CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Multilingual Digital Humanities
Authors	Nilsson-Fernàndez, Pedro;Dombrowski, Quinn
Publication date	2022-12-01
Original Citation	Nilsson-Fernàndez, P. and Dombrowski, Q. (2022) 'Multilingual Digital Humanities', in O'Sullivan, J. (ed)., The Bloomsbury Handbook to the Digital Humanities, New York: Bloomsbury Publishing, pp. 81-90. isbn: 9781350232112
Type of publication	Book chapter
Link to publisher's version	https://www.bloomsbury.com/uk/bloomsbury-handbook-to-the-digital-humanities-9781350232112/
Rights	© 2023. Access to this chapter is restricted until 6 months after publication by request of the publisher.
Download date	2026-11-17 09:55:52
Item downloaded from	https://hdl.handle.net/10468/14084

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

CHAPTER EIGHT

Multilingual Digital Humanities

PEDRO NILSSON-FERNÀNDEZ (UNIVERSITY COLLEGE CORK) AND
QUINN DOMBROWSKI (STANFORD UNIVERSITY)

Digital humanities is contextual, shaped by the identities and interests of the people who engage with it, and the communities in which they are situated. The singular origin story featuring a meeting of minds between Father Busa and IBM has been complicated (Terras and Nyhan 2016) and called into question (Earhart 2015) and challenged by narratives that illustrate the diverse paths along which digital humanities has evolved (e.g., O'Sullivan 2020; Wang et al. 2020). The reasons motivating any one particular course of development vary: funding availability, cultural priorities, and the individual interests and resources available to influential early figures all play a role in shaping how a particular kind of DH came to be. Articulating the long-standing nature and diverse trajectories of DH communities beyond the mono-myth of Busa is an important ongoing project in the field, but even the Busa-linked lineage of "institutional DH" has its under-explored facets.

From IBM-sponsored events in the 1960s, to the formation of European and North American DH organizations in the 1970s, to the establishment of the Alliance of Digital Humanities Organizations (ADHO) in the 2000s, this thread of mainstream institutional DH has explicitly adopted an international scope. In practice, "international" has often meant North America, Western Europe, and occasional others. Furthermore, the global hegemony of US and UK universities, the early engagement of scholars from the US and UK (such as Jerome McGann and Willard McCarty), and the need for a lingua franca at these international events laid the groundwork for treating the English language as a default. This is not to suggest the universal ubiquity of English—see, for example, the long and unbroken tradition of Italian-language DH work within the *Informatica Umanistica* community,[1] or Cristal Hall's study on the intersections and oppositions between traditional and digital Italian Studies (2020). But as noted in Fiormonte (2016), the locus of control for key components of digital infrastructure, from ICANN domain registration, to Unicode, to the most ubiquitous operating systems and software, is situated in the Anglo-American world. The cultural biases and priorities of the Anglo-American world permeate these technologies (Noble 2018), even where they are intended for global use. This has downstream implications for DH.

There is no conscious, deliberate plot to permanently install the English language at the pinnacle of institutional power within DH. Rather, it is most often a matter of monolingual-Anglophone obliviousness with regard to language. Language is a source of friction familiar to anyone whose primary language is not English. From software interfaces, to keyboards, to the availability of subtitles or dubbed audio for the latest global media franchise, the friction of language is a part of everyday life, and digital humanities is no exception to that general rule. For an Anglophone DH

scholar living in primarily Anglophone countries whose research uses English-language materials, weeks or months can pass without language imposing any kind of barrier. Perhaps there are fleeting moments of curiosity—overhearing an unintelligible conversation or encountering an unreadable sign—but at almost no point is the scholar forced to engage with another language, lest they be fundamentally impeded in their goals. Other languages, when encountered, become ignorable, and that ignoring becomes a mental habit.

The accumulation of these mental habits across an entire community of monolingual Anglophone scholars renders language functionally invisible as a feature. When encountered, it is ignored, fading into a kind of background noise. This invisibility by normalization makes it even more difficult to recognize language as an axis of diversity in need of intervention, when considered alongside more visible and violent acts of discrimination on the basis of race and gender, as described in Earhart (2016).

