogy for digital humanities can also inform the use of digital tools in conducting linguistic analysis. Voyant tools (Sinclair & Rockwell 2016) is a collec-
tion of corpus tools often mentioned in the pedagogical literature for digital humanities (see Cro 2020, Varner 2016). Focusing specifically on language majors, Cro (2020) recommends attending to students’ digital proficiency as well as their language proficiency when designing digital activities. In brief, corpus tools are becoming more frequent in programs for language majors, both in language courses and in content courses addressing linguistics and literature/culture topics.

This literature review has elaborated on the need to use active pedagogies when teaching linguistics, and has shown some of the challenges for linguists teaching in language departments. Moreover, it has discussed why learning about linguistic variation is especially relevant for modern language majors and how this topic can be taught using corpus tools. While interventions using DDL methodologies support the idea that working with corpus tools can contribute to language learning, research on student perceptions of the use of corpora and corpus tools in linguistics and other content courses is still lacking.

3. Research questions. The study presented here explores learners’ views on the effectiveness of corpora and corpus tools in helping them learn about variation: that is, learning about (i) the structure of language and the choices speakers make within a variety of the target language (i.e. diastratic and diaphasic variation), and (ii) the choices that speakers make across varieties of language (i.e. geographic variation). Data collected in four iterations of the linguistics course under examination provides information to answer the following questions.

1. What are learners’ reactions to learning about linguistic variation in a Spanish L2 linguistics course?
2. How do learners perceive the effectiveness of corpora in learning about variation?

The data presented here is part of a larger data-collection effort that analyzed learners’ interactions with corpus-based tasks (see Marcos Miguel 2021).

4. Methodology. This study incorporates principles of action research: each iteration of the project served as a basis for the following iteration. Thus, the study procedure follows the steps of planning, collecting data, analyzing/reflecting, hypothesizing/speculating, intervening, collecting data, analyzing/reflecting, hypothesizing/speculating, and writing (see Burns 2005). This cycle will not end with the writing of this article, however, as the course is taught regularly. Figure 1 illustrates the cycle that was repeated four times in this study.
Using the data collected, this article focuses on student reactions to learning about variation and to using corpus tools in doing so. Similar to research on teacher cognition, which oftentimes shows discrepancies between what teachers believe and what they do in the classroom (see Borg 2006), students’ perceptions cannot be equated with what they do in the classroom or with what they learn. Nevertheless, learner perceptions offer a glimpse into student motivation. This motivation ‘influences the direction, intensity, persistence, and quality of the learning behaviors in which students engage’ (Ambrose et al. 2010:68–69).

4.1. Classroom context. This research was conducted in four sections of a third-year college Spanish course, with sixteen, fifteen, fifteen, and six learners, respectively. The sections were taught approximately one year apart at an LAC in the US. In the first iteration, the course was called ‘Grammar in Context’. The subtitle ‘An Introduction to Linguistic Analysis’ was added in the second iteration. In this course, students adopted a linguistic perspective to analyze grammar concepts learned during their language courses, moving from an understanding of pedagogical grammar to an understanding of descriptive grammar.

In the syllabus of the second iteration, the following description of the course was given in Spanish.

In this class, we will carry out linguistic analysis of oral and written texts. We will follow four main approaches:

1. corpus linguistics,
2. linguistic landscape,
3. variation analysis and
4. syntax analysis.

In sum, we will discuss topics related to linguistics and understand that grammar is a polysemic word. This class will complement other courses in the major/minor by focusing on language as the object of study.

The course goals were stated in the syllabus. They are copied below from the second iteration of the syllabus; the original was in Spanish.

1) You will learn the differences between prescriptive, pedagogical, and descriptive grammars.
2) You will learn basic terms in the fields of grammar and linguistics.
3) You will learn about corpus tools so that you can analyze a corpus and even create your own corpus.
4) You will be able to recognize social and geographical factors that affect grammar and other linguistic elements.
5) You will be able to illustrate linguistic phenomena with corpus data.
6) You will also work on your written and oral production. Spanish will be the language of communication and the object of study in this class.

The syllabus and calendar for the fourth iteration of the course are included as supplementary materials A and B. Classroom activities aimed to raise language awareness through ongoing investigation of language, encouraging students to talk analytically about language and involving ‘learners in exploration and discovery’ (Svalberg 2007: 291, based on Borg 1994) (see Marcos Miguel 2020:38–41, 2021:227–28, and see supplementary material C for sample lessons).

In addition to helping students raise their language awareness, a main goal of the course was to normalize technology (Bax 2003, 2011), seamlessly integrating corpus-based activities into the classroom so that they became ‘invisible, embedded in everyday practice’ (Bax 2003:23). As Pérez-Paredes (2019) has shown, corpus technology is not yet normalized in most L2 classrooms.

2 This study underwent IRB review at Denison University.
3 All supplementary materials can be accessed at http://muse.jhu.edu/resolve/151.
In the first iteration, the textbook (Lunn & DeCesaris 2007) served as the basis for building the course. The primary variation topics were aspect, ser/estar/haber, mood, voice, and lexical variation. Class discussion and activities focused on grammatical variation both within and across varieties. For example, whereas lexical variation offers an opportunity to analyze variation across communities of speakers and genres, voice variation (e.g. the use of passive versus active voice) can be explored within different genres used by the same community of speakers. The *Corpus del Español* (Davies 2002) was used as a main resource in the class; however, other corpora were presented to the learners, such as the *Corpus de Referencia del Español Actual* (CREA) (Real Academia Española n.d.).

In the second iteration, the corpus tool AntConc (Anthony 2017) was used with the *Corpus of Oral de Referencia del Español Contemporáneo* (CORLEC) (Marcos Marín 1992). Although every subsequent iteration of the course was modified based on student feedback and the instructor’s perceptions, the *Corpus del Español* was utilized in all iterations.

