Digital approaches to translation history

Judy Wakabayashi
Kent State University
jwakabay@kent.edu

Abstract: Digital translation history is defined here as a methodological approach that uses digital technologies to produce, enhance or disseminate research on translation history. This can help translation historians pose fresh questions and answer new and old ones. It entails mastering technical competencies in varying degrees while remaining grounded in the fundamentals of the historian's craft. This paper outlines the main affordances of digital approaches as applied to the study of translation history (how these can help translation historians do things better and/or differently in some respects), as well as the limitations. It introduces relevant techniques of text analysis (such as distant reading, topic modelling and stylometrics) and data visualization, which can help tease out patterns and relationships (e.g. textual, conceptual, geographic and personal networks) in dynamic ways that potentially create new knowledge and facilitate public engagement with scholarship.

Keywords: digital translation history, digital humanities, historiography

1. Relevance of digital humanities for translation historians

The digital humanities (DH) investigate traditional humanities questions and questions made newly possible by applying computing tools and techniques to digitized and born-digital materials.[1] Although translation historians make wide use of digital media to facilitate and enhance conventional research (e.g. information retrieval and management; software for presenting and disseminating research), many have not fully explored how information technologies can help pose and/or answer research questions that might otherwise be difficult to even envisage.

DH methods have been applied to both historical and textual studies, which suggests their relevance to studying the history of translated texts. Digital translation history is defined here as a methodological approach that uses digital technologies to produce, enhance or disseminate research on translation history, including the study of digitized texts, born-digital texts, and other digital artefacts (e.g. images, audio) relevant to translation history. The goals include:

- supporting *conventional* research agendas, by saving time and effort and allowing more thorough and extensive investigations,
- *revising* previous assumptions and findings on the basis of new and more data and newly revealed patterns and connections,
- generating *unanticipated* research questions and facilitating new kinds of research and new presentation modes,
- facilitating teamwork and public engagement.

[1] Born-digital texts are "authored to use affordances of screen-based interactions and new media technologies and are neither digitizations of print-based materials nor reproducible in print forms" (Eyman & Ball, 2015, p. 65).

Examples of databases relevant to translation history include the Perso-Indica database of Persian works on Indian learned traditions, which identifies the proportion of translations in relation to original works in India between the thirteenth and nineteenth centuries[2]; the Renaissance Cultural Crossroads project,[3] which has served as a basis for research by Barker and Hosington (2013) and others; the French Book Trade in Enlightenment Europe database[4] of book trade-based cultural transfers in late eighteenth-century francophone Europe; and the TETRA (*Teatro e Tradução*) project, which focuses on the history of theatre translation in Portugal (1800 – 2009).[5] Databases are not, however, the only useful tool for translation historians, as outlined later.

Digital translation history requires not only the skills and insights of any historian (including source evaluation and comparison, contextualization, critical interpretation, the imagination to envisage new questions and approaches), but also those of a data analyst. It involves considering who created the digital material, for what purpose, when, what was excluded, whether the digital source is "a coherent body of materials" since its origin or an assembly from diverse sources (Cohen & Rosenzweig, 2006, p. 25), and whether non-digital materials were altered during digitization, possibly without readers being notified. Although digitized texts might seem to be a textual and visual facsimile, they are often decontextualized. Hence Weller (2013, p. 7) stresses the importance of noting "the original experience, the original medium", particularly because material is often shifted from medium to medium nowadays.

So how can translation historians use DH to complement non-computational methods in historiographically valid ways? What are the advantages, implications and potential pitfalls of a partial shift from documents to data (or documents as data)?

2. Advantages and potential

Digital media allow us to do history *better* in several respects:

Capacity and *comprehensiveness*: digital media make more of the historical record available because of the low costs in saving it, while massive data sets allow more extensive investigations than relying on random or 'representative' cases.

Accessibility: "once the initial expenses are met, reaching an additional person costs almost nothing" (Cohen & Rosenzweig, 2006, p. 4). Moreover, digitalization can include "details that are otherwise unavailable, forgotten, ignored, or impossible to extract" (Jockers, 2013, p. 27) and can "conserve fragile/precious objects while presenting surrogates in more accessible forms" (Deegan & Tanner, 2002, p. 32).

Time-saving: "text-mining methods allow us to direct our scarce attention to those materials in which we already have reason to believe we will find relevant information" (Wilkens, 2012, p. 255).

Flexibility: digital media can handle sounds, images and moving pictures, opening up translation and interpreting history beyond the textual medium.

Diversity: digital media enable more public engagement, allowing "experts and users alike to comment on original source material" (Terras, 2012, p. 49). Digital historians can also do things *differently* because of the following features:

Manipulability: electronic tools allow searches not otherwise (readily) possible, particularly across documents. Nevertheless, Jockers (2013, p. 9)

---

[2] http://www.perso-indica.net/index.faces
[3] https://www.hrionline.ac.uk/rcc/
[4] http://fbtee.uws.edu.au/main/
[5] http://tetra.letras.ulisboa.pt/tetra/en

argues that "the sheer amount of data now available makes search ineffectual as a means of evidence gathering. […] What are required are methods for aggregating and making sense out of both the nuggets and the tailings." DH also provides tools for this.

Interactivity: user-generated content is a feature of Web 2.0 interfaces, which facilitate "multiple forms of historical dialogue – among professionals, between professionals and nonprofessionals, between teachers and students, among students, among people reminiscing about the past" (Cohen & Rosenzweig, 2006, p. 6).