The annual DH international conference organized by ADHO provides a venue for scholars with financial means to meet and build relationships with colleagues from across the world.[2] The talks are most often given in English, and informal socializing among participants from diverse linguistic backgrounds often settles into the common denominator of English. The international pool of authors who attend this conference and showcase their research output in venues such as *Digital Scholarship in the Humanities* provide Anglophone scholars with a lens onto DH shaped by other cultural and linguistic contexts. While this engagement with multicultural scholars is crucial, there is also very important and significantly different work presented and published in non-Anglophone venues. This work is widely unknown to monolingual Anglophone scholars, who often have little experience working with scholarship in languages other than English. In practice, a decreasing number of US and UK[3] humanities graduate programs focusing on language and literature recommend a reading knowledge of at least one other "scholarly" language (e.g., German, French, or Russian), but even where such a recommendation exists, actually reading and citing scholarship in those languages is not required as part of graduate training. Whether or not they find it easy to navigate academic prose in English, scholars who work in other languages are still far more likely to cite work in English than vice versa (Fiormonte 2015).

Attending an international conference such as ADHO's may itself be a step closer to breaking out of the national frame of reference that circumscribes many disciplinary conferences, but that conference and its associated journal have a culture that is distinct from a simple amalgamation of the local DH cultures represented by the participants (Fiormonte 2014a). In this chapter, we lay out a set of recommendations for Anglophone scholars to expand their engagement with DH beyond the Anglophone world. We treat this value as a given; to understand DH (or any discipline in the humanities by extension—be it literature, history, philosophy, etc.), it is essential to look beyond the scope of any single language, although many previous efforts to "define DH" have failed to do so (Fiormonte 2014b). However, even if an Anglophone scholar is skeptical of the value other national and linguistic digital humanities practices can bring to their own work, overlooking non-Anglophone DH will exact a toll on those around them. In most institutions, DH courses and workshops are interdisciplinary, drawing students from a wide range of disciplines—including area studies, history, and non-English literatures. DH syllabi designed to be applied to English text may fail if applied without additional cultural contextualization—and pre-processing in the case of textual digital tools—to works produced in other languages. Not being aware of this challenge, or what steps to take to begin adapting English-oriented methods, means failing students who may consider working in languages other than English. Similarly, a scholar who develops and

promotes a tool for analyzing text, and hard-codes certain English-oriented decisions (e.g., around stopwords), may be inadvertently limiting the usability and impact of that tool. Furthermore, we encourage Anglophone scholars who hold the reins of infrastructure to engage more seriously with language in contexts including scholarly publishing, even if the result leads to increased linguistic friction.

LANGUAGE IN DH JOURNALS

The spirit shown in fragments from Digital Humanities manifestos originating from THATCamp Paris (Dacos 2011) or UCLA's Mellon Seminar in Digital Humanities (UCLA 2008), stating that "interdisciplinarity, transdisciplinarity and multidisciplinarity are empty words unless they imply changes in language, practice, method, and output" (UCLA 2008), falls short when it comes to multilingualism. Later, this same manifesto articulates that "traditional Humanities is balkanized by nation, language, method, and media" and that "Digital Humanities is about convergence."

The price of prioritizing this convergence at any cost is that we tend to perceive the duality between global and local in very absolute terms. The "global" becomes a synonym for forward-thinking, endless, and borderless collaborative practices, whereas the "local" is associated with opposite values. Claiming that "nation" acts as divider of a discipline when writing from a monolingual and Anglo-centric position is hardly neutral when the lingua franca that will mediate the "global" is the one spoken in your nation.

However, for nations where English is not the first language, the sacrifices involved in being part of the international academic network are considerable. The imposition of English as lingua franca brings an intellectual disadvantage to scholars in the humanities who produce high-quality research in languages other than English but are not able, or simply choose not to, render that research output in English. When research is not translated into English, its dissemination and impact beyond the territories in which its language is spoken officially is strictly dependent on a factor that bears little to no connection with the quality of the research itself: the number of international academics in the field who are able to read and review work in that specific language.[4] Part of this dissemination occurs in the form of citing such output; interestingly enough, if these international peers choose to direct-quote fragments of the work rather than paraphrasing its ideas, most English-language journals will still ask for these fragments to also be translated into English. This results in additional unpaid labor that becomes another burden on the multilingual researcher. Without removing any traces of the original language in which the article was written, scholarship written in other languages is put at a disadvantage, regardless of its quality or relevance to the field. English-language articles are more convenient sources to quote—particularly in terms of labor and visibility—only because of the immediate accessibility granted by the lingua franca they are written in.