### 4.2. Participants.

Most of the learners were in their third year or later of university study and were expected to be at an intermediate-high level of Spanish proficiency (American Council on the Teaching of Foreign Languages 2012). However, as is typical in post-secondary institutions (see Tschirner 2016), there is a range of proficiency levels from intermediate-low to advanced-mid in the third year of study. Study participants also included heritage speakers. In general, most of the participants had taken over four years of Spanish prior to this course (see Table 1). Even though years of study are not equal to proficiency, this information helps establish a linguistic profile of the learners. Note that a total of six students were enrolled in the fourth iteration, but only four of them completed the questionnaire.

<table>
<thead>
<tr>
<th></th>
<th>1st iteration</th>
<th>2nd iteration</th>
<th>3rd iteration</th>
<th>4th iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 16)</td>
<td>(n = 15)</td>
<td>(n = 15)</td>
<td>(n = 4/6)</td>
</tr>
<tr>
<td>Years of Spanish before university</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 and over</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>Less than 4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Heritage speakers</td>
<td>1≤</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Year at the university</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th year</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3rd year</td>
<td>7</td>
<td>5</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>2nd year</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

*Table 1. Student profile.*

*Although there were two heritage speakers in the class, only one reported this status in the questionnaire: one of them reported years of formal instruction, while the other focused on naturalistic exposure.*

The classroom was the main source of exposure to Spanish for most of the learners. Given the nature of the program, some students might have been simultaneously enrolled in other courses for the Spanish major.

### 4.3. Classroom instruction.

In all iterations, the class met three times per week for fifty minutes each session. Corpus activities were carried out in a computer laboratory in groups or pairs. The instructor would give general instructions about the activity, projecting a handout on the screen and modeling a sample search. Learners had access to a digital copy of this handout in their institution’s LMS, and usually received a hard copy of the handout in class as well. The instructor was available to help students complete the activities. Most of the assignments required learners to compare structures within a language variety or across different varieties. If students did not finish the assignments...
in class, they were given additional time to complete them on their own. Students could use the lab, which was available after hours, or work on their own computers. The assignments were graded for completion.

4.4. Instruments.

Questionnaire. Participants took an anonymous online questionnaire during the last week of the semester. A total of fifty questionnaires were collected across all four iterations. The questions explored students’ reactions to learning about variation and using corpus tools toward that end. The second, third, and fourth iterations included two additional questions related to learning about variation, which were added based on the questionnaire and interview results of the first iteration. The questions were written in English (see Appendix A).

Interviews. Two learners from the first iteration were interviewed at the beginning of the following semester. The number of interviews was determined by the number of learners who expressed an interest in participating and who were available at the time of the interviews. In the interviews, which lasted around half an hour each, a set of main interview questions (see Appendix B) and the questionnaire results were discussed. The interviews served to triangulate the data from the questionnaire as they provided more in-depth discussion of some points raised by the other learners. The interviews were transcribed verbatim. Because of time constraints, interviews were carried out only after the first iteration.

4.5. Data analysis. Data from the questionnaires (see Appendix A) was categorized according to the two main research questions that this study addresses: (i) What are learners’ reactions to learning about linguistic variation? (see Table 5 below for the list of questions analyzed), and (ii) How do learners perceive the effectiveness of corpora in learning about variation? (see Table 6 for the list of questions analyzed). For the open-ended questions, a content-based approach to analyzing the data was followed. That is, the data was read several times to create codes, from which patterns were identified (see Hsieh & Shannon 2005, Saldaña 2016). These codes took into consideration prior research, since some themes mentioned by the participants were already present in existing literature on the use of corpora in L2 courses. The questionnaire data was color-coded in an Excel sheet.

The questionnaires from the first two iterations and the interviews were first analyzed together, and similarities and differences between the iterations were identified. Next, data from the third and fourth iterations was added to the analysis using the initial coding scheme. The differences noted in this second analysis with a larger data set prompted a revision to the initial coding scheme, which is presented in the results section.

5. Results. This results section contains three subsections. First, a qualitative description of the codes and categories is given (§5.1). The quantitative summary (§5.2) tallies the codes, highlighting the most frequent and quantifying the answers to the yes/no questions. Given that there were only two student interviews, interview data is not tallied and is discussed separately from the questionnaires (§5.3). Due to the small sample sizes, no statistical analyses were carried out.

Student comments are labeled with the letter C, for example, ‘C1’. This label is based on the order in which the comments appeared in the manuscript. Students are given a unique ID, for example, ‘S1’, and are also identified by the iteration in which they participated, namely 1st, 2nd, 3rd, or 4th. Thus, the code ‘S2, C1, 1st’ must be read as stu-
dent 2, comment 1, 1st iteration. Additional references to any comment in the article are
given using the comment label, for example, ‘C1’.

Although the four iterations are discussed together for this analysis, the question-
naires (and the interviews in the first iteration) were analyzed after every semester in
order to assess what changes were necessary for the following iteration.

5.1. Codes and categories utilized in the questionnaire. In this section, a rep-
resentative example from each iteration is given verbatim for each code. If an iteration
does not appear among the examples, it means no participant in that iteration produced a
comment related to that code and/or the question was not included in the questionnaire.

Reactions to learning about variation. Table 2 describes the main content the
students perceived themselves to have learned about variation during the semester
(question 2 in Table 5 below). Establishing students’ perceived learning enables us to
draw conclusions about their reactions to learning about content. Although the third and
fourth questions in this category were open-ended (see Table 5), responses to these
questions were categorized as positive (yes) or negative (no). This binary coding was
chosen since the answers did not provide much nuanced information beyond a positive
or negative reaction and/or provided responses similar to those in Table 2.

