RiCL Research in Corpus Linguistics

The *Colonial Texts Corpus* for the *Digital Library of Old Spanish Texts*

Sonia Kania[a] – Francisco Gago Jover[b]
University of Texas at Arlington[a] – College of the Holy Cross[b] / United States

Abstract – This article offers a detailed description of the *Colonial Texts Corpus*, one of eleven subcorpora of the *Digital Library of Old Spanish Texts* published by the Hispanic Seminary of Medieval Studies. Launched in 2018, the corpus allows interactive access to semi-paleographic transcriptions of texts produced in the Americas during the colonial period, a textual type that is under-represented in existing electronic corpora. The rationale of the project is provided, as well as the criteria for the selection of texts to be included and their method of preparation. Finally, the interface of the corpus is illustrated, and its functionality is exemplified.

Keywords – electronic corpus; *Digital Library of Old Spanish Texts*; colonial texts; Colonial Spanish

1. INTRODUCTION

The *Colonial Texts Corpus* is one of eleven subcorpora of the *Digital Library of Old Spanish Texts* published by the Hispanic Seminary of Medieval Studies.[1] This paper provides an overview of the *Corpus of Colonial Texts* project, including the rationale behind its inception, the criteria established for the selection of texts, and the methodology employed in their preparation. Likewise, a brief history of the construction of the corpus is provided, as well as an illustration of its interface and examples of its functionality. Before describing the present project, it would be beneficial to contextualize it within the framework of other digital projects undertaken by the Hispanic Seminary of Medieval Studies.

[1] See http://www.hispanicseminary.org

2. BACKGROUND OF THE *DIGITAL LIBRARY OF OLD SPANISH TEXTS*[2]

The *Digital Library of Old Spanish Texts* (DLOST) is an online resource prepared by the Hispanic Seminary of Medieval Studies (HSMS, or the Seminary), a non-profit publisher that grew out of the Seminario de Estudios del Español Medieval. The latter was founded at the University of Wisconsin-Madison in 1931 by Professor Antonio García Solalinde, a renowned medieval philologist and disciple of Ramón Menéndez Pidal. HSMS has been a trailblazer in the use of digital technology in the humanities. In the early 1970s, then HSMS directors, Lloyd A. Kasten and John J. Nitti, began using computers as an important tool for the compilation of dictionaries and the analysis of texts. For their *Dictionary of the Old Spanish Language* project, they eschewed the use of modern editions of medieval texts as the source material, demanding that the primary sources be as free from editorial bias as possible. They created a data bank with machine-readable transcriptions of all the texts that would eventually be incorporated into the dictionary. In 1978, the HSMS published its first texts on microfiche, in what was to become the well-known *Texts and Concordances* series.

By 1997, HSMS had begun publishing the *Texts and Concordances* on CD-ROM. Although the new physical support allowed for easier access to the transcriptions (e.g. dedicated microfiche readers were no longer needed), the texts and concordances were still non-interactive flat files, which did not allow scholars to take advantage of their full range of possibilities. In 2005, the Seminary began exploring the possibility of offering all of its textual archives in an online format. These efforts culminated in the *Digital Library of Old Spanish Texts*, launched in 2011 with the publication of the *Prose Works of Alfonso X el Sabio*. This open-access repository preserves the original structure of the HSMS texts, but allows for a truly interactive access to the semi-paleographic transcriptions, as well as to a series of indexes (alphabetical, frequency, reverse alphabetical), and concordances in KWIC format.[3] It is to be noted that DLOST is not a digital corpus like the *Corpus Diacrónico del Español* (CORDE), for example, but rather a digital library organized into subcorpora, grouped according to author, subject, dialect, geographic region, or literary genre. Researchers are able to perform some basic linguistic

[2] This overview is based on Gago Jover (2011, 2015). Other sources are cited where appropriate.

[3] A Key Word in Context (KWIC) concordance is a listing of all the words that occur in a text; each key word is shown within its immediate context, i.e. with forms both to the left and to the right of the key word, with a reference to where it appears (folio and line).

searches of the contents of the texts, within individual texts or within each subcorpus.[4] The principal aim of DLOST is to facilitate access to the more than 400 transcriptions published by the Seminary since 1978, with the indices and concordances being the principal means of access to the texts. By 2017, ten subcorpora had been published on DLOST, representing a total of 346 texts with nearly twenty-eight million tokens of data.[5]

3. The *Colonial Texts Corpus*

The *Corpus of Colonial Texts* (CCT) project represents the logical next step for the *Digital Library of Old Spanish Texts*. Given the constraints of time and resources, only Peninsular medieval and early modern texts had been converted to the online format prior to the inception of the present project. HSMS' *Colonial Spanish American Series*, which includes some nine works, had not been incorporated into the repository. With the *Colonial Texts Corpus*, we intend to greatly expand the Seminary's publications related to colonial Spanish America.[6] We describe in detail the parameters of the corpus below and provide a brief history of its construction.