Choosing the translation route is not as straightforward a solution as one might initially expect. While several companies provide academic translation services to English catering to source languages such as Chinese, Portuguese, and Spanish (at an approximate cost of 700 USD), the price and availability changes the moment you select a smaller language such as Catalan as the source language (1,700 euro).[5] Provided that the scholar can afford to produce such a translation, the translation adds an additional layer to the already highly competitive process of academic publishing; we also need to take into account that additional translation work will be required at

every stage of the publication process, as manuscript revisions are a natural step in the process of publishing in any peer-reviewed journal.[6] Both the economic barriers and the complexity of the process explain why many scholars prefer to attempt an exercise of self-translation. It is precisely in these contexts that the dismissive attitudes towards multilingualism in academic publishing have a more devastating effect on non-English-speaking scholars, whose efforts to reduce that linguistic gap go highly unnoticed in English-language scholarship. As part of *The Digital Humanities Long View* series[7] Riva Quiroga noted a recent online controversy initiated by a tweet[8] from a co-editor of the *Journal of Contemporary History*, Sir Richard Evans, who unapologetically describes imperfect English as a major reason for rejecting papers.

In the absence of translation or work written originally in English, the odds of Anglophone scholars encountering articles from scholars who work primarily in other languages when exploring major international DH journals are low (Fiormonte 2015). Such articles tend to be published in special issues in a specific non-English language, and are unlikely to be connected into the citation networks of English-language articles. The emergence of these special issues (such as the Spanish-language *Digital Humanities Quarterly* vol. 12, issue 1 in 2018) is a positive step towards multilingual DH. However, this linguistic compartmentalization reinforces the power imbalance between languages. It isolates these articles, reducing serendipitous encounters for scholars browsing the journal. Furthermore, when we look specifically at the non-English languages used for these special issues, they are unequivocally global languages: Spanish, French, Portuguese. Scholarship in other languages—including the languages of other ADHO constituent organizations—is pushed further to the periphery.

As a discipline that praises itself for breaking down the barriers preventing free access to knowledge and information, DH routinely fails to acknowledge the barrier posed by language. Following the rationale used by Laura Estill when she claimed that Digital Humanities had a "Shakespeare problem" (Estill 2019), we may well say that Western Digital Humanities has an English problem.

LANGUAGES IN THE DH COMMUNITY

The web, writ large, is a much richer source of information about DH praxis in other languages and cultural contexts than international journals, and web-based content is among the easiest to access via machine translation thanks to numerous browser-based plugins. Despite the dominance of the English language in most search engines, searching for the translation of "digital humanities" in different languages will surface projects, articles, and discussions that would remain completely invisible otherwise. It is a matter of linguistic awareness; in the case of Google, for example, adjusting options like region and language in which to conduct the search will produce completely different results. Likewise, there are different layers of linguistic bias when it comes to global languages. A search involving languages and cultures in the Iberian Peninsula beyond Spanish and Portuguese—i.e., Catalan, Euskera, or Galician—is likely to produce more Spanish and Portuguese language results than in those three minoritized languages struggling for space with their global counterparts.

Exploring blog posts on digital humanities written in other languages, following scholars on social media who post about DH in their native tongues, or looking at abstracts from conferences held outside of English-speaking countries offers a multilingual and multicultural insight into the

discipline that is blurred in English-only contexts. These academic communities may not only interpret DH concepts in a different way (e.g., see Berra et al. 2020), but may draw upon a long and unique history, and may bring fresh approaches to an Anglophone-dominated discipline.

Most importantly, some of these academic communities may be at earlier stages in the incorporation of digital humanities methodologies in their research. This affords them a unique opportunity to reflect on initial practices of the discipline, to reconsider the validity of some paradigms that have been established in "mainstream" Anglophone DH—both in theory and praxis—and to question them from beyond an Anglophone perspective. The creation and or modification of infrastructures in order to adjust to the multidisciplinary nature of DH will take different forms when responding to specific cultural challenges in every case. As Isabel Galina shows when she chronicles the creation of *RedHD* (2014) and Puthiya Purayil Sneha when mapping the emergence of DH in India (2017), we cannot assume that DH is received and practiced identically in different communities globally.