<table>
<thead>
<tr>
<th>Codes</th>
<th>Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Formal levels of structural variation</td>
<td>there is a lot of it, lexically and morphologically (S2, C1, 1st)</td>
</tr>
<tr>
<td></td>
<td>I learned about how language varies in terms of sintaxis, phonetics, and</td>
</tr>
<tr>
<td></td>
<td>morphology (S18, C2, 2nd)</td>
</tr>
<tr>
<td></td>
<td>Language variation consists of lexicon, morphology and syntax. (S33, C3,</td>
</tr>
<tr>
<td></td>
<td>3rd)</td>
</tr>
<tr>
<td>2. Social and geographical aspects</td>
<td>every region of every Spanish-speaking country has its own flourishes to</td>
</tr>
<tr>
<td></td>
<td>the language that may change how words are pronounced, what words</td>
</tr>
<tr>
<td></td>
<td>are used, and even the meaning of the words themselves (S5, C4, 1st)</td>
</tr>
<tr>
<td></td>
<td>… variation is pivotal to understanding how people in different regions and</td>
</tr>
<tr>
<td></td>
<td>contexts speak (S27, C5, 2nd)</td>
</tr>
<tr>
<td></td>
<td>That language really depends on geographical, social, political context and</td>
</tr>
<tr>
<td></td>
<td>is extremely variant, even within a specific country or region. It can also</td>
</tr>
<tr>
<td></td>
<td>vary in how it used—through media, speaking, etc. (S42, C6, 3rd)</td>
</tr>
<tr>
<td></td>
<td>I have learned that there are a lot of factors that lead to the variation of</td>
</tr>
<tr>
<td></td>
<td>language which I actually think was very very interesting. I used to think</td>
</tr>
<tr>
<td></td>
<td>that it was geographic location that led to small differences but now I</td>
</tr>
<tr>
<td></td>
<td>know that era, pronunciation, community, outside influences, etc. can also</td>
</tr>
<tr>
<td></td>
<td>contribute just as much to these differences within the Spanish lan-</td>
</tr>
<tr>
<td></td>
<td>guage. (S49, C7, 4th)</td>
</tr>
<tr>
<td>3. Rule/patterns</td>
<td>The main points of variation that I learned this semester are the differ-</td>
</tr>
<tr>
<td></td>
<td>ences between prescriptive rules of language versus how real people use the</td>
</tr>
<tr>
<td></td>
<td>language and their variations. (S11, C8, 1st)</td>
</tr>
<tr>
<td></td>
<td>I learned about the specifics behind many grammatical rules and how these</td>
</tr>
<tr>
<td></td>
<td>rules are formed/applied to different populations (S17, C9, 2nd)</td>
</tr>
<tr>
<td>4. A general increase in knowledge</td>
<td>I feel like I learned a lot. (S16, C10, 1st)</td>
</tr>
<tr>
<td></td>
<td>That is so much more dense and interesting than I originally thought. (S21,</td>
</tr>
<tr>
<td></td>
<td>C11, 2nd)</td>
</tr>
<tr>
<td></td>
<td>What factors influence a language. (S36, C12, 3rd)</td>
</tr>
</tbody>
</table>

Table 2. What are the main points you have learned about variation this semester? (1st–4th iterations).

For the third question in Table 5, ‘How do you think doing research on variation has
helped you (or not) to understand the field of linguistics?’, most of the answers referred
to the content of the class and/or specific methodologies used in the class, already seen in Table 2, such as the following:

- doing research on variation has helped me to understand what people mean when they say they are learning about language. After doing this research, I understand what it means to analyze a corpus, find patterns, and come to conclusions about linguistic variation. (S18, C13, 2nd)

- I think that doing research on variation has helped me understand the field of linguistics by teaching me important factors that determine a language. (S33, C14, 3rd)

- I used to think that it was the geographic location that led to these many small differences across not just the Spanish language but other languages as well but now I know that era, pronunciation, community, outside influences, etc. can also contribute just as much to these differences within the Spanish language. (S49, C15, 4th)

Only the following answer was considered negative:

- i don’t think it has because you can learn a language perfectly well without knowing about variations (S35, C16, 3rd)

For the fourth question in Table 5, ‘Do you think paying attention to variation can be a way to keep learning Spanish (learning language)?’, most answers were positive. The negative responses are particularly illuminating, as they focused on the same tensions between content and language practice in upper-level language courses as C16. The following responses were classified as negative:

- This question is hard to answer because I feel as if my spoken/written Spanish itself has not improved with this class, but my knowledge of Spanish as a language (in a more practical sense) has improved. (S17, C17, 2nd)

- it could be but it might interfere because focusing on things other than the language itself may get confusing (S35, C18, 3rd)

- I don’t necessarily know if paying attention to variation would help one learn Spanish as I do think that in focusing on the variation it can get confusing as certain words, especially slang change by location and time. (S49, C19, 4th)

**Reactions to using corpus tools to learn about variation.** This section presents participants’ reactions to using corpus tools. Positive reactions are given in Table 3, and negative reactions in Table 4.

As seen in Table 3, student responses showed how they understood the multiple types of information that corpus searches could give about historical, social, and cultural changes. In the third iteration, students were more specific in naming their difficulties with the *Corpus del Español*, so it was easier to target those questions in the fourth iteration. Table 4 provides student insights into the negative aspects of using corpora.

Negative feelings were most prevalent in the first iteration, where some students expressed them even when discussing the positive elements of using corpus tools. Comments seem to grow more positive across iterations, but reservations about the tools were communicated in all.

Across iterations, proficiency issues were barely mentioned in participant responses. However, some of the comments about the overwhelming nature of the data could point to difficulties with proficiency rather than difficulties with the tools. Proficiency issues are discussed in §5.3 with the interview participants.

**5.2. Quantitative summary of the questionnaires across iterations.** This section quantifies responses across the four iterations. Tables 5 and 6 summarize the number of times each code was mentioned in each iteration. In the third iteration, not all questionnaire takers answered all of the questions, so the total number of responses is different from the participant number reported in Table 1. Additional numerical discrep-
ancies are due to learners discussing more than one topic in some open-ended questions, so the total number of codes in a given question may be higher than the total number of participants.