3.1. Rationale and objectives

The goal of our project is to produce a corpus of philologically rigorous transcriptions of Spanish colonial texts and incorporate them into the Seminary's DLOST, a publication medium that will enable open, interactive access to the texts in an online format. The overarching impetus of the project is to provide reliable primary sources to inform the history of the Spanish language during the colonial period. Despite the recent advances in the availability of electronic corpora from which to extract empirical data to perform such studies, the low number of texts from Latin America included in these corpora is

[4] A lemmatized database with advanced search capabilities, which will include all HSMS texts, is in preparation. This is the *Old Spanish Textual Archive*, or OSTA (see Gago Jover and Pueyo Mena 2018a, 2018b).

[5] These are, in order of publication: *Prose Works of Alfonso X el Sabio*; *Spanish Medical Texts*; *Navarro-Aragonese Texts*; *Spanish Legal Texts*; *Spanish Biblical Texts*; *Spanish Poetic Texts*; *Early Celestina Texts*; *Spanish Chronicle Texts*; *Lazarillo de Tormes (1554) Texts*; *Fuero General de Navarra Texts*. Full bibliographic information can be found in Gago Jover (2011).

[6] As is the prevalent practice in the United States and elsewhere, we use the term 'colonial' as a descriptor relating to the territories of Latin America that maintained political ties with Spain during the period 1492 to 1898. Our use of the term is in no way pejorative, but rather a means of encompassing the wide variety of administrative structures that existed during the time period, including viceroyalties, captaincies, etc. (see Bethell 2002).

striking. For example, the Real Academia Española's CORDE, a corpus which spans the beginning period of the language until 1974, contains a textual archive in which only 6% of texts are from Latin America. The texts of the *Corpus Hispánico y Americano en la Red: Textos Antiguos* (CHARTA) network, a project aimed at publishing texts from Spain and Latin America from the twelfth to the nineteenth centuries, has 8%. While we recognize that temporal and geographic criteria limit the pool of Latin American texts, even in Davies' (2002–) *Corpus del Español* only 16% of the texts dated 1500–1900 are from Latin America.[7] Considering the fact that 90% of Spanish speakers reside in the Americas, the lack of representative texts needs to be addressed.

In the area of Colonial Spanish studies, we are fortunate that Spain's colonizing enterprise has left us with a plethora of primary documentation. Nevertheless, many of the seminal texts from the period only reach the public via modern editions or, as is the case of the documentary record of the U.S. Hispanic Southwest, in the form of English translations (cf. Craddock 2015). This has subjected the original texts to biases, including misreadings and mistranslations. A case in point can be drawn from one of the texts of our corpus, *Relación de la Jornada de Cíbola* by Pedro de Castañeda de Najera, which offers an eyewitness account of the Coronado expedition of 1540–1542. The *Relación* survives in a copy from 1596; the classical rendition of the text is Winship (1896). While the latter's edition and translation are of obvious historical interest, the transcription conflicts with current standard philological practice in several respects. For instance, no indications are given for folio numbers in the original manuscript, abbreviations are not adequately explained, and punctuation is not included. Most importantly, there are also numerous instances of transcription errors. In the first paragraph of the text alone, there are three mistakes: *las cosas e casos* (fol. 1r6) 'the things and cases' is transcribed as *las cosas acasos* (1896:414), *aquella* (fol. 1v16) as *aquello* (415), and *no le faltara de que dar relaçion* (fol. 2r10–11) 'will not be lacking [material] about which to provide an account' as the nonsensical no le *faltara de quedar relaçion* (415).

For historical linguists, who must find their evidence in orthographic cues, even more benign editorial interventions, such as spelling modernizations, can render the texts virtually useless for their purposes. Cortés' *Cartas de Relación* provide illustrative examples of the importance of scrupulously maintaining the orthography of the primary

[7] The data presented above are taken from Company Company 2019.

text.[8] One of the authoritative editions of Cortés' texts is Delgado Gómez (1993). It is based on the Vienna Codex with variants noted, except those of a phonetic nature. Delgado Gómez (1993: 100–102) loosely interprets what is considered phonetic, modernizing much of the spelling, including variations between /e/ and /i/, whereby *seguio* is represented as *siguió*, between *b* and *v* (*biven* becomes *viven*), and between *ç* and *z* (*dezir* > *decir*). Likewise, the use of *h* is regularized (*artos* becomes *hartos*), double *ss* is modernized to *s*, whereby all imperfect subjunctive verbs in *-sse*, for example, are spelled *-se*, and even *x* becomes *j* (*dixeron* > *dijeron*). These changes obscure data related to some of the most important phonological developments of the language during the fifteenth and sixteenth centuries, including variation between atonic vowels, the merger of /b/ and /β/, the devoicing of the sibilants, the loss of /h/ in words that descended from Latin F-, and the retraction of the articulation of Old Spanish /ʃ/ to Modern Spanish /x/ (see Lapesa 1981; Penny 2002; Torrens Álvarez 2018). For this reason, paleographic editions, which faithfully represent the language of the originals, are more reliable.[9]

Equally important is the issue of accessibility—Old Spanish texts are usually preserved in libraries and archives that require special access. Even when open access to texts is provided through digital means, non-specialists are not often equipped to decipher the handwriting or typescript of the text. There is thus a critical need for faithfully edited primary sources of colonial Spanish America that can be accessed by a variety of users. In the absence of such documentary sources, we will be unable to further our knowledge of the language of the period, of its concomitant cultural manifestations, and of the history it tells.

