

Developing Geographically Oriented NLP Approaches to Sixteenth-Century Historical Documents: Digging into Early Colonial Mexico

Diego Jiménez-Badillo <diego_jimenez_at_inah_dot_gob_dot_mx>, Museo del Templo Mayor, Instituto Nacional de Antropología e Historia

Patricia Murrieta-Flores <p_dot_a_dot_murrieta-flores_at_lancaster_dot_ac_dot_uk>, Digital Humanities Hub–History Department, Lancaster University

Bruno Martins <bruno_dot_g_dot_martins_at_tecnico_dot_ulisboa_dot_pt>, The Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa, INESC-ID, University of Lisbon

Ian Gregory <i_dot_gregory_at_lancaster_dot_ac_dot_uk>, Digital Humanities Hub–History Department, Lancaster University

Mariana Favila-Vázquez <mariana_dot_favila_at_gmail_dot_com>, Museo del Templo Mayor, Instituto Nacional de Antropología e Historia

Raquel Licerias-Garrido <r_dot_liceras_at_lancaster_dot_ac_dot_uk>, Digital Humanities Hub–History Department, Lancaster University

Abstract

This article introduces an ongoing Digital Humanities project aimed at leveraging the benefits of Natural Language Processing, Corpus Linguistics, Machine Learning, and Spatial Analysis for advancing the computational analysis of vast historical corpora. As a case study, the project concentrates on the *Relaciones Geográficas de la Nueva España* (1577–1585), one of the key corpora for understanding the early colonial period of Mexico.

Using a computer-assisted methodology called Geographical Text Analysis (GTA), the project offers automatic means for parsing historical texts and the markup of words referring both to place names (toponyms) and analytical concepts that are then linked to their geographic locations. Adding geospatial intelligence to the parsing of texts allows exploring hidden geographies and narratives in the historic corpus. The article provides a general overview of the corpus, describes the GTA methodology step by step, and reports on the progress achieved so far.

1. Introduction

This article introduces the project *Digging into Early Colonial Mexico: A Large-Scale Computational Analysis of Sixteenth-Century Historical Sources*, and describes the progress achieved so far.^[1] The project focuses on developing novel methods and tools for mining new knowledge from textual and pictorial historical records to facilitate the study of early Latin American history. It concentrates on one of the most important corpora related to the geography, culture, religion, economy, and history of native and Spanish peoples in Mexico and Guatemala, namely the *Relaciones Geográficas de la Nueva España* (Geographic Reports from New Spain), and specifically those compiled between 1577 and 1585, hereafter referred to as RGs.

The proposed methodology — called Geographical Text Analysis (GTA) — enables the automatic markup of words referring both to place names (toponyms) and analytical concepts relevant to historical research, and adds geospatial intelligence to the parsing of texts. This process is useful not only to find the most salient terms in textual corpora or to identify and disambiguate personal and geographical names, but also to facilitate the discovery of unsuspected data patterns and relationships, such as spatial correlation of events, interaction of people, links of persons and themes to places, and many others.

One of the main contributions of this project has been the development of an analytical model based on deep-learning

1

2

3

and adapted to the specific challenges found in texts like the RGs. In order to achieve named entity recognition, disambiguation and classification, GTA employs methodologies from the fields of Natural Language Processing (NLP), Corpus Linguistics (CL) and Machine Learning (ML). Although NLP, CL and ML techniques are used frequently in textual analysis today, they are most commonly applied to contemporary texts composed in English. Little work has been done to adapt their use to historical corpora written in old language forms like the sixteenth-century Spanish.^[2] The RG corpus additionally contains other features that complicate the use of such tools, which makes very difficult the automatic “reading” of the documents. These include the presence of multi-lingual expressions in Nahuatl, Zapotec, Otomi, and many other indigenous languages [Harvey 1972], as well vocabulary only used in certain regions at the time.

A second contribution of our project has been to facilitate, using Geographical Information Systems (GIS), the automatic linking of the results of these processes of textual analysis with locational data. This combination of NLP, CL and ML techniques with spatial analysis methods creates powerful resources to discover unsuspected geographies in the textual reports. Our model will allow scholars of colonial Latin America to learn new information, extract it, cross-link it to other document resources, and, most importantly, study significant historical topics, such as those we discuss in section 3.5 below.

Apart from the GTA model itself, we foresee several other products of this project. The first is a digital version of the RGs corpus, annotated in accordance with analytical categories that can be applied to other corpora through an interoperable ontology. The second is a digital gazetteer containing geographic coordinates as well as political, cultural, administrative and physical data, not only of the sixteenth-century sites mentioned in the RGs, but from other relevant sources and historical periods. The third is a set of software tools that link the spatial information contained in the gazetteer with the textual and pictorial contents of the RGs in order to facilitate the discovery and analysis of geographies not immediately apprehended with normal reading.

In writing this paper, our goal is to raise awareness about the possibilities of applying computer-assisted methods to the analysis of historical texts. Thus, we describe every step of GTA in the context of our specific study case, hoping that this will encourage more Humanities researchers to adopt this kind of approach.

The rest of the article is organized as follows: Section 2 provides an overview of the *Relaciones Geográficas* corpus; Section 3 describes our analytic methodology and reports on our progress; Section 4 describes future applications to the study of history; and the final section offers a general conclusion.

2. The Sixteenth Century Relaciones Geográficas de la Nueva España

The compilation of the *Relaciones Geográficas de la Nueva España* was part of a large, long-standing project, sponsored first by Emperor Charles V and then by King Philip II, who sought to collect geographic, economic, historic and cultural information from all the kingdoms and territories under their rule. The initiative emerged not only from the need to control the exploitation of resources and resolve important issues such as land possession, management of labor and the transfer of privileges to colonists, but also to compile scientific knowledge and fulfill intellectual curiosity [Delgado López 2010] [De la Garza 1983]. This involved both collecting descriptions and making maps [Mundy 1996, 1–27]. The whole project included Castile, Aragon, Italy, the Netherlands, Portugal (since its annexation in 1581), and the viceroyalties of Peru and New Spain.

For the New World enterprise, the Spanish king, through the Council of Indies, commissioned his leading cosmographers to gather information by means of questionnaires. These were conceived by Juan de Ovando and sent to the Indies in 1569 (37 questions), 1570 (200 questions) and 1573 (135 questions). The responses that came back to Madrid, however, were scarce, probably because the burden of collecting data was put on high-level ecclesiastical and government officials who lacked the necessary knowledge or were not willing to answer so many questions.

The project was revived by Juan López de Velasco, cosmographer-chronicler of Philip II, who in 1577 devised more feasible enquiries. The first one attempted to gather observations about moon eclipses from various parts of the Indies [Cline 1964]. This would allow solving the elusive problem of determining longitude coordinates for the American territories, which in turn would make it possible to establish correctly the position of the Indies with respect to Europe in

a universal map [Edwards 1969] [Mundy 1996, 17–9]. A second independent enquiry sought to complete the scarce descriptions received in the previous years, sending only fifty carefully formulated questions to be answered not by high-level authorities, as in previous attempts, but by *corregidores* and *alcaldes mayores*, the local administrative officials with direct knowledge of the situation in towns, villages and hamlets, who also received printed instructions on how to answer the survey. Figure 1 shows the first page (out of three) of the printed *Instrucción y Memoria* sent as a guide to the respondents.

The questionnaire included very detailed information about the life in New Spain. One single query might cover topics such as demography, type of settlements, intercultural opinions, and ethnolinguistics.^[3] There are also questions about local history, including the names and settlement of towns, and about agricultural and mineral resources. See Appendix 1 for a complete translation of the questionnaire.

Responses to the questionnaire were produced between 1579 and 1585 and are now held in Spain, Scotland, and the United States. In total there are 168 written reports corresponding to an equal number of main *cabeceras* (“head-towns” or main districts). These also contain descriptions of 248 subordinate jurisdictions, 414 towns and many smaller villages and hamlets [Cline 1964]. Together they cover more than half the political subdivisions existing in New Spain circa 1580 [Cline 1972c].^[4] Figure 2 reproduces one of the manuscript pages of the *Relación de Atlatlauhca* (present-day Atlatlauhacán, Morelos).

Alongside the texts, the corpus contains 78 maps. Some were painted by indigenous artists following native representations of places and landscape (Figure 3 and Figure 4), while others were clearly drawn by creoles or peninsular Spaniards (Figure 5). In the words of Robertson (1972: 243), they represent “the largest single group of interrelated early colonial pictorial manuscripts which has come down to us.”^[5] Of these maps, 23 came from Mexico; 3 from Michoacán; 1 from Nueva Galicia; 21 from Oaxaca; 18 from Tlaxcala; 3 from Yucatán and 2 from Guatemala. There are also accounts of 26 manuscripts and 16 maps lost centuries ago [Cline 1972c, Tables 6 and 10] [Robertson 1972], as well as an additional map that disappeared more recently [Mundy 1996, 226].

Despite the fact that some authors failed to answer several questions, both the texts and the maps offer favorable conditions to test new approaches for the analysis of large historical corpora. The corpus of RGs offers highly structured information — due to the standardized questionnaire — which gives the reports a degree of homogeneity rarely seen in other sixteenth-century collections. The geographic descriptions also cover more than half the territory of New Spain and the level of detail is remarkable not only in terms of orography, landscape, and natural resources, but also the way of life of native peoples and colonists, as well as the interactions between different social groups. Such richness, consistency, and span facilitate comparative research across small, medium, and large areas on topics of social, historical, ethnographic, linguistic, economic, and political interest.

A further reason for selecting the RGs is a desire to improve the accessibility and valorization of the corpus as a main source for historical enquiry. Seminal research has been conducted by Cline (1972), Bravo-García (2018), Gerhard (1972), Gruzinski (1991), Moreno Toscano (1968), Mundy (1996) and a few other scholars [Cáceres Lorenzo 2013] [Delgado López 2010]. The publications by Cline (1972) and Gerhard (1972), for example, were ground-breaking works of Mexican historical geography. Moreno Toscano (1968), on the other hand, used the complete textual information of the RGs to perform a matrix analysis oriented to reconstruct the economic geography of New Spain, whilst Mundy (1996) analyzed all the surviving maps to identify continuities and changes in the indigenous cartographic tradition during the colonial period. For his part, Gruzinski (1991: 77-103) has approached the corpus by studying how the contrasting — sometimes conflicting — worldview of conquistadors and native peoples shaped the responses to the questionnaire, pointing out interesting details about the colonization of the imaginary.

Single documents of the *Relaciones* have also been the focus of historical research [Afanador Pujol 2015] [Bellesteros García 2005] [Barlow 1949] [Fernández Christlieb 2006] [Morato-Moreno 2017] [Mundy 2013] [Nuttall 1926] (among others), but in general the textual and cartographic documents have served as ancillary sources in regional studies, rather than as a principal mine of data [Dahlgren de Jordan 1954] [Brand 1960] [Gibson 1967]. This is due perhaps to the difficulty of cross-linking information among the many documents or to a lack of tools for uncovering non-explicit

knowledge hidden in the subtext of the answers to the questionnaire.

3. Methodology and current progress

As discussed in the introduction, we are developing advanced tools for unveiling new information from the RGs corpus by applying a methodology known as Geographical Text Analysis (GTA). The resulting digital resources and software tools will be released at the end of the project under an open-access license. The methodology was originally produced by some members of this project to reveal new, unsuspected patterns of information by focusing on the geographic components of the historical narrative [Cooper and Gregory 2011] [Murrieta-Flores et al. 2015] [Donaldson et al. 2015] [Gregory et al. 2015] [Porter et al. 2015] [Martins and Murrieta-Flores 2017] [Murrieta-Flores et al. 2017] [Gregory et al. 2018].

We are now in the process of refining and extending this methodology with contributions from all members of this project and from researchers engaged in parallel initiatives (e.g., Santos et al. 2017; Santos et al. 2018; Won et al., 2018). In this particular project we followed four phases: obtaining a computer-readable corpus, the creation of a gazetteer, the semantic annotation and geoparsing of the corpus, and data mining and historical analysis. In the following sections, we discuss each stage in detail.