Hypertextuality: this allows non-linear movement through data or narratives. Hyperlinks to other texts can enhance digitally published translations. Calhoun (2017, p. 139) says "a digital edition might incorporate supplementary material such as definitions, textual variants, and bibliographic references, at a hypertext level; render primary sources searchable for specific tokens and metalanguage; or enable users to define, isolate, and then save sub-corpora."

Time analyses: Robertson and Mullen (2017, p. 20) observe that "computing affords a view of the *longue durée* otherwise obscured by individual examples", with the potential to problematize existing periodizations. Time- and date-stamping of digitally created documents allows "a new form of temporaneous comparison and analysis" (Weller, 2013, p. 8).

Challenge to canonicity: by minimizing bias in text selection, digital approaches can supplement, even undermine, existing canons, which can be somewhat arbitrary and self-perpetuating.

Identification of the typical, anomalous, (dis)continuities and clusters: shifting away from the canonical allows greater focus on the 'mundane' translations that constitute the bulk of translation history. Software can help identify the typical *and* the exceptional, cluster items into categories, and reduce big data to a small dataset that represents the corpus more comprehensively than standard sampling. If something of interest appears in the smaller set, the computer can retrieve similar items. Researchers can go back and forth between the two sets, "experimenting with new categories and groupings" (Manovich, 2012, p. 469).

Patterns: digital corpora can reveal systematicity – e.g. through corpus approaches focusing on keywords (e.g. their different frequency from in other corpora) and collocations.[6] These patterns might not otherwise be apparent or sufficiently delineated. Interpreting their significance, however, requires human judgment.

Repurposing: with little time or effort, datasets can be "adapted, supplemented and transformed" (Mussell, 2013, p. 87) or placed in new contexts that can reveal "unexpected properties and relationships" (p. 90). Metadata also offer a source for mining (although translations and translators are not always assigned a field in databases).

Virtual unification: digital collections can bring scattered sources together.

3. New concepts of 'text', 'author' and 'language'

Digital media have expanded textual notions to include multimedia forms that differ in some respects from oral, manuscript and print texts. Websites, wikis, blogs, email and tweets are subject to translation and can constitute historical sources (sometimes with untraceable authors). Many sources are already available *only* in digital form. This requires rethinking our concept of archives,

---

[6] McEnery and Baker (2016, p. 4) note, however, that corpora "used to explore the past … are typically small" – a problem when examining low- or moderate-frequency words.

the connection between medium and knowledge production, and how we preserve, access and interpret these artefacts.

Digital texts are also affecting models of authorship and readership. Web tools facilitate collaborative writing, so the meaning of *author* is changing. The fact that "all digital work can be easily manipulated and remixed" undermines textual authority (Eyman, 2015, p. 72). Readers can also "customize the presentation of data to isolate issues of particular interest to them, rather than depending on the author" (Theibault, 2013, p. 180).

The growing perception of programming languages as *language* and of programming as *writing* acknowledges source code as a semiotic system with its own stylistic elegance and as a signifying cultural object. This arguably places computer programs within the purview of translation research, particularly in terms of intersemiotic translation. The field of Critical Code Studies applies literary analysis methods to computer code, and this can be done within a historical context. Although studying the history of translation between programming languages or between natural and computer language lies beyond the interests and expertise of most translation historians, these possibilities suggest how digital media broaden our object of study.

4. Building digital resources

Although the consensus seems to be that designing or building digital archives, tools or methods – not just digitizing material, but knowing how to code – is not necessary for qualifying as a digital humanist, "sensitivity to the capacities and possibilities of working in a digital environment" is essential (What is Digital Humanities, 2012).

The verbal, visual and structural design of resources can affect their argument and use. If one is creating a website, for example, it is essential to decide on its main purpose – "to share knowledge, to educate the public, to appeal to donors, to connect to a wider research community, etc." (Potts, 2015, p. 259) –, its audience (e.g. translation historians, interdisciplinary researchers, the public), and whether to take a hands-off approach, interpret the materials, or mix archival materials with interpretive essays. Sample features for a translation history website include biographical sketches, oral histories (audio- or videotaped interviews, with or without transcripts), primary documents (preferably searchable both within and across texts), background essays, historic photographs, zoomable and pannable maps, a bibliography, links to relevant websites, and a glossary.[7] Cohen and Rosenzweig caution, however, that

> topical sites … sometimes lack focus and wind up being a hodgepodge of materials centered on a particular theme. Often, it makes more sense to try to excel at one thing – at providing access to a rich archive, offering an intriguing interpretive exhibit, or supplying effective classroom tools or resources. (2006, pp. 49-50).

Preparing critical editions is one approach. Boyle (2015, p. 134) suggests considering "as one corpus, the evolving relations between primary texts, secondary scholarship, and tertiary commentary". For instance, *The Quintilian Project*[8] aims to compile "all the English translations alongside secondary scholarship" regarding the classical Roman rhetorician, so as to offer "a unique vantage point from which to visualize how Quintilian is taken up over time, determine which passages are cited most frequently, and discover which translations instigate the most responses" (Boyle, 2015, p. 134). Boyle also mentions digital editions that emphasize contexts of text production and

---

[7] For a website on Iraqi warzone interpreters developed by some of my students as a class project, see http://www.translationhistory.com/iraqinterpreters/.
[8] http://caseyboyle.net/project/the-quintilian-project/

reception (p. 130) – e.g. by including the notebooks, manuscript fragments, prose essays, letters and journalistic articles of a translator or theorist from the past. With digitalized manuscripts, Calhoun (2017, p. 147) stresses the importance of quality images, faithful transcription, and the inclusion of annotations about "lineation, hand changes, scribal emendations and abbreviations" – details whose omission hinders access to "the underlying manuscript reality".