3.2. Scope: Temporal, geographic, and typological

Texts to be included in the *Colonial Texts Corpus* will be those written in any area of the Americas during the colonial period, 1492 to Independence. Given the varied chronology of the independence movements by country, the end date will depend on the area involved, for example, 1821 for Mexico but 1898 for Cuba. Texts with an original

[8] Cortés is said to have written five *cartas de relación*, or official reports that he sent to Charles V regarding the conquest of Mexico. The first *carta* was either lost or never existed; in editions of the *Cartas de Relación*, the *Carta de Veracruz*, written by members of the town council in 1519, takes its place. The *cartas* survive in the Vienna Codex, which includes all five letters, and the Madrid codex, which includes the four *relaciones*. See Delgado Gómez (1993).

[9] For other examples of why we need reliable editions of colonial texts, see Craddock and Polt (2008).

production date (OPDT) and a specific production date (SPDT) that both fall within the colonial period are preferred.[10] Until the arrival of the printing press in Mexico in 1539 and its subsequent spread to other areas of the Americas, many early colonial texts were printed in Spain. Therefore, place of composition will be loosely construed as 'American' for texts that are closely related to colonial Latin America but which may have been copied or published elsewhere. This is especially relevant for texts from the sixteenth century. For example, Cortés' *Cartas de Relación* were written in Mexico. Although the originals are lost, the texts are extant in manuscript copies (see Section 3.1). Three survive in early imprints published in Spain.[11] Likewise, the *Relación de la Jornada de Cíbola* was composed in San Miguel de Culiacán, Mexico, but survives in a copy produced in Seville in 1596.

Texts to be included in the corpus will be of a wide variety, both verse and prose. Although we recognize the value of archival materials for studying the historical development of the language, brief notarial documents will not form part of the corpus.[12] Our focus is on texts of a more extensive narrative nature, which will serve as source material not only for DLOST, but also for OSTA. The following serve as examples of the ideal types of texts to be included in the corpus: chronicles, *memoriales*, *relaciones*, official letters, travel narratives, as well as works of a religious or literary nature. Legal texts that form part of a larger whole will also be included, for example, judicial proceedings, as will personal letters forming part of a larger narrative bundle.

3.3. Methodology

The texts of the corpus will be transcribed according to the guidelines established by the Seminary in Mackenzie (1997). HSMS' semi-paleographic transcription system attempts to replicate, to the extent possible, various details related to the format and appearance of the text: folio and column number, original spelling, abbreviations and their resolution,

[10] The OPDT refers to the date that the text was originally produced while the SPDT refers to the date of the production of the specific manuscript copy or imprint. For example, internal evidence shows that the *Relación de la Jornada de Cíbola* was written sometime after the death of Joanna of Castile, so its OPDT is 1555 *a quo*; its SPDT is 1596, the date of the extant copy. See Faulhaber (1997–) regarding the dating of texts.

[11] These are the second, third, and fourth *relaciones*, published in 1522, 1523, and 1525, respectively (2CR, 3CR, and 4CR of the *Colonial Texts Corpus*; see Appendix).

[12] A noteworthy project that includes texts of this type is the *Corpus Diacrónico y Diatópico del Español* (CORDIAM), which deals exclusively with texts from Latin America. Archival documents are included in the subcorpus CORDIAM-Documentos.

upper- vs. lower-case letters, rubrics, glosses, headings, catch words, scribal errors and emendations, as well as editorial interventions (Gago Jover 2015). This allows the reader to reconstruct the format and appearance of the original text, ensuring philological integrity.

Contributors to the CCT project will edit their texts following philological best practices. Typically, the scholar will work from a digital facsimile and, when feasible, will correct the initial transcription by comparing it to the original text in the library or archive in which it is housed. The publication will follow the *Texts and Concordances* framework of the HSMS, with optional introduction, the transcribed text, the indices, and the concordances. These will be published in an open-access format on DLOST in the *Colonial Texts Corpus*. A link to digital images of the text will also be provided when available.

