frontiers in Digital Humanities

ORIGINAL RESEARCH
published: 09 March 2018
doi: 10.3389/fdigh.2018.00002

Ensemble Named Entity Recognition (NER): Evaluating NER Tools in the Identification of Place Names in Historical Corpora

Miguel Won[1], Patricia Murrieta-Flores[2] and Bruno Martins[1]*

[1] Instituto Superior Técnico, INESC-ID, Universidade de Lisboa, Lisbon, Portugal, [2] History Department, Digital Humanities Hub, Lancaster University, Lancaster, United Kingdom

The field of Spatial Humanities has advanced substantially in the past years. The identification and extraction of toponyms and spatial information mentioned in historical text collections has allowed its use in innovative ways, making possible the application of spatial analysis and the mapping of these places with geographic information systems. For instance, automated place name identification is possible with Named Entity Recognition (NER) systems. Statistical NER methods based on supervised learning, in particular, are highly successful with modern datasets. However, there are still major challenges to address when dealing with historical corpora. These challenges include language changes over time, spelling variations, transliterations, OCR errors, and sources written in multiple languages among others. In this article, considering a task of place name recognition over two collections of historical correspondence, we report an evaluation of five NER systems and an approach that combines these through a voting system. We found that although individual performance of each NER system was corpus dependent, the ensemble combination was able to achieve consistent measures of precision and recall, outperforming the individual NER systems. In addition, the results showed that these NER systems are not strongly dependent on preprocessing and translation to Modern English.

Keywords: Spatial Humanities, Digital Humanities, natural language processing, historical corpora, toponym recognition, Early-Modern English, Republic of Letters

OPEN ACCESS

Edited by:
Arianna Ciula, King's College London, United Kingdom

Reviewed by:
Adam Crymble, University of Hertfordshire, United Kingdom
Sara Tonelli, Fondazione Bruno Kessler, Italy

****Correspondence:***
Patricia Murrieta-Flores
p.a.murrieta-flores@lancaster.ac.uk

Specialty section:
This article was submitted to Digital History, a section of the journal Frontiers in Digital Humanities

Received: *24 November 2017*
Accepted: *12 February 2018*
Published: *09 March 2018*

Citation:
Won M, Murrieta-Flores P and Martins B (2018) Ensemble Named Entity Recognition (NER): Evaluating NER Tools in the Identification of Place Names in Historical Corpora. Front. Digit. Humanit. 5:2. doi: 10.3389/fdigh.2018.00002

INTRODUCTION

The exploration of place in texts within the field of Spatial Humanities has advanced substantially in the past years. The combination of geographic information systems (GIS), natural language processing (NLP), and Corpus Linguistics has enabled new ways of identifying and analyzing the mention of place names in literary and historical corpora (Dross, 2006; Bailey and Schick, 2009; Hyun, 2009; Grover et al., 2010; Gregory and Hardie, 2011; Piotrowski, 2012; Silveira, 2014; Gregory et al., 2015; Murrieta-Flores and Gregory, 2015; Murrieta-Flores et al., 2015; Porter et al., 2015; Cooper et al., 2016). Although successful so far, the majority of the geographical research carried out in Digital Humanities has mainly relied on relatively simple techniques: the identification and annotation of place names either by hand or through customized Named Entity Recognition (NER) techniques for geoparsing, where places in the text can be automatically identified and

then disambiguated by matching them to a particular gazetteer, normally relying on rules or preexisting statistical NER models. The problems derived from the adequate identification of place names are well known in the field of geographic information retrieval (GIR) (Purves and Jones, 2011; Santos et al., 2015a,b; Melo and Martins, 2016). However, it has been only recently that the field of Spatial Humanities has started to address these problems in a more systematic way (see, for instance, the DH2016 workshop dedicated to historical gazetteers[1]). In the case of historical documents, the challenges in the identification of places in documents for their later analysis are the same as in other domains considered in GIR, and these can be mainly divided in issues of (a) place reference identification and (b) place reference disambiguation. Issue (a) can involve challenges related to language changes over time, spelling variations, OCR errors, sources written in multiple languages, and general ambiguity in language use (e.g., words that, depending on the context, can refer either to places or to other types of concepts), among others. On the other hand, issue (b) refers to the correct association of geospatial information (e.g., gazetteer entries) or coordinates to a particular place (Santos et al., 2015a,b), also involving challenges related to ambiguity in language use (e.g., places sharing the same name), or related to changes in administrative geography (e.g., boundary changes over time, places that change names, etc.). At the moment, the identification of geographical places written in historical documents is not carried out in a uniform fashion across the Spatial Humanities and, as said before, it usually relies on manual annotation or the use of one customized NER tool and one main gazetteer. The reasons for this are varied, and they might be related to the fact that the Spatial Humanities is still a young field of research that remains highly experimental, and that the challenges involved in creating fully automated processes are not easy to tackle (Gregory et al., 2015; Purves and Derungs, 2015; Wing, 2015). These might range from the necessary creation of tailored rules and/or applications for the case of languages other than English, to various problems derived from the fact that NER systems were created with modern contents in mind, in most cases relying on statistical learning from annotated datasets of modern news articles (Grover et al., 2008, 2010; Rupp et al., 2013, 2014; Batjargal et al., 2014; Cneudecker, 2014). A necessary step to move toward a possible standardization in the methodologies we employ in the Spatial Humanities, to automatically resolve place names in corpora, is the assessment of the performance of different NER systems when used in historical sources.

The work that we present in this article derives from a Short Term Scientific Mission carried out in the context of the EU COST Action IS1310 "Reassembling the Republic of Letters" project.[2] The so called *Respublica literaria* makes reference to an extraordinary network of letters that enabled men and women to share and exchange all sorts of information and knowledge between 1500 and 1800, supported by the revolution in postal communications at the time. This exchange knitted together civil circles and allowed not only intellectual breakthroughs but also the discussion of economic, political, religious, and cultural ideas among others, which would see the emergence of many of the modern European values and institutions (Goodman, 1996; Feingold, 2003; Hamilton et al., 2005; Shelford, 2007; Berbara and Enenkel, 2011). The extent of this network is enormous, and the materials and documentation available related to it are, not only scattered around the world but also in many cases difficult to access (Ostrander, 1999; Dalton, 2004; Furey, 2006). The "Reassembling the Republic of Letters" project envisions a highly interdisciplinary approach combining transnational digital infrastructures, as well as historiographical and computational methods aiming to collect, standardize, analyze, and visualize unprecedented quantities of this epistolary data. In this context, diverse groups within the Action have been working on different aspects of this aim. *Working Group 1: Space and Time* is looking to analyze the spatial and temporal dimensions of these very large and disparate historical epistolary datasets, finding the most optimal ways of resolving place names for their later integration and analysis with GIS.

The research related to the scientific mission we present in this article had its main objective in exploring and evaluating modern tools that could be used to automate the identification and retrieval of geographic information from historical texts, particularly from the Republic of Letters (RofL) datasets. This study also aimed to bring new insights giving shape to the future research agenda of this project, looking to address the multiple problems involved in place reference identification and disambiguation in historical datasets. As part of this work, we carried out an evaluation of a selection of NER systems freely available and commonly used in historical datasets, and we implemented a voting system that combines all the considered NER systems, aiming to improve results. For our particular case study, we used two letter collections: the *Mary Hamilton Papers* and the *Samuel Hartlib Papers*. These sets enabled us to test the systems in different scenarios, including their performance with Early and Modern English as well as diverse levels of data cleanliness.

The rest of the article is organized as follows: Section "NER in Historical Documents" presents an overview on previous work related to NER over historical documents, providing a context to our approach. In "The Datasets" Section the historical background regarding both letter collections is described as an introduction, and then we explain the current format of these sources, and specify the parts of the collections that were used. The "Methodology" Section reports the tasks that were carried out regarding the preprocessing of the collections and details the experimental methodology including the rationale and the approaches that were taken for the use and assessment of the NER systems. The "Results" Section explains the results, providing also an in-depth discussion. The "Conclusions and Future Work" are considered in the final section.