Table 5 summarizes participants’ perceptions of their learning and its value. The third and fourth questions were not included in the first iteration’s questionnaire. Table 6 summarizes the frequency of positive and negative responses to using the corpus tool to learn about variation.

Table 6 indicates that students in the second, third, and fourth iterations had somewhat more positive reactions to the use of corpus tools than students in the first iteration. Students in the third and fourth iterations were markedly more positive in explaining what they liked about corpus tools. However, technological difficulties with the corpus tools caused frustration for students in all iterations.

Responses to the first question in Table 5 and the first question in Table 6 reveal similarities across iterations, in that students claimed not to have studied variation or to have used corpora prior to this class. Given students’ lack of previous experience with corpus tools, their answers to the fourth question in Table 6 are especially relevant as they could indicate whether the use of corpus tools can be normalized into students’ regular learning practices. This normalization does not seem to have happened, as most

<table>
<thead>
<tr>
<th>CODE</th>
<th>QUOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Examples/ information</td>
<td>It allows us to see examples of how to use certain phrases, words, etc. (S7, C20, 1st) I like having the opportunity to see examples of Spanish conversations/written works at different levels of formality and complexity. (S28, C21, 2nd) I liked that it brought context and real world examples to linguistic differences. (S41, C22, 3rd) It can give you very useful information about what you are trying to find. (S47, C23, 4th)</td>
</tr>
<tr>
<td>2. Comparison</td>
<td>I like that it allows us to see many comparisons at one time, including a way to see trends over time. I thought that was the most interesting part of using the corpora. (S9, C24, 1st) I found it interesting to look at different registers and study language contextually. I also had never looked at grammar like this before. (S19, C25, 2nd) … i enjoyed the way the corpora compared the use of different words in a variety of dialects and showed what type of work they were used most in. It was very easy to use and easy to read the results. (S46, C26, 3rd) … I think it’s interesting to see how patterns of speech differ by time, location, context, etc. (S50, C27, 4th)</td>
</tr>
<tr>
<td>3. Nothing</td>
<td>Nothing. The online corpora was difficult to use and understand. I felt as though it did not add much to my learning experience. I still do not see how the use of the corpora related to our coursework. (S6, C28, 1st)</td>
</tr>
<tr>
<td>4. Exploring words</td>
<td>(see S7, C20, 1st) … In addition, it was interesting to be able to compare words from different regions of the world. (S18, C29, 2nd) (see S46, C26, 3rd) Provides interesting information and it makes you think about words in different ways. (S48, C30, 4th)</td>
</tr>
<tr>
<td>5. Interesting</td>
<td>It can be interesting to see the different statistics (S3, C31, 1st) It is so interesting! I really like the analysis we did in class with it. (S21, C32, 2nd) i have enjoyed the grammar aspect of this class much more than the history spanish classes i have taken at XX. i have always been interested in linguistics and this class definitely highlighted that for me (S46, C33, 3rd) (see S50, C27, 4th)</td>
</tr>
</tbody>
</table>

Table 3. What do you like about using corpora? (1st–4th iterations).
<table>
<thead>
<tr>
<th>CODE</th>
<th>Proficiency issues</th>
<th>QUOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>I feel like I don’t know the language well enough to benefit from seeing real examples. Like I can’t identify what makes the examples notable. (S1, C34, 1st)</td>
<td>I feel like I don’t know the language well enough to benefit from seeing real examples. Like I can’t identify what makes the examples notable. (S1, C34, 1st)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. Issues with searches</th>
<th>QUOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Specific comments from the 3rd iteration are given here as they refer to the same corpus.)</td>
<td>A veces me confuso porque hay mucho información. También, necesito pasar mucho tiempo a leer todos los ejemplos para encontrar algo significante [Sometimes I get confused because there is a lot of information. Moreover, I also need to spend a lot of time reading the examples to find something noteworthy] (S12, C36, 1st)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. Difficulties in understanding corpus tools’ purposes</th>
<th>QUOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Everything. I still do not understand what a corpus is for or why I would ever use it outside of the context of this class. (S5, C41, 1st)</td>
<td>Everything. I still do not understand what a corpus is for or why I would ever use it outside of the context of this class. (S5, C41, 1st)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>QUESTION</th>
<th>ANSWERS</th>
<th>1ST (n = 16)</th>
<th>2ND (n = 15)</th>
<th>3RD (n = 9)</th>
<th>4TH (n = 4)</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Have you studied language variation before this class?</td>
<td>1. Yes</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2. No</td>
<td>13</td>
<td>13</td>
<td>9</td>
<td>4</td>
<td>39</td>
</tr>
<tr>
<td>2. What are the main points you have learned about variation this semester?</td>
<td>1. Formal levels of structural variation</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>2. Social and geographical aspects</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>3. Rule/patterns</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>4. A general increase in knowledge</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>3. How do you think doing research on variation has helped you (or not) to understand the field of linguistics?</td>
<td>1. Positive</td>
<td>N/A</td>
<td>15</td>
<td>8</td>
<td>4</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>2. Negative</td>
<td>N/A</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4. Do you think paying attention to variation can be a way to keep learning Spanish (learning language)?</td>
<td>1. Yes</td>
<td>N/A</td>
<td>12</td>
<td>7</td>
<td>3</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>2. No</td>
<td>N/A</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 4. What do you not like about using corpora? (1st–4th iterations).
learners were not planning to use the corpus tools further. Most learners who reported that they planned to use corpora and corpus tools beyond the class were intending further study in linguistics, though personal curiosity was also a factor.