DH practitioners who work, write, and research in and about minoritized languages and cultures have valuable lessons to share, drawing upon the constant struggle for visibility of their research output. In many cases, the choice to use their native language in academia is not brought by the inability to express themselves in English or in another global language, but is a question of linguistic activism. Their resilience against the imposition of global languages in their discipline responds to a constant fight for the survival of their mother tongue and the conservation of a digital cultural record that is disrupted by colonialism and neocolonialism (Risam 2019).

This determined struggle for visibility may also resonate with current discourse in DH around sustainability of DH projects; sustainability of non-English DH projects becomes a more pressing issue, as in some cases, the loss of digital presence or the interruption of these projects can have devastating consequences such as loss of digital cultural record. While English-language projects and their data are likely to be supported by a stronger digital infrastructure and often managed, stored, updated, and secured by more than one institution, projects and datasets related to knowledge written in a smaller language may be in a different state of conservation and is often at risk of disappearing altogether. This situation should be taken into account when prioritizing resources, such as funding and labor, for long-term archiving of digital projects.

LANGUAGES IN DH TUTORIALS AND TOOLS

Acknowledging the importance of building DH tools whose pipelines are flexible enough to adapt to multilingual environments is vital to drive change of attitude towards linguistic diversity in the discipline. While some DH tools are specifically built for the purpose of catering to the needs of a single project, the tools that have made the biggest difference in the field are those whose application extends beyond individual research questions and benefit the wider DH community. If we observe the reception and use of some of these key tools through the lenses of a project such as *The Programming Historian*—which provides volunteer-written tutorials for such tools in English, Spanish, French, and Portuguese—the potential for growth of non-English-speaking users of these tools is significant. The presence of multilingualism in the platform is not limited to the translated tutorials; *The Programming Historian*'s English-language tutorials create awareness of the multilingual applications of the tools referenced. The "Introduction to Jupyter Notebooks" tutorial (Dombrowski, Gniady, and Kloster 2019) uses *Harry Potter* fanfic metadata in Italian as the example file to be used in the tutorial, while also providing a link to a Jupyter notebook created

for an article on stylometry published in Spanish (Calvo Tello 2016). A similar example can be seen in Andrew Akhlaghi's tutorial on "OCR and Machine Translation with Tesseract," using Russian texts as source material.

The community of multilingual DH scholars would benefit from the broad adoption of what has become known in linguistics as the "Bender Rule," after University of Washington professor Emily Bender (Bender and Friedman 2018). Put simply, the Bender Rule is "always name the language you are working on," and it responds to a situation where, much like in DH, English exhibits extreme unmarkedness in everything done in the field of linguistics. If a new dataset, resource, tool, project, article, book, or chapter appears, and no specific language cue is associated with it, we automatically assume its language to be English. The presence and dominance of the language in the field is so obvious that it simply becomes the standard. Naming language is a first step towards reasserting the role of English *as a language*, and one among many possible options.

In a DH tool development context, the Bender Rule could be adapted as "always include a 'languages' section in your tool documentation." A great example of what such a section may look like can be found within *Voyant Tools* documentation, where users can learn about the ten different languages its interface is available in, on the one hand, and the different languages that the tool can analyse, on the other. The developers are very honest in admitting the shortfalls of the wide generic support of the tool, admitting that it reduces language-specific functionality, but are also proud to mention the language-specific awareness brought by their document tokenizer, and their interface readiness for RTL[9] scripts. Looking at the list of translators of the interface, we can see that it is composed by practitioners who work on DH through each of the particular languages, this volunteer or community-driven effort can be seen in most of what we do to expand the linguistic horizons of DH (see Ghorbaninejad et al. 2022, for a description of the process by which the interface was translated to Arabic).

Against the evident imbalance between the availability of data (corpora, datasets) in English language and languages other than English, and the reported issues with existing online datasets in languages other than English (Caswell et al. 2021), tool development should explore ways to counteract this data inequality by prioritizing algorithms that work better with less data. In the same way that minimal computing initiatives work on the basis that there is a scarcity of hardware resources, multilingual DH tool building should bear in mind data scarcity in their design. In parallel, tools could be designed proactively to assume a lack of reliable datasets, providing plugins or packages that can easily incorporate human-supervised data validation instruments working parallel to its main purpose. While practices such as crowdsourcing may be considered less effective in English-language DH (due to this need for data validation), it is a vital resource for multilingual DH.

