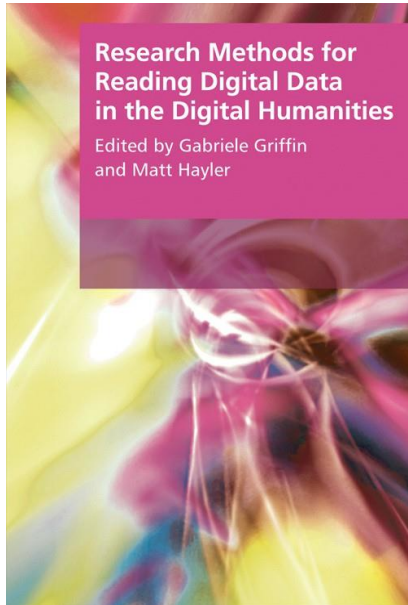# Research methods for reading Digital Data in the Digital Humanities

Edited by Gabriele Griffin and Matt Hayler

Edinburgh, Edinburgh University Press, 2016, pp. 214

Review by Valeria Franceschi[*]

This volume collects a number of essays dealing with different methodological issues pertaining to the multiple aspects of working with data from a digital humanities perspective. It explores widely different types of data and methods of analysis in different fields of study, highlighting the common questions and matters that may emerge when dealing with digital or digitized data. As per the editors, Gabriele Griffin and Matt Hayler, this collection "is aimed broadly at Humanities scholars, especially from the arts, literature and history, who wish to engage with digital humanities research methods but do not necessarily know much about such methods or how they might be put into practice" (introduction, 3). By focusing on methodology, authors present case studies rooted in digital humanities, highlighting the benefits of working with digital data to work with bigger datasets, obtain additional perspectives and means of interpreting data as well as the potential limitations and downsides of focusing on the digital perspective to the detriment of traditional methods.

The volume opens with a contribution by Hyler himself, "Matter Matters: The Effects of Materiality and the Move from Page to Screen." Through his work with the European research network EREAD (Evolution of Reading in the Age of Digitisation), the author illustrates the effects of the change from the physical object in print to a digital form, with digital texts being as meaningful as print ones despite their lack of materiality, but requiring new ways of reading: one example mentioned in the chapter is constituted by Kindle texts, which contain hyperlinks (dictionaries, Wikipedia, etc.), allowing the reader access to the full range of connotations and relations of words. The experience of reading the same novel changes when the print text is digitized.

Natalie M. Houston's contribution, "Reading the visual page in the Digital Archive," focuses on digitizing images and other things that are not transcribable in usual plain text. Houston starts from the idea that

---

[*] *Valeria Franceschi is currently a junior researcher in English Language and Translation at the Department of Foreign Languages and Literatures of the University of Verona. She has recently published* Exploring Plurilingualism in Fan Fiction: ELF Users as Creative Writers *(Cambridge Scholars Publishing, 2017). In addition to English as a Lingua Franca, her research areas of interest include digital communication and corpus linguistics.*

transcription into plain texts involves the loss of layers of meaning encoded in the visual layout of the page, including historical meaning, as the visual code can change across cultures and time. The author shows how technology can come to help: she and Neal Audenaert have developed a prototype software application called VisualPage, applied in this case to 19th century poetry, that demonstrates the possibilities for exploratory analysis of the visual codes of digitized printed materials (47), extracting quantitative measures for text density, line length, margin size, line spacing and line indentation.

In "Paratextual Navigation as a Research Method: Fan Fiction Archives and Reader Instructions," Maria Lindgren Leavenworth builds upon Genette's work on the context of paratext in fan fiction, illustrating how the notion of paratext can change in digital contexts, specifically in fan fiction archives. Her study of fan fiction archived on FanFiction.net shows that while certain elements can still be ascribed to Genette's delineation - fanfic filing and tagging, categorising and summaries, which create expectations in the reader – cross-referencing is here made possible through hyperlinks as well as paratextual communication with the reader. This occurs via Author's Notes (A/N), where the writer addressed the readers directly, sometimes in each published chapter, explaining the reasons for their narrative choices and their creative processes.

In "Data Mining and Word Frequency Analysis," Dawn Archer exemplifies some of the uses that can be made of Corpus Linguistics to carry out linguistic analysis, namely popular techniques such as frequency profiling, concordancing, collocations, n-grams, etc. (73), adopting both descriptive and significance statistical frequency analysis (74-75). In her essays, she presents different examples of these techniques, as the use of keywords and keyness in ediscovery, author distribution, and creation of the *other*; the use of uni-grams and bi-grams for authorship distribution; collocation extraction to reveal ideological uses of the language; She also mentions how a combination of corpus linguistics and psychology has been employed in a study to determine whether the language of psychopaths has something distinctive.