I will use to help me understand how to use different words in the right context. (S10, C44, 1st)

If something interests me about language patterns, I’ll probably use the corpus del español (S23, C45, 2nd)

Just for curiosity. Showing friends cool stuff we can do with it. (S21, C46, 2nd)

if I’m interested in the use of a word.. most likely use the english corpus (S46, C47, 3rd)

If I see a word that I don’t know in a reading I’ll probably get curious about where it’s used most and how it’s used. (S50, C48, 4th)

Results for the fifth question in Table 6 indicate that the Corpus del Español and CORPES XXI (Corpus del Español del Siglo XXI [Corpus of 21st Century Spanish]; Real Academia Española 2018) were learners’ preferred tools. The data in the third iteration is skewed, however, as the Corpus del Español was used almost exclusively. Despite some criticism of AntConc (Anthony 2017), some students did name it as their preferred corpus tool.

5.3. Interviews.
Reactions to learning about variation. Two learners from the first iteration participated in the follow-up interviews (Appendix B). In general, the learners saw variation as a topic worthy of study. S51 pointed out that they were already aware of variation since they had been taught the second-person singular distinction between informal tú and formal usted. These pronouns can be regarded as examples of both structural and social variation. As a form of structural variation, they are used to express degrees of

<table>
<thead>
<tr>
<th>QUESTION</th>
<th>ANSWERS</th>
<th>1ST (n = 16)</th>
<th>2ND (n = 15)</th>
<th>3RD (n = 13)</th>
<th>4TH (n = 4)</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Have you used corpora before this class?</td>
<td>1. Yes</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2. No</td>
<td>15</td>
<td>14</td>
<td>12</td>
<td>3</td>
<td>44</td>
</tr>
<tr>
<td>2. What do you like about using corpora?</td>
<td>1. Examples/information</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>2. Comparison</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>3. Nothing</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>4. Exploring words</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>5. Interesting</td>
<td>1</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>3. What do you not like about using corpora?</td>
<td>1. Proficiency issues</td>
<td>1*</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2. Issues with searches/navigation</td>
<td>11</td>
<td>9</td>
<td>13</td>
<td>4</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>3. Difficulties in understanding corpus tools’ purposes or usefulness</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>4. Will you keep using corpora after this class?</td>
<td>1. Yes</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>2. No</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>3</td>
<td>37</td>
</tr>
<tr>
<td>5. From the corpora we have seen in class, and others you might know, which one is your favorite corpus?</td>
<td>Corpus del Español</td>
<td>11</td>
<td>7</td>
<td>10</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>AntConc</td>
<td>N/A</td>
<td>6</td>
<td>N/A</td>
<td>N/A</td>
<td>6*</td>
</tr>
<tr>
<td></td>
<td>CREA</td>
<td>2</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>2*</td>
</tr>
<tr>
<td></td>
<td>Corpes XXI</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>2</td>
<td>2*</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 6. Reactions to corpus tools.
* A discussion about difficulties related to language proficiency arose during the interview with S52.
* = Not actively used in all iterations.
formality within a variety. However, the social relationships they express can vary from one variety of Spanish to the next; thus, the use of these pronouns can serve as a form of social variation. S51 pointed out how tú and usted function as a form of structural and social variation. S52 additionally pointed to some dialectal variations that had been presented prior to this class, such as

the ‘z’ sound or they drop the ‘s.’ You knew the little things here and there, but it was never concrete or explicit with it. Everybody knows about vosotros and stuff like that, but nobody really knew all the differences or sometimes why they happen. That was something, and I thought that was interesting, that I was able to explore that. (C49)

S52 had a solid grasp of grammar before starting this class and thought it was necessary for learners to have a strong command of the language before discussing grammatical variation. S52 also felt that it was counterintuitive to discuss variation using the rules from the textbook as a starting point, since the corpus and handouts could suffice to infer rules. That is, S52 preferred to move from the examples to the rule rather than the other way around. For S52, working with variation was a clear way to expand grammar knowledge.

Reactions to learning about corpus tools. One of the main questions discussed during the interview was how to make the tools more appealing to students as devices for learning about variation and for continued language learning beyond the class. S52 pointed out that it was difficult for learners to see how these explorations would relate to language learning and how they could be important beyond this class. Furthermore, S52 commented on classmates’ difficulties when reading concordances because of their lack of proficiency. S51 additionally acknowledged that it took a bit of adjustment to get used to employing corpus tools due to difficulties in understanding their purpose and the layout of the Corpus del Español.

6. Discussion.

6.1. What are learners’ reactions to learning about linguistic variation in a Spanish L2 linguistics course? In order to know whether students reacted positively or negatively to the topic of linguistic variation, it is important to examine first what they thought they had learned about this topic.

Even though learners had been exposed to examples of variation since their beginner courses, students in all iterations considered variation to be a new topic. In all iterations, describing variation patterns and finding examples that did not fit with the pattern learners had previously learned was a new activity for most, to which they did not always react positively. In the first iteration, even the fact that the textbook (Lunn & DeCesaris 2007) discussed variation and offered rules from a descriptive perspective was something new. Accepting that not everything is true or false is a fundamental part

---

4 Vosotros is a pronoun used for second-person plural informal in Spain. In other countries, ustedes is the pronoun used for both second-person plural formal and informal. The use of the pronoun vosotros implies the use of a corresponding verbal suffix.

5 In addition to some nonauthentic examples, the textbook included real examples and references to literary readings in which learners could observe variation in language patterns caused by a variety of geographic and social factors. The textbook also encouraged students to consult with native speakers about language patterns as well as to find real examples on the internet and in other materials. As the authors indicated in the preface, ‘As linguists, we are not satisfied with the así se dice [it is said so] approach to grammar … [This textbook] is an introductory linguistic analysis of those points of Spanish grammar that nonnative speakers find hard to use and native speakers find hard to explain’ (Lunn & DeCesaris 2007:xi).
of the learning process in general; thus, learning about language variation can be a beneficial component of the general education curriculum at the college level.

