_____

Using a Corpus Effectively in a Spanish Grammar Course

Carlos Benavides, Ph.D.
Associate Professor
Department of Foreign Literature and Languages
University of Massachusetts Dartmouth
cbenavides@umassd.edu

Introduction[1]

There is a growing body of literature on the use of corpora for language teaching. Most of the studies address the use of corpora in English-language classes, and within those that deal with Spanish instruction, quite a few focus on the teaching of vocabulary, pragmatics, writing, or on gathering data from a corpus for use in course preparation or the creation of course materials. Some studies also offer ideas as to how students and instructors can use corpora in a class (e.g. explore collocations to determine the positions of certain words), but not all of the studies provide concrete examples or activities. Only few deal specifically with how students could use corpora in Spanish (or English) grammar courses as a tool to enhance learning of grammatical concepts and structures, and no studies, to our knowledge, investigate the use of corpora in advanced Spanish grammar courses.

In order to fill this gap, the present exploration examines the use of a large corpus—the Corpus del Español[2]—in an advanced Spanish grammar course. Our focus is on having students search the corpus with the goal of finding concordances containing particular types of collocations and tokens, in order to better understand and use the grammatical concepts covered in class (e.g. preterite vs. imperfect, *ser* and *estar*). We believe that through corpus searches, students can verify for themselves the accuracy of grammar explanations and examples provided in their course textbook, formulate grammatical rules and exceptions to those rules, find alternate or divergent uses of a range of grammatical structures, and perform other related tasks. Our study describes the results of the use of the Corpus del Español and reports on the effects on learning of integrating a corpus in our advanced Spanish grammar course.

Spanish Course and Textbook

The study was carried out in a third-year advanced Spanish grammar course. The textbook for the course consists of 14 chapters, six of which were covered during the semester. Each chapter consists of one reading, one to four grammar topics (such as preterite vs. imperfect, *ser* and *estar*, special verb constructions such as with *gustar*, and articles), a lexical section, which deals mainly with word formation processes (use of prefixes and suffixes), and a section on writing that covers topics such as spelling, punctuation and use of connectors.

As part of the grammar explanations to the students, the textbook provides a series of sample sentences. Even though some of these sample sentences are taken from the reading at the beginning of

_____

each chapter, most of them are isolated, invented, and presented out of context. Furthermore, what we found is that the examples are frequently not enough for students to gain a full understanding of the usage of each grammatical structure. In contrast to a relatively small set of sentences, we believe a corpus offers large amounts of authentic, contextualized examples of language use and can provide more language examples.

Participants

Nine students who were enrolled in the course participated in the study. The students were trained by the instructor in the use of the corpus both in class and at a computer lab. Students conducted most of the corpus searches at home as part of their assignments, but a number of the interactive corpus-based activities were also conducted in class (see below).

Corpus Employed

The Corpus del Español (CDE) is an online corpus that consists of more than 100 million words in more than 20,000 Spanish texts from the 1200s to the 1900s. The corpus texts appear in four registers beginning in 1900, namely, spoken, fiction, newspaper, and academic (for a total of over 20.5 million words, with approximately equal amounts of words per register). The vast majority of 20th century texts date from 1970 onwards, and roughly half the texts come from Latin America and the other half from Spain.

The user-friendly interface of the CDE allows for a wide range of searches, including searches for exact words or phrases, words surrounding those items (i.e. collocates), wildcards, lemmas, parts of speech and other grammatical features (e.g. finding adjectives ending in –*udo*; finding nouns accompanied by plural, feminine articles; finding verbs in different tenses).

When we compare the CDE with other corpora, we get a sense of the power and versatility of the CDE itself. The CREA (Corpus de referencia del español actual: Real Academia Española 2011; http://corpus.rae.es/creanet.html) is a Spanish corpus containing over 154 million words from written and oral texts published between 1975 and 2004. The written sources of the CREA consist of books, newspapers, magazines, and miscellaneous texts. As is well-known, Google is a search engine that provides access to billions of pages and documents. Even though Google is not a corpus per se, searches of collocations can be made with the search engine and therefore it can be compared to a corpus. However, neither CREA nor Google allow the types of sophisticated queries that CDE enables.

For example, even though Google allows searches with a wildcard in between words (e.g. *ellos * caminar*), no useful searches are produced when the wildcard is placed inside words (e.g. *des*ado*) or attached to the edges of words (e.g. *camin**). Conversely, CREA only allows the wildcard to be attached directly before or after the target word, not between words. Furthermore, unlike the CDE, neither CREA nor Google are tagged corpora, so searches for lemmas, parts of speech and grammatical features, or any combination of these cannot be made. In addition, the CDE allows queries for words or expressions with a given number of collocates on either side of the key word, a very useful feature since a word of interest that is associated with the key word is not always adjacent to the key word. Moreover, the CDE

_____

_____

keeps a history of all searches done by a user, and customized search lists can be created as well. These are only some of the advantages of CDE over the other two sources.