An example of digital tools designed to compare retranslations over time is the Version Variation Visualization project,[9] where researchers have built language-neutral tools for analysing parallel multi-translation corpora to "uncover patterns relating to different types of translation, historical periods and genetic relations and patterns relating to different sub-sets of segments" (Geng et al. 2015, p. 274).

5. Distant reading

An alternative to *creating* digital resources is to make more effective use of *existing* ones. Although the immersive reading long applied to print texts can be used with digital texts, the extensiveness of big data can offer a different, more comprehensive and representative picture. Franco Moretti (2005, 2013) advocates 'distant reading' of massive numbers of canonical *and* unexceptional texts, through text analysis methods such as word frequencies, sentiment analysis[10] (systematically identifying and classifying a writer's attitudes on a particular topic and comparing the results with norms identified in other texts; this can be used to trace attitudinal changes over time), topic modeling, pattern recognition, and visualization in the form of graphs, maps, trees and clouds. The focus is on quantitative breadth rather than qualitative, interpretive depth, but it is possible to drill down to more granular levels.

Although distant reading "can flatten the particularity and ambiguity of the objects and processes that literary critics often seek to capture" (Long, 2015, p. 289), it complements close reading that focuses on singularities, facilitating back-and-forth movement between the micro- and macro-scales.

Some reader-related websites of potential interest to translation historians include the Reading Experience Database (RED)[11] and The Archaeology of Reading in Early Modern Europe website (focusing on manuscript annotations).[12]

5.1 Text analysis tools
One place to start looking for useful software is DIRT (Digital Research Tools[13]), which helps with choosing a tool based on one's aims – e.g. annotation, collaboration, network analysis, publishing, statistical analysis, text cleaning or visualization.

Corpus researchers already use text analysis software, and a corpus-informed approach (e.g. concordancing; retrieving lexical clusters) can be applied to certain aspects of translation history, such as analysing translated works or paratexts, oral history transcripts, or "changes and constants in language and vocabulary use" (Hudson, 2000, p. 241).

Textual analysis packages have four broad functions (Hoffman & Waisanen, 2015):

---

[9] www.tinyurl.com/vvvex
[10] E.g. DICTION; www.dictionsoftware.com
[11] http://www.open.ac.uk/Arts/reading/UK/index.php
[12] http://archaeologyofreading.org/
[13] http://dirtdirectory.org/

"[G]enerate basic statistics about a text, such as word count, average sentence length, number of adjectives" (p. 171), to gauge lexical richness, frequent syntactical patterns and readability indexes. This allows "simple but substantiated generalizations" (p. 171) and comparisons of these features between source and target texts and also over time. Features such as frequency do not, however, necessarily correlate with (historical) significance. Nor does the absence of a term in surviving texts necessarily mean it was never used or that the *concept* was not in play (p. 172).

"[C]reate indexes and concordances", showing expressions in context (Hoffman & Waisanen, 2015, p. 170). This reveals usage patterns and, for instance, positive or negative valences of culturally or theoretically important conceptual words and how these have changed or spread over time and/or space.[14] The mass digitization of (mostly Western) books now under way – as well as newspaper databases, which are disproportionately prominent – offers rudimentary concordances, but these holdings are not representative of commercially available works or the works of interest to translation historians. Nor are they amenable to proper corpora searches such as those possible with specialized software (e.g. Antconc[15] or WordSmith Tools[16]).

Use preprogrammed or user-generated dictionary-based programs to indicate "how common or deviant a text's language is in comparison with other texts" (Hoffman & Waisanen, 2015, p. 176). These programs cannot, however, indicate "how the actual locations of various terms relate and link with other terms" (p. 177).

"[D]o cluster analyses […] to determine the most important concepts in a given text or group of texts and how they are related to each other" (pp. 170-171) – e.g. not just how terms tend to collocate linguistically but also how they are related conceptually, which might change over time. Pinpointing conceptual clusters could be particularly useful, for instance, in examining historical texts discussing translation theory. Automated semantic analysis can identify classes of comments in paratexts, revealing patterns in how translators have conceptualized the act of translating.

Topic modeling is a related technique for identifying recurring themes in a corpus (rather than searching for predetermined keywords).[17] A sample project might involve exploring (changes in) the preoccupations and discursive framework in a translation journal or theorist's writing over time.