NER IN HISTORICAL DOCUMENTS

Named Entity Recognition is a subtask of NLP aiming to identify real-world entities in texts, such as names of persons, organizations, and locations, among others (Nadeau and Sekine, 2007).

[1] http://aplace4places.github.io/program.html.
[2] http://www.republicofletters.net/.

In the past few years and with the advent of massive digitization of textual sources [see, for instance, the report by Gerhard and van den Heuvel (2015)], NER has become of increasing interest in Digital Humanities due to its potential for information extraction and analysis at large scale from historical and literary documents. Although NER technologies can achieve impressive results with modern corpora, historical documents pose multiple challenges. In comparison with the NER bibliography produced in the NLP field each year, the number of examples dealing with datasets of historical character is still small. Despite this, it can be said that research in this area seems to be growing, and there are several examples of Digital Humanities projects using NER systems over historical datasets (Crane and Jones, 2006; Borin et al., 2007; Byrne, 2007; Grover et al., 2008; Brooke et al., 2015; Mac Kim and Cassidy, 2015; van Hooland et al., 2015; Ehrmann et al., 2016; Sprugnoli et al., 2017). In general, these works have not only experimented with NER applied to historical materials such as newspapers and other text collections but also many of them have addressed some of the most pressing challenges involved in the use of current state-of-the-art NER systems with historical materials. This is the case of research using NER for materials with disparate qualities of digitization and OCR, non-European or classical languages, or collections featuring spelling variations (Alex et al., 2012; Batjargal et al., 2014; Neudecker et al., 2014; Nagai et al., 2015; Erdmann et al., 2016; Kettunen et al., 2016). Previous research within Digital Humanities has also tackled the related problem of text geoparsing, leveraging NER methods for recognizing place references in text, often together with other heuristics for the complete resolution of place references into gazetteer entries and/or geographical coordinates (Rayson et al., 2006; Baron and Rayson, 2008; Pilz et al., 2008; Grover et al., 2010; Freire et al., 2011; Gregory and Hardie, 2011; Brown et al., 2012; Alex et al., 2015; Gregory et al., 2015; Murrieta-Flores et al., 2015; Santos et al., 2015a,b; Wing, 2015; Clifford et al., 2016).

Although the increase in this kind of research is greatly encouraging, for research related to historical collections a NER system should ideally be able to work with historical languages, and geoparsing techniques should be able to work with historical gazetteers (Goodchild and Hill, 2008; Manguinhas et al., 2008; Berman et al., 2016). Nevertheless, these are still early days in the Spatial Humanities and although such systems are not yet widely available, approaches integrating historical or tailored gazetteers with NER and Open Linked Data technologies are gradually emerging (Gregory et al., 2015; see, for instance, Alex et al., 2015; Simon et al., 2015). In the meantime, it can be said that the majority of these approaches make use of only one NER system. In the case of place names and if geographic disambiguation is also of interest, modern gazetteers are usually used. In addition, the wide variation of historical material available and differences in terms of languages and quality of capture (e.g., different OCR qualities, OCR versus transcription, etc.) means that comparisons between NER in different corpora might be very difficult if not impossible. To date, there have been only few attempts to evaluate different NER systems within Digital Humanities research, primarily with modern historical material (i.e., nineteenth- and twentieth-century newspapers) (Ehrmann et al., 2016). Therefore, an evaluation of available NER systems and a possible comparison between Early-Modern and Modern English were considered invaluable to understand not only the performance of these technologies with earlier datasets but also to identify and consider the most efficient direction that research related to these technologies in the context of the Spatial Humanities should take in the next few years.

THE DATASETS

The datasets chosen to carry out our assessments were the Mary Hamilton Papers and the Samuel Hartlib collection. These were selected because (a) they are considered datasets of historical significance and (b) they present different issues and challenges. While the Samuel Hartlib Papers were written in Early-Modern English (EME), the Mary Hamilton Papers were written in Modern English (ME). This difference allowed us to test the performance of NER in both versions, and check, in the case of the Hartlib Papers, whether transforming EME into a ME version would make a difference in the NER systems' performance. In addition, the Hamilton Papers were annotated in TEI, including manual annotations for person and place names, allowing us to use this dataset further to evaluate the final results. In the case of the Hartlib Papers, the available annotations correspond exclusively to the editorial process and they do not include any annotated named entities. Therefore, for the evaluation of final results in this case, 50 letters were selected, and place names were first manually annotated for this set.

The Mary Hamilton Papers

Mary Hamilton (1756–1816) was courtier and governess of the daughters of George III. Although she is not considered a particularly prominent figure, she stood at the center of the intellectual, aristocratic, literary, and artistic circles of London during the late eighteenth-century (Prendergast, 2015). As such, she had direct contact with many members of the royal family, as well as many significant figures of the time. The collection now called the "Mary Hamilton Papers" includes a vast set of correspondence between Hamilton and her husband John Dickenson, the royal family including the queen and princesses, her friends at the court, and multiple members of the Bluestocking circle such as Elizabeth Montagu, Frances Burney, and Mary Delany, among others. Written in ME, in addition to 2,474 letters, the collection also contains 16 diaries and 6 manuscript volumes, all of which are part of the University of Manchester Library's Special Collection.[3] This set is regarded as an important resource not only for the study of the social circles of Britain but also the intellectual elites, as well as the political, economic, and cultural environments at the time. Although most of this collection is already digitized, the vast majority is not transcribed, and currently, as part of the "Image to Text: Mary Hamilton Papers (c.1750–c.1820) Project" directed by David Denison and Nuria Yáñez-Bouza, only 161 letters dated between 1764 and 1819 have been transliterated (Denison and Yáñez-Bouza, 2016). Our work

[3] http://archiveshub.jisc.ac.uk/features/maryhamilton/.

used this section of the collection (161 letters), which contains over 70,000 words of text and is available for research in plain text format, as well as annotated in TEI.

The Samuel Hartlib Papers

Samuel Hartlib (1600–1662) has been regarded as one of the greatest intelligencers of the seventeenth century. From a mercantile family and grandson of the head of the English trading company in Elbing, Hartlib settled during the late 1620s in England after fleeing from the war taking place in Central Europe and would become an accomplished scholar as well as one of the most active reformers and connected intellectuals at the time (Webster, 1970). His archive is considered one of the richest in Europe due to the insight it provides in terms of the intellectual advancements provoked by the dissemination of ideas, gathering of information, technical discoveries, and theological discussions taking place in the network he was part of, and fueled by the displacements resulting from the turbulent period he lived in. Written mainly in EME, but also German, French, Dutch, and Latin, and reaching all Europe, as well as Great Britain, Ireland, and New England, his surviving archive is not only extant, running over 25,000 folios but it is also considered highly complex in terms of its geographic, chronologic and prosopographical span (Greengrass et al., 2002, 2013). The first initiative to digitize this material was the Hartlib Papers Project, which finalized in 1996 and created a complete electronic edition with full-text transcriptions and facsimile images of these texts. Later projects such as Cultures of Knowledge[4] and Early-Modern Letters Online (EMLO)[5] have contributed to the enhancement of this collection. Today, an enlarged edition with additional material is available in an HTML annotated format through the Digital Humanities Institute (previously the Humanities Research Institute) at the University of Sheffield,[6] and a selection of 4,718 records is available through EMLO. Although the material from the collection made available by the DHI is large, consisting of around 3,165 letters, the Hartlib corpus, unlike the Hamilton Papers, does not contain annotations that might allow the evaluation of the NER systems with the full corpus. For this reason, we have selected a set of 54 letters and manually annotated, within the documents, all mentioned locations. This set was used to measure the performance of each NER system in respect to its ability to identify the mentioned locations in the Hartlib letters.