3.1 First phase: Obtaining a computer-readable corpus

The first phase of the methodology is obtaining computer-readable files from the more comprehensive editions of the RGs. We began with Francisco del Paso y Troncoso's *Papeles de Nueva España*, which contains a significant portion of the corpus. Transcriptions from the original manuscripts were already available in electronic format, but it was necessary to identify and correct significant errors caused by the Optical Character Recognition (OCR) software. This is a common problem with old documents, to which Alpert-Abrams offers an approach in her 2016 article on the corpus known as *Primeros Libros*.

After completing six volumes of *Papeles de Nueva España* (Del Paso y Troncoso 1905), the work continued with the RGs from Yucatán, published in volume 11 of *Colección de documentos inéditos de las posesiones de España en ultramar* (1898). Additionally, thanks to a generous permission from the National Autonomous University of Mexico (UNAM), we were also able to digitize a more recent version of the 54 RGs from Yucatán transcribed and edited by De la Garza (1983) and to acquire in a digital format 114 RGs published in ten volumes by Acuña (1982–1988). In the end, we obtained two versions of the whole corpus in machine-readable format: the first one includes the notes, references and commentary from the scholars who transcribed and studied the documents, while the second contains only the transcription of the original RGs texts. A third version that includes all the NLP annotations was produced during the third phase of the project.

3.2 Second phase: Creation of a gazetteer

The second phase starts with a compilation of all the place names that exist in the corpus as well as many others from relevant sources, and then solving any linguistic and locational uncertainties that prevent their unequivocal identification.

3.2.1 Compilation of place names.

This is done by extracting toponym indices from published editions of the RGs, as well as from comprehensive studies on the historic geography of New Spain (Gerhard 1972a, 1972b, 1991; Cline 1972a, 1972b, 1972c; Moreno Toscano, 1968; Tanck de Estrada et al., 2005). Through this procedure, the UK and Mexican teams were able to compile a list of 14,654 place names (Liceras-Garrido et al. 2019).

3.2.2 Linguistic disambiguation.

The resulting list was then subjected to a linguistic disambiguation process, during which the most difficult challenges were dealing with place names found in different native languages, as well as resolving name variations and spelling inconsistencies. Important names of cities, like Tlaxcala, have remained in original form since the sixteenth century, but

many others have changed significantly over time for a variety of reasons, including: historic causes;^[6] renaming due to changes in the boundaries of original settlements;^[7] and addition of words such as the name of a patron saint to an indigenous toponym.^[8] Another problem relates to spelling variations of the same toponym,^[9] the application of Castilian transliterations to the names of ancient settlements,^[10] or dealing with homonymies.^[11]

State-of-the-art methods to tackle these challenges are discussed in Nick and Tent (2017) and Santos et al. (2015, 2017, 2018). Additionally, there is a valuable body of research on the subject of Mexican toponymy, so we have benefited from publications by Anaya Monroy (1960), Guzmán Betancourt (1987, 1989, 1998), Lefevre and Paredes Martínez (2017), León Portilla (2002), Márquez and Ramos Navarro Wold (1998), Mundy (2014), Muntzel and Villegas Molina (2010), etc. This adds to the exhaustive research done by the editors of the RGs, especially Acuña (1982–1988), De la Garza (1983) and Del Paso y Troncoso (1905).

3.2.3. Geographic disambiguation.

Another challenge at this stage is finding latitude and longitude coordinates of towns and villages. The solution involves, first, detecting and merging duplicate entries by comparing the place name list with records found in online geographical databases, such as the catalogue of indigenous localities, produced by Instituto Nacional de Estadística Geografía e Informática (INEGI) and Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO);^[12] the geographic dataset created for the cartography of Mexico by INEGI; the general registrar of archaeological sites compiled by Instituto Nacional de Antropología e Historia (INAH);^[13] the GeoNames geographical database;^[14] and the Getty Thesaurus of Geographical Names.^[15] Through both linguistic and locational disambiguation, we were able to obtain 12,748 correct names. Of these, we have so far been able to pin-point 3919 physical locations (Figure 6), which are the entries used for the gazetteer.

The structure of the gazetteer complies with Linked Open Data standards to guarantee its interoperability between different systems and repositories and facilitate its application into other projects. We have relied on the experience of massive reference resources like the Getty Thesaurus of Geographical Names^[16] and especially the Alexandria Digital Library Gazetteer (ADL), from which we adopted the relational database schema [Hill 2000] [Hill et al. 1999]. For importing and exporting data, we adopted the file format JSON, created by the World Historical Gazetteer^[17], which is now an accepted standard in many textual annotation and analysis platforms, such as Recogito. The database of our gazetteer will be made available through the World Historical Gazetteer, the Pelagios Network, and through Recogito. In addition, the gazetteer will be also available in a standalone downloadable GIS format from the project's website and repositories.

3.3 Third phase: Semantic annotation of the corpus

The third stage comprises four tasks: defining an annotation schema by selecting the most relevant analytic categories for the intended historical analysis; annotating manually, and in accordance with the previously defined annotation schema, two different samples of documents: one for training the NLP deep-learning model and another one to validate it; tuning the deep-learning model to automatically perform name entity recognition, disambiguation and classification, in order to obtain a fully annotated corpus; and assessing the performance of the model in accordance with the gold-standard validated by experts.

3.3.1 Defining the annotation schema (ontology).

The selection of the ontology categories can be done by consensus among expert linguists and historians, or with the help of automatic computer methods that extract a “seed ontology” directly from the corpus [Eynard et al. 2012]. The first method is more time-consuming, but it is simpler and on many occasions yields better results, so this was the approach we followed.

The RGs ontology contains 21 entity classes and many more sub-categories relevant to social, political, territorial and economic terms appearing recurrently through the RGs (Table 2). For example, the general entity class “architecture”

includes sub-categories such as “domestic” and “religious,” which offers the possibility to distinguish between house and convent buildings described in texts. Another class “cosmogony” includes sub-categories such as “ritual,” “festivity,” “deity,” etc. Each entity type follows the definitions and linked-data standards of DBpedia to ensure the interoperability of our dataset.

3.3.2 Manual annotation of text samples for training the NLP model.

Using such entity classes we carried out a manual annotation of two samples of 18 out of 168 documents (10% of the corpus) from the RGs using the online tagging tools developed by Tagtog, a commercial company that generously provided a free license to our project.^[18] In order to ensure the quality of the annotated samples, five experts tagged the same documents independently. We then obtain statistics on the level of consensus among the experts by using the inter-annotator agreement tool from Tagtog. Figure 7 illustrates an annotated document.

3.3.3 Development of the NLP deep-learning model.

Once the training and gold-standard sample data are available, the actual development of the NLP model can begin. Our team in Portugal, led by Bruno Martins, combined two state-of-the-art approaches of deep-learning: neural networks and word embedding.

3.3.3.1 Neural networks.

A neural network is a computational model, loosely inspired on the neural structure of the brain, that is able to “learn by example” how to perform repetitive complex tasks. In the NLP and CL fields, it is applied, among other things, to perform word classification, or the assigning of a label to each word in an input text, indicating whether it belongs to a certain semantic category (e.g. toponym, person, animal, plant, institution, cultural concept), or to a “part-of-speech” syntactic category (i.e. noun, pronoun, adjective, determiner, verb, adverb, preposition, conjunction, and interjection).

Modern NLP software libraries leverage such statistical models [Goldberg 2017], making probabilistic decisions on the best annotations for previously unseen texts. One of the best is SpaCy. To run a deep-learning model in SpaCy we need to conduct a training phase during which the software is fed with the training data sample (in our case 10% of the corpus). This annotated training sample contains many known examples of the kind of words expected in the unannotated text (the remaining 90% of the corpus).

Because we know the correct answers — thanks to the gold-standard data sample — , we can give the model feedback on its predictions in the form of an error gradient or the loss function that calculates the difference between the training example and the expected output. If the accuracy of the annotations is not sufficient, then an iterative refining process can be carried out until accomplishing satisfactory results.

Then, to link the gazetteer with the corpus texts, it is necessary to assign correct longitude and latitude coordinates to each toponym recognized by the neural network. The Mordecai geoparsing library [Halternamn 2017] is a good tool for that job. It takes as input the place names recognized by SpaCy, and uses an index over the contents of the gazetteer to find the potential coordinates of extracted place names. In other words, the names recognized in the text are matched against similar names in the gazetteer. Through this process the gazetteer and the corpus are finally linked and ready for the implementation of query tools.

3.3.3.2 Word-embedding.

Complementing the above procedures, the word-embedding algorithm makes it easier for the neural network to distinguish all of the different semantic contexts in which a word might be used. This process follows the idea that “a word is characterized by the company it keeps” [Firth 1957]. For example, in the phrase “[t]he order of San Francisco was established in these lands in 1524,” the term “San Francisco” refers to an institution, but in “[t]he mendicant friars helped the poor following the teachings of San Francisco,” the same term refers to a person. In another context, it could mean a geographic location, as for instance in the sentence “San Francisco Tlalcalalcalpan is inhabited at the present by 18,721 people.”

We specifically pre-trained a word-embedding algorithm in an unsupervised way to statistically predict the meaning of words in the context of real natural language utterances. Thus, we needed training data sufficiently big and varied to tackle the specific challenges of the RGs. One is that the spelling in the corpus varies considerably among the many authors. This is sometimes, in part, because words are spelled in accordance with their Latin origin, rather than their actual pronunciation. Another is that the corpus borrows heavily on vocabulary from indigenous languages, especially place names, and terms to identify flora, fauna, cultural concepts, etc.

37

Given the fact that the structure of modern Spanish still resembles to a high degree the language recorded in the RGs, we decided to combine modern and historical data sources to train our model. Contemporary data came from several Spanish corpora such as ANCorra [Taulé et al. 2008], IMPACT-es [Sánchez-Martínez et al. 2013], or the *Corpus del Español*,^[19] while numerous examples of sixteenth-century Spanish texts from different archives and ten dictionaries of indigenous languages [García Cubas 1888-1891] [Peñafiel 1987] [Robelo 1897] [Robelo 1900] [Robelo 1902a] [Robelo 1902b] [Robelo 1904] [Robelo 1905] [Starr 1920] provided resources to deal with the language complexities particular to New Spain.

38

By thus leveraging modern and historical data in a combined training dataset, we can simultaneously benefit from having large amounts of contextual information from which the model can learn more easily, together with some data representing the specificities of historical Spanish. Furthermore, to avoid problems with out-of-vocabulary words, such as those brought forward by the inconsistent spelling or cultural expressions, we used a word-embedding method that also creates representations for sub-words, and which is able to reconstruct entire word representations from such sub-words [Bojanowski et al. 2017].

39

4. Future Work: Data mining and historical analysis

At this stage, several data mining and analysis procedures can be carried out. The first is an adaptation of collocation analysis — a technique from Corpus Linguistics (CL) — which allows for extracting the most statistically significant phrases that are found next to a term or keyword of interest (the “collocates,” in CL jargon). Our variant of the technique is *geographic collocation analysis*, originally developed by the research group at Lancaster [Gregory et al. 2015] [Murrieta-Flores et al. 2015] [Murrieta-Flores et al. 2017] [Murrieta-Flores and Gregory 2017]. This facilitates the recognition and retrieval, across the corpus, of all place names that collocate or are close to a keyword of interest according to a predefined proximity threshold, usually defined in terms of the maximum number of words between the search-term and the collocate. In this manner, if a researcher is interested in investigating economic aspects related to agriculture, they could retrieve, for instance, all phrases in the context of words like *maíz*, *siembra*, *subsistencia*, *mantenimientos*, *comida* (i.e. “maize”, “sowing”, “subsistence”, “sustenance”, “food”), which would appear linked to a map showing the geographic locations where such terms appear.

40

Furthermore, by putting words in context through collocation, a historian could extract in a few seconds — as opposed to the weeks that it would take to access the information without GTA tools — interesting details about the first colonists in New Spain. Such a query, for instance, might involve running a collocate search of the word “conquistador” (Figure 8). Another type of search can be conducted using general categories, such as employing the term “food” to retrieve words like “beans,” “fish,” or “bread” (Figure 9).