Stylometrics software such as the Java Graphical Authorship Attribution Program could help ascribe translatorship of anonymous translations, based on translations of known provenance.[18] Hung, Bingenheimer and Wiles (2010) used a digital approach to show that 24 Buddhist sutras, traditionally attributed to different Chinese translators, were translated by the same translator or group of translators. Other uses include textual dating, verifying the authenticity of historical documents and examining the relationship (e.g. stylistic diversity over time) among different translations by the same translator or among translators from different periods, genders, locations, classes or educational backgrounds, or among translations of the works of the same author. The assumption is that a translator's stylistic habits remain detectable through the style of the different authors translated. As Jockers (2013, p. 63) points out, external factors (e.g. genre, register, age, ethnicity, nationality, time period) might "influence or even overpower the latent … signal". Research suggests that features such as articles,

---

[14] The Genealogies of Knowledge project at the University of Manchester (http://genealogiesofknowledge.net/genealogies-knowledge-corpus/) seeks to "explore the evolution and contestation of key political and scientific concepts as they have travelled across centuries, languages and cultures".

[15] http://www.laurenceanthony.net/software/antconc/

[16] http://www.lexically.net/wordsmith/

[17] E.g. Overview: https://blog.overviewdocs.com/; MALLET: http://mallet.cs.umass.edu/topics.php. See Da (2019, pp. 625-629) for a critique of topic modeling.

[18] https://evllabs.github.io/JGAAP/.

conjunctions and pronouns are most indicative of individual style (p. 64). Forsyth and Lam (2014) found that inter-translator discriminability was possible in their digital study of nineteenth-century French translations (i.e. the translators' 'handprints' were present, although less so than the authors'). Historians might use stylometrics to explore questions such as the nature of the differences between canonical and marginal translators, or whether women translators have historically been more likely to use sentence fragments, for instance, and how any such tendencies have changed over time.

Another relevant function is automatic extraction of places and names (people, organizations) through named entity recognition (NER). Place names identified in a corpus of texts about Translation Studies, for example, might trace the shifting 'balance of power' in the discipline, or personal names in translators' correspondence might point to social networks. Other useful tools are image-processing techniques and handwritten text recognition (HTR) technology that facilitate the reading of old documents. For instance, SMART-GS[19] is a tool for transcribing and studying digitized historical manuscripts (mainly Japanese).

6. Information visualization

Grossman (2015, p. 42) points out that "Ironically, an excess of information resists analysis and comprehension in much the same way a lack of it does." Data do not necessarily equate with knowledge and understanding. One aid here is data visualization, the intersemiotic 'translation' of statistical or other information into visual representations. Beyond merely displaying findings more efficiently than in print, it can help tease out patterns and relationships in ways that create new knowledge and facilitate public engagement.

Historians have long made use of tables, graphs, dynastic and genealogical charts, timelines, maps and cartograms, but less static possibilities are now available, such as animated maps or interactive timelines. Visualization packages include Wordle, Many Eyes and Phrase Net, but simple word-cloud tools can lead to erroneous conclusions. As noted above, frequency does not always equate with significance, and word length and the space around words can distort relative importance. Other possible problems with visualization software include unclear legends, "false visual cues" and "unnecessary clutter and contrived images that [make] visualizations confusing" (Theibault, 2013, p. 177). Ironically, complex visualizations can require textual explanations and argumentation for historians lacking visual literacy.

Google's Ngram Viewer is a search engine that helps chart the trajectory of words and phrases in Google's text corpora (8 languages) between 1500 and 2008. It could be used, for instance, to trace the changing interest value of particular translators or theorists. However, "The only metadata provided are publication dates, and even these are frequently incorrect. Different printings, different editions, and the unaccounted-for presence of duplicate works in the corpus complicate matters even further." (Jockers, 2013, p. 120). Jockers concludes that Ngram Viewer

> cannot tell us *why* a particular word was popular or not; it cannot address the historical *meaning* of the word at the time it was used …, and it cannot offer very much at all in terms of *how* readers might have perceived the use of the word." (p. 122)

Moreover, the corpus changes over time; there is no way to find "words near other words" or search for synonyms; and the interface is poor (Shea, 2014, para. 39).

---

[19] http://en.sourceforge.jp/projects/smart-gs

One alternative is Bookworm,[20] which "makes it easy to turn any collection of texts into a richly searchable database; you can visualize trends, but with many more ways to slice data than Ngram Viewer allows" (2014, para. 42). Although word frequency-based conclusions about themes or significance are open to error [21] and frequency results do not explain underlying causal mechanisms, they might challenge existing ideas or narratives and trigger questions or hypotheses for follow-up by other means.

Another use of visualization software is to show historical networks – textual, conceptual, geographical and personal, as exemplified, for instance, through ties and communications among translators, authors and stakeholders such as publishers. Network analysis can be used to explore correlations between position within a network and "strategies of translation and selection", as in Long (2015). The *possibilities* are suggested by network analysis software such as Gephi[22] and sites such as *Mapping the Republic of Letters*,[23] while the *challenges* are noted by Theibault (2013, pp. 182-183) and Da (2019, pp. 630-631). Despite potential drawbacks, visualization tools help generate questions and test hypotheses (e.g. about centrality and marginality). The translation historian can then explore the underlying causes.

6.1 Spatial analysis

Visualization is particularly helpful with geographical data. Historical materials often contain location information, and historians have long paid attention to how space and place shape historical experiences and processes. Recent years have witnessed a focus on "themes of region, diaspora, colonial territory, and contact zones and rubrics such as 'border' and 'boundary'" (Bodenhamer, 2013, p. 24) – all relevant to translation history, as are questions of core and periphery. Maps support spatially embedded arguments and narratives, and computer-based spatial analysis helps historians formulate questions and identify patterns that textual sources alone might not readily suggest. Putnam (2016, p. 398) adds that "Visualizations of geotagged data can free us from reliance on predetermined spatial units" (e.g. nation-states).