In all iterations, only a handful of topics related to variation were discussed, and the number of topics was reduced as the iterations progressed. Specifically, after the first iteration the explicit language-learning component was eliminated, and after the second iteration some classical topics in Hispanic linguistics, such as the use of the pronoun se (see Fernández López n.d.), were no longer included, as they were deemed to be too challenging for inclusion in an introductory linguistics course taught in the students’ non-dominant language. As Shin and Hudgens Henderson (2017:204) put it, even for language-focused grammar classes, ‘the sociolinguistic approach [to teaching grammar] privileges depth over quantity of grammatical concepts and thus necessarily leaves out some topics’.

Learners’ lack of proficiency (see Klee & Tedick 1997) might be a factor contributing to the challenging nature of learning about variation in a language major and, consequently, to learners’ negative reactions (e.g. C34, interviews). This challenge is consistent with some of the literature discussed in Macaro’s (2018) review of research on content courses taught in students’ L2 and with Roussel et al.’s (2017) cognitive-load hypothesis, in that (i) L2 content learning can be less conducive to L1 content learning, and (ii) L2 content-based instruction can contribute less to language learning than L2 language-development classes. In this research line, the difficulties of learning content in an L2 are reduced to a cost-benefit perspective.

As discussed in the literature review, however, the context of linguistics courses in a language major at most LACs poses a challenge to this cost-benefit construct. That is, linguistics courses may not be offered outside of language departments. It seems advisable to compromise, striving for a certain level of proficiency in both language development and content knowledge, rather than sacrificing one for the other.

Because of the need to combine language and content instruction, strategies for coping with the language proficiency challenge experienced by the students must be considered. For example, including examples of variation in a language that is more familiar to most students, such as English in this study, might be helpful in facilitating learners’ understanding of variation. Including readings in the dominant language can also be helpful. Given the heterogeneity of proficiency levels in any content course, the scaffolding provided will be classroom- and program-dependent (see Tedick & Lyster 2020).

It is worth noting that research on content learning and language learning in higher education is scarce, especially on learning languages other than English (see Macaro 2018, Polio & Zyzik 2009). In the data examined for the current study, a consistent pattern regarding L2 language learning was observed. Namely, students equated language learning with language production, and specifically with speaking. Yet as noted in §1, the World-Readiness Standards (The National Standards Collaborative Board 2015) remind instructors that language learning is much more comprehensive than this limited conception, encompassing the five Cs of communication, cultures, connections, comparisons, and communities. As discussed in the literature review, working with structural and/or social variation is a way to bring the five Cs into the classroom. But as some students (C18, C19) pointed out, working on linguistic variation can lead to a certain level of insecurity. Similarly, a learner in the fourth iteration, after having discussed patterns of third-person pronouns, asked ‘how do I use them now?’.

Students’ insecurities could be addressed through class discussions. For instance, it might help to emphasize not only that language learning encompasses more than speaking but also that the study of variation is likely to contribute more to receptive skills
than to productive skills. In a course that combines content and language learning, learners do not need to master all of the different patterns, but they need to be aware of them. In fact, raising awareness of L2 patterns and including reflection on L1 patterns contribute to language learning (see McManus & Marsden 2019, Svalberg 2007).

In general, learners in this study were positive about the content. Although students pointed to the same themes across all iterations, the second and fourth groups (see Table 2 and Table 5) tended to provide more in-depth responses to the questionnaire. These differences are probably due to the stronger focus on research in the second iteration and the inclusion of targeted practice with several user-friendly corpora in the fourth iteration. That is, the second iteration explored more fully the topic of linguistic variation in Spanish, whereas the first focused equally on language learning and learning about language variation. Meanwhile, the third iteration included more targeted practice with only one corpus, whereas the fourth offered targeted practice with additional corpora. Consequently, the second and fourth iterations had more opportunities for research and deeper exploration of the topics about language, which may have prompted those students to provide more comprehensive and targeted feedback on the questionnaire.

In particular, the focus on research in the second iteration seems to have prompted students to provide more comprehensive feedback on the research process itself. Learners in the second iteration were the only participants to differentiate between the questions ‘How do you think doing research on variation has helped you (or not) to understand the field of linguistics?’ and ‘What are the main points you have learned about variation this semester?’. Students in the third and fourth iterations responded to both of these questions by listing the points they had learned about variation, while students in the second iteration responded to the first question by listing the ways in which their understanding of the field of linguistics had (or had not) grown. Students’ ability to distinguish between these two questions suggests that the focus on research in the second iteration helped them develop a deeper understanding of research in the field of linguistics.

Despite this positive outcome in the second iteration, the focus on research was reduced in the third and fourth iterations in response to negative feedback. As Weimer (2013) discusses, active learning pedagogies are not always easy to implement and require several trials, as instructors need to adapt the activities to the level of student readiness. Moreover, instructors may have expert blind spots, causing them to underestimate the difficulty involved in their instructional practices (see Ambrose et al. 2010). Furthermore, student interests, backgrounds, and language proficiency are different in every classroom, requiring the instructor to constantly adapt to new classroom settings.

6.2. How do learners perceive the effectiveness of corpora in learning about variation? In the analysis of the second research question, it was clear that learners could see how corpus tools offer a way to find real language examples and draw comparisons between varieties. Moreover, learners’ perceptions of corpus tools and corpora became increasingly positive across iterations. Only in the first iteration were there categorical dislikes, and the word ‘interesting’ was used with increasing frequency and consistency in participant responses to the question ‘What do you like about using online corpora?’ (see Table 3).