41

In the following subsections, we highlight several topics of historic analysis that we plan to examine through this research.

42

4.1 Historical phenomena changing the face of New Spain at the end of the sixteenth century

One area of application for the techniques detailed in this study is understanding the coexistence of competing systems of land ownership and labor in New Spain during the colonial period. In New Spain during the short time of compilation of the RGs, there was a strong debate about the system of *encomiendas* and the desire to replace it with the so-called *repartimientos*. The *encomienda* was the right to exploit the labor of indigenous populations, granted by the Spanish Crown to conquistadors, who at the beginning of the colonization held this privilege in perpetuity and could, therefore,

43

pass it to their descendants with little intervention from the royal government in Madrid. In contrast, the *repartimiento* only allocated the labor force of native workers for a fixed period of time in exchange for a meager remuneration. The latter gave the royal government more control over its subjects and resources. Neither the holders of *encomiendas* nor *repartimientos* were allowed to appropriate lands, as private property was reserved for *haciendas*. These three systems coexisted in the 1570s despite the enactment of the so-called New Laws in 1542, which attempted to protect the rights of the indigenous populations and prevent the dispossession of their lands by abolishing the *encomiendas*. Throughout the texts of the RGs, therefore, it is common to find claims by conquistadors' descendants of their *encomienda* rights, and others defending the benefits of *repartimientos*. A comprehensive study of the claims and contentions found in the RG corpus would improve our understanding of the period.

4.2 The local government organization of New Spain toward the third quarter of the sixteenth century

A second application of these methodologies has to do with disambiguating local structures of governance in New Spain. The early colonial period was a time when the functions and spheres of influence of the four branches of colonial government were not well defined, and consequently the spatial boundaries of the judicial, administrative, military and ecclesiastical jurisdictions overlapped (Figure 10). This was especially true at the municipal level. Consequently, the corpus contains terms associated with several kinds of jurisdictions whose meanings were not consistent throughout the vast territory of New Spain (e.g. *partidos de clérigos*, *distrito de mision*, *custodia*, *guardianía*, *vicaría*, *presidencia*, etc.). The subject has been a matter of ongoing debate among historians [Gerhard 1972b, 66], but we believe that applying collocation analysis for extracting the contextual use of such words would shed light on how those institutions functioned in parallel to *alcaldías mayores* and *corregimientos* and this would be an important contribution to the knowledge of this period.

44

4.3 Settlement pattern analysis to elucidate the territorial organization of New Spain

A third area for future study has to do with demarcating the boundaries of territories in New Spain. By reading the RGs, we could attempt to identify the names and boundaries (*linderos*) of towns belonging to *corregimientos* (territorial demarcations for administrative and judicial purposes) mentioned in the corpus. Circa 1570, there were 200 of such units whose frontiers have been outlined by Gerhard (1972a). However, the smaller demarcations (towns, villages) suffered adjustments in boundaries and jurisdictions. As Gerhard points out, at the time of the RGs, "[e]arly civil and ecclesiastical divisions lost importance and were eliminated; others came into being as a result of conquest, proselytization, mining activity, and other causes" [Gerhard 1972b, 66–9]. The application of a mathematical model that analyzes the maximum available territory around each settlement within a region, like the one proposed by Ducke and Kroefges (2008), could provide preliminary hypotheses on such demarcations. These would then be tested against the actual historical information extracted from the corpus through geographic collocation analysis. One of the many ways that such a study could begin is by extracting references on the routes that local administrators take for their visits to towns within their jurisdiction, the distances they traversed, the geographic features and topography of the landscape, etc. (Figure 10) This, and much other information, is available throughout the corpus, but more specifically in the answers to questions 7, 8, 11, and 12 (see Appendix 1).

45

4.4 Analysis of historical diseases

A fourth research question has to do with our understanding of the outbreak of disease (1576-1581) that took the lives of half the population of New Spain during the early colonial period [Acuña-Soto et al. 2004]. This was the second time that *cocoliztli* (as the epidemic was called) had decimated the native population, causing the disappearance of whole communities, shortages in labor-force for the *encomiendas* and *repartimientos*, the displacement of people over different regions, or, on the contrary, the congregation of survivors (sometimes of different ethnic communities) in new Indian towns, all of which altered dramatically the traditional settlement pattern [Gerhard 1972a, 24]. How did this phenomenon vary over different regions? How many people died in each locality? How did this epidemic compare with others in the past? How did the new settlement change in comparison with Prehispanic times? These are all questions that GTA could help to answer by collocating words and phrases related to health (such as diseases, symptoms, and

46

remedies) and analyzing them with spatial mathematical models of disease contagion.

4.5 The spatial distribution of natural resources (flora, fauna, minerals, etc.) and exploitation practices at the time of the RGs

A fifth application of our methodology has to do with the study of natural resources. GTA will allow the extraction and analysis of regional historical information on natural resources such as plants. There are already interesting archaeological, historical and ethnographic research related to plants and their use in Mesoamerica, medicinal and otherwise [Albores Zarate 2015] [Chávez Mejía et al. 2020] [Hersch 1999]. Some further research has been also dedicated to assessing the importance of the trees and edible plants introduced by Europeans. With the annotation of the RGs in this project and the development of GTA, it is now possible not only to extract information regarding native and European plants introduced in the different regions of New Spain, but also to contribute to further debates. It is known, for instance, that Mesoamerican cultures had an extended knowledge of plants for medicinal use and most of the research about this has been carried out using colonial primary sources. We could now create a map of all the plants mentioned in the RGs and compare this dataset with archaeological data and other primary sources that were dedicated to the recording of medicinal plants, such as the *Código Badiano* or the *Historia Natural de Nueva España*. This would allow researchers, for instance, to assess the continuation in the use of specific plants for remedies and consumption in specific regions of New Spain. Other interesting avenues of possible research with GTA analysis are the availability of food and animals both native and introduced, the use of natural resources, the existence of cultural artefacts, the presence of priests and religious orders in particular regions, as well as types of architecture.

47

4.6 Exchange networks, trading practices and communication routes

An additional topic is the study of exchange networks and trading practices, as well as communication routes, both terrestrial and aquatic, among the populations described in the RGs. One such study has already been conducted by Favila-Vázquez (2019), who successfully reconstructed aquatic communication networks in the Rio Balsas region by extracting information from the responses available to queries 38 to 47 of the RGs (Figure 11). These contain crucial data regarding navigation by sea, including location of ports and coastal landscape. Together with data from question 19 (focused on fluvial networks), the extraction of data from the RGs corpus allowed Favila-Vázquez to reconstruct a riverine and coastal system of communication that explains the spatial connectivity of the Valley of Mexico to the Pacific coast.

48

5. Conclusion

Geographical Text Analysis can make an important contribution to the study of massive corpora, like the RGs, by expediting the extraction and analysis of specific information that would be too slow or cumbersome to find in big datasets through traditional reading methods. It also can provide tools to uncover patterns behind the narrative of historical sources and link them to their geographic locations. Thus, GTA creates a way for historians to apply what in computational literary studies is called “distant reading.”^[20]

49

At the same time, the use of geographic collocations and the concept of “keywords in context” permit a more detailed, granular comprehension of the corpus by identifying those parts that deserve a “close reading.” This flexibility has been demonstrated in previous research projects employing GTA [Donaldson et al. 2017] [Murrieta-Flores et al. 2017] [Taylor et al. 2017]. Having a multiplicity of “place names in context” associated with keywords of interest opens an unprecedented range of possibilities in terms of identifying information patterns and hypothesis-testing across large corpora, which in turn allows answering historical questions in a way that would be impossible otherwise.

50

In such a context, the proposal of GTA must not be seen as an attempt to replace the careful micro study of historical sources practiced by modern historians. More exactly, GTA is a methodology for parsing and crosslinking information at a macro scale throughout a whole corpus in a relatively short period of time, so historians can use their efforts and time more efficiently to focus on the more important documents or fragments of texts.

51

One of the main contributions of our project specifically has been adapting named entity recognition, disambiguation,

52

and classification techniques to text collections that either belong to specific historical periods or might pose a further challenge due to their multi-language nature. We have succeeded in developing a new deep-learning model to process the language of the RGs despite the orthographic variance across the corpus and the high frequency of indigenous vocabulary. Nevertheless, there is still much work to be done, especially for the indigenous languages of the Americas [Mager et al. 2018].

Finally, we have already produced a significant number of digital datasets of value, such as the gazetteer, which includes not only sixteenth-century localities but place names from other Mexican historic periods. Hopefully, these will prove valuable resources for humanists engaged in studying the Latin American historical and archaeological past. The methods and tools being developed by the project have the advantage of being generic, and therefore, applicable to other corpora.

53

We are committed to releasing the digital products and software tools through several repositories, including GitHub and our own online platform. Our hope is that scholars from diverse fields will be able to identify the desired information, extract it, cross-link it to other document resources, and most importantly, answer significant research questions: Where do certain phenomena happen more often? Who was responsible for certain events? What was happening in specific places at certain times? Such kind of enquiry will take historical analysis a step further, contributing significantly to advancing research on the colonial period in Latin America, and humanities research more generally.

54

Appendix 1. The Questionnaire of the Relaciones Geográficas^[21]

Memorandum Of The Things That Are To Be Answered, And Of That Which Shall Be Taken Into Account.

1. Firstly: For towns of Spaniards, state the name of the district or province in which it lies. What does this name mean in the native language, and why is it so called?
2. Who was the discoverer and conqueror of this province? By whose order was it discovered? In what year was it discovered and conquered, as far as is readily known?
3. What is the general climate and character of the province or district? Is it very cold or hot, humid or dry? Does it have much or little rain, and when, approximately, does it fall? How violently and from where does the wind blow, and at what times of the year?
4. Is the terrain flat or rugged, clear or wooded, with many or few rivers or springs, abundant in or lacking water? Is it fertile or without pasture, abundant or lacking in fruits and sustenance?
5. Are there many or few Indians? Were there more or fewer at other times, and what are the known causes of this? Are they presently settled in planned and permanent towns? Describe the degree and quality of their intelligence, inclinations, and way of life. Are there different languages in the province or a general language that all speak?
6. What is the latitude, or the altitude of the Pole Star, at each Spanish town, if it has been taken and is known, or if anyone knows how to observe it? Or on which days of the year does the sun not cast a shadow at midday?
7. What are the league distances and the direction of each Spanish city or town in the district to the city of its *Audiencia*, or to the town where the governor to whom it is subject resides?
8. Likewise, what are the league distances and the direction of each Spanish city or town to adjacent ones? Are these leagues long or short, through flat or hilly land, over straight or winding roads, easy or difficult to travel?
9. What are the present and former names and surnames of each city or town? Why are they so called, if known? Who gave each its name, who was its founder, by whose order was it settled? What year was it founded? How many residents did the settlement first have, and how many does it have now?
10. Describe the sites upon which each town is established. Is each upon a height, or low-lying, or on a plain? Make a map of the layout of the town, its streets, plazas and other features, noting the monasteries, as well as can be sketched easily on paper. On it show which part of the town faces south or north.
11. For native towns, state only how far they are from the town in whose corregimiento or jurisdiction they are,

and how far they lie from their cabecera de doctrina.