Geographic Information Systems (GIS) software highlights aspects such as scale and proximity.[24] It "captures, stores, manages, displays, and analyses information linked to a location on earth. […] It also is an intelligent or interactive map that allows users to query the database and see the results visualized" (Bodenhamer, 2013, p. 25), including in terms of temporal change. GIS software integrates and interrelates not just quantitative data, but also textual, image, audio and other qualitative data that share a location. For instance, it would be possible to link population, publication and employment statistics, oral histories, videos, or images of historical texts and translators related to a particular site of translation. Information can be viewed separately or together and at different scales, and different layers can represent different themes.

An example of a text-to-map move would be georeferencing source text publications in a given language and the site of their translation in one or more target languages to highlight 'hot spots' or 'blank spaces'. Mapping could also be used, for instance, to identify patterns in translators' locations. Other tasks

---

[20] http://bookworm.culturomics.org/

[21] For instance, a search for "Lawrence Venuti" would miss references to "Venuti" and "Larry Venuti" (false negatives) or might include people with the same name who are *not* the translation theorist (false positives). See Da (2019, p. 605) for a critique of word frequency-based studies.

[22] https://gephi.org/

[23] http://republicofletters.stanford.edu/

[24] ESRI ArcGIS is the most widely used GIS software. It is expensive, but many universities have licenses. Free GIS software includes QGIS (https://qgis.org/en/site/). Sample mapping software includes eSpatial (https://www.espatial.com) and iMapBuilder (https://www.imapbuilder.com/). A helpful bibliography about historical GIS can be found at http://www.hgis.org.uk/bibliography.htm.

might involve creating a translation history layer for Google Earth or mapping translation theorists' institutional affiliations using Neatline.[25]

Bodenhamer (2013) presents several valid criticisms of GIS as a tool for historians. There is now a trend toward simpler mapping software, such as databases with mapping capabilities and the even simpler web mapping (Google Maps, etc.). Some governments make digitized maps available to researchers. Despite the drawbacks of spatial approaches, translation historians can benefit from giving greater consideration to spatial relationality. Although this concept underpins connected history, translation historians have been slow to explore digital tools that help to reveal such connections and construct spatial arguments.

7. Digital oral history

Oral histories can offer embodied, unmediated voices from people involved in recent translation history, thereby sharing authorship/authority in generating knowledge. Digital technologies can enhance oral history through improved recording and new engagement modes, such as allowing listeners to add their voices to online oral histories in an evolving 'conversation'. The Internet has opened up access to oral histories in terms of distribution, archiving and content management. Boyd and Larson note that

> Media outlets such as YouTube or SoundCloud offer near instant and free distribution of audio and video oral histories, while digital repository and content management systems like Omeka or CONTENTdm, or even Drupal or Wordpress, provide powerful infrastructure for housing oral histories in a digital archive or library. (2014, p. 4)

Although *creating* an online oral history database is a major undertaking,[26] translation-related searches of existing oral history repositories can prove beneficial.

Nevertheless, digital oral history raises issues such as the "increased vulnerability of narrators, infrastructure obsolescence, and a host of other ethical issues, particularly with heritage collections" (Boyd & Larson, 2014, p. 5), so it is important to balance availability with an ethical approach. Oral recordings are also difficult to search or navigate, so descriptive metadata in textual form are necessary. Boyd and Larson (2014, pp. 4-5) note that systems such as OHMS (Oral History Metadata Synchronizer) "enhance access to oral histories online, connecting a textual search of a transcript or an index to the correlating moment in the online audio or video interview." Transcription – an expensive process – raises issues such as whether to correct grammatical errors, which affects the reliability and unmediated nature of accounts. Preservation costs are another aspect.

8. Collaboration and publicly engaged scholarship

DH makes information more freely sharable and lends itself to participatory, multi-authored forms of knowledge production with other researchers and the public. Translation historians wishing to build digital resources will find it helpful, even essential, to collaborate with information sciences colleagues and can in turn contribute "qualitative and interpretive perspectives" (Grossman, 2012, para. 5), not to mention linguistic and area studies expertise. Digital translation history, particularly large data-driven studies, can benefit from

---

[25] http://neatline.org/about/

[26] A useful resource is the Oral History in the Digital Age website at http://ohda.matrix.msu.edu/

collaboration, since it is difficult for single researchers to 'cover' the relevant materials and skill sets.

Another focus of DH is scholarship that engages the public more, as well as more directly. This can help break down barriers between translation researchers, professional translators and the community by making research more relevant, personalized and accessible (e.g. blogs and podcasts). For instance, CommentPress, a WordPress plug-in, "allows users to read a document and comment on specific paragraphs, thus forming communities of discourse around discrete zones of text" (Liu, 2013). Potts (2015, p. 256) argues that rather than data-driven experiences, what is needed is more user-centred experiences and design. Even without direct interaction, DH projects typically encourage readers to interpret the information for themselves.

