



Artificial imagination, imagine: new developments in digital scholarly editing

Dirk Van Hulle¹

Published online: 25 April 2019
© Springer Nature Switzerland AG 2019

This special issue on Digital Scholarly Editing introduces several new developments, both in terms of the theory and practice of textual scholarship that are taking shape in the digital medium. But instead of enumerating or summarizing them, this introduction is meant as a reflection on some of the possible ways in which our work in digital scholarly editing could be useful to the broader field of digital humanities. After all, we tend to work on a microscale compared to the forms of macroanalysis and ‘distant’ reading that are currently dominant in digital literary studies. This raises the question: how can our research be relevant to these other sub-disciplines in digital humanities?

The title of this introduction is inspired by a short text by Samuel Beckett, called *Imagination Dead Imagine*, which can be read as a literary investigation into the workings of the human imagination. At first sight, the link with scholarly editing may seem far-fetched, but what many literary editorial projects have in common is a fascination with the creative process. After all, the human mind is at the core of ‘humanities’ research and scholarship or *Geisteswissenschaft*.

Within the framework of this bigger picture, the issues discussed in this volume present us not only with challenges but also with opportunities. If editions are “machines of knowledge” (see the contribution by Susan Schreibman and Costas Papadopoulos), and if they are “machines of simulation” (McGann 2014, 124; qtd. in the contribution by Julia Flanders, Ray Siemens et al.), this knowledge and simulation might be combined to simulate not just a *product*, such as a handwritten document—as in a digital facsimile (see the contribution by Mats Dahlström)—but also a *process*, such as the creative and imaginative process of a literary work.

✉ Dirk Van Hulle
dirk.vanhulle@uantwerpen.be

¹ Department of Literature, University of Antwerp, Antwerp, Belgium

Developments in AI and computational linguistics enable us to create writing bots that facilitate a writer's creative process, as a recent experiment in collaboration with the Dutch writer Ronald Giphart illustrates (see Manjavacas et al. 2017).¹ Giphart wrote a story, making use of 'Asibot', a writing bot that offered him the possibility to continue a sentence (adding a passage of ca. 100 characters) at any time in any one of eight different styles. These styles were based on the works of a few Dutch and Flemish writers (such as Gerard Reve and Kristien Hemmerechts), on the Dutch translations of Isaac Asimov's writings and, most interestingly, on the published works of Ronald Giphart himself. The programme also included keystroke logging, which enabled us to trace the entire creative process. Making use of this software, Giphart started writing and in the middle of his first sentence he activated the style of Gerard Reve: the bot offered a syntactically correct continuation consisting of 100 characters, after which Giphart finished the sentence. In the middle of the second sentence he activated his 'own' style, asking the bot to suggest a 100-character continuation in the Giphart style.

This is not the place to analyse the result of this particular writing experiment, but the point is that this basic form of artificial imagination is a form of *imitatio*. The bot is very good at imitating or simulating the style of a particular writer, based on his or her texts published so far. So long as the bot only offers recombinations of words used in the texts that have already been published he only serves as a supplier of words, comparable to, say, the notes in Joyce's *Finnegans Wake* notebooks, filled with verbal pillage, plundered from hundreds of source texts.

But what if we can teach the bot that writing is actually to a large extent *re-writing* and revising, not just a recombining of words the writer already used in his previous works, but a complex dialectic of composition and decomposition, writing, undoing and rephrasing? If the bot were able to simulate this *process*, it would be able not merely to imitate but to emulate the writer. This *aemulatio* would be a step in the direction of artificial imagination.

To start making this step, we need training data, which is what we are to some extent already producing in digital scholarly editions today. All the encoded transcriptions of a digital edition such as the Charles Harpur Critical Archive (discussed by Desmond Schmidt and Paul Eggert in this issue), the Charles Chesnutt Digital Archive (discussed by Stephanie P. Browner and Kenneth M. Price) or the Beckett Digital Manuscript Project (www.beckettarchive.org) – including all the tagged deletions, additions and substitutions – can be used as data to train the algorithm to simulate a particular writer's creative process. By means of state-of-the-art techniques of natural language processing and sentiment analysis it should be possible to analyse and visualise the differences in tone between versions of the same work, thanks to the detailed transcriptions of each separate version.