12. In addition, state how far the native towns lie from surrounding native or Spanish towns. Declare for all their direction from these other towns. Are the leagues long or short, the roads through level or hilly land, straight or winding?
13. What does the name of this [native] town mean in the indigenous language? Why is it called this, if known? What is the name of the language spoken by the natives of this town?
14. Who were their rulers in heathen times? What rights did their former lords have over them? What did they pay in tribute? What forms of worship, rites, and good or evil customs did they practice?
15. How were they governed? With whom did they wage war? How did they battle? What was their battle dress and clothing like, both formerly and now? What was their former and is their present means of subsistence? Were they more or less healthy than now and what are reasons for this that you may know?
16. For all towns, both Spanish and native, describe the sites where they are established. Are they in the mountains, in valleys, or on open flat land? Give the names of the mountains, valleys, and districts, and for each, tell what the name means in the indigenous language.
17. Is the land or site healthy or unhealthy? If unhealthy, why (if it is known)? What illnesses commonly occur, and what cures are commonly used for them?
18. How far or near, and in what direction does each town lie from a nearby prominent mountain or range? Supply its name.
19. What major river or rivers flow nearby? How distant and in what direction do they lie? How great is their flow? Is there anything notable known about their sources, water, orchards and other growth along their banks? Are there or could there be irrigated lands of value?
20. What are the significant lakes, lagoons, or springs within the town boundaries? Note anything remarkable about them.
21. What are the volcanoes, caves, and all other notable and remarkable aspects of nature in the district worthy of being known?
22. Which wild trees and their fruits are commonly found in the district? What are the uses of them and their woods, and to what good are they or could they be put?
23. Which cultivated trees and fruit orchards are in the region? Which were brought from Spain and elsewhere? Do these grow well?
24. What are the grains and seed plants, and other garden plants and vegetables that are or have been used as sustenance for the natives?
25. Which were brought from Spain? Does the land yield wheat, barley, wine, or olive oil, and in what quantities? Is there silk or cochineal in the region, and in what quantities?
26. What are the herbs or aromatic plants that the natives use for healing? What are their medicinal or poisonous properties?
27. What animals and birds, both wild and domestic, are in the region? Which ones were brought from Spain, and how well have they bred and multiplied?
28. In the region or within the town lands, are there any gold or silver mines or sources of other metals, or black or colored pigments?
29. Are there any quarries of precious stone, jasper, marble, or other notable ones? What value might they have?
30. Are there sources of salt in the town or nearby? If [residents] lack salt and other items necessary for their sustenance or clothing, where do they procure all these things?
31. How are the houses built and what is their form? What building materials are found in the town, or brought from elsewhere?
32. What are the towns' fortifications? Are there barracks, or any fortified or impregnable places within the boundaries of the district?
33. Through what dealings, trade, and profits do both Spaniards and Indians live and sustain themselves? What are the items involved and with what do they pay their tribute?
34. In which diocese of the archbishopric, bishopric, or abbey does each town lie? In which district does each

lie? How many leagues and in which direction does each town lie from the town of the cathedral and the cabecera of the district? Are the leagues long or short, along straight or winding roads, and through flat or hilly land?

35. In each town, what are the cathedral and parish church or churches? What is the number of endowed church offices and allotments for clergymen's salaries in each? Do any have a chapel or significant endowment, and, if so, whose it is and who established it?
36. What are the monasteries or convents of each Order? By whom and when were they founded? How many notable things do they contain? And what is the number of religious?
37. What are the hospitals, schools, and charitable institutions in the towns? By whom and when were they founded?
38. If the towns are on the seacoast, in addition to the above, report in your account the nature of the sea in the vicinity, whether it is calm or stormy, the nature of the storms and other dangers. What times, approximately, are they frequent?
39. Is there beach or cliffs along the coast? What are the prominent reefs and other dangers to navigation along it?
40. How great are the tides and tidal ranges? On which days and at which hours do they occur? In which season are they greater or lesser?
41. What are the notable capes, points, bights, and bays in the district? Note their names and sizes, as well as can be determined.
42. What are the ports and landings along the coast? Make a map showing their shape and layout as can be drawn on a sheet of paper, in which form and proportion can be seen.
43. What is their size and capacity? Note their approximate width and length in leagues and paces, as well as can be determined. Also note how many ships they will accommodate.
44. What is the depth in fathoms of each? Is the bottom clear? Are there any shallows and shoals? Indicate their locations, and whether the port is free of shipworm and other impediments.
45. What directions do their entrances and exits face? With which winds must one enter and depart?
46. What are their advantages or lack of them in the way of firewood, water, and provisions? What are the favorable or unfavorable considerations for entering and remaining?
47. What are the names of the islands along the coast? Why are they so named? Make a map, if possible, of their form and shape, showing their length, width, and lay of the land. Note the soil, pastures, trees, and resources they may have, their birds and animals, and the notable rivers and springs.
48. Where are the abandoned Spanish towns located in the region? When were they settled and abandoned? What is known about the reasons for abandonment?
49. Describe any other of the notable aspects of nature, and any notable qualities of the soil, air, and sky in any part of the region.
50. Having completed the account, the persons who have collaborated on it will sign it. It must be returned without delay, along with these directions, to the person from whom it was received.

Appendix 2. Figures

Instrucción, y memoria, de las relaciones que

se han de hazer, para la descripción de las Indias, que su Magestad mandó hazer, para el buen gouerno y ennoblecimiento dellas.



Primeramente, los Gouernadores, Corregidores, o Alcaldes mayores, a quien los Vireyes, o Audiencias, y otras personas del gouerno, embiaren estas instrucciones, y memorias impresas, ante todas cosas haran lista, y memoria de los pueblos de Españoles, y de Indios, que viuiere en su jurisdicción, en que solamente se pongan los nombres de ellos escriptos de letra legible, y clara, y luego la embiaran a las dichas personas del gouerno, para que juntamente, con las relaciones que en los dichos pueblos se hizieren, la embien a su Magestad, y al consejo de las Indias.

Y distribuyan las dichas instrucciones, y memorias impresas por los pueblos de los Españoles, y de Indios, de su jurisdicción, donde viuiere Españoles, embiando las a los concejos, y donde no, a los curas si los viuiere, y sino a los religiosos, a cuyo cargo fuer la doctrina, mandando a los concejos, y en cargando de parte de su Magestad, a los curas y religiosos, que dentro de vn breue termino, las respondan, y satisfagan como en ellas se declara, y les embien las relaciones que hizieren, juntamente con estas memorias, para que ellos como fueren recibiendo las relaciones, vayan embiandolas a las personas de gouerno que se las viueren embiado, y las instrucciones y memorias las bueluan a distribuir si fuer menester por los otros pueblos a dōde no las viueren embiado.

Y en los pueblos, y ciudades, dōde los Gouernadores, o Corregidores, y personas de gouerno residieren, haran las relaciones de ellos, o encargar las a personas inteligentes de las cosas de la tierra, que las hagan, segun el tenor de las dichas memorias,

Las personas a quien se diere cargo en los pueblos de hazer la relacion particular de cada vno dellos, responderan a los capitulos de la memoria, que se sigue por la orden, y forma siguiente:

Primeramente, en vn papel a parte, podrán por caueça de la relacion que hizieren, el dia, mes, y año de la fecha de ella: con el nombre de la persona, o personas, que se hallaren a hazerla, y el del Gouernador, y otra persona que les viuer embiado la dicha instruccion.

Y leyendo atentamente, cada Capitulo de la memoria, screuirá lo que huuiere q dezir a el, en otro capitulo por si, respondiendo a cada vno por sus numeros, como van en la memoria, vno tras otro y en los que no huuiere que dezir, de xarlos há sin hazer mención de ellos, y pasaran a los siguientes, hasta acabar los de leer todos, y responder los q tuuieren que dezir: como queda dicho, breue y claramente, en todo: afirmando por cierto lo que lo fuere, y lo que no, poniendolo por dudoso: de manera que las relaciones venganciertas, conforme a lo contenido en los capitulos siguientes.

Figure 1. First page of the printed *Instrucción y Memoria* sent by the Council of Indies to guide the compilation of the *Relaciones Geográficas*. Photograph courtesy of the Benson Latin American Collection. The General Libraries, The University of Texas at Austin (JGI xxiii-13).

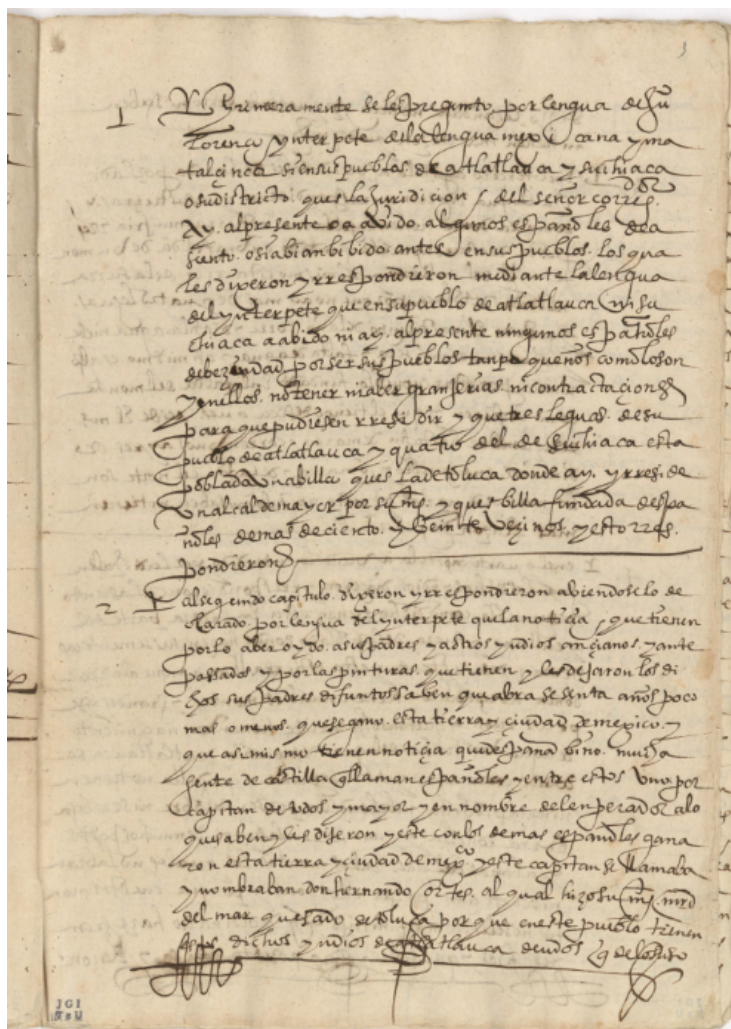


Figure 2. Manuscript from the *Relación of Atlatlahuca* (Morelos) containing responses to the first two questions of the enquiry.



Figure 3. Map of Ixtapalapa (Ciudad de Mexico) included in the *Relación Geográfica de Ixtapalapa* (1580), belonging to the corregimiento of Mexicaltzingo. Size of the original: 43 x 31 cm. The author is Martin Cano, a native local artist, who, as Mundy (1996: 63) points out, is one of the uses pictographs to represent the town's name, which in Nahuatl means "water near the flagstones": "Below the church, a hexagonal flagstone is surrounded by a blue ribbon of water. Ixtapalapa's church dominates one axis of the composition, while its twin community buildings, at left, dominate the other". Photograph courtesy of the Benson Latin American Collection. The General Libraries, The University of Texas at Austin (JGI xxiv-8).