METHODOLOGY

Data Preprocessing

Although both datasets are digitized, preprocessing them was required to optimally perform the NER tasks with the tools that were available for us to evaluate. This process consisted mainly in cleaning the datasets, i.e., translating the original markup into textual contents that can be taken as input by the NER tools, while at the same time taking care of problems such as word hyphenation. As said before, the Mary Hamilton set contains a total of 161 letters in XML files annotated in TEI. The annotations contain metadata such as authorship, date, information about the transliteration project, context of the letter according to previous research, corrections and suggestions made by the transliterator, and particular words and/or phrases annotated within the body text, including place names (**Figure 1**). In this case, the preprocessing consisted in the extraction of the body texts from the XML code, followed by a tokenization and tagging process, where each word, digit, and punctuation symbol was labeled with the tags "LOC" for location, and "O" for other. The TEI XML annotations were used for the labeling process of locations, and we also used them to address issues such as expanding abbreviations or removing word hyphenations.

In the case of the Hartlib corpus, it is coded in a set of HTML files that show a faithful representation of the original letters (i.e., they present the original text, together with a series of comments related to the editorial process and written by the transcribers) (**Figure 2**).

While there are words being suggested within angle brackets to clarify the meaning behind some phrases, all the comments are written within square brackets. These comments can consist in simple notes that should not be part of the main text, or suggestions from the transcribers about, for instance, words that cannot longer be read in the original manuscript. In this latter case, we considered that the suggestion should be included in the main text. See, for instance, the next two examples:

1. [word/s deleted]
2. [another hand?: Mr Williamsons]

The first case is a note stating that there was a deleted word in the original manuscript, while the second example is referring to a case where the manuscript has the name "Mr Williamsons" written in the text, although written by a different hand. For the present work, the cleaning process had to consider that the closest form of a meaningful text is needed, without the notes from the transcriber but with the suggestions of what should be incorporated in the final text. For the equivalent cases to the first example, a simple deletion from the original HTML file can be performed because it constitutes the transcriber comment and not part of the original text. However, in the case of the second example, "Mr Williamsons" should be kept. Usually, this type of preprocessing and cleaning can be implemented in an automated fashion, by following the definition of the patterns of annotation that were followed. However, we were not able to find a universal pattern for all the comments that would allow the implementation of an automatic cleaning process. Therefore, to study how the performance of NER tools depended on these comments, two cleaning processes were defined: *full* and *fast clean*. We have studied the impact of each cleaning process in the same set of 54 annotated letters mentioned earlier.

In the *full clean* process, we manually identified common patterns and exceptions to the rule in the transcriber's notes within the 54 letters. Based on these findings, all square brackets were then correctly removed and replaced by the appropriate text. This cleaning process creates the closest text form to the original.

[4]http://www.culturesofknowledge.org/.
[5]http://emlo.bodleian.ox.ac.uk/.
[6]https://www.dhi.ac.uk/projects/hartlib/.