This methodology distinguishes the *Colonial Texts Corpus* from other corpora in important ways. First, all texts in the corpus are transcribed using the same editorial criteria. Other corpora, such as CORDE and Davies (2002–), incorporate texts that were edited using a wide variety of criteria—from paleographic transcriptions of a single manuscript or imprint to critical editions that reconstruct evidence from multiple extant versions of a text. Moreover, the present corpus eschews the inclusion of modern editions in which orthography is regularized, contrasting in this way with the two corpora cited above, as well as with CORDIAM.[13] The *Colonial Texts Corpus* provides access to a specific manuscript or imprint, with minimal editorial intervention. The corpus also employs uniform chronological criteria, giving preference to the SPDT over the OPDT. In contrast, other corpora prioritize the OPDT. In Davies (2002–), for example, fifteenth-century copies of Alfonsine texts are included in the database as thirteenth-century source material. The features of the *Colonial Texts Corpus* highlighted above allow researchers to extract reliable data with which to perform contrastive analyses, comparing apples to apples, as it were.[14]

[13] CORDE, for example, uses the modern edition by Hernández (1988) of Cortés' *Cartas de Relación*. CORDIAM makes use of modern editions in the subcorpus CORDIAM-Literatura, which includes chronicles as well as other textual types.

[14] See Gago Jover (2015: 10) for references to projects that use data from DLOST. To these can be added three lexical studies in progress whose data regarding indigenous loanwords, semantic extensions, and Arabisms largely derive from the *Colonial Texts Corpus*.

3.4. Current status of project

Preparation of the corpus began in 2017. After the parameters above had been determined, the principal investigators began to construct the beta version of the webpage. The initial nucleus of texts consisted of existing transcriptions from the Colonial Spanish Series of the HSMS which fit the established typological criteria. These were CIB, PMZ, and RVC (see Appendix). With these three, COL was included (this transcription was among the HSMS textual archives but had not been published), which brought the initial nucleus to four texts representing 200,799 tokens of data. The first texts that were added to the *Colonial Texts Corpus* were 2CR and VCC. When the project was launched in 2018,[15] the textual archive consisted of six texts (305,510 tokens). At present, the corpus consists of eleven texts (512,590 tokens) and will be continuously expanded. Collaborators in the project currently have eight additional texts in preparation, with another dozen in the planning stages.

3.5. Interface

As seen in Figure 1, the initial window displays the **navigation menu** ① and the name of the corpus of texts.

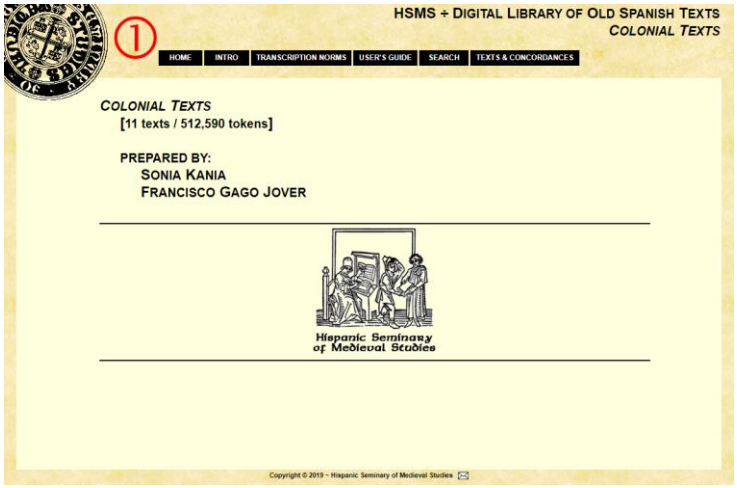


Figure 1: Initial window

The **navigation menu** ① provides quick access to:

a. **Home:** The initial page, with general information on the texts and concordances.

[15] The project was formally launched at the *XI Congreso Internacional de Historia de la Lengua Española* in Lima, Peru, in August 2018.

b. **Intro:** A brief overview of the CCT project.
c. **Transcription norms:** A brief summary of the transcription coding used in the texts.
d. **User's guide:** A brief explanation of the different parts of the interface.
e. **Search:** The corpus search page.
f. **Texts & concordances:** The interactive indexes, concordances, and texts described below.

Clicking on the **texts & concordances** button in the **navigation menu** ① brings up a list of all the works included ② (see Figure 2).

Figure 2: List of texts

As Figure 3 illustrates, clicking on one of the works opens up the **information menu** ③ of the selected item, which provides detailed information on the text and its concordances:

a. **Title.**
b. **Author.**
c. **Translator.**
d. **Specific Production Date.**
e. **Original Production Date.**
f. **Place of Production .**
g. **Library:** Current location of the manuscript or imprint.
h. **Printer:** Name of printer.
i. **Transcribed by:** Name of the person(s) who transcribed the work.
j. **Corrected by:** Name of the person(s) who corrected the transcription.
k. **Lexical Studies:** Link to information in the *Lexical Studies of Medieval Spanish Texts* database (Dworkin and Gago Jover 2004–2018).

l. **BETA manid:** Manuscript ID number and link to the *Bibliografía Española de Textos Antiguos* (Faulhaber 1997–).