LANGUAGES IN THE DH CLASSROOM

The institutionalization of digital humanities in the United States, Canada, and United Kingdom through degree and certificate programs has typically taken method as a major organizing principle, with discipline as a secondary or aggregating factor. Students accrue credits by taking courses, for instance, on network analysis, geospatial methods, computational text analysis, or TEI text encoding. There is often a survey "Intro to DH" course, and many programs have one or more courses that centers the work and debates within a discipline, such as "digital history" or

"computational literary studies." With few exceptions,[10] language is never an organizing principle for courses. Considering the staffing costs involved in offering a course, this may be justifiable: there may not be enough students working at the intersection of DH and any one specific non-English language to draw the necessary enrollment for a dedicated course. But the alternative of weaving linguistic awareness into methods courses that explicitly welcome all, regardless of what language they work in—has not been the expected standard for DH pedagogy. While some recent courses have placed greater emphasis on issues such as Unicode text encoding and lemmatization for languages with more inflection than English, it is easy to overlook these steps when working exclusively with English examples.

On a practical level, teaching a class using shared, English-language data makes a great deal of sense. Pre-prepared examples make it possible to deliberately plan for the complications and challenges that the material will introduce. It allows instructors to create assignments with correct answers identified in advance, facilitating grading. But problems arise when students learn (or assume) that the workflow taught in class for the English-language example text is, in fact, the workflow for *analyzing text*. Students who wish to apply these methods to other languages may conclude that text analysis "doesn't work" for their language when they apply the workflow used in class to non-English text. The result is that only the students who work on English-language texts leave the classroom with practical skills they can apply to their own research. This problem is most visible in text analysis courses, but many other methods—including network analysis and geospatial work—build on datasets that are themselves the result of text analysis. Teaching with a pre-prepared dataset that was created using tools that perform much better for English than other languages obscures the startup labor required to apply the methods in a new context—and hides the tool gap that may make these methods impractical for some languages (McDonough et al. 2019).

The Bender Rule is, once again, a step to begin with when addressing linguistic gaps in curricula. While it may seem trivial, stating clearly that a workshop on text analysis that only uses English text and tools is, in fact, a workshop on *English* text analysis is a step towards making language more visible as an important variable when doing this work. In some cases, acknowledging this limitation—and leaving students who are interested in other languages to accept it or look elsewhere—is all that is feasible. Librarians, graduate students, and people in assorted staff roles often teach DH workshops or courses on the margins of other jobs, with little or no time or financial support for reworking pre-prepared material to be more linguistically inclusive, or expanding their own knowledge about how to work with a range of commonly requested languages. In cases where an instructor does have support to develop a workshop or course that can serve a broader range of students, there are a number of meaningful steps that they can take. Particularly in the context of a multi-week or multi-month course, instructors can ask upfront about what language(s) students are interested in working on. The next step is to learn something about the orthography and grammar of any language(s) with which you are unfamiliar. Is the language written with spaces between words? Is it written right-to-left (RTL)? Some tools have problems processing or displaying RTL text (Wrisley et al. Forthcoming). Is there a lot of inflection? Most of these basic linguistic facts can be ascertained as simply as skimming the Wikipedia page for the language.

Having gathered this information, consider the impact on the methods you plan to teach. If your students want to work with *analytic languages* (which use "helper words" rather than inflection to convey relationships between words; languages include Chinese, Vietnamese, and English), and your methods are based on word counts (like TF-IDF[11] or topic modeling), you can simply mention

that the methods taught in class should also perform well for those languages. If, however, you are using word-count methods and some of your students want to work with highly inflected languages such as Ukrainian, or highly agglutinative languages such as Turkish, you should, at a minimum, mention the challenges associated with applying the methods taught in class to such languages. Better yet would be to equip the students with the tools they need to address these challenges, such as a lemmatizer.[12] Similarly, many natural language processing tools are language-specific. If you are teaching a method that involves POS tagging[13] or NER,[14] it is worth searching for what resources are available to perform similar tasks in the languages your students are interested in working with. The set of languages represented by students in your class may also shape your choice of tools.