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI>
 <teiHeader...>

<text xml:id="HAM/1/1/2/4">
        <body>

        <pb n="1"/>
        <lb/><dateline rend="align-right"><address><abbr expan="Queen's">Q.</abbr>
Lodge <placeName>Windsor</placeName></address>. <date when="1780-08-30">30
August. 1780</date>.</dateline>
        <lb/><salute type="opening">My dear <name role="addressee">Miſs
Hamilton.</name></salute> What can I
        <lb/>have to say? not much indeed! but to
        <lb/>wish You a good Morning, in the pretty
        <lb/>Blue and white Room where I had
        <lb/>the pleasure to sit and read with
        <lb/>You <note resp="#DAM" comment="&amp;quot;The Hermit&amp;quot;, a poem by
Rev. Thomas Parnell (d. 1717/8)."/>the <hi rend="underlined">Hermit</hi> a Poem which
is such
        <lb/>a Favorite with me that I have read
        <lb/>it twice this Summer, Oh what a <choice n="hyp"><orig>ble<g ref="#sm-long-
s">s</g>-
        <lb break="no"/><hi rend="align-
right">sing</hi></orig><reg>bleſsing<lb/></reg></choice>
```

FIGURE 1 | TEI XML sample taken from the Hamilton Letters.

Ref: 1/1/1A-3B
Notes: Turnbull (HDC p272) suggests just after April 1653.

[1/1/1A]

Worthy freind

your discourse with me the last weeke, & the earnestnes of your desire to see the grounds of my resolution made out; which is to bee quiet, & do my duty under the present power, as acquiescing in this reuolution of gouvernment hath put me upon a designe to giue you satisfaction, at least so farre as the opening of my thoughts will yeeld it, & as it may please God to blesse it ~~unto you;~~ <to cleer> ~~to both you see~~ <unto you> your duty in this iuncture of time wherin many are staggered; & murmur at the proceedings of those *that* are in power. but I haue beene taught ~~by~~ <of> God, to quiet mine ~~owne~~ affections; & <still my> thoughts at the Changes w*hi*ch ~~fall out of~~ <hee brings upon> this <present> world ~~as~~ <chiefly when> they relate to the ~~great~~ wheele of <a nationall> Gouvernment, w*hi*ch as it is supreme so <it> stands immediatly under the hand of the most high, who ruleth in the Kingdome of men & giueth it to whomsoeuer he will. I say I haue been taught to quiet my spirit at ~~the~~ <such> Changes, because I cannot find *that* it doth belong unto me, to iudge definitiuely of the rights w*hi*ch the supreme powers ~~ouer us in the world~~ <have>, <or> pretend to haue unto their places. <for> first I said, *that* it is no part of my duty as a Christian; to ~~subiect~~ <burden> my Conscience, ~~& burden it~~ with the affaires of state <which are> intrusted to the management of other men: those whom God doth engage into public places, it is their proper worke to Charge their Consciences with the care thereof; & if they do not lay their trust Conscionably to heart for the ende for w*hi*ch God hath put them in their places; but turne it to their priuat aduantages, I think it is a happines unto them, soone to bee put out of the same; because the lesse time they stay therin, the lesse guilt they contract unto their soules, & the lesse iniurie they do unto the public. [*left margin:* 1. Petr. 4. 15] the Apostle ~~forbids~~ <bids> all Christians ~~to~~ looke to themselues *that* they bee not found <as> busie bodies in other mens matters, & this sinne he ranketh ~~in *that* place under~~ <with> the Generall head of euill doing, & sets <it> parallel to the particular sinnes of Murther & Theft; now if I being a priuat man; should take upon me to Charge my Conscience w*i*th the iudicature of Public affaires; to determine by w*hat* right those *that* manage them take upon them their public places; I should as I conceiue go beyond my line, & shew my self a busie body in ~~other mens~~ <their> matters. <for> my Christian aime & profession, doth oblige me <~~only~~ first> to worke out mine owne Saluation in mine ~~owne~~ <priuat> calling, with feare & trembling, & <then to> hold forth the word of life <to others> *that* is the rule of harmlesnes & unblameablenes; as it becommeth a Child of God, without rebuke [*left margin:* Phil 2. 15, 16] in the midst [*word deleted*] of a crooked & perverse nation <therefore> a

FIGURE 2 | Rendered HTML sample taken from the Hartlib Papers. [Copyright: Greengrass et al. (2013). The Hartlib Papers. Published by HRI Online Publications, Sheffield. Available at: http://www.hrionline.ac.uk/hartlib.]

However, as discussed, its creation was not done automatically. In the case of the *fast clean* method, all square brackets and its content were automatically removed. Therefore, in this version, the final texts have missing words, in particular the suggestions written from the transcriber. The advantage of this preprocessing is that it can be automated and applied to the full set. We have additionally applied standard cleaning operations to both sets: HTML code, page breaks, and non-alphanumeric characters, except punctuation, were removed. Equivalently to the Mary Hamilton set of letters, tokenization and tagging with "LOC" and "O" tags was also applied.

In addition to the cleaning process, the performance of NER tools is also language dependent. Therefore, to study the potential impact of language stage difference, in the case of the Hartlib Collection we have translated documents from the original EME to ME, afterward considering both language stages for comparing NER results.

All scripts used to preprocess and parse the analyzed texts were written in Python 2.7, making use of the NLTK package (NLTK, 2017). The translation from EME to ME was performed using two tools: MorphAdorner (Burns, 2013) and VARD (Baron and Rayson, 2008; Archer et al., 2015). To test possible translation tool bias, we used each tool for performing the translation separately. Comparative results are shown in Section "Results."

Ensemble NER

Using the annotated epistolary corpora described earlier, we evaluated the performance of readily available NER tools in the recognition of place references. To ensure the independence of each prediction in the ensemble system, we considered multiple NER systems that (a) have been used (e.g., Edinburgh Geoparser, Stanford NER, etc.) or could be used with historical corpora, (b) are representative of the different approaches that are commonly used (e.g., rule-based systems and systems based on supervised machine learning, either considering linear sequence prediction models leveraging extensive feature engineering or models leveraging word embeddings and neural network architectures), and (c) are simple in terms of user interface and achieve a good performance on standard corpora of modern newswire text (e.g., the corpus used on the CoNLL-2003 evaluation on NER methods) (Tjong Kim Sang and De Meulder, 2003).

Most of the NER tools that were considered for our study are based on supervised machine learning, but we nonetheless did not experiment with using the epistolary corpora for training new models, instead focusing on evaluating the performance of the models that are directly distributed with these tools and that were trained and optimized for processing ME. Existing NER methods based on machine learning achieve remarkable results over modern newswire text [e.g., the system described in Lample et al. (2016) reports on an F1 score of 90.94 when recognizing names for persons, locations, and organizations, over the CoNLL corpus] but, given that the English language has changed significantly over time, even since the start of the early-modern period, it should be expected that the performance of these tools degrades significantly when processing historical contents. For instance, the modern practice of restricting capitalization to names, name-like entities, and certain emphatic uses, is only about two centuries old. In earlier English, nouns were freely capitalized, and capitalization is thus not a reliable way of picking out proper nouns. Although proper nouns have usually been capitalized in all forms of written English since about 1550, before that names could appear in lowercase. In an attempt to handle these issues, and as stated in the previous section, some of our experiments leveraged existing tools for mapping variant spellings to their standard modern forms, namely, MorphAdorner (Burns, 2013) and VARD (Baron and Rayson, 2008; Archer et al., 2015). These two tools leverage rules, word lists, and extended search techniques (e.g., spelling correction methods and other heuristics) for standardizing and modernizing spelling.