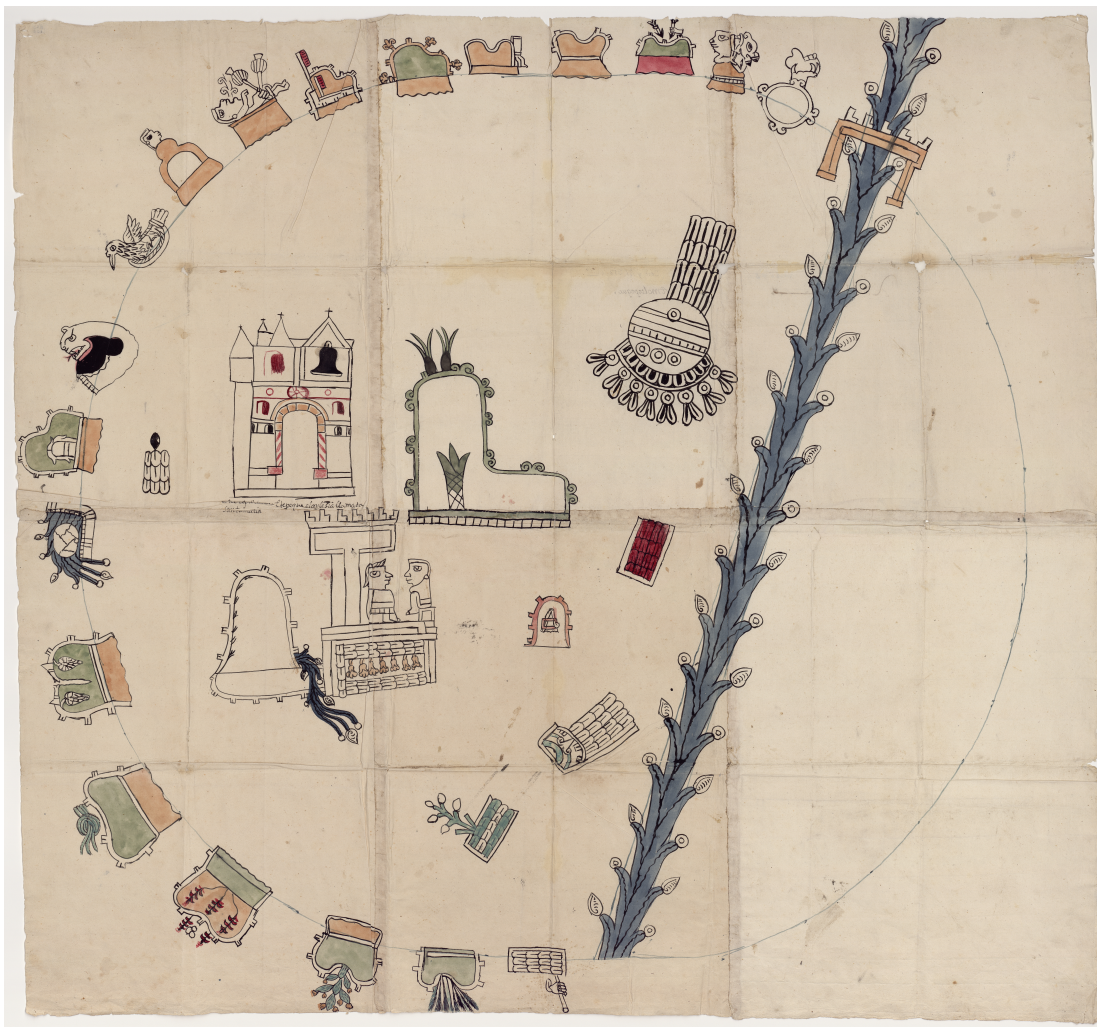


Figure 4. Map of the *Relación de Amoltepec*. This is another example of the indigenous tradition to represent a geographic region. The place names of neighbor and subject towns are drawn with logographs organized in a semi-circle, surrounding the church and the ruler's palace (Mundy, 1996: 113). Photograph courtesy of the Benson Latin American Collection. The General Libraries, The University of Texas at Austin (JGI xxiv-5).



Figure 5. Map of the *Relación de Tlacotalpa*. One of the best examples of charts in the European tradition that can be found in the RGs corpus. This was produced by Francisco Stroza Gali, a professional mariner in 1580 (Mundy, 1996: 51). A detailed study of the document reveals the extraordinary precision of the geographic measurements achieved by Gali [Morato-Moreno 2017].



Figure 6. Location of the 3919 sites georeferenced so far (April 2020) through linguistic and locational disambiguation, which are added to the project's gazetteer

The screenshot shows the Tagtog software interface for annotating the 'Relación de Merida' document. The document text is displayed with various words highlighted in different colors to represent different entity types. A search bar is visible at the top left, and a table of annotations is shown on the right side of the interface.

Entity Type	Count	IAA (%)
Food	156	80%
Climate	25	52%
Institution	10	59%
Kinship	7	74%
Date	17	49%
Cosmogony	53	65%

Figure 7. Annotation of the *Relación de Merida* with the Tagtog software tool. Notice the color code to distinguish different annotation categories: red is used for words referring to persons, either by name, profession or roles; green to locations (either toponyms or geographic features); light blue to dates, and so on.

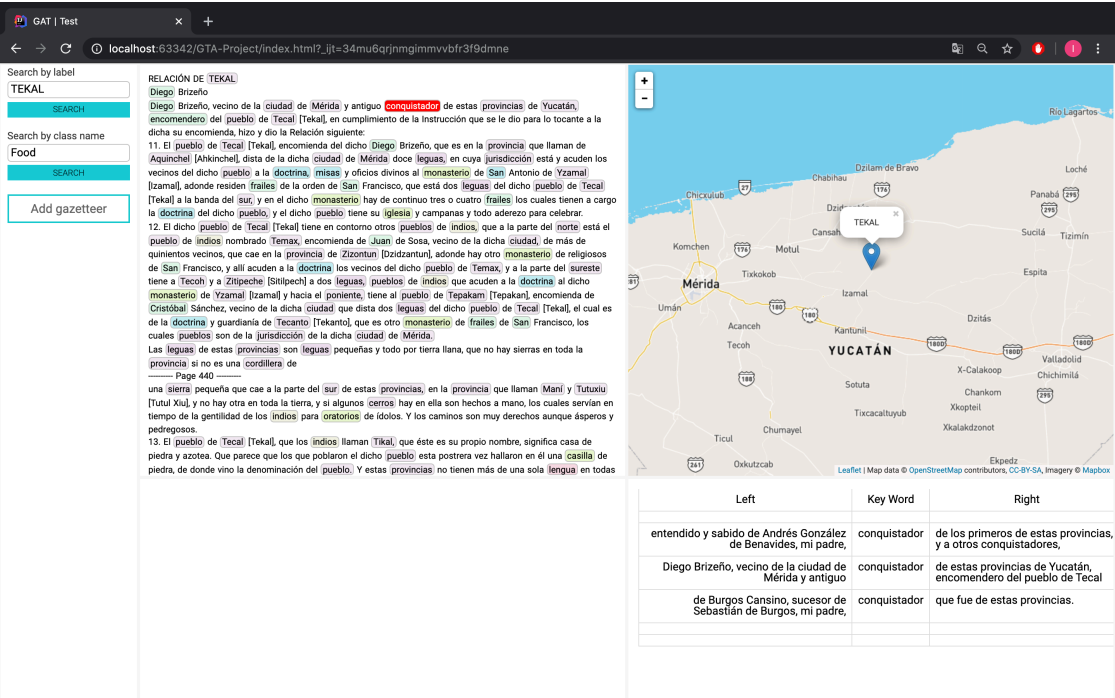


Figure 8. Extracting information with GTA tools (which are still a work in progress). In this example, the keyword “conquistador” is used to recover data from the colonists mentioned in the *Relación de Tekal*. The map shows the location of this town in Yucatán. Queries like this can be apply to a single document or to the complete corpus.

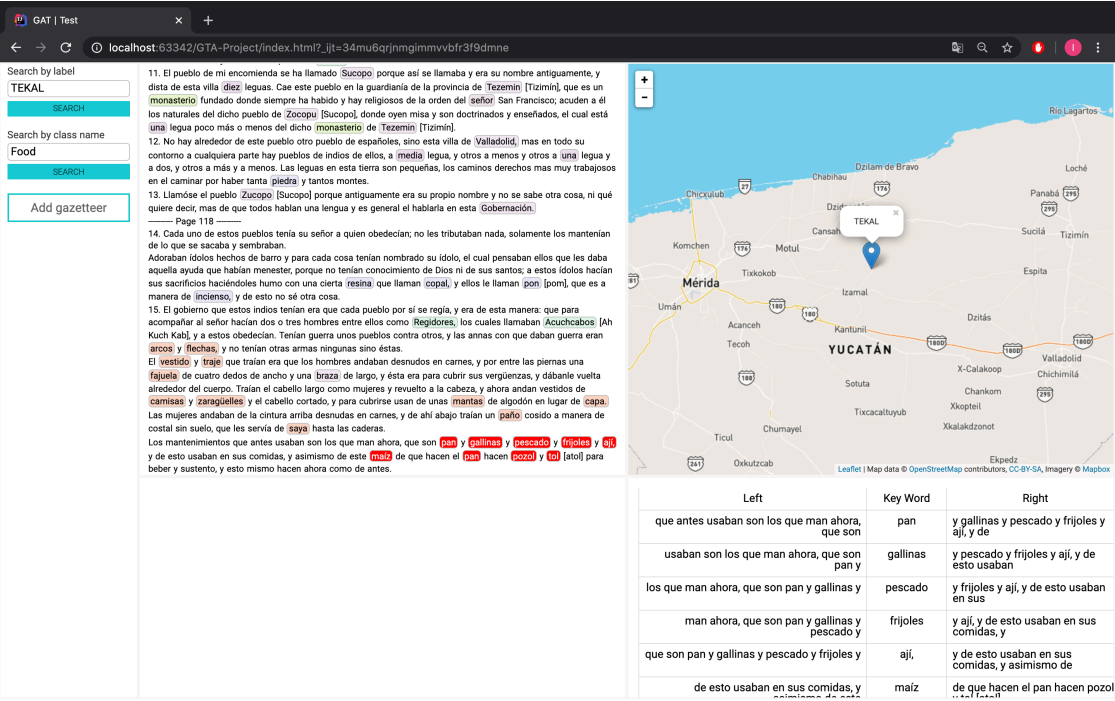


Figure 9. Another type of query that retrieves works in context belonging to the food category. In this example, the threshold for the collocation has been set 10 words before and after the key category.

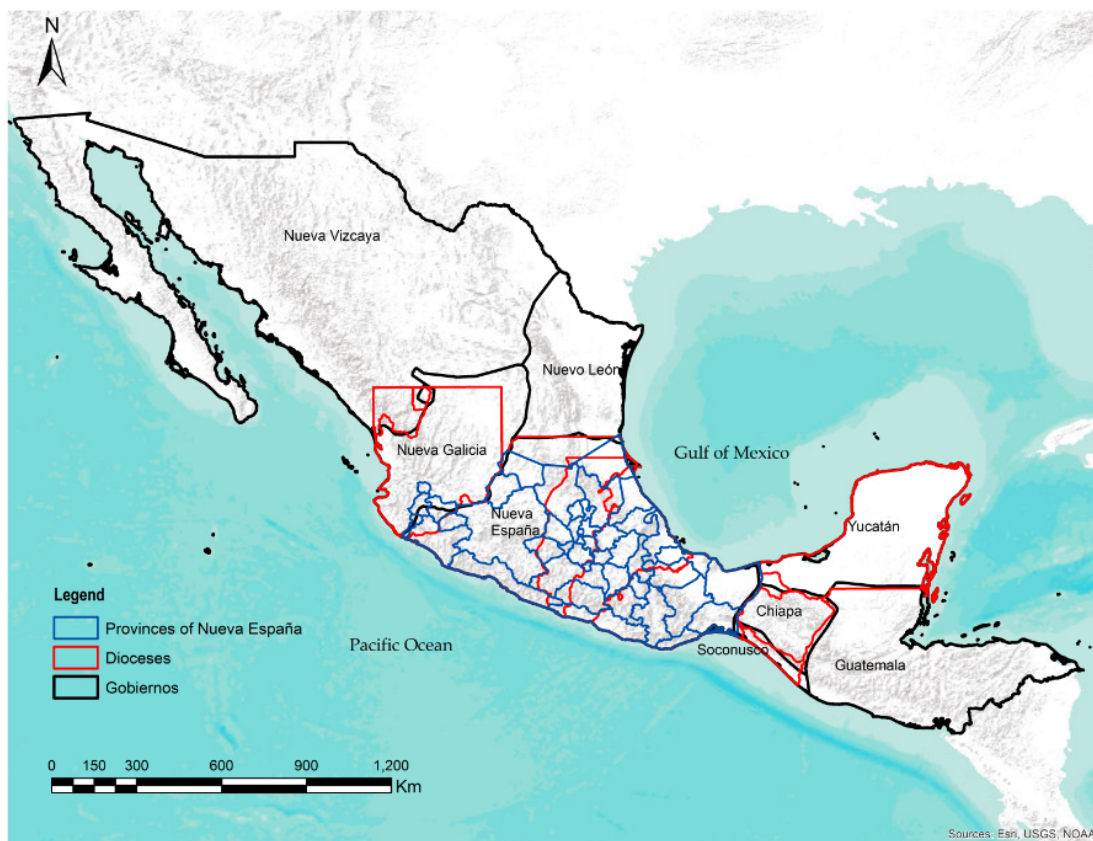


Figure 10. Map illustrating the overlapping of three types of colonial jurisdictions: political (*gobiernos*), ecclesiastic (dioceses), and administrative (*provincias*).