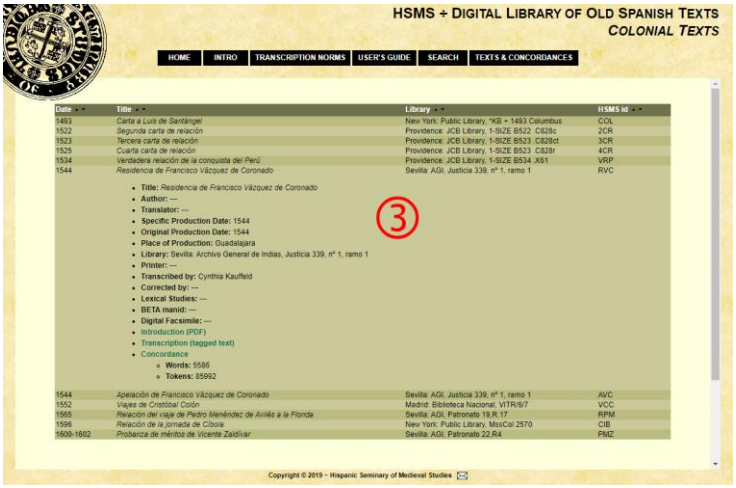
m. **Digital Facsimile:** Link to the digital facsimile of the work.

n. **Introduction (PDF):** Link to the introduction.

o. **Transcription (tagged text):** Link to the text with HSMS transcription tags.

p. **Concordance:** Link to the interactive indexes, concordances, and text.

- **Words:** Total number of unique word-forms in the text.
- **Tokens:** Total number of words in the text.



- Figure 3: Information menu

As shown in Figure 4, clicking on the **transcription (tagged text)** link brings up the following window, with the original transcription:

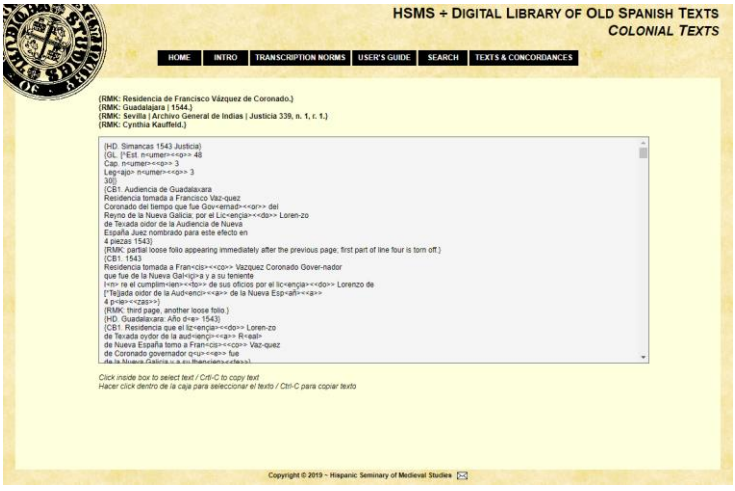


Figure 4: Tagged transcription

Clicking on the **concordance** link brings up the following window (Figure 5), made up of three sections:

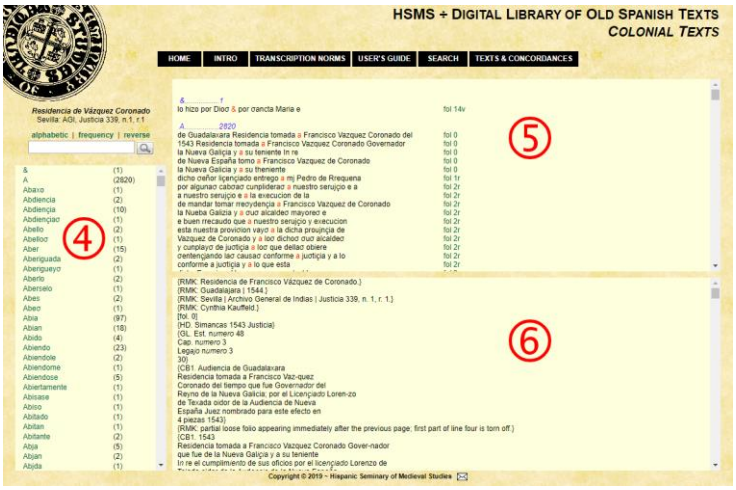


Figure 5: Wordlist, concordance, and text frames

At the left of the browser window is the **wordlist frame** ④, containing an **alphabetic list** of all words which are used in the source text. Clicking on a headword in the wordlist will make the **concordance frame** ⑤ scroll automatically to display all the instances of that headword, together with a line of context. The user can also select to see a **frequency** or a **reverse alphabetic** list. The **search box** allows the user to search for any word or combination of letters within each of the lists. The **concordance frame** ⑤ appears in the upper of the two large frames to the right of the wordlist. Beside each headword is a count of the number of times it occurs, and below it are all the occurrences, each in a line of context. To the right of each context line are the folio references. Clicking on a reference will make the **text frame** ⑥ scroll automatically to display the relevant part of the source text.