In the case of NER tools leveraging supervised machine learning, the recognition of the named entities is typically modeled as a sequence prediction task, where the objective is to assign a specific tag to each word in an input sentence of text (e.g., tag "LOC" for location and "O" for other). From these word tags, it is then possible to retrieve the spans of text that correspond to the named entities.

The following five different NER systems have been used in our tests: Stanford NER, NER-Tagger, the Edinburgh Geoparser, spaCy, and Polyglot-NER.

The Stanford NER software package[7] provides a general implementation of an entity recognition method based on supervised machine learning (Finkel et al., 2005), specifically leveraging sequence models based on the formalism of linear chain conditional random fields (CRFs). In brief, linear chain CRFs are discriminative probabilistic graphical models that work by estimating the conditional probability of a tag sequence given a sequence of words. The input sequence of words is modeled through features that are restricted to depend locally on the output tags (e.g., in first-order chain CRFs, features may only depend on pairs of output tags, although they can also consider the words in the entire input sequence), but the probabilities that are assigned to specific tagging decisions are normalized over the entire input sequence. After inferring the parameters (i.e., the weights of the different features) of a CRF model with supervised learning, an efficient search algorithm can be used to efficiently compute the best tagging (i.e., the most probable sequence of tags) for new input sequences of words. The software is distributed with a model for recognizing named entities (i.e., persons, locations, organizations, and other miscellaneous entities) in English text, trained with a mixture of data from previous NER competitions focusing on modern newswire documents (i.e., the CoNLL-03, MUC, and ACE named entity corpora). The project website mentions that the model that is directly provided with the tool is similar to the baseline local + Viterbi model described in Finkel et al. (2005), although adding new features based on distributional similarity on top of the standard features described in the paper (i.e., features based on word identity, capitalization, word suffixes and prefixes, lexicons of common nouns, etc.), which make the model more robust and domain independent.

The NER-Tagger software package[8] implements another approach based on supervised machine learning, in this case

[7] http://nlp.stanford.edu/software/CRF-NER.shtml.
[8] http://github.com/glample/tagger.

corresponding to one of the deep neural network architectures [i.e., the long short-term memory network (LSTM)-CRF approach] described by Lample et al. (2016). Specifically, this state-of-the-art tool leverages a type of recurrent neural network architecture known in the literature as LSTMs, in combination with the idea of modeling tagging decisions globally, as in the aforementioned CRFs. Instead of leveraging extensive feature engineering, this approach uses pretrained representations for the words (i.e., word embeddings) as the sole input features, building the embeddings through a procedure that considers co-occurrences in large corpora as well as word order, together also with the characters that compose the individual words (Ling et al., 2015) (i.e., even in the case of words that were not present in the corpus that was used for model training, this procedure can generate word embeddings with basis on the individual characters). A LSTM-CRF English model for assigning NER tags to words in English sentences, trained with data from the CoNLL competition (Tjong Kim Sang and De Meulder, 2003), is made available with this tool.

The Edinburgh Geoparser[9] is a ruled-based system that automatically recognizes place names and person names in text and that also disambiguates place names with respect to a gazetteer (Grover et al., 2010; Alex et al., 2015). The NER component is made up of a number of subcomponents that perform lexical lookups (i.e., words or sequences of words are looked up in various lexicons, for instance, listing common English words, person forenames, or geographic locations), or that apply rules that leverage linguistic context (e.g., matching titles for persons or words denoting place types). The disambiguation stage performs a gazetteer lookup for retrieving candidate disambiguations, and it then applies heuristics to rank the candidates (e.g., prefer bigger places, or prefer populated places to facilities). The original version of the Edinburgh Geoparser was a demonstrator configured for modern text but, since then, the system has been applied in numerous projects, and it has been adapted to georeference historical text collections, as well as modern-day newspaper text.

The spaCy[10] *software package* offers a fast statistical NER approach, based on a pastiche of well-known methods that is not currently described in any single publication. The default English model that comes with this tool identifies various named and numeric entities, including companies, locations, organizations, and products. In the website, the authors mention that their NER model corresponds to a greedy transition-based parser where the transition system is equivalent to the NER tagging scheme [i.e., an approach similar to the second model that is described in the paper by Lample et al. (2016)] guided by a linear model whose weights are learned using the averaged perceptron loss, *via* the dynamic oracle imitation learning strategy.

The Polyglot software package[11] implements the language-independent technique described by Al-Rfou et al. (2015), leveraging word embeddings [i.e., vectorial representations for words which encode semantic and syntactic features, pretrained from co-occurrence information on large amounts of text according to the procedure described by Al-Rfou et al. (2013)] as the sole features within a word-level classifier based on a simple neural network (i.e., a model with a single hidden layer that assigns NER tags for each word, taking as features the embeddings for the words in a window of text centered around each word that is to be classified). Polyglot recognizes three categories of entities (i.e., persons, locations, and organizations) and it currently supports 39 major languages besides English, with models trained on datasets extracted automatically from Wikipedia. When building the training datasets, the authors processed sentences from Wikipedia articles in multiple languages, looking also at the corresponding hyperlink structure. If a link in a Wikipedia sentence pointed to an article identified by Freebase as an entity, then the anchor text was considered as a positive training example for a particular entity type. Moreover, because not all entity mentions are linked in Wikipedia due to style guidelines, the authors also used oversampling and surface word matching to further improve model training.

Besides experimenting with the aforementioned five NER systems, we also made tests with an ensemble method based on voting that combines the results from the different systems. An ensemble of models can, in principle, perform better than any individual model, because the various errors of the models will average out. Also, we included in this process, the gazetteer built by EMLO to take advantage this set of information already available and collected from multiple letters from the Republic of Letters. The implemented voting system works as follow:

1. Each NER system that tags a span of text as a location counts as a vote for that span of text (e.g., if the word "London," in a given position, is tagged as a location by the Edinburgh Geoparser and Spacy, but not by any other NER tool, then this particular span of text receives two votes). Note that only the spans of text that have been recognized as locations by any of the NER systems are considered as candidates.
2. After all votes from the NER tools are assigned, a query is made for each candidate span of text in the EMLO gazetteer. If there is an entry for that location name in the gazetteer, then an additional vote is given to the corresponding span of text.
3. Each candidate span of text is tagged as a true location if it collected a minimum number of votes.