There is democratic potential in crowdsourcing (e.g. of text transcription[27]) and user-generated content. Online platforms for collaborative volunteer research (e.g. annotating and tagging documents for projects at Zooniverse.org) have similar potential. Wikipedia-like approaches can augment professionally written or archived sources. Davidson (2012, p. 480) suggests that "users might contribute information about the projects in which they are using the archive ..., or engage in theoretical debates in an open forum, or even contribute digitized content to the archive itself." Wikis offer an opportunity for dialogue between researchers and the (professional translation) community. Digital outreach projects can go beyond knowledge production and knowledge-sharing to collective activism, participating in broader cultural debates driven by a social purpose. Nevertheless, despite the potential of more publicly engaged scholarship, public participation in online translation history projects is likely to be low even among translators, and it might hinder innovative research that runs counter to accepted norms.

9. Limitations

Digital possibilities are seductive, but translation historians need to consider the following limitations and adopt an informed approach complemented by non-digital historical procedures and arguments. The "technical problems, logical fallacies, and conceptual flaws" in computational *literary* analysis – many of which are also relevant to computational historical analysis – are detailed in Da (2019).

Complexity: many meaningful aspects of translation history (e.g. causality) are too 'messy' for the quantitative approaches underpinning many (not all[28]) digital tools. Digital history also tends to rely on *homogenous* sources (Robertson & Mullen, 2017, p. 18), rather than the range of sources typically used by historians. Another challenge is the fluidity of categories over time. Country names and borders shift, and social changes mean that labels (e.g. socioeconomic labels) from one period might not reflect realities at other times. Although this fluidity also presents challenges in non-digital approaches, it makes it "difficult to insert any kind of authority control" into database fields (Crone & Halsey, 2013, p. 104).

Quality (and authenticity): all historians face questions of how and where to source reliable material, the completeness, accuracy and impartiality of sources, and how much constitutes an adequate sample. Apart from the possibility of digitally forged or manipulated documents, many digital materials do not exactly match the archival materials (e.g. in terms of selection, presentation or completeness) [29], and optical character recognition errors

---

[27] E.g. Scripto at http://scripto.org/.

[28] Information technology can handle not just quantitative data and structured textual information, but also unstructured texts such as books, web pages, sounds, and images.

[29] For instance, the physical properties of manuscripts and printed media – signifiers in their own

(particularly with older texts) or human input errors can lead to incorrect conclusions.[30] Borgman (2010, p. 217) concludes that page images (rather than digitized texts) are "better for comparing features of the original artefact". The archivist's selection of keywords can skew searches. Large datasets might be collected on an ad hoc basis and contain gaps and errors, often inherited from smaller datasets, but users might be unaware of this unless already knowledgeable about the topic. Nor might they realize how interpretive decisions – the selection (and exclusion), collation, structuring, and presentation of resources – shape their understanding or privilege particular ways of interacting with the materials (Crone & Halsey, 2013, p. 96).

Failure to exploit the potential: data collection and storage modes can limit the kinds of analysis possible, and some modes of online interaction can be rather passive or foster unnuanced responses (Cohen & Rosenzweig, 2006, p. 12). Users usually need to know in advance what they are looking for, and this must be describable in a search query, which is not always easy with the interpretative research typical of the humanities. Research questions need to be scaled appropriately, and the data needs to be organized using useful conceptual frameworks.

Durability: Terras (2012, p. 50) notes that digitalizing historical texts is "not a substitute for proper preservation" and might even "damage or compromise fragile or rare original materials". Moreover, there are challenges as to which aspects of the digital present to preserve for *future* translation historians. The ephemerality and sheer quantity of digital evidence (e.g. email correspondence between translators, authors and publishers) has implications for archiving born-digital material. Translators' successive drafts might not be available unless efforts are made to retain each electronic iteration. Similarly, online texts have multiple instantiations, so stable data capture becomes important. "Version control systems such as Git or Subversion trace changesets, or iterative development histories of live digital projects. All these forms (and many others) contain metadata that may be mined for research purposes." (Kennedy & Long, 2015, p. 142). There will be an ongoing need to recopy digital materials to new storage media and convert them into new formats to ensure continued accessibility. Another problem is link rot, so it is good practice to use permanent links.[31]

Culture blindness: since text production is in part a social process, cross-cultural differences are to be expected. Robertson and Mullen (2017, p. 20) point out that "Text analysis algorithms, for example, rely on cultural assumptions regarding language and its use that have repercussions for historical analysis." Anglo-American and European languages and cultures are over-represented in digitalized sources. Differences in access to technology in different parts of the world also risk perpetuating imbalances between scholars from the North and South.

Ethics: DH raises issues of privacy, cultural heritage, interpretive control and the right of representation. Relevant here are the Association of Internet Researchers 2012 guidelines on ethics and the 2006 Protocols for Native American Archival Materials, for example. It is possible to give varying levels of access to different groups (e.g. not allowing non-Aboriginals access to sensitive Aboriginal sources).[32]

---

right – are easily lost in digital versions unless precautions are taken (e.g. specifying the dimensions). Other facts might also be obscured (e.g. a book's borrowing history) or altered (e.g. how readers navigate through the work).

[30] Standardizing spellings before input affects source integrity. "If it becomes necessary to code or standardize in order to speed processing or create algorithms, this is added (rather than substituted for column fields) at a later stage." (Hudson, 2000, p. 231).

[31] For instance, see https://perma.cc/.