The text from which the concordance was made appears in the **text frame** ⑥, to the right of the wordlist. The scroll bars can be used to navigate in the text. To facilitate reading, the text is shown stripped of all transcription tags, with abbreviations resolved in italics, the combinations ***c'***, ***n~***, ***s'***, and ***z'*** as ***ç***, ***ñ***, ***σ***, and ***s***, respectively, and the *calderón* as ¶. In this way, the tagged transcription of the fragment in (1) of the *Carta a Luis de Santángel* (COL, fol. 1r) is shown with stripped tags (2):

(1) puse nonbre la isla de santa maria de[ ]concepcion ala tercera ferrandina ala quarta la isla [isa]bella | ala qui<n>ta la Jsla Juana e asi a cada vna nonbre nueuo Quando yo lleg($u)[u]e ala Juana seg- | ui io la costa della al poniente yla falle tan grande q<ue> pense que seria tierra firme la proui<n>cia de | catayo y como no falle asi villas y luguares enla costa dela mar saluo pequen~as poblaciones [...]

(2) puse nonbre la isla de santa maria de concepcion ala tercera ferrandina ala quarta la isla isabella | ala qui*n*ta la Jsla Juana e asi a cada vna nonbre nueuo Quando yo llegue ala Juana seg-ui | io la costa della al poniente yla falle tan grande q*ue* pense que seria tierra firme la proui*n*cia de | catayo y como no falle asi villas y luguares enla costa dela mar saluo pequeñas poblaciones [...]

3.6. Functionality

It is possible to search within a text or the entirety of the corpus. For the first type of search (within a single text) use the **text box** ⑦ as displayed in Figure 6. The search is performed in the selected index (alphabetic, frequency, or reverse); it is possible to anchor the search string to the beginning or the end of a word by using a bar (/), for example, **/aceit**, **ndos/**.

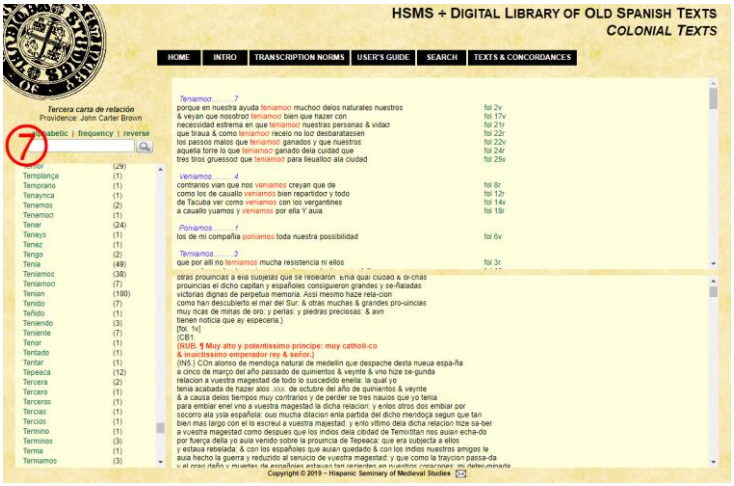


Figure 6: Search within a single text

As shown in Figure 7, to search the entirety of the corpus, click on the **search** button in the **navigation menu** ① to bring up the **search window** ⑧.

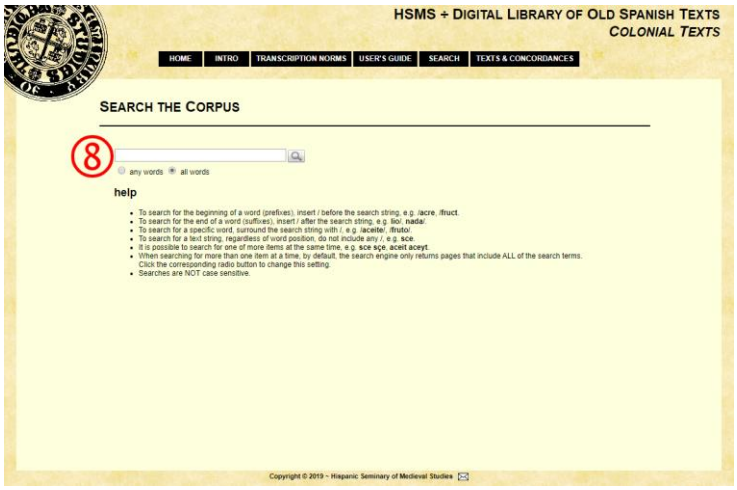


Figure 7: Search within the corpus

To search, type the search string in the text box.