These encoded transcriptions already contain quite a bit of information that can help us detect patterns of textual change. Thus, the encoded transcripts provide information on deletions and substitutions. This kind of information can be modelled and charted in plots showing the percentages of added, deleted,

¹ The experiment was a collaboration between the Meertens Institute (KNAW, Amsterdam) and ACDC (the Antwerp Centre for Digital humanities and literary Criticism, University of Antwerp), involving Folgert Karsdorp, Mike Kestemont, Dirk Van Hulle, Enrique Manjavacas, Benjamin Burtenshaw, Vincent Neyt and Wouter Haverals.

modified and unchanged words. The result can serve as a tool to detect patterns in terms of an author's poetics. In the case of a writer such as Samuel Beckett, the overall pattern (www.beckettarchive.org/statistics) corresponds with the author's self-proclaimed poetics of "less is more". The statistics indeed show that Beckett cut more than he added, with a relatively stable ratio of one added word for every three deleted words on average – both on the level of the separate work and on the level of the oeuvre as a whole (Beckett 2018). We may dismiss this kind of indirect reading as merely confirming what we already knew, or too "unambitious"² in scope to be relevant. But we can also see it as a move, no matter how modest, in the direction of more complex (macro)analyses. If combined with other techniques such as part of speech tagging, sentiment analysis and computational semantics, experiments such as these and the ones discussed in the present volume do suggest new ways in which digital scholarly editing can contribute to forms of distant reading.

So far, distant reading is usually applied to one version of a text. What digital scholarly editing can offer is a way to enable distant reading *across versions*, which would be a necessary step in the development of artificial imagination in our discipline. Such a panoramic form of genetic reading enables readers to examine not only a work in progress, but also an oeuvre in progress, and as more and more digital genetic editions become available, possibly even literary periods in progress, including macroanalyses across versions.

As any scholarly editor knows, literary imagination is not only a matter of individual mental power, but often an interaction between an intelligent agent and his or her material and cultural environment. This includes a writer's interaction with her library, with her editor, with her friends, with social media, with her laptop, with old files, with websites, with her own earlier drafts and with the physical space of notebooks. Scholarly editors are in an optimal position to analyse especially the creative potential of the interaction with what in writing studies is called the 'text produced so far' (TPSF). If we manage to find suitable ways to digitally map this interaction, digital scholarly editions may serve as valuable sources of information providing training data for research into artificial imagination.

This special issue of *Digital Scholar*, therefore, investigates the state of the art in digital scholarly editing by raising questions such as: How should we frame concepts such as 'copy' and 'facsimile' in the age of digital reproduction? How does a digital scholarly edition differ from print editions? Can we further develop the notion of the hybrid edition? How do we conceive of the scholarly edition in 3D? Can we combine the notions of a digital archive and a digital edition? How can we reappraise textual collation in a digital paradigm? How do we adjust or rethink editorial theory to cope with born-digital works of literature?

According to Matthew Jockers, "we have reached a tipping point, an event horizon where enough text and literature have been encoded to both allow and, indeed, force us to ask an entirely new set of questions about literature and the literary record" (2013: 4). I believe that in digital scholarly editing we may not have reached that tipping point yet, and that it may still take a while before panoramic reading of entire periods in progress

² According to Franco Moretti, 'the ambition is now directly proportional to the distance from the text: the more ambitious the project, the greater must the distance be' (Moretti 2013, 48).

and macro-analyses across versions will be operative. But this is precisely why we need to keep investing in the genetic *microanalyses* of drafts, typescripts and other versions, marking up variants as a necessary step in the direction of artificial imagination and new ways of macroanalysis applied to more than one version.

References

- Beckett, S. (2018). In D. van Hulle, S. Weller, V. Neyt (Eds.), *Fin de partie / Endgame: A digital genetic edition*. Brussels: University Press Antwerp. Retrieved from www.beckettarchive.org. Accessed 26 June 2018.
- Jockers, M. L. (2013). *Macroanalysis: Digital methods and literary history*. Urbana/Chicago/Springfield: University of Illinois Press.
- Manjavacas, E., Karsdorp, F., Burtenshaw, B., Kestemont, M. (2017). Synthetic literature: Writing science fiction in a co-creative process. *Proceedings of the Workshop on Computational Creativity in Natural Language Generation (CC-NLG 2017)*, Santiago de Compostella, 4 September 2017. Association for Computational Linguistics, 2017, pp. 29–37. Retrieved from <http://aclweb.org/anthology/W17-3904>. Accessed 26 June 2018.
- McGann, J. (2014). *A new republic of letters: Memory and scholarship in the age of digital reproduction*. Cambridge: Harvard University Press.
- Moretti, F. (2013). *Distant reading*. London/New York: Verso.