[32] The Mukurtu project (http://www.mukurtu.org) is a "platform built with indigenous communities to manage and share digital cultural heritage" (Sano-Franchini, 2015, p. 161). It

Other issues are that intellectual property gates hamper access, rights to reproduce material from archives and books are expensive, and books still in copyright cannot be subjected to large-scale data-driven investigation. Large datasets relevant to translation historians' concerns, particularly with 'minor' languages or cultures, might not exist, and research on social media sites (e.g. networks of translation activists) might face bans on "scraping" material. A lack of interoperability with other interfaces is another constraint on access.

In addition, computational history "tends to work on a scale that elides individual historical actors" (Robertson & Mullen, 2017, pp. 18-19). Conley et al. (2015) point out such "big-data pitfalls" as reverse causality (Y causing changes in X, rather than the expected direction of X causing a change in Y), unobserved heterogeneity (relevant variables that correlate with observed variables but are unobserved), sample-selection issues, aggregation bias (inappropriate extrapolation to a sub-group or individual from data aggregated for a group), or "spatial or temporal autocorrelation" (similarity between nearby observations as a function of spatial or temporal proximity). More fundamentally, DH risks a reductionist, positivist or uncritical approach with banal results. It is important to avoid fetishizing big data, which needs to be complemented by case studies and conventional sources. Lara Putnam (2016, p. 392) points out that digitized sources make it possible to bypass contextual browsing, which can lead to negative results. Adequate theorization is also essential if the data are not to seem trivial. Although digital approaches create new intellectual possibilities, they risk occluding others.

Digital translation history also presents practical challenges. One involves the necessary skills, although not all projects require advanced computing skills. Another is the sheer work involved in digitalizing and describing items in an existing collection or creating digital projects. Labour and infrastructure costs make DH challenging for researchers with little funding.

10. Closing thoughts

Digital resources and methods offer additional tools for exploring historical experiences of translation. Naturally, the tool must fit the purpose, and not all research projects or paradigms lend themselves to digital approaches. Nevertheless, in the early stages of any project it is worth considering such possibilities. If appropriate and implemented thoughtfully, DH can add a dimension to how we understand translation history. In addition, Gibbs and Owens (2013, p. 159) argue that

> [T]he new methods used to explore and interpret historical data demand a new level of methodological transparency in history writing. Examples include discussions of data queries, workflows with particular tools, and the production and interpretation of data visualizations. At a minimum, historians' research publications need to reflect new priorities that explicate the process of interfacing with, exploring, and then making sense of historical sources in a fundamentally digital form – that is, the hermeneutics of data. This may mean de-emphasizing narrative in favor of illustrating the rich complexities between an argument and the data that supports it. It may mean calling attention to productive failure – when a certain methodology or technique proved ineffective or had to be abandoned.

Although digital tools (no matter how carefully chosen) do not replace 'analogue' research or critical thinking, I hope this preliminary examination of the transformative potential of digital translation history will encourage further explorations. Ultimately, however, what is of interest is the *results* of research

---

uses cultural protocols that allow users to "define a range of access levels for digital heritage objects and collections".

enabled by these tools, rather than the platform or methodology or unsubstantiated promises.

References

Barker, S. K., & Hosington B. M. (Eds.). (2013). *Renaissance cultural crossroads: Translation, print and culture in Britain, 1473 – 1640*. Leiden and Boston: Brill.

Bodenhamer, D. J. (2013). The spatial humanities: Space, time and place in the new digital age. In T. Weller (Ed.), *History in the digital age* (pp. 23-38). London and New York: Routledge.

Borgman, C. L. (2010). *Scholarship in the digital age: Information, infrastructure, and the internet*. Cambridge, Mass.: MIT Press.

Boyd, D. A., & Larson, M. A. (Eds.). (2014). *Oral history and digital humanities: voice, access, and engagement*. New York: Palgrave Macmillan.

Boyle, C. (2015). Low fidelity in high definition: Speculations on rhetorical editions. In J. Ridolfo & W. Hart-Davidson (Eds.), *Rhetoric and the digital humanities* (pp. 127-139). Chicago and London: The University of Chicago Press.

Calhoun, D. (2017). What gets lost in the digital (re-)presentation of older linguistic texts? Digital editions, manuscript reality, and lessons from the digital humanities for the history of linguistics. *Beiträge Zur Geschichte Der Sprachwissenschaft, 27*(1), 137-166.

Cohen, D. J., & Rosenzweig, R. (2006). Digital history: A guide to gathering, preserving, and presenting the past on the Web. *History and technology, 23*(3), 316. https://doi.org/10.1080/07341510701396393

Conley, D., Aber, J. L., Brady, H., Cutter, S., Eckel, C., Entwisle, B., … Scholz, J. (2015). Big data. Big obstacles. *Chronicle of Higher Education*. Retrieved from http://m.chronicle.com/article/Big-Data-Big-Obstacles/151421

Crone, R., & Halsey, K. (2013). On collecting, cataloguing and collating the evidence of reading: The "RED movement" and its implications for digital scholarship. In T. Weller (Ed.), *History in the digital age* (pp. 95-110). London and New York: Routledge.

Da, N. Z. (2019). The computational case against computational literary studies. *Critical inquiry, 45*(3), 601-639.

Davidson, C. N. (2012). Humanities 2.0: Promise, perils, predictions. In M. K. Gold (Ed.), *Debates in the digital humanities* (pp. 476-489). Minneapolis and London: University of Minnesota Press.

Deegan, M. & Tanner, S. (2002). *Digital futures: Strategies for the information age*. New York: Neal-Schuman Publishers, Inc.