CONCLUSION

There is vastly more to digital humanities than what is reflected from an Anglophone outlook, even with the glimpses into the DH of other contexts offered by international venues. Too often, the work that appears in these "international" journals and conferences is less a representation of the diversity of local digital humanities as it is experienced by scholars around the world, than a predictable and limited piece of that work that is likely to be more comprehensible to "mainstream" Anglophone DH. Even scholars who are famous for not holding back in their criticism of Anglophone DH find themselves expressing these opinions even more forcefully when writing for non-Anglophone audiences. Language gives scholars a way to play, for instance, with the concept of (implicitly Anglophone) "digital humanities" set against the concept of "humanités numériques," which grows from and is shaped by the Francophone world. Engaging with these other linguistic worlds is easier than ever with the help of machine translation, but tools are limited by scholars' willingness to make the effort to use them. It requires some effort to find and follow scholars who write in languages other than English, in formal and informal online spaces, hitting the "translate" button on Twitter or in the browser rather than skipping over their work. Monolingual Anglophone scholars may be comfortable within the bounds of that linguistic sphere, but they owe it to their audience with broader horizons to at least be aware of and clear about that limitation in their knowledge of the field. The Bender Rule of ubiquitously naming language is a place to start; if your awareness is limited to the Anglophone world, defining your expertise as English-language DH rather than DH at large would go a long way in making non-English DH a more visible type of valued expertise. But even small steps into a world of greater linguistic diversity—like considering the impact of language on the tools you build, or browsing the programs of conferences held in other languages—can have a significant impact on your own understanding of the field, and your ability to support students and colleagues whose scholarship extends beyond English.

NOTES

1. See the website of the *Associazione per l'Informatica Umanistica e la Cultura Digitale* for examples, http://www.aiucd.it/.
2. The exclusionary costs associated with attending the ADHO conference have been sharply criticized, particularly by scholars from the Global South (Earhart 2018).
3. In the particular case of the UK, Modern Language departments have seen the consequences of several government reforms, the latest a controversial one in 2006 deeming a second language optional in secondary school programs.

4. Taking Catalan Studies and the Catalan language as examples of field of study and minoritized language, international dissemination of this output will very likely depend on members of academic organizations such as *The Anglo-Catalan Society* (ACS), a network connecting academics in the British Isles and Ireland, and the *North American Catalan Society* (NACS) to connect with the other side of the Atlantic.
5. These approximate figures are based on online quote inquiries made to two academic translation agencies: AJE (https://www.aje.com/) and Transistent (https://www.transistent.com/).
6. The same academic translation companies mentioned above provide premium packages (at an additional cost) that cover a second translation and review of subsequent versions of the submitted manuscript.
7. Riva Quiroga's talk "Multilingual Publishing in Digital Humanities," as part of the virtual seminar series *The Digital Humanities Long View* (co-hosted by UCL Centre for Digital Humanities and the Center for Spatial and Textual Analysis at Stanford University) was written up in the *Stanford Daily*.
8. Evans's tweet reads: "206 articles were sent in to the Journal of Contemporary History in 2020, 39 were accepted (17%). In 2019 the figure was 136 and 31 accepted (26%). I don't think we rejected any that we should have accepted. Main reason for rejection: poor English!" (https://twitter.com/RichardEvans36/status/1358831507197018113?s=20). As part of the same conversation, and referring to the possibility of accepting work in languages other than English, his reply reads: "The JCH has only published in English since its foundation (by two non-native speakers) half a century ago. It would be folly to do anything else"(https://twitter.com/RichardEvans36/status/1359187632580927494?s=20).
9. RTL stands for Right-to-Left written, read and processed languages, i.e., Arabic, as opposed to Left to Right, i.e., English.
10. Such as "DLCL 204: Digital Humanities across Borders" at Stanford University, where each student brings their own non-English language to work with. Molly des Jardin's "East Asian DH" course at UPenn and the summer 2021 UPenn Dream Lab workshop on "East Asian Studies & Digital Humanities" are more similar to discipline-scoped DH classes than language-scoped ones, though the impact of language on method is discussed in both.
11. Term Frequency-Inverse Document Frequency.
12. A lemmatizer is a tool from Natural Language Processing that identifies the lemma in a word via full morphological analysis.
13. POS or part-of-speech tagging is a process in corpus linguistics that marks words up in a corpus as corresponding to a particular part of speech, taking into account its context and meaning.
14. NER or named-entity recognition is a method of recognizing and classifying text into predefined categories (i.e., person names, locations, etc.) from within unstructured text-based data.

REFERENCES

Bender, Emily and Batya Friedman. 2018. "Data Statements for Natural Language Processing." *Transactions of the Association for Computational Linguistics* 6: 587–604.