The performance of the NER systems is usually evaluated in terms of the precision, recall, and F1 measures. Precision is the percentage of correct entities in respect to all entities tagged as a location by the NER system. Therefore, if only one entity is tagged, and if that entity is in fact a true location, we would have a precision of 100%. Recall, on the other hand, is the percentage of correctly tagged entities in respect to the total absolute true number of entities. If 10 entities are correctly identified as a location but in fact there are 100 true locations mentioned in the texts, then we would have a recall of 10%. Finally, the F1 measure is a combination of these two parameters, equally balancing between precision and recall. The F1 measure is computed as the harmonic mean of precision and recall.

[9] http://www.ltg.ed.ac.uk/software/geoparser/.
[10] http://spacy.io.
[11] http://polyglot.readthedocs.org.

Notice that the aforementioned three metrics are computed with basis on the individual entity references (i.e., the spans of text that are recognized as locations) in the text. Therefore, if the island of "Great Britain" is mentioned several times in a document, the NER system should identify all these citations as locations, tagging both words as part of a location and recovering the correct spans of text. The different mentions of "Great Britain" would all be accounted for, in the computation of precision, recall, and F1.

RESULTS

Tables 1–**5** show the results obtained from each individual NER system, as well as their combination with the voting approach explained in the previous section. We considered five different voting thresholds (i.e., the minimum number of votes that a span of text should receive, to be considered a location), ranging from two to five votes. **Tables 1** and **2** show the evaluation carried out with the letters preprocessed with the *full clean* preprocessing rules followed by a translation from the original text to ME using the MorphAdorner and VARD systems, respectively. **Table 3** shows the equivalent results for the letters that have been fully cleaned, but considering the original EME text. **Table 4** shows the results with the *fast clean* preprocessing, again with EME text. Finally, **Table 5** shows the results for the Mary Hamilton letters. In all tables, we have signaled (in bold and different color) the results corresponding to the best F1 score.

From the results, we can see that the combination of multiple systems through voting, with a minimum of two to three votes, was able to consistently outperform the individual NER systems. In addition, all experiments resulted in a best minimum F1 score of approximately 70, which gives consistency to the analysis.

The best F1 score was obtained for the Hartlib Letters set with *fast clean* preprocessing and without translation to ME (**Table 4**). In connection to this outcome, we did not observe a significant difference either between full and fast cleaning or between EME and ME, where the gain in terms of the F1 measure between *fast clean* preprocessing and *full clean* preprocessing, together with EME to ME translation (with MorphAdorner), is only from 72.8 to 73.3. These results lead us to consider that when dealing with letters from the Samuel Hartlib corpus, carrying out the fast cleaning process might be enough as preprocessing, and the translation to ME does not seem to bring any considerable benefit. In fact, the lowest F1 score was obtained with *full clean* preprocessing and translations to ME using the VARD system, corresponding to an F1 score of 69.1 (**Table 3**).

Another observation is that the F1 scores for the Hartlib set of letters are higher for all scenarios (**Tables 1**–**4**) in comparison

TABLE 1 | Precision, recall, and F-measure for the 50 selected letters of the Samuel Hartlib corpus, with full clean preprocessing and translated from the original to Modern English using MorphAdorner.

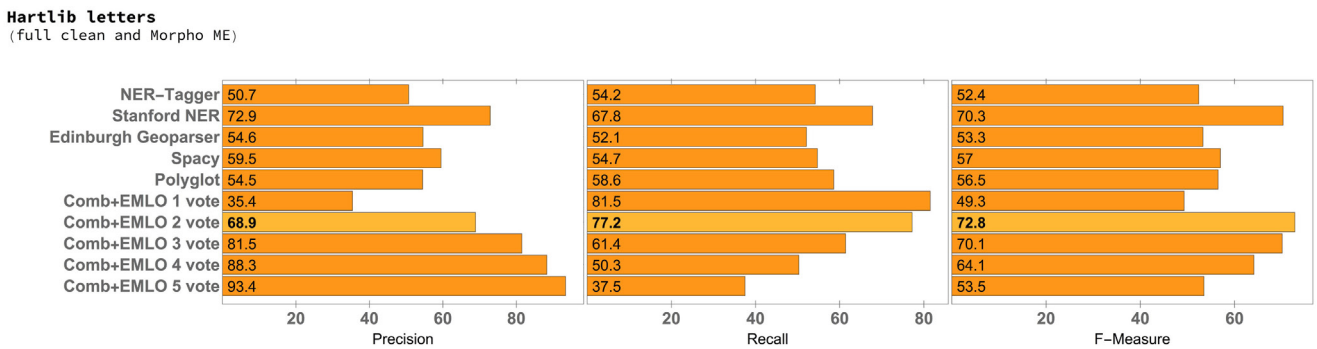


TABLE 2 | Precision, recall, and F-measure for the 50 selected letters of the Samuel Hartlib corpus, with full clean preprocessing and translated from the original to Modern English using VARD.

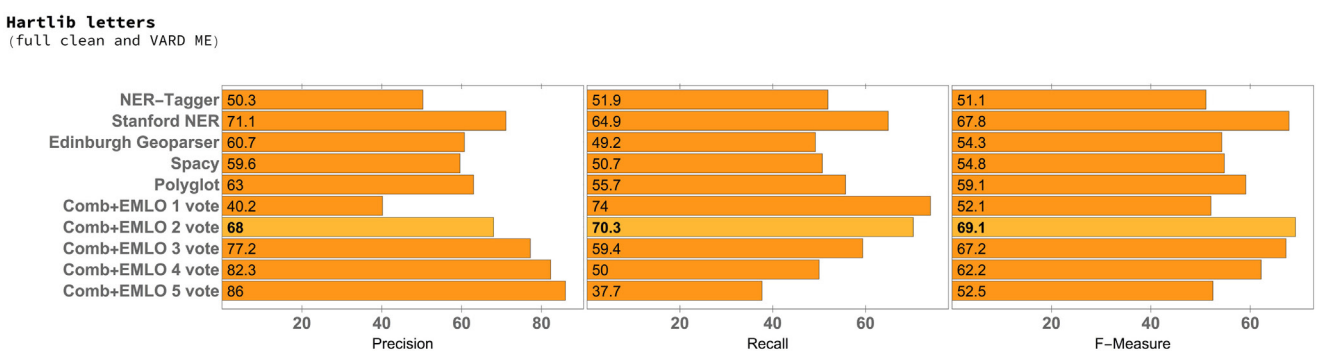


TABLE 3 | Precision, recall, and F-measure for the 50 selected letters of the Samuel Hartlib corpus with full clean preprocessing.

Hartlib letters
(full clean and EME)

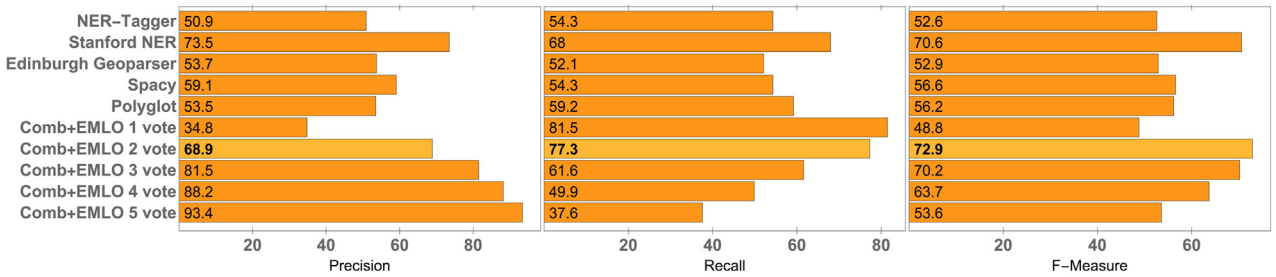


TABLE 4 | Precision, recall, and F-measure for the 50 selected letters of the Samuel Hartlib corpus, with fast clean preprocessing.

Hartlib letters
(fast clean and EME)

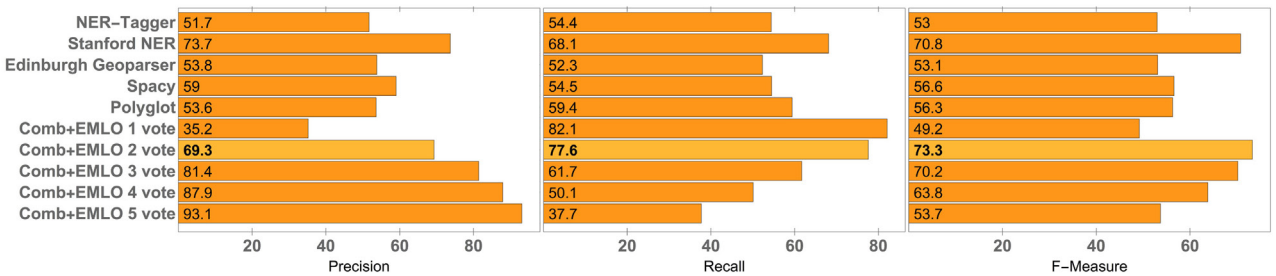
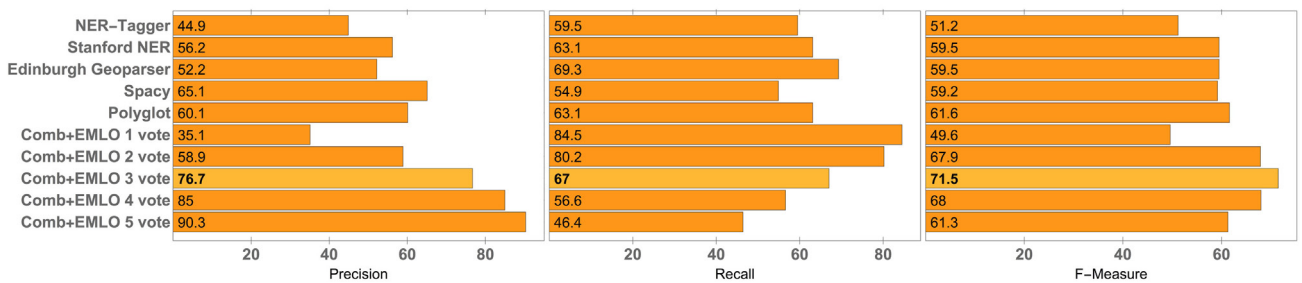


TABLE 5 | Precision, recall, and F-measure over the collection of Mary Hamilton letters.

Hamilton letters



with the Hamilton letters, except when the VARD system was used. This result was surprising, because the original files from the Hamilton set of letters are cleaner and written in ME.

In respect to the individual NER systems, while Polyglot gave the best F1 score in the case of the Hamilton papers, for all Hartlib scenarios it was the Stanford NER system that gave the best results. A somewhat surprising result was the fact that Stanford NER clearly outperformed the Edinburgh Geoparser, i.e., a rule-based system specifically considering historical documents, or the NER-Tagger software package, which leverages a more advanced statistical model that also considers character-based embeddings for addressing the problem of out-of-vocabulary words. In the particular case of the Hartlib Papers, the performance of Stanford NER in terms of the F1 score was also constantly close to the one obtained by the voting system. However, simply combining Stanford NER with the EMLO gazetteer (i.e., requiring for the places recognized by Stanford NER to also be present in the EMLO gazetteer), as shown in the last row of **Table 4**, would produce worse results than those achieved by the voting system. In the case of the experiments with the Hartlib Papers, the recall scores that are obtained with

the voting system, when considering a minimum of two votes, are always significantly higher than those obtained with the individual NER methods.

CONCLUSION AND FUTURE WORK

Although the Spatial Humanities has advanced very quickly in terms of the analysis of place names mentioned in corpora, as the field grows and scholars look for expedite techniques to support the exploration of large digital datasets, more research is needed to solve particular problems related to the correct identification of place names in historical datasets. While one pathway might be to look for specialized techniques that perform the identification and resolution of place names in historical documents, another possibility is to refine and fine tune NER systems/models that are already available. With this experiment, we aimed to shed light on the performance of commonly used NER technologies over historical documents.

The development of this particular experiment allowed us to (1) test individual NER systems that are readily available with datasets from the seventeenth and eighteenth centuries; (2) identify whether steps of preprocessing in terms of cleaning and translation to ME affect substantially the results of NER in a given corpus; and (3) to test the performance of an ensemble system based on voting with historical data. While it was observed that from all the individual systems Stanford NER was the one that outperformed all the others in all tests with the Hartlib Papers, the lower result obtained with the Hamilton letters makes clear that further experimentation with different corpora is needed, and that there are still many challenges to overcome. For instance, the low scores obtained in comparison with other studies performed with modern datasets might be in part due to the long recognized issue in historical corpora with spelling variations (e.g., Sueden, Sweden; Canterburie, Canterburry, Canterbury), which many of the NER systems do not recognize. We are currently already working on possible solutions toward these problems (Santos et al., 2017a,b), and one of the systems that was considered in our tests [i.e., the NER-Tagger software package implementing the ideas described by Lample et al. (2016)] uses character-based word embeddings to avoid problems with out-of-vocabulary words. Datasets such as the EMLO gazetteer, which record the variations in place name spelling as they occur in the period covered by the letters, might prove useful for the disambiguation of place names and for the creation, tailoring, and testing of NER systems for historical datasets. Another important issue, still to fully address in the case of these corpora, is the successful identification of complex entity types such as addresses. Another important matter to focus in future work is the fact that many of these letters can have combinations of different languages, or be written majorly in one language although referring to entities in a different language (e.g., Latin and English; German and Latin; etc.).