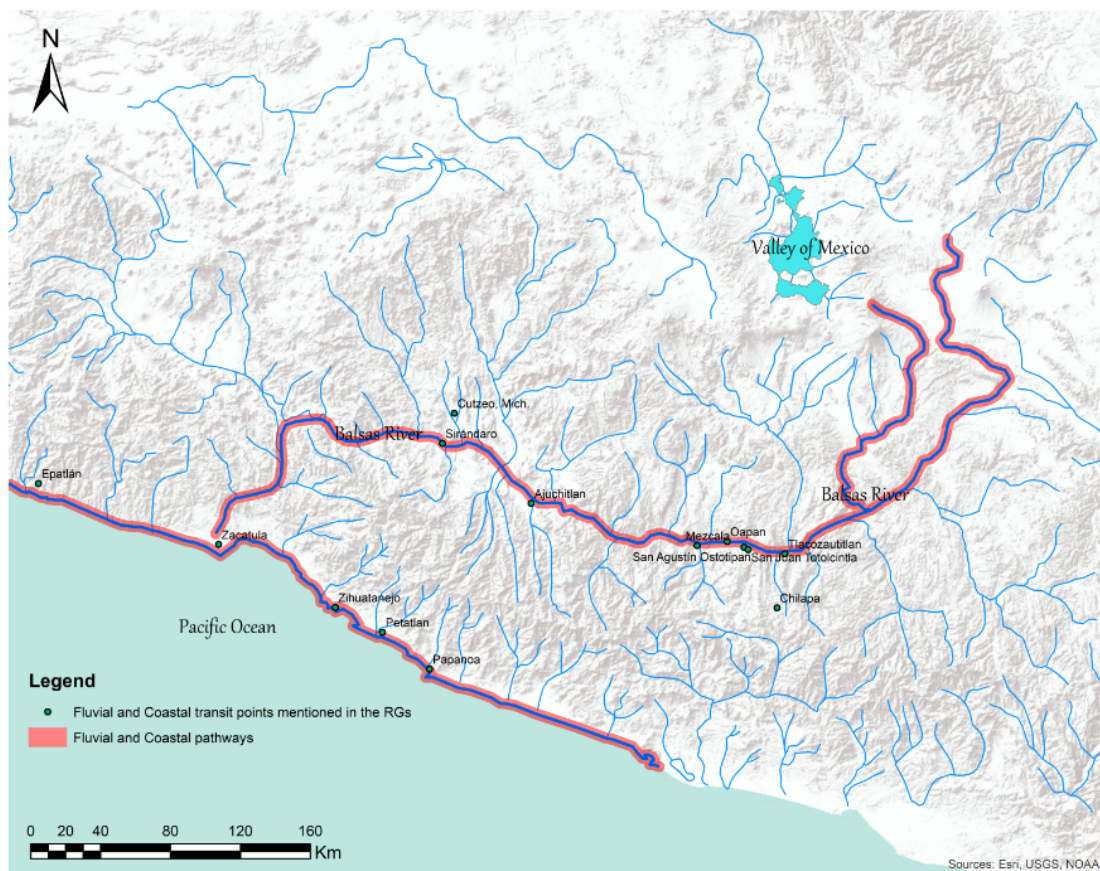


Figure 11. Map showing the location of towns that were part of the indigenous network of navigation through the Pacific coast and the Balsas river towards the third quarter of the 16th century. Photograph courtesy of Favila-Vázquez (2019).

Entity	Ontological definition	Labels
Person	http://dbpedia.org/page/Person	female/male, title, profession
Date	http://dbpedia.org/ontology/date	
Institution	http://dbpedia.org/page/Institution	civil, ecclesiastical
Location	http://dbpedia.org/ontology/location	settlement type, generic location, geographic feature type, toponym, address, imaginary, ecclesiastical jurisdiction, civil jurisdiction
Activity	http://dbpedia.org/ontology/activity	agriculture, warfare, economy, mining, domestic, female/male
Animal	http://dbpedia.org/ontology/animal	insect, mammal, reptile, bird, amphibian, aquatic, domesticated
Plant	http://dbpedia.org/page/Plan	
Food	http://dbpedia.org/page/Food	
Natural Resource	http://dbpedia.org/page/Natural_resource	
Cultural artefact	http://dbpedia.org/page/Cultural_artifact	house goods, commodities, clothing, weapon, tool
Architecture	http://dbpedia.org/page/Architecture	religious, civil, domestic
Cosmogony	http://dbpedia.org/page/Cosmogony	ritual, festivity, activity, deities, saints, object
Health	http://dbpedia.org/page/Health	disease, remedy
Route of Transportation	http://dbpedia.org/ontology/RouteOfTransportation	terrestrial, aquatic, route direction, distance
Kinship	http://dbpedia.org/page/Kinship	
Climate	http://dbpedia.org/page/Climate	
Ethnic Group	http://dbpedia.org/page/Ethnic_group	
Social Class	http://dbpedia.org/page/Social_class	
Language	http://dbpedia.org/page/Language	
Event	http://dbpedia.org/page/Event	historical, disasters
Measurement	http://dbpedia.org/page/Measureme	value, tribute, weight, population

Table 1. Annotation schema: entities, DBpedia definitions and labels per entity.

Acknowledgements

We would like to thank the generous grant awarded by the 2017 T–AP initiative of the Trans–Atlantic Platform in Social Sciences and Humanities Research, as well as the support received in each country by the Economic and Social Research Council (ESRC, UK) (ES/R003890/1), the Foundation for Science and Technology (FCT, Portugal), and the Consejo Nacional de Ciencia y Tecnología (project 275015). Likewise, we would like to thank the Red Conacyt de Tecnologías Digitales para la Difusión del Patrimonio Cultural (project 294971), also supported by Conacyt.

55

Notes

[1] This project operates thanks to a grant received from the 2017 “Digging into Data” initiative of the Trans–Atlantic Platform for Social Sciences and Humanities Research (T-AP). T-AP is a consortium supported by research institutions in 16 countries, among them the Economic and Social Research Council (ESRC, UK), the Foundation for Science and Technology (FCT, Portugal), and the National Council of Science and Technology (CONACYT, México). This project is conducted by three teams of historians, archaeologists, geographers, linguists and computer scientists from the following institutions: The Digital Humanities Hub and the History Department of Lancaster University, UK; The Instituto de

[2] NLP techniques have often been applied to annotated texts written in modern English, and a few have been tried with historical documents composed in earlier forms of speech in several languages [Borin et al. 2007] [Brooke et al. 2015] [Byrne 2007] [Crane and Jones 2006] [Ehrmann et al. 2016] [Grover et al. 2008] [Sprugnoli et al. 2017] [Van Hooland et al. 2015] [Won et al. 2018].

[3] To give a complete picture of the enquiry, we reproduce the questions list in Appendix 1. The questionnaire in Spanish can also be found in De la Garza (1983: 7-12) and Acuña (1982-1988, vol. 1: 73-78), as well as in English publications by Nutall (1926), Cline (1964, 1972c: 234-237), Mundy (1996: 227-230), and Bravo-García (2018: 417-422).

[4] The census of these RGs published by Cline (1964, 1972d) and updated by Acuña (1982-1988) enumerates 34 RGs from Antequera; 17 from Michoacán; 15 from Tlaxcala; 54 from Yucatán; 2 from Guatemala; 34 from Mexico and 12 from Nueva Galicia (i.e. Guadalajara), though Cline (1972d) counts only 33 from Mexico. The collection is spread across four repositories: 80 documents are in the Archivo General de Indias in Seville; 46 in the Real Academia de la Historia in Madrid; 41 in the Latin American Collection of the Benson Library, University of Texas; and one in the Glasgow University Library. Two handmade transcriptions of the RGs, made by Joaquín García Icazbalceta, are now in the National Library of Mexico City [Cline 1964, 359] [Acuña 1982-8, 9].

[5] The latest comprehensive analysis of these maps is by Mundy (1996).

[6] For instance, Antequera is now Oaxaca.

[7] In one example, sixteenth-century Cempoala (Hidalgo) moved from a hill to a valley to guarantee access to water, which caused severe alterations on its limits.

[8] Tetlapayac, for instance, is the modern Santiago Teltlapayac.

[9] Quauhyocan is also written as Quacyocan, for example.

[10] For example, ancient Quahnahuac is modern Cuernavaca.

[11] Tetela, for instance, refers both to a town in modern Morelos and to a settlement in Puebla, and towns called Atlatlauhca can be found both in Morelos — origin of an RG — and Oaxaca.

[12] <http://www.cdi.gob.mx/localidades2010-gobmx/>

[13] <http://registropublico.inah.gob.mx/index.php>

[14] <http://www.geonames.org/>

[15] <http://www.getty.edu/research/tools/vocabularies/tgn/index.html>

[16] <http://www.getty.edu/research/tools/vocabularies/tgn/>

[17] <http://whgazetteer.org>

[18] www.tagtog.net

[19] <https://www.corpusdelespanol.org>

[20] The premise of distance reading is that there are simply too many documents for anyone to read and study simultaneously and within the limits of the human brain and under the time imposed by contemporary academic schedules.

[21] Reproduced from Mundy, Barbara. 1996. The mapping of New Spain. Indigenous cartography and the maps of the Relaciones Geográficas, Appendix B, pp. 227-230. Chicago: The University of Chicago Press.

Works Cited

Acuña 1982-8 Acuña, René. 1982-1988. *Relaciones Geográficas del siglo XVI*, 10 vols. México: Universidad Nacional Autónoma de México. Instituto de Investigaciones Antropológicas.