Eyman, D., & Ball, C. (2015). Digital humanities scholarship and electronic publication. In J. Ridolfo & W. Hart-Davidson (Eds.), *Rhetoric and the digital humanities* (pp. 65-79). Chicago and London: The University of Chicago Press.

Forsyth, R. S. & Lam, P. W. Y. (2014). Found in translation: To what extent is authorial discriminability preserved by translators? *Literary and Linguistic Computing, 29*(2), 199-217.

Geng, Z., Cheesman, T., Laramee, R. S., Flanagan, K., & Thiel, S. (2015). ShakerVis: Visual analysis of segment variation of German translations of Shakespeare's Othello. *Information Visualization, 14*(4), 273-288.

Gibbs, F. & Owens, T. (2013). The hermeneutics of data and historical writing. In J. Dougherty & K. Nawrotzki (Eds.), *Writing history in the digital age* (pp. 159-170). Ann Arbor: The University of Michigan Press.

Grossman, J. (2012). Big Data: An opportunity for historians? *Perspectives on History, 50*(3) (March). Retrieved from https://www.historians.org/publications-and-directories/perspectives-on-history/march-2012/big-data-an-opportunity-for-historians

Grossman, L. (2015). What's this all about? *Time Magazine*, July 6-13, p. 42.

Hoffman, D., & Waisanen, D. (2015). At the digital frontier of rhetoric studies: An overview of tools and methods for computer-aided textual analysis. In J. Ridolfo & W. Hart-Davidson (Eds.), *Rhetoric and the digital humanities* (pp. 169-183). Chicago and London: The University of Chicago Press.

Hudson, P. (2000). *History by numbers: An introduction to quantitative approaches*. London: Arnold.

Hung, J.-J., Bingenheimer, M., & Wiles, S. (2010). Quantitative evidence for a hypothesis regarding the attribution of early Buddhist translations. *Literary and Linguistic Computing, 25*(1), 119-134.

Jockers, M. L. (2013). *Macroanalysis: Digital methods and literary history*. Urbana, Chicago, and Springfield: University of Illinois Press.

Kennedy, K., & Long, S. (2015). The trees within the forest: Extracting, coding, and visualizing subjective data in authorship studies. In J. Ridolfo & W. Hart-Davidson (Eds.), *Rhetoric and the digital humanities* (pp. 140-151). Chicago and London: The University of Chicago Press.

Liu, A. (2013). From reading to social computing. *Literary studies in the digital age: An evolving anthology*. Online: Modern Language Association of America. https://doi.org/10.1632/lsda.2013.0

Long, H. (2015). Fog and steel: Mapping communities of literary translation in an information age. *Journal of Japanese Studies, 41*(2), 281-316.

Manovich, L. (2012). Trending: The promises and the challenges of big social data. In M. K. Gold (Ed.), *Debates in the digital humanities* (pp. 460-475). Minneapolis: University of Minneapolis Press.

McEnery, A., & Baker, H. (2016). *Corpus linguistics and 17th-century prostitution: Computational linguistics and history*. London: Bloomsbury Academic.

Moretti, F. (2005). *Graphs, maps, trees: Abstract models for literary history*. London and New York: Verso.

Moretti, F. (2013). *Distant reading*. London: Verso.

Mussell, J. (2013). Doing and making: History as digital practice. In T. Weller (Ed.), *History in the digital age* (pp. 79-94). London and New York: Routledge.

Potts, L. (2015). Archive Experiences: A vision for user-centered design in the digital humanities. In J. Ridolfo & W. Hart-Davidson (Eds.), *Rhetoric and the digital humanities* (pp. 255-263). Chicago and London: The University of Chicago Press.

Putnam, L. (2016). The transnational and the text-searchable: Digitized sources and the shadows they cast. *American Historical Review, 121*(2), 377-402. doi:10.1093/ahr/121.2.377

Robertson, S., & Mullen, L. (2017). *Digital history and argument*. Roy Rosenzweig Center for History and New Media. Retrieved from https://rrchnm.org/argument-white-paper

Sano-Franchini, J. (2015). Cultural rhetorics and the digital humanities: Toward cultural reflexivity in digital making. In J. Ridolfo & W. Hart-Davidson. (Eds.), *Rhetoric and the digital humanities* (pp. 49-64). Chicago and London: The University of Chicago Press.

Shea, C. (2014, January 13). Erez Aiden Contains Multitudes. *Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/Erez-Aiden-Contains-Multitudes/143871/

Terras, M. (2012). Digitization and digital resources in the humanities. In C. Warwick, M. Terras, & J. Nyhan (Eds.), *Digital humanities in practice* (pp. 46-70). London: Facet Publishing.

Theibault, J. (2013). Visualizations and historical arguments. In J. Dougherty & K. Nawrotzki (Eds.), *Writing history in the digital age* (pp. 173-185). Ann Arbor: The University of Michigan Press.

Weller, T. (2013). *History in the digital age*. London and New York: Routledge.

What is Digital Humanities? – A Symposium (2012). Retrieved February 27, 2016 from http://www.emsah.uq.edu.au/digitalhumanities

Wilkens, M. (2012). Canons, close reading, and the evolution of method. In M. K. Gold (Ed.), *Debates in the digital humanities* (pp. 249-258). Minneapolis and London: University of Minnesota Press.