A significant find in the experiments reported on this article was that substantial preprocessing and translation to ME, for the Hartlib corpus and possibly other seventeenth-century datasets, might not be needed. This is of great interest due to the fact that, to simply get to the point where the extraction of this kind of information can be of use in terms of analysis (e.g., when creating visualizations and maps from this information), much of the time spent by the scholars is usually cleaning or trying to standardize the datasets. Our study shows that the differences observed in terms of performance with the so-called *full clean* and *fast clean* preprocessing strategies is not as sharp as it would be expected and, therefore, a fast cleaning of the dataset might be enough. Equally, the differences between Early and Modern English are minimal even between each of the individual NER systems and, therefore, the original versions can be used for these tasks. We additionally suggest that if a clean and human annotated corpus is needed, the CoNLL-2003 corpus size (946 news articles) could be considered as a good measure regarding the performance expectations versus annotated corpus sizes, since it is a traditional corpus used to train NER tools.

Finally, it must be noted that although this research accomplished the evaluation of the performance of these NER tools, further research is needed to deeply understand how the underlying models work with historical corpora and how they differ. We will devote part of our forthcoming efforts to this. In the meantime, we consider that the ensemble method proposed here not only provides better results than to simply use one tool on its own, but also a more stable and therefore reliable performance.

AUTHOR CONTRIBUTIONS

PM-F and BM proposed the challenge to solve. MW and PM-F carried out the preprocessing of the corpus. MW carried out the research and proposed the NLP methodology. PM-F elaborated the case study and the application of the methods. BM oversaw the technical aspects of the research.

ACKNOWLEDGMENTS

We want to thank David Denison and Nuria Yáñez-Bouza from the University of Manchester, who kindly shared their work and the Hamilton corpus with us. Thanks are also due to the HRI at the University of Sheffield for providing access to the Hartlib corpus, as well as to Howard Hotson, Arno Bosse, and Mark Greengrass for their guidance and help with it. We would also like to thank Barbara McGillivray and Robin Buning for sharing their experiences and code related to the preprocessing of their own datasets. Thanks are also due to Thomas Wallning and Vanda Anastácio for their support with the STSM. The work carried out in this WG1 Short Term Scientific Mission was sponsored by the EU COST Action IS1310 "Reassembling the Republic of Letters" (http://www.cost.eu/COST_Actions/isch/IS1310).

REFERENCES

Alex, B., Byrne, K., Grover, C., and Tobin, R. (2015). Adapting the Edinburgh geoparser for historical georeferencing. *International Journal of Humanities and Arts Computing* 9: 15–35. doi:10.3366/ijhac.2015.0136

Alex, B., Grover, C., Klein, E., and Tobin, R. (2012). Digitised historical text: does it have to be mediOCRe? In *Proceedings of KONVENS 2012*, Edited by J. Jancsary, 401–409. Vienna, Austria: ÖGAI.

Al-Rfou, R., Kulkarni, V., Perozzi, B., and Skiena, S. (2015). POLYGLOT-NER: massive multilingual named entity recognition. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, Edited by S. Venkatasubramanian and J. Ye, 586–594. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Al-Rfou, R., Perozzi, B., and Skiena, S. (2013). Polyglot: distributed word representations for multilingual NLP. In *Proceedings of Conference on Computational Natural Language Learning CoNLL'2013. Presented at the Computational Natural Language Learning CoNLL'2013*, Vancouver, Canada.

Archer, D., Kytö, M., Baron, A., and Rayson, P. (2015). Guidelines for normalising early modern English corpora: decisions and justifications. *ICAME Journal* 39: 5–24. doi:10.1515/icame-2015-0001

Bailey, T.J., and Schick, J.B.M. (2009). Historical GIS: enabling the collision of history and geography. *Social Science Computer Review* 27: 291–6. doi:10.1177/0894439308329757

Baron, A., and Rayson, P. (2008). VARD2: a tool for dealing with spelling variation in historical corpora. In *Postgraduate Conference in Corpus Linguistics*, Lancaster.

Batjargal, B., Khaltarkhuu, G., Kimura, F., and Maeda, A. (2014). An approach to named entity extraction from historical documents in traditional Mongolian script. In *2014 IEEE/ACM Joint Conference on Digital Libraries (JCDL)*, 489–490. London, UK.

Berbara, M., and Enenkel, K.A.E. (2011). *Portuguese Humanism and the Republic of Letters*. Leiden, Boston: BRILL.

Berman, M.L., Mostern, R., and Southall, H. eds. (2016). *Placing Names: Enriching and Integrating Gazetteers, the Spatial Humanities*. Bloomington: Indiana University Press.

Borin, L., Kokkinakis, D., and Olsson, L.-J. (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)*, 1–8. Prague, Czech Republic: Association for Computational Linguistics.

Brooke, J., Hammond, A., and Hirst, G. (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. Presented at the North American Association for Computational Linguistics*, 1–6. Denver, Colorado.

Brown, T., Baldridge, J., Esteva, M., and Xu, W. (2012). The substantial words are in the ground and sea: computationally linking text and geography. *Texas Studies in Literature and Language* 54: 324–39. doi:10.7560/TSLL54303

Burns, P.R. (2013). *Morphadorner v2: A Java Library for the Morphological Adornment of English Language Texts*. Evanston, IL: Northwestern University.

Byrne, K. (2007). Nested named entity recognition in historical archive text. In *International Conference on Semantic Computing (ICSC 2007). Presented at the First IEEE International Conference on Semantic Computing (ICSC)*. Irvine, CA: IEEE.

Clifford, J., Alex, B., Coates, C.M., Klein, E., and Watson, A. (2016). Geoparsing history: locating commodities in ten million pages of nineteenth-century sources. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 49: 115–31. doi:10.1080/01615440.2015.1116419

Cneudecker. (2014). *Named Entity Recognition for Digitised Historical Newspapers*. KB Research.

Cooper, D., Donaldson, C., and Murrieta-Flores, P. eds. (2016). *Literary Mapping in the Digital Age, Digital Research in the Arts and Humanities*, 31. New York: Routledge.

Crane, G., and Jones, A. (2006). *The Challenge of Virginia Banks: An Evaluation of Named Entity Analysis in a 19th-Century Newspaper Collection*, 31. Chapel Hill, NC: ACM Press. doi:10.1145/1141753.1141759

Dalton, S. (2004). *Engendering the Republic of Letters: Reconnecting Public and Private Spheres in Eighteenth-Century Europe*. Quebec: McGill-Queen's Press—MQUP.

Denison, D., and Yáñez-Bouza, N. (2016). *Image to Text: Mary Hamilton Papers (c.1750-c.1820)*. Available at: https://www.research.manchester.ac.uk/portal/en/publications/image-to-text(8d4335c7-1b07-4bcc-ac96-20e884011ca5)/export.html (accessed February 13, 2017).

Dross, K. (2006). Use of geographic information systems (GIS) in history and archeology. *Historical Social Research/Historische Sozialforschung* 31: 279–87.

Ehrmann, M., Colavizza, G., Rochat, Y., and Kaplan, F. (2016). Diachronic evaluation of NER systems on old newspapers. In *Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016)*, Edited by S. Dipper, F. Neubarth, and H. Zinsmeister, 97–107. Bochum, Germany: Bochumer Linguistische Arbeitsberichte.

Erdmann, A., Brown, C., Joseph, B.D., Janse, M., and Ajaka, P. (2016). Challenges and solutions for Latin named entity recognition. In *COLING*, 9. Osaka: Association for Computational Linguistics.

Feingold, M. (2003). *Jesuit Science and the Republic of Letters*. Massachusetts: MIT Press.

Finkel, J.R., Grenager, T., and Manning, C. (2005). *Incorporating Non-Local Information into Information Extraction Systems by Gibbs Sampling*. Michigan: Association for Computational Linguistics. 363–70. doi:10.3115/1219840.1219885

Freire, N., Borbinha, J., Calado, P., and Martins, B. (2011). *A Metadata Geoparsing System for Place Name Recognition and Resolution in Metadata Records*. ACM Press. 339.

Furey, C.M. (2006). *Erasmus, Contarini, and the Religious Republic of Letters*. Cambridge: Cambridge University Press.