As suggested by Knouse (2018) and Knouse et al. (2015), small research projects should be a basic element in any linguistics course. While the corpus-based exploratory assignments introduced learners to the topic of linguistics, students’ final research papers and presentations, not discussed here for reasons of space, showed how students were able to satisfactorily discuss structural and sociolinguistic variation using corpus data across iterations.
Still, the interviews and questionnaire responses revealed that some students perceived the corpus tasks to lack connection with out-of-class needs and even course objectives. In all iterations, some students were skeptical about the idea of using corpora to learn about variation (see C42). Some students’ expectations about what a grammar course entailed did not fit with the corpus approach—that is, a disconnect between learner expectations for a class versus the goals set in the syllabus can explain the rejection of incorporating corpus tools in the classroom (e.g. C41, C42). Indeed, the instructor-researcher’s experiences suggest that Spanish L2 learners in similar contexts tend to think of pedagogical grammar when they think of a grammar course. Learners might be more receptive to working with corpora and corpus tools if instructors discussed in depth the role of linguistics courses in the major. In general, sharing the goals of and reasons for using certain approaches contributes to student acceptance of these new approaches, which in turn contributes to student learning (see Weimer 2013).

As Weimer (2013) discusses, integrating active teaching methodologies implies preparing students to face new approaches to learning. Weimer identifies four reasons for student resistance to active learning: (i) they may be required to do more work, (ii) they may fear the changes involved in the implementation of active learning, (iii) they may become less certain of their own skills and knowledge, and (iv) they may not be ready for active learning. Although each of these four factors likely played a role in the students’ skepticism about using corpora and corpus tools to learn about language variation, the third factor seems the most relevant in this case. While pedagogical grammars are infallible, descriptive grammars are dependent on frequencies, which are fallible. The fact that descriptive grammars are driven by context, rather than prescriptive rules, may undermine students’ sense of confidence that they ‘know’ grammar. Such concerns were suggested by students who indicated that learning about variation can be confusing and that the use of the corpus tools seemed disconnected from the goals of the course.

To overcome negative perceptions such as these, it is important to raise students’ awareness of the utility of corpus tools, especially of their value in illustrating variation patterns with real examples. It is likewise important to raise students’ awareness of the fact that searching for patterns can help them both in understanding the language and in identifying varieties. Whereas students seemed to recognize the benefit of using corpus tools to locate examples in corpora for the purpose of learning about variation, they did not see any benefit in this exercise with regard to increasing their productive proficiency. However, analyzing and reflecting on examples found in corpora can contribute to improving students’ written and oral output (see Kennedy & Miceli 2010, O’Sullivan & Chambers 2006, Vyatkina 2013, 2016a,b as examples of successful classroom interventions). Moreover, despite the lack of targeted production practice in the examined classroom, these corpus-based activities can still contribute to communicative practice (Marcos Miguel 2021). Nevertheless, if students are not ready, the instructor must meet them where they are. Integrating new activities and approaches to learning in a classroom should be done gradually (see Weimer 2013).

Despite these challenges, some of the participant comments (C44–47) suggest that corpus tools can become normalized, which implies that the technology is no longer seen as a novelty but as a part of the classroom (see Bax 2011), in much the same way that LMSs, online dictionaries, and online textbooks have become ‘normal’ rather than ‘exceptional’ within the past few decades. Once they are fully normalized, it will be easier for corpora and corpus tools to find their place across the curriculum. It should not be forgotten that corpus tools have become part of digital humanities pedagogy (e.g. Cro 2020, Varner 2016), which helps to situate them not only in linguistics courses, but also across the humanities curriculum.
Another issue discussed by the learners is that corpus tools and corpora are not user-friendly, for example, when searching for patterns. Furthermore, the interface usually provides many examples at once, which can cause the learners some uncertainty. To address this concern, the search assignments were simplified in the third and fourth iterations, requiring students to carry out mostly lexical searches rather than searching for phrasal structures. Lexical searches included the exploration of a fixed pattern of specific verbs and pronouns, the search for anglicisms in Spanish, and the investigation of the use of gender markers in professions across time. Whereas these types of searches were helpful in reducing complexity, they seemed to diminish learners’ capacity for discovering patterns, as the term ‘perceiving patterns’ no longer appeared in their responses to the questionnaire. Nevertheless, these students were able to comment in great depth about variation observed within and across words, such as genre, historical, and geographical changes (C26, C27). Thus, lexical searches appear to have given students a solid foundation of corpus work on which to build more complex syntactic searches.

Student comments about technological difficulties are consistent with results reported by Paz (2013) and Benavides (2015) in advanced grammar courses. Still, the user-friendliness of the tools is beyond most teachers’ abilities to fix. As Godwin-Jones (2017:21) says, ‘[t]he vast majority of DDL studies are from researchers in the field, not from teachers. That is only likely to change if user interfaces to corpora improve’. In order to normalize the technology, there needs to be more collaboration between researchers and instructors so that most learners feel competent enough in utilizing corpus tools and researchers better understand how instructors and learners want to use the tools.

Fortunately, corpus tools keep evolving. Lemmatized Spanish corpora, such as the CORPES XXI and Corpus del Español, have been evolving along with the iterations of the course in this study. In a lemmatized corpus, learners can find all forms of a word without having to look for each specific example. As all of these lemmatized corpora offer similar outputs, the learning curve in the fourth iteration was smoother than in the second iteration when AntConc was used with a nonlemmatized corpus. Nevertheless, recognizing word categories is necessary for using any lemmatized corpus, and learners in this context have a limited awareness of such categories in both their L1 and L2. In fact, the concept of the lemma can be confusing for some learners (see C38). One of the benefits of taking a linguistics course is that such courses invite further reflection on language as an object of study, so that students can utilize this metalanguage for their own learning. With knowledge of this metalanguage, corpus tools can promote learner independence beyond the classroom and across disciplines.

Furthermore, given that some of the difficulties of implementing corpus use in the classroom can be attributed to students’ lack of understanding of the role of linguistics courses within the major, the importance of learning about variation and about language should be emphasized in elementary and intermediate language courses. Again, the topic of linguistic variation and the use of corpora can be aligned with the five Cs, which are a main reference in most language programs. Moreover, a study of language variation can help students look beyond binary patterns of correct versus incorrect thinking. The development of such critical thinking skills is an overarching goal of higher education.