- Acuña–Soto et al. 2004** Acuña–Soto, Rodolfo, David W. Stahle, Matthew D. Therrell, Richard D. Griffin, and Malcolm K. Cleaveland. 2004. "When half of the population died: The epidemic of hemorrhagic fevers of 1576 in Mexico". *FEMS Microbiology Letters* 240(1): 1–5. doi:10.1016/j.femsle.2004.09.011.
- Afanador Pujol 2015** Afanador Pujol, A. (2015). *The relación de Michoacán (1539-1541) and the politics of representation in colonial Mexico*. University of Texas Press. <https://doi.org/10.7560/771383>.
- Albores Zarate 2015** Albores Zarate, B. (Ed.). 2015. *Flor-flora: Su uso ritual en Mesoamérica*. Toluca de Lerdo, Estado de México : Fondo Editorial Estado de México; El Colegio Mexiquense.
- Alpert-Abrams 2016** Alpert–Abrams, Hannah. 2016. "Machine reading the *Primeros Libros*", *DHQ: Digital Humanities Quarterly* 10(4). <http://www.digitalhumanities.org/dhq/vol/10/4/000268/000268.html>.
- Anaya Monroy1960** Anaya Monroy, Fernando. 1960. "Presencia espiritual de la cultura náhuatl en la toponimia". *Estudios de Cultura Náhuatl* 2: 9–25.
- Barlow 1949** Barlow, Robert H. 1949. "Relación de Zempoala y su partido, 1580". *Tlalocan. Revista de Fuentes para el Conocimiento de las Culturas Indígenas de México* 3(1): 29-41.
- Bellesteros García 2005** Ballesteros García, Víctor Manuel. 2005. *La Pintura de la Relación de Zempoala de 1580*. México: Universidad Autónoma del Estado de Hidalgo.
- Bojanowski et al. 2017** Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov. "Enriching Word Vectors with Subword Information." *Transactions of the Association for Computational Linguistics* 5 (2017): 135-46.
- Borin et al. 2007** Borin, Lars, Dimitrios Kokkinakis, and Leif-Jöran Olsson. 2007. "Naming the past: Named entity and animacy recognition in 19th century Swedish literature". In *Proceedings of the Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2007)*, pp. 1–8. Prague, Czech Republic: Association for Computational Linguistics. <http://www.aclweb.org/anthology/W/W07/W07-0901.pdf>.
- Brand 1960** Brand, Donald. 1960. *Coalconan and Motines del Oro, an ex-Distrito of Michoacan, México*. Austin: University of Texas, Institute of Latin American Studies.
- Bravo-García 2019** Bravo-García, Eva. 2019. *Las voces del contacto. Edición y estudio de Las Relaciones Geográficas de México (Siglo XVI)*. Warsaw: University of Warsaw.
- Brooke et al. 2015** Brooke, Julian, Adam Hammond, and Graeme Hirst. 2015. "GutenTag: An NLP-driven tool for digital humanities research in the project Gutenberg Corpus". In *Proceedings of the North American Association for Computational Linguistics*, pp. 1–6. Denver, Colorado. <http://www.cs.toronto.edu/pub/gh/Brooke-etal-2015-CLfL.pdf>.
- Byrne 2007** Byrne, Kate. 2007. "Nested named entity recognition in historical archive text". In *International Conference on Semantic Computing (ICSC 2007)*. Irvine, CA, USA: IEEE. Doi:10.1109/ICSC.2007.107.
- Chávez Mejía et al. 2020** Chávez Mejía, M.C., L. White Olascoaga, S. Moctezuma Pérez, F. Herrera Tapia. 2020. "Prácticas curativas y plantas medicinales: Un acercamiento a la etnomedicina de San Nicolás, México". *Cuadernos Geográficos*, 56(2): 26. <http://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=124569245&lang=es&site=eds-live>.
- Cline 1964** Cline, Howard F. 1964. The "Relaciones Geográficas of the Spanish Indies, 1577–1586". *The Hispanic American Historical Review* 44(3): 341–374. doi:10.2307/2511856.
- Cline 1972a** Cline, Howard F. 1972a. "Introductory notes on territorial divisions of Middle America". In *Handbook of Middle American Indians, Volume 12: Guide to Ethnohistorical Sources, Part One*, edited by Howard F. Cline, pp. 17–62. Austin: University of Texas Press.
- Cline 1972b** Cline, Howard F. 1972b. "Ethnohistorical regions of Middle America". In *Handbook of Middle American Indians, Volume 12: Guide to Ethnohistorical Sources, Part One*, edited by Howard F. Cline, pp. 166–182. Austin: University of Texas Press.
- Cline 1972c** Cline, Howard F. 1972c. "The Relaciones Geográficas of the Spanish Indies, 1577–1648". In *Handbook of Middle American Indians, Volume 12: Guide to Ethnohistorical Sources, Part One*, edited by Howard F. Cline, pp. 183–242. Austin: University of Texas Press.
- Cline 1972d** Cline, Howard F. 1972d. "A census of the *Relaciones Geográficas* of New Spain, 1579–1612". In *Handbook of Middle American Indians, Volume 12: Guide to Ethnohistorical Sources, Part One*, edited by Howard F. Cline, pp. 364–369. Austin: University of Texas Press.

- Cooper and Gregory 2011** Cooper, David, and Ian N Gregory. 2011. "Mapping the English Lake District: A literary GIS". *Transactions of the Institute of British Geographers* 36(1): 89–108. doi:10.1111/j.1475–5661.2010.00405.x.
- Crane and Jones 2006** Crane, Gregory and Alison Jones. 2006. "The challenge of Virginia Banks: An evaluation of named entity analysis in a 19th-century newspaper collection". In *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '06)*, Chapel Hill, NC, USA: ACM Press. doi:10.1145/1141753.1141759.
- Cáceres Lorenzo 2013** Cáceres Lorenzo, María Teresa. 2013. "Tipos de Relaciones Geográficas en el siglo XVI". *Crítica Hispánica* 35(1): 45-66.
- Cáceres Lorenzo 2016** Cáceres Lorenzo, María Teresa. 2016. "Heterogeneidad léxica en las Relaciones Geográficas de Yucatán". *Alpha. Revista de Artes, Letras y Filosofía*, 42 (online) <http://dx.doi.org/10.4067/S0718-22012016000100007>; https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0718-22012016000100007.
- Dahlgren de Jordan 1954** Dahlgren de Jordan, Barbro. 1954. *La Mixteca: su cultura e historia prehispánicas*. México: Universidad Nacional Autónoma de México.
- De la Garza 1983** De la Garza, Mercedes (Ed.). 1983. *Relaciones histórico-geográficas de la gobernación de Yucatán: (Mérida, Valladolid y Tabasco)*. First edition, 2 vols. México: Universidad Nacional Autónoma de México.
- Del Paso y Troncoso 1905** Del Paso y Troncoso, Francisco (Ed.). 1905. *Papeles de Nueva España*. Segunda Serie: Geografía y Estadística, vols. 1, 3–7. Madrid: Sucesores de Rivadeneyra.
- Delgado López 2010** Delgado López, Enrique. 2010. "Las Relaciones Geográficas como proyecto científico en los albores de la modernidad". *Estudios Mesoamericanos* 2(9): 97–106.
- Donaldson et al. 2015** Donaldson, Christopher, Ian N. Gregory, and Patricia Murrieta-Flores. 2015. "Mapping 'Wordsworthshire': A GIS study of literary tourism in Victorian Lakeland". *Journal of Victorian Culture* 20(3): 287–307. doi:10.1080/13555502.2015.1058089.
- Donaldson et al. 2017** Donaldson, Christopher, Ian N. Gregory, and Joanna E. Taylor. 2017. "Locating the beautiful, picturesque, sublime and majestic: Spatially analysing the application of aesthetic terminology in descriptions of the English Lake District". *Journal of Historical Geography* 56 (April): 43–60. doi:10.1016/j.jhg.2017.01.006.
- Ducke and Kroefges 2008** Ducke, Benjamin and Peter C. Kroefges. 2008. "Layers of Perception". *Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA2007)*, Berlin, Germany, April 2–6, edited by Axel Posluschny, Karsten Lambers, and Irmela Herzog, pp. 246-251. (Kolloquien zur Vor- und Frühgeschichte, Vol. 10). Bonn: Dr. Rudolf Habelt GmbH. https://proceedings.caaconference.org/paper/78_ducke_kroefges_caa2007/.
- Edwards 1969** Edwards, Clinton R. 1969. "Mapping by questionnaire: An early Spanish attempt to determine New World geographic positions". *Imago Mundi* 23: 17–28.
- Ehrmann et al. 2016** Ehrmann, Maud, Giovanni Colavizza, Yannick Rochat, and Frédéric Kaplan. 2016. "Diachronic evaluation of NER systems on old newspapers". In *Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016)*, edited by Stephanie Dipper, Friedrich Neubarth, and Heike Zinsmeister, pp. 97–107. Bochum, Germany: Bochumer Linguistische Arbeitsberichte. <http://infoscience.epfl.ch/record/221391>.
- Eynard et al. 2012** Eynard, Davide, Matteo Matteucci, and Fabio Marfia. 2012. "A modular framework to learn seed ontologies from text". In *Semi-automatic ontology development processes and resources*, edited by Maria Teresa Pazienza and Armando Stellato, pp. 22-47. Hershey, PA: IGI Publishing.
- Favila Vázquez 2019** Favila Vázquez, Mariana. 2019. "La navegación prehispánica: Un sistema de conectividad del paisaje mesoamericano. Modelo de interacción entre la costa del Pacífico y el Altiplano Central (Postclásico Tardío-Siglo XVI)". PhD Thesis, México: Universidad Nacional Autónoma de México. Facultad de Filosofía y Letras, Instituto de Investigaciones Filológicas. Programa de Maestría y Doctorado en Estudios Mesoamericanos.
- Fernández Christlieb 2006** Fernández Christlieb, Federico and Gustavo Garza Merodio. 2006. "La pintura de la Relación Geográfica de Metztitlán, 1579". *Secuencia. Revista de historia y ciencias sociales* 66: 160-186.
- Firth 1957** Firth, J.R. *A synopsis of linguistic theory Studies in linguistic analysis*. Oxford: Blackwell (1957).
- García Cubas 1888-1891** García Cubas, Antonio. 1888-1891. *Diccionario geográfico, histórico y biográfico de los Estados Unidos Mexicanos*. 2 vols. México: Antigua Impr. de Murguía. <https://catalog.hathitrust.org/Record/010416157>.
- Gerhard 1972a** Gerhard, Peter. 1972a. *A Guide to the historical geography of New Spain*. First edition. Cambridge, UK: Cambridge University Press.

- Gerhard 1972b** Gerhard, Peter. 1972b. "Colonial New Spain, 1519–1786: Historical notes on the evolution of minor political jurisdictions". In *Handbook of Middle American Indians, Volume 12: Guide to Ethnohistorical Sources, Part One*, edited by Howard F. Cline, pp. 63–137. Austin: University of Texas Press.
- Gerhard 1991** Gerhard, Peter. 1991. *La frontera sureste de la Nueva España*. México: Universidad Nacional Autónoma de México.
- Gibson 1967** Gibson, Charles. 1967. *Tlaxcala in the sixteenth century*. California: Stanford University Press.
- Goldberg 2017** Goldberg, Yoav. 2017. "Neural network methods for natural language processing". *Synthesis Lectures on Human Language Technologies* 10(1): 1-309. doi:10.2200/S00762ED1V01Y201703HLT037.
- Gregory et al. 2015** Gregory, Ian, Christopher E. Donaldson, Patricia Murrieta-Flores, and Paul E. Rayson. 2015. "Geoparsing, GIS, and textual analysis: Current developments in spatial humanities research". *International Journal of Humanities and Arts Computing* 9(1): 1–14. doi:10.3366/ijhac.2015.0135.
- Gregory et al. 2018** Gregory, Ian Norman, Christopher E. Donaldson, Andrew Hardie, and Paul E. Rayson. 2018. "Modelling Space in Historical Texts". In *The shape of data in Digital Humanities: Modeling texts and text-based resources*, edited by Julia Flanders and Fotis Jannidis. London: Routledge. <http://eprints.lancs.ac.uk/81427/>.
- Grover et al. 2008** Grover, C., S. Givon, R. Tobin, and J. Ball. 2008. "Named entity recognition for digitised historical texts". In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. Paris: ELRA.
- Gruzinski 1991** Gruzinski, Serge. 1991. *La colonización de lo imaginario. Sociedades indígenas y occidentalización en el México español. Siglos XVI-XVIII*. México: Fondo de Cultura Económica (original in French, *La colonisation de l'imaginaire, Sociétés indigènes et occidentalisation dans le Mexique espagnol XVI-XVIII e siècle*. Paris: Editions Gallimard, 1988).
- Guzmán Betancourt 1987** Guzmán Betancourt, Ignacio (Ed.). 1987. *De toponimia y topónimos. Contribuciones al estudio de los nombres de lugar provenientes de lenguas indígenas de México*. México: Instituto Nacional de Antropología e Historia.
- Guzmán Betancourt 1989** Guzmán Betancourt, Ignacio. 1989. *Toponimia mexicana. Bibliografía general*. México: Instituto Nacional de Antropología e Historia.
- Guzmán Betancourt 1998** Guzmán Betancourt, Ignacio, ed. 1998. *Los nombres de México*. México: Secretaría de Relaciones Exteriores.
- Gómez 2004** Gómez, Raúl. 2004. *Los Virreinos americanos*. Madrid: Dastin Export.
- Halternamn 2017** Halterman, Andrew. 2017. "Mordecai: Full text geoparsing and event geocoding". *The Journal of Open Source Software*, 2 (January): 91. doi:10.21105/joss.00091.
- Harvey 1972** Harvey, H. R. 1972. "The Relaciones Geográficas, 1579-1586: Native Languages". In *Handbook of Middle American Indians, Volume 12: Guide to Ethnohistorical Sources, Part One*, edited by Howard F. Cline, pp. 279–323. Austin: University of Texas Press.
- Hersch 1999** Hersch, Martinez, P. (1999). "De hierbas y herbolarios en el Mexico actual". *Arqueología Mexicana*, 39:60-65.
- Hill 2000** Hill, Linda L. 2000. "Core elements of digital gazetteers: Placenames, categories, and footprints". In *Research and advanced technology for digital libraries*, pp. 280–290. Lecture Notes in Computer Science 1923. Berlin: Springer. doi:10.1007/3-540-45268-0_26.
- Hill et al. 1999** Hill, Linda, James Frew, and Qi Zheng. 1999. "Geographic names: The implementation of a gazetteer in a georeferenced digital library". *D-Lib Magazine* 5(1). doi:10.1045/january99-hill.
- Lefebvre and Martínez 2017** Lefebvre, Karin and Carlos Paredes Martínez, eds. 2017. *La memoria de los nombres: la toponimia en la conformación histórica del territorio. De Mesoamérica a México*. Morelia, Michoacán: Universidad Nacional Autónoma de México, Centro de Investigaciones en Geografía Ambiental. <http://www.h-mexico.unam.mx/node/19835>.
- León Portilla 2002** León Portilla, Miguel. 2002. "El destino de las lenguas indígenas de México". In *El despertar de nuestras lenguas (Queman tlachixque totlahtolhuan)*, edited by Natalio Hernández. México: Editorial Diana, Fondo Editorial de las Culturas Indígenas.
- Liceras-Garrido et al. 2019** Liceras-Garrido, Raquel, Mariana Favila-Vázquez, Katherine Bellamy, Patricia Murrieta-Flores, Diego Jiménez-Badillo, et al. 2019. "Digital Approaches to Historical Archaeology: Exploring the Geographies of 16th Century New Spain". *Open Access Journal of Archaeology & Anthropology* 2(1). OAJAA.MS.ID.000526. DOI:

- López de Velasco 1894** López de Velasco, Juan. 1894. *Geografía y descripción universal de las Indias*. Madrid: D. Justo Zaragoza: Establecimiento tipográfico de Fortanet. <http://www.cervantesvirtual.com/obra-visor/geografa-y-descripcin-universal-de-las-indias-recopilada-por-el-cosmografocronista-juan-lpez-de-velasco-desde-el-ao-de-1571-al-de-1574-publicada-por-d-justo-zaragoza-0/html/>.
- Mager et al. 2018** Mager, Manuel, Ximena Gutierrez-Vasques, Gerardo Sierra, and Ivan Meza. 2018. "Challenges of language technologies for the indigenous languages of the Americas". In *Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018)*. <https://arxiv.org/abs/1806.04291>.
- Martins and Murrieta-Flores 2017** Martins, Bruno, and Patricia Murrieta-Flores. 2017. "GeoHumanities 2017 workshop report". In *Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities*, Redondo Beach, California, USA.
- Morato-Moreno 2017** Morato-Moreno, Manuel. 2017. "The Map of Tlacotalpa by Francisco Gali, 1580: An early example of a local coastal chart in Spanish America". *The Cartographic Journal* 55(1): 3-15.
- Moreno Toscano 1968** Moreno Toscano, Alejandra. 1968. *Geografía económica de México (siglo XVI)*. México: El Colegio de México.
- Mundy 1996** Mundy, Barbara. 1996. *The mapping of New Spain: Indigenous cartography and the maps of the Relaciones Geográficas*. Chicago: University of Chicago Press.
- Mundy 2013** Mundy, Barbara E. 2013. "Mapping Babel: A 16th Century Indigenous Map from Mexico". *The Appendix: A New Journal of Narrative & Experimental History*. Online publication available at: <http://theappendix.net/issues/2013/10/mapping-babel-a-sixteenth-century-indigenous-map-from-mexico>.
- Mundy 2014** Mundy, Barbara E. 2014. "Place-names in Mexico-Tenochtitlan". *Ethnohistory* 61 (2): 329-355. doi:10.1215/00141801-2414190.
- Muntzel and Molina 2010** Muntzel, Martha C., and María Elena Villegas Molina, (Eds.). 2010. *Itinerario toponímico de México: Ignacio Guzmán Betancourt*. México: Instituto Nacional de Antropología e Historia.
- Murrieta-Flores and Gregory 2017** Murrieta-Flores, Patricia, and Ian N. Gregory. 2017. "Cruzando Fronteras en humanidades digitales: Análisis geográfico de textos de interés histórico y arqueológico con sistemas de información geográfica". In *Arqueología computacional. Nuevos enfoques para la documentación, análisis y difusión del patrimonio cultural*, edited by Diego Jiménez-Badillo, pp. 199-212. México: Instituto Nacional de Antropología e Historia.
- Murrieta-Flores et al. 2015** Murrieta-Flores, Patricia, Alistair Baron, Ian Gregory, Andrew Hardie, and Paul Rayson. 2015. "Automatically analyzing large texts in a GIS environment: The Registrar General's Reports and cholera in the 19th century". *Transactions in GIS* 19(2): 296-320. doi:10.1111/tgis.12106.
- Murrieta-Flores et al. 2017** Murrieta-Flores, Patricia, Christopher Donaldson, and Ian N Gregory. 2017. "GIS and literary history: Advancing digital humanities research through the spatial analysis of eighteenth-century travel writing". *Digital Humanities Quarterly* 11(1). <http://www.digitalhumanities.org/dhq/vol/11/1/000283/000283.html>.
- Márquez and Wold 1998** Márquez, Ofelia, and Lilian Ramos Navarro Wold. 1998. *Compilation of colonial Spanish terms and document related phrases*. Midway City, CA: SHAR Press (Society of Historical and Ancestral Research). <http://www.somosprimos.com/spanishterms/spanishterms.htm>.
- Nick and Tent 2017** Nick, I. M., and Jan Tent. 2017. "Guest editorial on indigenous names and toponyms". *Names, A Journal of Onomastics* 65(4): 190-193.
- Nuttall 1926** Nuttall, Zellia, (Ed.). 1926. *Official reports on the towns of Tequizistlan, Tepechpan, Acolman, and San Juan Teotihuacan, sent by Francisco de Castañeda to His Majesty, Philip II, and the Council of the Indies, in 1580. Papers of the Peabody Museum of Archaeology and Ethnology*, 11(2): 45-84. Cambridge, Massachusetts: Harvard University.
- Pedrote Romero and Bravo-García 2019** Pedrote Romero, Antonio and Eva Bravo-García. 2019. "La autoría de las Relaciones Geográficas mexicanas: Las voces náhuatl a través de los redactores". *Anuario de Estudios Americanos*, 76(1): 123-153. Doi: 10.3989/aeamer.2019.1.06.
- Peñafiel 1897** Peñafiel, Antonio. 1897. *Nomenclatura geográfica de México, etimologías de los nombres de lugar correspondientes a los principales idiomas que se hablan en la República*. 3 vols. México: Oficina tipográfica de la Secretaria de Fomento. <http://archive.org/details/nomenclaturageog00pe>.
- Porter et al. 2015** Porter, Catherine, Paul Atkinson, and Ian Gregory. 2015. "Geographical Text Analysis: A new approach to understanding nineteenth-century mortality". *Health & Place*, 36 (November): 25-34.

- Robelo 1897** Robelo, Cecilio Agustín. 1897. *Nombres geográficos indígenas del estado de Morelos. Estudio crítico de varias obras de Toponomatología nahoa*. Cuernavaca: Luis G. Miranda impresor. <http://archive.org/details/nombresgeogrifi00robeuoft>.
- Robelo 1900** Robelo, Cecilio Agustín. 1900. *Nombres Geográficos Indígenas Del Estado de México: Estudio Crítico Etimológico*. Cuernavaca: Luis G. Miranda impresor. <http://cd.dgb.uanl.mx/handle/201504211/14251>.
- Robelo 1902a** Robelo, Cecilio Agustín. 1902a. *Nombres geográficos mexicanos del estado de Veracruz: Estudio crítico etimológico*. Cuernavaca: Luis G. Miranda impresor.
- Robelo 1902b** Robelo, Cecilio Agustín. 1902b. *Toponimia tarasco-hispano-nahoa*. Cuernavaca: Impr. de J. D. Rojas. <http://archive.org/details/toponimiatarasc00robegoog>.
- Robelo 1904** Robelo, Cecilio Agustín. 1904. *Diccionario de aztequismos, ó sea, catalogo de las palabras del idioma nahuatl, azteca ó mexicano, introducidas al idioma castellano bajo diversas formas*. Cuernavaca: Printed by the author. <https://catalog.hathitrust.org/Record/009024985>.
- Robelo 1905** Robelo, Cecilio Agustín. 1905. *Diccionario de mitología náhuatl*. México: Ediciones Fuente Cultural; Librería Navarro. <https://catalog.hathitrust.org/Record/001940449>.
- Robertson 1972** Robertson, Donald. 1972. "The pinturas (maps) of the *Relaciones Geográficas* with a catalog". In *Handbook of Middle American Indians, Volume 12: Guide to Ethnohistorical Sources, Part One*, edited by Howard F. Cline, pp. 243–278. Austin: University of Texas Press.
- Santos et al. 2015** Santos, João, Ivo Anastácio, and Bruno Martins. 2015. "Using machine learning methods for disambiguating place references in textual documents". *GeoJournal* 80(3): 375–392. doi:10.1007/s10708–014–9553–y.
- Santos et al. 2017** Santos, Rui, Patricia Murrieta–Flores, and Bruno Martins. 2017. "Learning to combine multiple string similarity metrics for effective toponym matching". *International Journal of Digital Earth*. doi:10.1080/17538947.2017.1371253.
- Santos et al. 2018** Santos, Rui, Patricia Murrieta–Flores, Pável Calado, and Bruno Martins. 2018. "Toponym matching through deep neural networks". *International Journal of Geographical Information Science* 32(2): 324–348. doi:10.1080/13658816.2017.1390119.
- Sprugnoli et al. 2017** Sprugnoli, Rachele, Giovanni Moretti, Bruno Kessler, Sara Tonelli, and Stefano Menini. 2017. "Fifty years of European history through the lens of Computational Linguistics: The De Gasperi Project". *Italian Journal of Computational Linguistics* 2(2): 89–100.
- Starr 1920** Starr, Frederick. 1920. *Aztec place-names, their meaning and mode of composition, selected from the Spanish of Agustin de la Rosa, Antonio Peñafiel and Cecilio A. Robelo*. Chicago: Privately printed by the author. <https://catalog.hathitrust.org/Record/001360976>.
- Sánchez–Martínez et al. 2013** Sánchez–Martínez, Felipe, Isabel Martínez–Sempere, Xavier Ivars–Ribes, and Rafael C. Carrasco. 2013. "An Open diachronic corpus of historical Spanish: Annotation criteria and automatic modernisation of spelling". *arXiv:1306.3692 [cs]*, June. <http://arxiv.org/abs/1306.3692>.
- Tanck de Estrada 2005** Tanck de Estrada, Dorothy. 2005. *Atlas ilustrado de los pueblos de indios*. México: El Colegio de México, El Colegio Mexiquense, Comisión Nacional para el Desarrollo de los Pueblos Indígenas, and Fomento Cultural Banamex.
- Taulé et al. 2008** Taulé, Mariona, M. Antònia Martí, and Marta Recasens. 2008. "AnCorà: Multilevel annotated corpora for Catalan and Spanish". *LREC 2008*. <https://aclanthology.info/papers/L08–1222/I08–1222>.
- Taylor et al. 2017** Taylor, Joanna E., Ian N. Gregory, and Christopher E. Donaldson. 2017. "Combining close and distant reading: A multiscalar analysis of the English Lake District's historical soundscape". *International Journal of Humanities and Arts Computing*, 12(2): 163–182. <http://eprints.lancs.ac.uk/89167/>.
- Van Hooland et al. 2015** Van Hooland, S., M. de Wilde, R. Verborgh, T. Steiner, and R. Van de Walle. 2015. "Exploring entity recognition and disambiguation for cultural heritage collections". *Digital Scholarship in the Humanities* 30(2): 262–279. doi:10.1093/llc/fqt067.
- Won et al. 2018** Won, Miguel, Patricia Murrieta–Flores, and Bruno Martins. 2018. "Ensemble named entity recognition (NER): Evaluating NER Tools in the identification of place names in historical corpora". *Frontiers in Digital Humanities* 5. doi:10.3389/fdigh.2018.00002.