Gerhard, J., and van den Heuvel, W. (2015). *Survey Report on Digitisation in European Cultural Heritage Institutions 2015*. PrestoCentre. Available at: https://www.prestocentre.org/library/resources/survey-report-digitisation-european-cultural-heritage-institutions-2015

Goodchild, M.F., and Hill, L.L. (2008). Introduction to digital gazetteer research. *International Journal of Geographical Information Science* 22: 1039–44. doi:10.1080/13658810701850497

Goodman, D. (1996). *The Republic of Letters: A Cultural History of the French Enlightenment*. New York, NY: Cornell University Press.

Greengrass, M., Leslie, M., and Hannon, M. (2013). *The Hartlib Papers*. Sheffield: HRI Online Publications. Available at: http://www.hrionline.ac.uk/hartlib

Greengrass, M., Leslie, M., and Raylor, T. (2002). *Samuel Hartlib and Universal Reformation: Studies in Intellectual Communication*. Cambridge: Cambridge University Press.

Gregory, I., Donaldson, C., Murrieta-Flores, P., and Rayson, P. (2015). Geoparsing, GIS, and textual analysis: current developments in spatial humanities research. *International Journal of Humanities and Arts Computing* 9: 1–14. doi:10.3366/ijhac.2015.0135

Gregory, I.N., and Hardie, A. (2011). Visual GISting: bringing together corpus linguistics and geographical information systems. *Lit Linguist Computing* 26: 297–314. doi:10.1093/llc/fqr022

Grover, C., Givon, S., Tobin, R., and Ball, J. (2008). Named entity recognition for digitised historical texts. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. Paris: ELRA.

Grover, C., Tobin, R., Byrne, K., Woollard, M., Reid, J., Dunn, S., et al. (2010). Use of the Edinburgh geoparser for georeferencing digitized historical collections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368: 3875–89. doi:10.1098/rsta.2010.0149

Hamilton, A., Boogert, M.H.V.D., and Westerweel, B. (2005). *The Republic of Letters and the Levant*. Netherlands: BRILL.

Hyun, Joong Kim (2009). Past time, past place: GIS for history. *Social Science Computer Review* 27: 452–3. doi:10.1177/0894439308329769

Kettunen, K., Mäkelä, E., Ruokolainen, T., Kuokkala, J., and Löfberg, L. (2016). *Old Content and Modern Tools – Searching Named Entities in a Finnish OCRed Historical Newspaper Collection 1771–1910*. CoRR abs/1611.02839, 124–135.

Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C. (2016). *Neural Architectures for Named Entity Recognition*. San Diego, California: HLT-NAACL.

Ling, W., Dyer, C., Black, A.W., and Trancoso, I. (2015). *Two/Too Simple Adaptations of Word2Vec for Syntax Problems*. San Diego, California: HLT-NAACL.

Mac Kim, S., and Cassidy, S. (2015). Finding names in trove: named entity recognition for Australian historical newspapers. In *Australasian Language Technology Association Workshop 2015*, 57. Melbourne.

Manguinhas, H., Martins, B., and Borbinha, J.L. (2008). A geo-temporal web gazetteer integrating data from multiple sources. In *2008 Third International Conference on Digital Information Management*, 146–153.

Melo, F., and Martins, B. (2016). Automated geocoding of textual documents: a survey of current approaches: automated geocoding of textual documents. *Transactions in GIS* 3–38. doi:10.1111/tgis.12212

Murrieta-Flores, P., Baron, A., Gregory, I., Hardie, A., and Rayson, P. (2015). Automatically analyzing large texts in a GIS environment: the registrar general's reports and cholera in the 19th century: automatically analyzing large historical texts in a GIS environment. *Transactions in GIS* 19: 296–320. doi:10.1111/tgis.12106

Murrieta-Flores, P., and Gregory, I. (2015). Further frontiers in GIS: extending spatial analysis to textual sources in archaeology. *Open Archaeology* 1: 166–5. doi:10.1515/opar-2015-0010

Nadeau, D., and Sekine, S. (2007). A survey of named entity recognition and classification. *Lingvisticae Investigationes* 30: 3–26. doi:10.1075/li.30.1.03nad

Nagai, N., Kimura, F., Maeda, A., and Akama, R. (2015). Personal name extraction from Japanese historical documents using machine learning. In *2015 International Conference on Culture and Computing (Culture Computing)*.

Neudecker, C., Wilms, L., Jan Faber, W., and van Veen, T. (2014). Large scale refinement of digital historical newspapers with named entities recognition. In *IFLA 2014 Newspaper Section Satellite Meeting*, Geneva.

Ostrander, G.M. (1999). *Republic of Letters: The American Intellectual Community, 1776-1865*. Indianapolis: Madison House.

Pilz, T., Ernst-Gerlach, A., Kempken, S., Rayson, P., and Archer, D. (2008). The identification of spelling variants in English and German historical texts: manual or automatic? *Literary and Linguistic Computing* 23: 65–72. doi:10.1093/llc/fqm044

Piotrowski, M. (2012). Natural language processing for historical texts. *Synthesis Lectures on Human Language Technologies* 5: 1–157. doi:10.2200/S00436ED1V01Y201207HLT017

Porter, C., Atkinson, P., and Gregory, I. (2015). Geographical text analysis: a new approach to understanding nineteenth-century mortality. *Health & Place* 36: 25–34. doi:10.1016/j.healthplace.2015.08.010

Prendergast, A. (2015). *Literary Salons across Britain and Ireland in the Long Eighteenth Century*. Berlin: Springer.

Purves, R., and Jones, C. (2011). Geographic information retrieval. *SIGSPATIAL Special* 3: 2–4. doi:10.1145/2047296.2047297

Purves, R.S., and Derungs, C. (2015). From space to place: place-based explorations of text. *International Journal of Humanities and Arts Computing* 9: 74–94. doi:10.3366/ijhac.2015.0139

Rayson, P., Archer, D., Baron, A., and Smith, N. (2006). Tagging Historical Corpora – the problem of spelling variation. In *Digital Historical Corpora, Dagstuhl-Seminar 06491, International Conference and Research Center for Computer Science*. Wadern, Germany: Schloss Dagstuhl.

Rupp, C.J., Rayson, P., Baron, A., Donaldson, C., Gregory, I., Hardie, A., et al. (2013). Customising geoparsing and georeferencing for historical texts. *Presented at the Conference on Big Data, 2013 IEEE International,* 59–62. Silicon Valley, CA: IEEE. doi:10.1109/BigData.2013.6691671

Rupp, C.J., Rayson, P., Gregory, I., Hardie, A., Joulain, A., and Hartmann, D. (2014). Dealing with heterogeneous big data when geoparsing historical corpora. *Presented at the Conference on Big Data (Big Data), 2014 IEEE International*, 80–83. Washington, DC: IEEE.

Santos, J., Anastácio, I., and Martins, B. (2015a). Desambiguação de Entidades Mencionadas em Textos na Língua Portuguesa ou Espanhola. *IEEE Latin America* 13.

Santos, J., Anastácio, I., and Martins, B. (2015b). Using machine learning methods for disambiguating place references in textual documents. *GeoJournal* 80: 375–92. doi:10.1007/s10708-014-9553-y

Santos, R., Murrieta-Flores, P., Calado, P., and Martins, B. (2017a). Toponym matching through deep neural networks. *International Journal of Geographical Information Science* 32: 324–48. doi:10.1080/13658816.2017.1390119

Santos, R., Murrieta-Flores, P., and Martins, B. (2017b). Learning to combine multiple string similarity metrics for effective toponym matching. *International Journal of Digital Earth* 1–26. doi:10.1080/17538947.2017.1371253

Shelford, A. (2007). *Transforming the Republic of Letters: Pierre-Daniel Huet and European Intellectual Life, 1650-1720*. Rochester, NY: University Rochester Press.

Silveira, L.E da (2014). Geographic information systems and historical research: an appraisal. *International Journal of Humanities and Arts Computing* 8: 28–45. doi:10.3366/ijhac.2014.0118

Simon, R., Barker, E., Isaksen, L., and de Soto Cañamares, P. (2015). Linking early geospatial documents, one place at a time: annotation of geographic documents with Recogito. *e-Perimetron* 10: 49–59.

Sprugnoli, R., Moretti, G., Kessler, B., Tonelli, S., and Menini, S. (2017). Fifty years of European history through the lens of computational linguistics: the De Gasperi Project. *Italian Journal of Computational Linguistics* 2: 89–100.

Tjong Kim Sang, E.F., and De Meulder, F. (2003). *Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition*. Association for Computational Linguistics. 142–7.

van Hooland, S., De Wilde, M., Verborgh, R., Steiner, T., and Van de Walle, R. (2015). Exploring entity recognition and disambiguation for cultural heritage collections. *Digital Scholarship in the Humanities* 30: 262–79. doi:10.1093/llc/fqt067

Webster, C. ed. (1970). *Samuel Hartlib and the Advancement of Learning*. Cambridge: Cambridge University Press.

Wing, B.P. (2015). *Text-Based Document Geolocation and Its Application to the Digital Humanities*. Ph.D. thesis, University of Texas, Austin.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.