Berra, Aurélien, Emmanuel Château-Dutier, Sébastien Poublanc, Émilien Ruiz, Nicolas Thély, and Emmanuelle Morlock. 2020. "Editorial. Donner à lire les humanités numériques francophones." *Humanités numériques* 2: 2–5.

Calvo Tello, José. 2016. "Entendiendo Delta desde las Humanidades," *Caracteres*, May 27. http://revistacaracteres.net/revista/vol5n1mayo2016/entendiendo-delta/.

Caswell, Isaac, Julia Kreutzer, Lisa Wang, Ahsan Wahab, Daan van Esch, Nasanbayar Ulzii-Orshikh, Allahsera Tapo et al. 2021. "Quality at a Glance: An Audit of Web-crawled Multilingual Datasets." *arXiv*, October 25. https://arxiv.org/abs/2103.12028.

Dacos, Marin. 2011. "Manifesto for the Digital Humanities." *Hypotheses*, March 26. https://tcp.hypotheses.org/411.

Dombrowski, Quinn, Tassie Gniady, and David Kloster. 2019. "Introduction to Jupyter Notebooks," *Programming Historian* 8. https://doi.org/10.46430/phen0087.

Earhart, Amy. 2015. *Traces of the Old, Uses of the New: The Emergence of Digital Literary Studies*. Michigan, MN: University of Michigan Press.
Earhart, Amy. 2016. "Digital Humanities Futures: Conflict, Power, and Public Knowledge." *Digital Studies/Le champ numérique* 6, no. 1.
Earhart, Amy. 2018. "Digital Humanities within a Global Context: Creating Borderlands of Localized Expression." *Fudan Journal of the Humanities and Social Sciences* 11 (September): 357–69.
Estill, Laura. 2019. "Digital Humanities' Shakespeare Problem." *Humanities* 8 (1): 45.
Fiormonte, Domenico. 2014a. "Dreaming of Multiculturalism at DH2014." *Infolet: Cultura e critica dei media digitali*, 6. https://infolet.it/2014/07/07/dreaming-of-multiculturalism-at-dh2014/.
Fiormonte, Domenico. 2014b. "Humanidades Digitales y diversidad cultural." *Infolet: Cultura e critica dei media digitali*, 6. https://infolet.it/2014/02/01/humanidades-digitales-y-diversidad-cultural/.
Fiormonte, Domenico. 2015. "Towards monocultural (digital) Humanities?" *Infolet: Cultura e critica dei media digitali*, 7. https://infolet.it/2015/07/12/monocultural-humanities/.
Fiormonte, Domenico. 2016. "Towards a Cultural Critique of the Digital Humanities." *Historical Social Research/Historische Sozialforschung* 37 (3): 59–76. http://www.jstor.org/stable/41636597.
Ghorbaninejad, Masoud, Nathan Gibson, David Joseph Wrisley. Forthcoming [2022]. *Debates in the Digital Humanities*. Minneapolis, MN: University of Minnesota Press.
Hall, Crystal. 2020 "Digital Humanities and Italian Studies: Intersections and Oppositions." *Italian Culture* 37 (2): 97–115. doi: 10.1080/01614622.2019.1717754.
McDonough, Katherine, Ludovic Moncla, and Matje van de Camp. 2019. "Named Entity Recognition Goes to Old Regime France: Geographic Text Analysis for Early Modern French Corpora." *International Journal of Geographical Information Science* 33 (12): 2498–522.
Noble, Safiya Umoja. 2018. *Algorithms of Oppression*. New York, NY: New York University Press.
O'Sullivan, James. 2020. "The Digital Humanities in Ireland." *Digital Studies/Le champ numérique* 10 (1).
Risam, Roopika. 2019. *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy*. Evanston, IL: Northwestern University Press.
Terras, Melissa and Julianne Nyhan. 2016. "Father Busa's Female Punch Card Operatives." In *Debates in DH 2015*, edited by Matthew K. Gold and Lauren F. Klein, 60–5. Minneapolis, MN: University of Minnesota Press.
UCLA 2008. "A Digital Humanities Manifesto." December 15.
Wang, Xiaoguang, Xu Tan, and Huinan Li. 2020. "The Evolution of Digital Humanities in China." *Library Trends* 69 (1) (summer): 7–29.