- To search for the beginning of a word, insert / before the search string, e.g. **/acre**, **/fruct**.
- To search for the end of a word, insert / after the search string, e.g. **lio/**, **nada/**.
- To search for a specific word, surround the search string with /, e.g. **/aceite/**, **/fruto/**.
- To search for a text string, regardless of word position, do not include /, e.g. **sce**.
- It is possible to search for one or more items at the same time, e.g. **sce sçe**, **aceit aceyt**.
- When searching for more than one item at a time, by default, the search engine only returns pages that include all of the search terms. Click the corresponding radio button to change this setting.
- Searches are not case sensitive.

Searches are performed in the entirety of the corpus, and the search results page ⑨ shows all the texts in which the search string appears (see Figure 8). Clicking on the title brings up the concordances of the corresponding text.

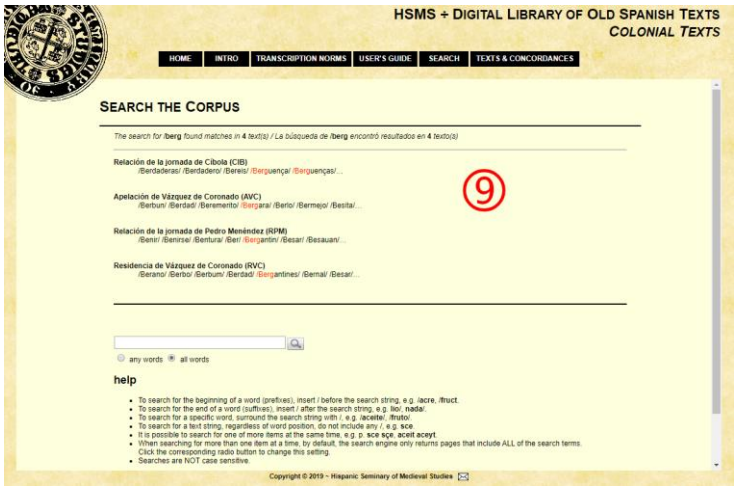


Figure 8: Search results page

Considering the nature of the *Colonial Texts Corpus*, more a repository of texts than a full-fledged linguistic corpus, and despite the limitations of the search engine, it is still possible to perform various types of queries. For example, looking at the distribution of forms such as *fijo(s)* / *hijo(s)*, one can observe that forms with initial *f-* only appear, in alternation with forms with *h-*, in 2CR, 3CR, and VCC, while the rest of the texts only present forms with *h-*. Looking at *fuesse* vs. *fuese* in all the texts, three of the eleven documents both *fuesse* and *fuese* (VCC, 4CR, PMZ), one (2CR) uses only *fuesse*, and four (AVC, RVC, CIB, RPM) use only *fuese*. By using the concordances, it is possible to study the use of any form in a specific text; for instance, in VCC, when studying the alternation of forms of the imperfect subjunctive in *-se* / *-sse*, a greater use of forms in *-se* (269 tokens: 94.1%) is observed compared to the forms in *-sse* (17 tokens: 5.9%). The preference for forms in *-se* can also be seen in the three verbs where both forms, *-se* / *-sse*, are present, as shown in (3).

(3) dixese (4) / dixesse (1)
fuese (48) / fuesse (5)
llegase (6) / llegasse (2)

4. CONCLUDING REMARKS

This article has presented a new subcorpus of the *Digital Library of Old Spanish Texts*, namely the *Colonial Texts Corpus*. The aim of the *Corpus of Colonial Texts* project is to provide open, interactive access to a corpus of texts that are transcribed using philologically rigorous criteria. This project therefore responds to the need for reliable primary sources related to colonial Latin America, a textual type that is under-represented

in existing electronic corpora. In this way, we will extend the reach of DLOST to a new group of users, i.e. scholars of colonial Latin America. Finally, this on-going project contributes to our broader goal of preserving the linguistic, cultural, and literary history of Spanish in the Americas.

References

Bethell, Leslie ed. 2002. *América Latina en la Época Colonial*. Vol. 1, *España y América de 1492 a 1808*. Barcelona: Crítica.

CHARTA = *Corpus Hispánico y Americano en la Red: Textos Antiguos*. http://www.corpuscharta.es

Company Company, Concepción. 2019. Voces e historia conceptual. Contribución a la construcción identitaria del español en América. Plenary talk given at the *Jornadas de Investigación El léxico americano en su historia: análisis y perspectivas de estudio*, Universidad de Querétaro, Querétaro, Mexico, October 2019.

CORDE = Real Academia Española: Banco de datos (CORDE) en línea. *Corpus Diacrónico del Español*. http://www.rae.es

CORDIAM = Academia Mexicana de la Lengua: *Corpus Diacrónico y Diatópico del Español*. http://www.cordiam.org

Craddock, Jerry R. 2015. The Cíbola Project: Mission statement and staff. *UC Berkeley: Research Center for Romance Studies*. https://escholarship.org/uc/item/3jt748vt (7 January, 2020.)