The following essay, "Reading Twitter: Combining qualitative and quantitative methods in the interpretation of twitter material" by Stefan Gelfgren, focuses on two case studies carried out on Twitter data in the field of digital religion. Both studies deal with the relationship between social media and the Church and how social media may be used to negotiate religious authority. The compilation of archives of two Twitter debates via the *twapperkeeper.com* and *tweetarchivist* services allowed for the visualization of networks, then analyzed with the Textometrica tool. These networks provide additional data than simple qualitative analysis of the texts, making it possible to "[detect and visualise] the relations between the different actors and their individual impact and influence within the conversation" (93). The author advocates then a dualistic approach, where quantitative and qualitative findings complement each other.

Coppélie Cocq's essay, "Reading small data in indigenous contexts: Ethical perspectives" focuses on the ethical implications of working with digital data. Working with Sámi languages and communities in a context of indigenous / minority studies, ethical concerns may emerge when engaging with online communities with few individuals that overlap with identified local groups (111). These may be related to increased access to data and subsequent issues of anonymity and confidentiality; difference between "right to access and the right to use, and between the right to use and the right to disseminate" (116). Cocq's highlights the importance of informed consent and data anonymization, even if that may limit the dissemination of results. In addition, she suggests that bringing people benefit should be a priority in this type of research. Putting people at the center of the research is also the main point of the following essay, "Knowing your crowd: an essential component to crowdsourcing research," in which Gabriel K. Wolfenstein deals with the complexities of doing successful crowdsourcing (126). The author agrees that crowdsourcing may be useful for research in the humanities, but he highlights, through the description of two case studies, the importance of knowing the community with which scholars want to engage, as well as what their priorities are. When adopting a crowdsourcing approach, scholars should be aware that establishing successful collaboration with a community is an ongoing project, and that understanding the needs of the community should be made a priority.

In "Fantasies of scientificity: Ethnographic identity and the use of QDA software," Anna Johansson and Anna Sofia Lundgren write about the importance of taking into account the "ideological underpinnings of software as well as their implications for research practice and results" (148). Their study focuses on the use of software for qualitative analysis, known as QDAS (Qualitative Data Analysis Software), which are usually either perceived as a solution to methodological issues in ethnography or rejected by ethnographers who prefer to employ traditional methods and processes. The authors draw "on the psychoanalytic notion of

'fantasy' to discuss ethnographic experiences of digital tools" (149) in order to shed light on how ethnographers perceive QDA software and the fantasies that are projected onto it as well as – and more than – the pros and cons of this type of software.

Robert Glenn Howard takes us back to the study of online communities and their networks in "Digital Network Analysis: Understanding Everyday Online Discourse Micro- and Macroscopically," where he investigated a website where parents skeptical of medicine gather and discuss their positions. His interest was specifically on the concept of vernacular authority, which is opposed to institutional authority. With the help of a programmer, software was created to download the posts in the forum and code them according to different fields; this allowed for the creation of visualizations using other network graphing software. This type of analysis can draw attention to the most and least influential members of the community, showing in this case that users engaging in anti-institutional discourse were considered authorities in the forum. The results provided by the software-based analysis show according to the author than computational, quantitative approaches may be necessary to make generalizations about the use of discourse online.

In the last essay in this collection, "Dealing with Big Data," authors Tobias Blanke and Andrew Prescott discuss the controversy around big data. Academia is enthusiastic about big data, thanks to the amount of data that was previously unavailable, and the tools that can be used to investigate such datasets. The controversial aspects lie around the potential replacement of sociological theories with mathematical models, the se of data for surveillance, and quantification, with statistical analysis unable to provide enough insights. However, the authors point out, "big data methodologies have more in common with traditional research preoccupations of the humanities than might at first be thought" (198). In addition to the broader big data perspective, it is also necessary to contextualize the data and perform a close reading of the single datum.

This collection accomplishes exactly what it set out to do; it offers scholars who are taking their first steps into the field of digital humanities a valuable overview of what it means to engage with digital humanities, providing a range of varied examples involving computing technology. The studies presented here illustrate successfully a variety of ways in which digital humanities can be exploited to contribute to data collection and interpretation, showing scholars taking their first steps in the field that engaging with such research methods, even in combination with more traditional methods, can highlight new patterns and provide additional insights.