On the other hand, there are advantages of CREA and Google over the CDE – and those have to do in part with corpus size. Both have token numbers beyond 100 million words (and well beyond the 20.5 million words CDE has for the 20th century). Also, while the CDE is free, students need to register to use it fully. In contrast, CREA, which is also free, requires no registration. Finally, CREA enables searches with variables such as author, work, and topic, a feature not available in CDE (or Google), but in our opinion, searches with these variables are typically not needed in grammar courses, although they might be useful for literature classes.

Activities, Tasks, and Assignments in the Course

Both individually and in groups, students conducted corpus searches in class, in the computer lab and as homework assignments to complete the following tasks:

> *1. Compare corpus findings with the explanations, descriptions and examples provided in the textbook.*
>
> > a) Find evidence in the corpus that confirms or contradicts the textbook rules and explanations. More specifically, find examples, exceptions and counterexamples to the rules and explanations provided in the textbook.
> >
> > b) Find additional examples and variant uses from the ones provided in the textbook, with the purpose of possibly expanding the rules and explanations given in the textbook.

In task 1a) students use the corpus to test the rules and patterns they have learned. This is an example of deductive learning (see sample exercise below). In task 1b) they use the concordance data to make generalizations, formulate rules, and discover new usage patterns and rules. This is an example of inductive learning (see sample exercise below).

> *2. Discover other linguistic phenomena that may be incidental to the searches.*

This task may be considered a form of intuitive learning and is compatible with Bernardini's (2004) suggested approach of guiding learners to browse corpora in open-ended, exploratory ways. As suggested in the corpus literature, corpus-based study is an aid to intuition.

_____

> ***3. Use corpus searches as an aid in completing the textbook exercises assigned as homework.***
>
> ***4. Answer questions*** *(prepared by the instructor)* ***based on or about concordances*** *(see sample exercises below).*

Some Sample Exercises

I present here two sample exercises, each with a very small number of examples from concordances, illustrating both inductive and deductive types of exercises. We found it important that students consistently practice both types of activities, and are encouraged to conduct open-ended, exploratory searches (task 2).

Inductive Exercise: *From the concordances to the rule*

Concordancia con el pretérito

Utiliza la concordancia como base para contestar las siguientes preguntas:
1. ¿En qué situaciones se usa el pretérito? Formula la regla para el uso de verbos en el pretérito.
2. ¿Cuáles son algunas expresiones (p. ej. *a veces*) con las cuales se usa comúnmente el pretérito?
3. ¿Hay alguna excepción a la regla del pretérito (1) en esta concordancia?

de pronto, una señora de unos cuarenta y cinco años pidió la palabra y **habló** explicando algo sobre unos cursos que a ella le habían dado de...
unos cursos que a ella le habían dado de... de alimentos. **Habló** unos diez minutos, con una propiedad extraordinaria. Entonces le pregunté al gerente quién
pero por razones mm... de tipo presupuestarias [sic], siempre se me **habló** de... una reducción de presupuesto, que... oscilaba.
, Universidad. Inf.?. - Yo creo que el año pasado se **habló** algo de eso con la OEA y el estado de la Florida. Inclusive yo
. No, " pasa adelante ". Siempre hablaba de tú. Nunca nos **habló** de usted ni de vos sino de tú a los alumnos. " Siéntate.

Center for Advanced Language Proficiency Education and research (CALPER)
The Pennsylvania State University
CALPER Corpus Community Report No. 4

_____

| **Deductive Exercise:** *From the rule to the concordances* |

Concordancia con el imperfecto

Utiliza la concordancia como base para contestar las siguientes preguntas:

1. El imperfecto se usa para hablar de acciones en curso (que no han sido completadas), descripciones, o acciones habituales, en el pasado. ¿Hay ejemplos en la concordancia que confirmen esta regla?

2. ¿Hay alguna excepción a la regla del imperfecto (1) en esta concordancia?

3. ¿Cuáles son algunas expresiones (p. ej. *en ese momento*) con las cuales se usa comúnmente el imperfecto?

de lo que hacía. Nunca estaba del todo contento con lo que conseguía. **Hablaba** mientras seguía dibujando, o pintando o esculpiendo. Nunca le oí decir que una
su vida. Era un hombre extraordinariamente sereno, con una intensidad difícilmente descriptible. **Hablaba** muy poco. Muy pálido y delgado, ya padecía de leucemia. Sus ojos
es su tópico favorito? - Uno que ha inventado Anson: « No se **hablaba** ayer de otra cosa en Madrid... » 43 - Otro articulista.
...] Inf.a - Bueno, mirá sobre eso de las publicaciones yo **hablaba** el otro día. A mí me encuentra - - - la gente o G
. y no era por echármela ni nada, sino que yo de repente te **hablaba** de una cosa y alienación y... y: " Bueno, ¿

Survey

In order to shed some light on the question of how students learned course content with carrying out corpus searches, we conducted a survey. The survey was adapted from Liu and Jiang (2009) and was administered in English. The nine students enrolled in the course completed the survey at the end of the semester. The survey had both a quantitative and a qualitative component.

The quantitative part consists of eight-questions with responses on a 1-5 scale (5 highest score), while the qualitative component consists of eight open-ended questions, plus an additional question on how students used the corpus to complete textbook exercises assigned as homework.