7. CONCLUSION. This project contributes to research on the teaching of language variation in linguistics courses, exploring how L2 students react not just to learning about language variation in the L2 but also to the use of corpora in studying this topic. Following an action-based research approach, this study has shown that learning about
social and structural variation not only aligns with the five Cs but also can be a well-received topic in content-level courses. However, the study has also illustrated several challenges that exist in utilizing corpus tools for teaching and learning about language variation. Many students encounter difficulties in using the tools; thus, teachers must scaffold students’ use of the tools carefully. Moreover, some students are resistant to or skeptical of frequency-based models of grammar instruction, which offer more than one ‘right answer’. These models contradict the pedagogical models of ‘correctness’ to which students are no doubt accustomed. Therefore, instructors must proceed thoughtfully when introducing students to the study of variation, a concept that complicates, rather than simplifies, language rules and encourages students to think beyond simple dichotomies of ‘right’ and ‘wrong’.

One of the most interesting points of our results is that so many learners thought of variation as a new topic. Given that learners are exposed to more than one variety of the L2 through their language instructors and materials, more explicit comments about variation could raise awareness of variation at an earlier stage of proficiency. If successful communication consists of ‘knowing how (grammar), when (context) and why (purpose), to say what (vocabulary) to whom (audience)’ (The National Standards Collaborative Board 2015:11), more emphasis on variation is needed at all levels of instruction in order to promote knowledge of social and structural variation that students can use and recognize for successful communication. Whereas speaking practice is fundamental for language learning, the skills and knowledge that learners acquire when studying variation and using corpus tools are also necessary for mastering a language.

APPENDIX A: QUESTIONNAIRE

1. How many years of Spanish have you had prior to the university?
2. How many semesters of Spanish have you had at the university? (Options: from 1 to 8 semesters)
3. Do you speak Spanish outside of class? (Options: yes/no)
4. If you answer yes to question 3, how often do you speak Spanish outside of class? (Options: less than once a month/once a month/2–3 times a month/once a week/2–3 times a week/daily)
5. Have you used corpora before this class? If so, explain.
6. What do you like about using online corpora?
7. What do you not like about using online corpora?
8. From the corpora we have seen in class, and others you might know, which one is your favorite corpus? Explain why.
9. How often have you used a corpus outside of class this semester? [First iteration; modified to How often have you read the articles/chapters assigned for this class this semester? in following iterations.] (Options: once a month/2–3 times a month/once a week/2–3 times a week/daily/less than once a month/never)
10. How often have you used a dictionary outside of class this semester? (Options: once a month/2–3 times a month/once a week/2–3 times a week/daily/less than once a month/never)
11. How often have you used your textbook outside of class this semester? (Options: once a month/2–3 times a month/once a week/2–3 times a week/daily/less than once a month/never)
12. Will you keep using corpora after this class? (Options: yes/no)
13. If you answer ‘yes’ to the previous question, for what purpose will you keep using corpora?
14. Have you studied language variation before this class? If so, how did you study it?
15. What are the main points you have learned about variation this semester?
16. How do you think doing research on variation has helped you (or not) to understand the field of linguistics? (learning about language) [Question added in the second iteration.]
17. Do you think paying attention to variation can be a way to keep learning Spanish (learning language)? Explain. [Question added in the second iteration.]

APPENDIX B: MAIN INTERVIEW QUESTIONS

1. Did you use the corpus tools for other classes and for your personal use?
2. Are you using corpus tools this semester? Are you taking any Spanish classes?
3. What do you like the most about using corpus tools?
4. What kinds of corpus-tasks do you like? (That we used in class or that you think can be designed to help students) Why?
5. What did you like the least about using corpus tools?
6. What do you think of using corpus tools for learning about variation? Does it work?
7. What other elements can be used to learn about variation?

REFERENCES


Davies, Mark. 2004b. Student use of large, annotated corpora to analyze syntactic variation. *Corpora and language learners*, ed. by Guy Aston, Silvia Bernardini, and Dominic Stewart, 245–70. Amsterdam: John Benjamins.


Fernández López, Justo. n.d. Valores y funciones de la partícula se. Online: http://hispanoteca.eu/Gram%C3%A1ticas/Gram%C3%A1tica%20espa%C3%B1ola/SE%20-20Valores%20de%20la%20part%C3%A9cula%20se.htm.


Freeman, Scott; Sarah L. Eddy; Miles McDonough; Michelle K. Smith; Nnadozie Okoroafor; Hannah Jordt; and Mary Patt Wenderoth. 2014. Active learning in-


O’Keeffe, Anne; Michael McCarthy; and Ronald Carter. 2007. From corpus to classroom: Language use and language teaching. Cambridge: Cambridge University Press.


ROUSSEL, STÉPHANIE; DANIELLE JOULIA; ANDRÉ TRICOT; and JOHN SWELLER. 2017. Learning subject content through a foreign language should not ignore human cognitive architecture: A cognitive load theory approach. Learning and Instruction 52:69–79. DOI: 10.1016/j.learninstruc.2017.04.007.


VARNER, STEWART. 2016. Library instruction for digital humanities pedagogy in undergraduate classes. Laying the foundation: Digital humanities in academic libraries, ed. by Heather Gilbert and John White, 205–22. West Lafayette, IN: Purdue University Press. Online: https://docs.lib.purdue.edu/purduepress_ebooks/33/.


[marcosn@denison.edu]

[Received 7 April 2020;
revision invited 27 August 2020;
revision received 25 January 2021;
revision invited 7 June 2021;
revision received 25 August 2021;
accepted pending revisions 1 November 2021;
revision received 16 November 2021;
accepted 16 November 2021]