Craddock, Jerry R. and John H. R. Polt. 2008. An object lesson: Why we need good editions of the documents of the Hispanic Southwest. *UC Berkeley: Research Center for Romance Studies*. https://escholarship.org/uc/item/6w33k9v5 (7 January, 2020.)

Davies, Mark. 2002–. *Corpus del Español: 100 million words, 1200s–1900s*. http://www.corpusdelespanol.org/hist-gen/

Delgado Gómez, Ángel ed. 1993. *Hernán Cortés, Cartas de Relación*. Madrid: Clásicos Castalia.

Dworkin, Steven N. and Francisco Gago-Jover. 2004–2018. *Lexical Studies of Medieval Spanish Texts*. Hispanic Seminary of Medieval Studies, *La Corónica: A Journal of Medieval Hispanic Languages, Literatures, and Cultures*. http://www.hispanicseminary.org/lsmst/index.htm (7 January, 2020.)

Faulhaber, Charles B. dir. 1997–. *BETA (Bibliografía Española de Textos Antiguos)*. The Bancroft Library. University of California, Berkeley. http://vm136.lib.berkeley.edu/BANC/philobiblon/beta_en.html (7 January, 2020.)

Gago Jover, Francisco ed. 2011. *Digital Library of Old Spanish Texts*. Hispanic Seminary of Medieval Studies. http://www.hispanicseminary.org/textconc-en.htm (7 January, 2020.)

Gago Jover, Francisco. 2015. La *Biblioteca Digital de Textos del Español Antiguo* (BiDTEA). *Scriptum Digital* 4: 5–36.

Gago Jover, Francisco and F. Javier Pueyo Mena. 2018a. El *Old Spanish Textual Archive*, diseño y desarrollo de un corpus de textos medievales: El corpus textual. *Cuadernos del Instituto Historia de la Lengua* 11: 165–209.

Gago Jover, Francisco and F. Javier Pueyo Mena. 2018b. El *Old Spanish Textual Archive*, diseño y desarrollo de un corpus de textos medievales: Lematización y etiquetado gramatical. *Scriptum Digital* 7: 25–35.

Hernández, Mario ed. 1988. *Hernán Cortés, Cartas de Relación*. Madrid: Historia 16.
Lapesa, Rafael. 1981. *Historia de la Lengua Española* (ninth edition). Madrid: Gredos.
Mackenzie, David. 1997. *A Manual of Manuscript Transcription for the Dictionary of the Old Spanish Language* (fifth edition by Ray Harris-Northall). Madison: Hispanic Seminary of Medieval Studies.
Penny, Ralph. 2002. *A History of the Spanish Language* (second edition). Cambridge: Cambridge University Press.
Torrens Álvarez, María Jesús. 2018. *Evolución e Historia de la Lengua Española* (second edition). Madrid: Arco Libros.
Winship, George P. ed. and trans. 1896. *The Coronado Expedition, 1540–1542*. Bureau of Ethnology, Smithsonian Institution, Annual Report, 14. Washington, D.C.: Government Printing Office.

APPENDIX[16]

COL (1493): *Carta a Luis de Santángel*. New York: Public Library, *KB + 1493 Columbus.
2CR (1522): *Segunda Carta de Relación*. Providence: JCB Library, 1-SIZE B522 .C828c.
3CR (1523): *Tercera Carta de Relación*. Providence: JCB Library, 1-SIZE B523 .C828ct.
4CR (1525): *Cuarta Carta de Relación*. Providence: JCB Library, 1-SIZE B523 .C828r.
VRP (1534): *Verdadera Relación de la Conquista del Perú*. Providence: JCB Library, 1-SIZE B534 .X61.
RVC (1544-1545): *Residencia de Francisco Vázquez de Coronado*. Sevilla: AGI, Justicia 339, nº 1, ramo 1.
AVC (1544–1545): *Apelación de Francisco Vázquez de Coronado*. Sevilla: AGI, Justicia 339, nº 1, ramo 1.
VCC (1552): *Viajes de Cristóbal Colón*. Madrid: Biblioteca Nacional, VITR/6/7.
RPM (1565): *Relación del Viaje de Pedro Menéndez de Avilés a la Florida*. Sevilla: AGI, Patronato 19, R.17.
CIB (1596): *Relación de la Jornada de Cíbola*. New York: Public Library, MssCol 2570.
PMZ (1600–1602). *Probanza de Méritos de Vicente de Zaldívar*. Sevilla: AGI, Patronato 22, R.4.

Corresponding author
Sonia Kania
Department of Modern Languages
230 Hammond Hall, Box 19557
701 Planetarium Place
Arlington TX 76019
e-mail: skania@uta.edu

received: January 2019
accepted: March 2020

[16] Texts are listed in chronological order. Information provided includes three-character HSMS id, SPDT, title, library (preceded by city), and call number.