

## Renaissance and Reformation Renaissance et Réforme



### Real Academia Española (Royal Spanish Academy). Corpus Diacrónico del Español (CORDE; Diachronic corpus of Spanish). Database.

Dale Shuger

Volume 43, numéro 4, automne 2020

Spaces of Power of the Spanish Nobility (1480–1715)  
Les espaces de pouvoir de la noblesse espagnole (1480–1715)

URI : <https://id.erudit.org/iderudit/1076839ar>  
DOI : <https://doi.org/10.33137/rr.v43i4.36395>

[Aller au sommaire du numéro](#)

Éditeur(s)

Iter Press

ISSN

0034-429X (imprimé)  
2293-7374 (numérique)

[Découvrir la revue](#)

#### Citer ce compte rendu

Shuger, D. (2020). Compte rendu de [Real Academia Española (Royal Spanish Academy). Corpus Diacrónico del Español (CORDE; Diachronic corpus of Spanish). Database.] *Renaissance and Reformation / Renaissance et Réforme*, 43(4), 239–243. <https://doi.org/10.33137/rr.v43i4.36395>

collection nevertheless provides an opportunity to discourse and reflect on the “world” that Don Quixote perhaps perceived as uniquely his own.

DAMIÁN ROBLES

Texas A&M University

<https://doi.org/10.33137/rr.v43i4.36394>

**Real Academia Española (Royal Spanish Academy).  
Corpus Diacrónico del Español (CORDE; Diachronic corpus of Spanish).  
Database.**

Madrid: Real Academia Española. Accessed 6 January 2021.  
[rae.es/banco-de-datos/corde](http://rae.es/banco-de-datos/corde).

The Corpus Diacrónico del Español (CORDE; Diachronic corpus of Spanish) and the complementary Corpus de Referencia del Español Actual (CREA; Reference corpus of current Spanish) together constitute a searchable database of the Spanish language from its first appearances in print to the present. CORDE covers texts prior to 1975, and CREA picks up from there. This review will focus exclusively on CORDE, although the design and methodology of the two are similar.

According to the Real Academia’s introductory page, CORDE draws on a database of 250 million words from texts published across the Spanish-speaking world, “in prose and verse, and within each modality, in narrative, lyric, drama, technical-scientific, legal, religious, and journalistic texts” (“Home”; [rae.es/banco-de-datos/corde](http://rae.es/banco-de-datos/corde)).<sup>1</sup> As should be evident, this is an incredibly promising resource for scholars in any number of endeavours, both those using the results as a bibliographic tool and those seeking statistical information about word and language usage over time. However, as with any database, the quality of results is limited by the quantity of data in the system and the ease of access to that data.

Before addressing the question of CORDE’s utility in practice, I will offer an outline of the interface and its functions. The home page allows the user to input under “Consulta” (“Search”) any word or phrase, and to limit the search

1. All translations of site material are my own. The site is exclusively in Castilian Spanish.

according to genre, date range, and nation. Basic word or phrase searches are intuitive; for more complex searches using Boolean operators, there is a guide under “Ayuda” (Help) with instructions. It should be noted that searches are case- and diacritic-sensitive, so users who wish to get more comprehensive results on a single word search should enter a search of “Palabra O palabra” in order to recover instances at the beginning of a sentence as well as within an oration. Once users input a word/phrase and any desired filters, they are directed to a second page, which is less intuitive.

The intermediate page offers three options. The first two seem designed for more statistically based searches: the first provides statistical summaries of the results by year, region, or genre; the second allows users to limit results via random sampling. The third option, “Obtención de ejemplos” (Get examples), provides access to the results themselves. These results can be organized according to various criteria, some of which (year, country) are self-explanatory, while others (“casos”) are less so. There are instructions for limiting and classifying results with these features on the “Ayuda” page. From a design point of view, the help functions are not particularly well highlighted; the “Ayuda” link goes directly to the entire user manual rather than cross-linking from where the user encounters a problem to a specific solution. Given that a beginning user will need to go back and forth between the manual and the site, it would have been preferable to make the “Ayuda” automatically open in a new tab.

The default option for “Obtención de ejemplos” is “Concordancias” (Concordances): a single line of text with the searched-for word or phrase in the centre. Clicking on a given line will provide a full paragraph of context and the bibliographical information for the source document. The user can also choose to access this information directly by changing “Concordancias” to “Párrafos” (Paragraphs) or “Documentos” (Documents) on the previous page.

Finally, at the bottom of the main page is a link to the “Nómina de autores y obras” (List of authors and works) which allows a user to see if a given author or work is included in CREA and/or CORDE, or to see what authors/works are included for a given geographical, generic, or date range search (genre-specific searches are available for CORDE only). For both works and authors, searches must be exact, including articles and punctuation for titles, or else one is told that there are no results; there are no suggestions for alternate searches or “nearby” results, and there is no keyword search option. This makes it very difficult to locate many pre-modern authors or works, given the variations in surnames

and titles across editions. For instance, when I searched “Lope de Vega” I was told there were zero records, with no suggestion to try “Vega Carpio, Lope de.”