_____

Quantitative results

The quantitative component of the survey yielded a wide range of scores, going from one end of the scale to the other, which is indicative of a high degree of polarization. Given this marked contrast in responses, the results do not provide a clear measure of student improvement in learning.

Qualitative results

On the other hand, the qualitative results, which are more nuanced and furnish richer detail, provide evidence that the use of the corpus promotes student learning in a number of important ways. First, the searches significantly enhance deeper learning of course content through hands-on discovery, which encourages both inductive and deductive learning (see sample exercises above). Second, the corpus searches have contributed to increasing student interest in the learning of grammar, an aspect of language learning that is usually considered tedious. In addition, corpus use has contributed to improving student language proficiency and has given an incentive to students to want to employ this tool in other courses to further their learning.

The following student comments taken from the survey, as well as from the textbook homework assignments, provide support for the observations made above.

Sample comments from the survey

"My interest in grammar has grown more."

"I now have a better understanding of how important it is to use the correct endings on verbs and what context to use them in."

"When we are asked to look up exceptions, you learn what verbs obey different rules and in what contexts."

"I would like to do more corpus searches in projects in other Spanish classes."

"The corpus has helped with understanding word order and what type of word comes after certain words."

"Better use of my grammar skills when it comes to writing."

"I believe that integrating the grammar explanations with corpus searches has helped improve my proficiency."

"Sometimes textbooks do not go into enough detail to cover certain aspects of grammar. The corpus is a great tool to find evidence to 'prove' the rules of grammar."

_____

Sample comments from the textbook homework assignments

Six sets of textbook exercises were assigned as homework, one set for each of the chapters covered in the course. The representative comments from students shown below give an idea of the range of grammatical topics for which students found the use of the corpus helpful. It also can be seen how students reached valid conclusions on the basis of their corpus searches. These conclusions in turn helped students complete the textbook exercises successfully. (Note: exercise numbers are identified as "X." Comments in square brackets have been added by the researcher.)

"In my opinion, both the preterite and imperfect of *sentir* would make sense. So I did a search for [*sentir*] *deseos* 'feel like doing something' in the corpus and I noticed that this expression is used more in the preterite (20 instances) than in the imperfect (13 instances), so I used the preterite in my answer [and that was the most appropriate answer]."

"For exercise X I did a search for both [*estar*] *antiguo* 'be very old' and [*ser*] *antiguo* 'be very old' in order to make a quick comparison between the two expressions. There were many examples with *ser* (37), but none with *estar*. This helped me find the correct answer for the exercise."

"I was having trouble with the two constructions. I used the corpus to find more examples [than those available in the textbook] with these verbs."

"For exercise X I looked up verbs and adjectives that indicate emotion and I saw that these forms are used in sentences with the subjunctive many times. This gave support to the explanation in the textbook and also helped me complete my homework."

"After finding the expressions in the corpus I think I understood why we need to use the subjunctive with those expressions."

"I have noticed that the corpus helps me in many ways. If I know some rules, but still have doubts about the situations when I need to apply those rules, I consult the corpus. And I also like to find evidence in the corpus of the use of a given rule in past centuries as well." [Note: this is an example of incidental learning.]

"The corpus helped me to distinguish between some phrases that use the subjunctive and others that do not ... and so many examples appeared, that they affirmed my ideas. I can sincerely say that the Corpus del Español really is quite useful."

"For exercise X I conducted searches to find out the frequency with which the definite article appears with the seasons. I found hundreds of results, both with and without the definite article, which was useful to understand the explanation in the textbook."

"My corpus search confirmed the textbook rule. On the basis of the corpus searches, I decided to omit the article in these cases."

_____

Concluding Remarks

Conducting this corpus study was well worth the effort. I now have first-hand evidence that corpus use improves student learning of grammatical content. I went on to incorporate the use of the corpus for a second semester in the same grammar course, and also integrated corpus use in an advanced course on the history of the Spanish language. Interestingly, both the quantitative and qualitative survey results for these two courses were similar to the ones for the first iteration of the grammar course (described above). The qualitative results in particular have encouraged me to make corpus use a permanent part of my teaching. I am eager to explore exciting new ways to use the corpus to enhance student learning, especially in courses where corpus use may be more amenable.

Endnotes

[1]In this CALPER Corpus Community report, I will describe how I used the Corpus del Español with my students in one of our advanced Spanish grammar classes. I presented the study which I conducted in an interactive session at the annual convention of the American Council on the Teaching of Foreign Languages (ACTFL) in Philadelphia, November 2012. In the session, I explored ways to enhance learning of course content in a Spanish grammar class through the use of the Corpus del Español and shared strategies and techniques of how to use a corpus effectively.

[2]The Corpus del Español, which was created by Mark Davies (Brigham Young University), is a freely accessible, user-friendly online corpus of oral and written texts that allows for highly sophisticated searches. Davis, Mark (2002). Corpus del Español. Accessible at: http://www.corpusdelespanol.org/

References

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future development. In J. McH. Sinclair (ed.), *How to use corpora in language teaching* (pp. 15-36). Amsterdam, Philadelphia: John Benjamins.

_____

_____