To gauge the depth and breadth of the collection, I thought I would begin with authors at the centre of the Spanish literary canon and work my way outward. As would be expected, the site contains most of the works of Cervantes and Quevedo; however, I was surprised to find only thirty-three of Lope’s *comedias* in the database, a tiny fraction of the author’s total output, although the database also contains many works by Lope in other genres. As can be imagined for a database with only thirty-three Lopean comedias, the holdings for less canonical authors are scant and arbitrary. More contemporary authors fare no better: across CORDE and CREA there are only five works by Camilo José Cela and two by Juan Goytisolo. The most canonical Latin American authors are represented only by their most famous works: two works for Mariano Azuela, one for Domingo Sarmiento, and one for José Eustasio Rivera. Canonical poets whose works have been gathered into anthologies fare better than prose authors: for example, Francisco de Quevedo and Rubén Darío both have a significant portion of their work on the site, but when I searched for less canonical—although still well-known and studied poets—I usually came up with nothing, particularly for women. Most of the legal, political, and notarial documents come from contemporary documentary anthologies, but there are many important documentary collections not included. For example, I searched for several of the major epistolary and Inquisitorial documentary collections I see on my bookshelf, all to no avail. A quick search of the scientific and popular culture categories produced similar results: thirteen documents under “matemáticas,” eleven results for “gastronomía y comida,” and twelve for “arquitectura y urbanismo.” Curiously, across the three subjects there is not a single document from the 1700s.

There is no information on the site as to the process of selection, cataloguing, and incorporation of works. Nor is there any indication as to whether the database is continually being expanded as new works are put in, although in the three or four years that I have been using CORDE, the introductory text has read “250 million entries,” suggesting that the database set is fixed, pending the release of a new version.

As far as I could tell, all works, regardless of original publication date or genre, are scanned from a contemporary edition. The bibliographical information for each edition is given, but no explanation is offered as to why

that edition was chosen. There is no way of knowing to what extent a given edition contains modernized orthography or syntax, which obviously impacts the reliability of searches that seek to track changes in orthography or grammar over time. Each distinct work is included only once, regardless of the number of editions published over the centuries, but there is no explanation of how determinations were made as to what constitutes a “distinct” text. This is not an insignificant question for those who work with the *romancero* or *cancionero*, for example. For all these reasons, it seems to me that CORDE is least useful for the type of search that the RAE offers as its *raison d’être*: “any diachronic study related to the Spanish language [...] to document words, classify them as archaic or in disuse, to know the origin of certain terms or the tradition of their usage, or the first appearance of words” (“Home”). I fear that the lacunae in the database contents, as well as the lack of clarity regarding criteria of editions included, make CORDE unsuited to any conclusive study of diachronic use. Although statistical breakdowns abound on the site, I am even more skeptical of the options for producing meaningful statistics. It seems impossible to treat the statistical results as anything more than a statistical breakdown of an arbitrarily (or mysteriously) curated subset of “works in the CORDE database.”

Some of these questions are endemic to the field of computational humanities, but the total opacity of the CORDE site’s inclusion criteria means that users cannot fully address the biases coded into the system. The utility of the site would be greatly enhanced by a much clearer presentation of editorial decisions. In its current form, CORDE would at best be able to confirm general tendencies and perhaps disprove a hypothesis of first usage.

Despite the focus on linguistic searches in the RAE presentation, the main value of CORDE would seem to be for literary scholars and historians using it as a bibliographical tool in investigations on a particular subject or expression. For these purposes, CORDE is a huge step forward. The only similar database is Google Books, which has a far larger content base; however, Google searches cannot be limited according to region, genre, or date range. That said, CORDE is a far cry from a comprehensive database of the Spanish language, and the gaps in material as well as the lack of transparency regarding editorial decisions could produce misleading results. The promise of this site is enormous, and this researcher hopes that CORDE will follow in the footsteps of Portal de Archivos Españoles (PARES; [pares.culturaydeporte.gob.es/inicio.html](http://pares.culturaydeporte.gob.es/inicio.html), also reviewed in this issue), the database for Spanish national archives, which also debuted with

significant glitches and gaps but was re-launched in a much improved second version. The Borgesian utopia of a universal library of human language, even one human language, is a monumental task, and it is understandable that this initial version of CORDE has not yet achieved it.

DALE SHUGER

Tulane University

<https://doi.org/10.33137/rr.v43i4.36395>

**Hernando Morata, Isabel, project lead.**

**Calderón en Red/Una Biblioteca Calderoniana (Calderón on the Web/A Calderón library). Other.**

Santiago de Compostela, Spain: Grupo de Investigación Calderón de la Barca (GIC) (Research group on Calderón de la Barca). University of Santiago de Compostela, 2016. Accessed 1 May 2020.

[calderonenred.wordpress.com](http://calderonenred.wordpress.com).

Writing this review in the spring of 2020 during lockdowns and enforced stay-at-home orders, I saw the value of a website on the works of the Spanish dramatist Pedro Calderón de la Barca (1600–81), designed by Isabel Hernando Morata in 2016. The digitized collection of libraries has always been considered laudable, as it prevents wear and tear on manuscripts. The increased searchability of authors is in keeping with the contemporary global encouragement of digital media. One day, a different media ecosystem for scholars will be privileged in which information created by individuals or groups can be replicated endlessly and distributed freely. Researchers will no longer have to secure prodigious grants to visit distant libraries. All too often these visits to ferret out precious resources are fruitless because the text has been moved elsewhere; thus, the convenience of having a reliable resource at hand is undeniable.<sup>1</sup>

The construction of this resource took place during Hernando Morata's years as a doctoral and post-doctoral student at the University of Santiago

1. As regards the growing trend of online resources, see for example, "Final Progress Report, California Open Educational Resources Council" (1 December 2015; [tinyurl.com/FPRCAOERC41516](http://tinyurl.com/FPRCAOERC41516)) and "OER Adoption Study: Using Open Educational Resources in the College Classroom" (1 April 2016; [tinyurl.com/WPOERAdoption040116](http://tinyurl.com/WPOERAdoption040116)) by the same council.