

# Coding editions. Computational approaches to the editing of pre-modern texts.

**Cugliana, Elisa**

elisa.cugliana@uni-koeln.de

CCeH - Universität zu Köln, Deutschland

## Introduction

The endlessness of the digital space allows scholarly editors to conceive edition projects on a grand scale. Indeed, the eternal mission of philology can be identified in the quest for the best way to represent and/or reconstruct primary sources, in order to save them from oblivion and grant the public the most informed access to them. Consequently, the lack of space boundaries characterising the digital environment is one of the first aspects making such an environment perfectly suitable for hosting ambitious edition projects. Considering then the possibilities disclosed by hypertextuality, multimodality and multimediality, one can hardly argue against the digital way of editing, despite the ongoing challenges it must face (cf. for instance Rosselli Del Turco 2016), of which sustainability is clearly in the forefront. However, there are reasons to believe that state-of-the-art digital editions have not yet overcome the limits of a still rather bookish paradigm, obeying a mostly representational logic (Van Zundert 2018, Cugliana and Van Zundert 2022).

In this contribution, I will argue that the key to the next level of digital scholarly editing is to be found in computation, that is, in the actual coding of the whole editorial workflow. I will present the theory, the advantages and the challenges of such a computational approach to the editing of pre-modern texts, backing up my claims with examples from the praxis and from my own scholarly work in the field of digital philology. Of course, the field of computational editing is still in its infancy, which means that fully computational editions have not been published yet. However, there are projects<sup>1</sup> which can already prove the validity of the arguments presented in this paper or that at least seem to move in the same direction.

## Theoretical background

While the digital paradigm has been widely described and commented upon (among others by Stella 2007 and Sahle 2016), there is a crucial aspect that has not been sufficiently underlined, which however could represent a turning point in the history of (digital) philology. It is the case of the use of programming code throughout the

different phases of the editorial process, aiming at the realisation of what Barabucci and Fischer (2017) defined as the formalisation of textual criticism. The authors, in their conclusions, state that a “shared formalization would lead to the semi-automatization of the editorial process”, where “the responsibility of the editors would be to describe their choices and decisions”, while that of the computers would be to “deal with applying these rules and decisions in the best way”.

The act of formalising the competence of the editor is to be seen as an achievement *per se*, for it can lead to more accountability and verifiability of the scholarly processes, which can be shared in its complexity with the community. As a matter of fact, such processes are of fundamental importance not only from a methodological perspective, but also for the proper understanding and contextualisation of the presented results. In this respect, Van Zundert (2018) points out how the current digital methods used in scholarly editing, instead, still aim at merely recording the results of the philological work in a representational fashion,<sup>2</sup> without keeping track of the decisions, analyses and actions that led to those results. In a computational edition, on the other hand, the scholarly competence and actions are not only expressed in a formal language, but they are also operationalised. This, in turn, brings about new affordances, such as the replicability and reusability of the editorial work, as well as the introduction of a degree of indirection allowing for highly controlled modelling of the edition process.

If methods are there to realise theoretical principles in the best possible way, at the same time they can in turn influence the principles themselves, opening up new perspectives uncovering, for instance, implicit biases and illogicalities. Indeed, the use of computation for the modelling and operationalisation of editorial knowledge can contribute to perfecting the editorial workflow in the way envisioned by McCarty (2005), who referred to the “meaningful surprise” often arising from the process of modelling and computing.<sup>3</sup>

## The praxis of the computational method

Some examples from the actual application of this approach will hopefully prove its potential for the field of Digital Scholarly Editing.

## Achievements

During my PhD, which I completed in February 2022, I edited an Early High German version of Marco Polo’s travel account, also known as “Version D1” of the *Devisement dou Monde*. One of the main goals of my research was to automatise relevant parts of the editorial workflow, such as the normalisation of the texts. As a matter of fact, bearing in mind the concept of “text as a wheel” of different possible facets, as theorised by Sahle 2013, I wanted to produce (at least) three different levels of normalisation of the text (diplomatic, semi-diplomatic and interpretative).

Together with my colleague Gioele Barabucci (Norwegian University of Science and Technology), we developed a method based on a rule-and-exception principle, featuring three XProc pipelines, one for each level of normalisation (Cugliana and Barabucci 2022). Each pipeline consists of a series of XSLT stylesheets which deal with the different steps of the normalisation process, such as the levelling of allographs, the expansion of abbreviations, the regularisation of capitalisation etc.<sup>4</sup> In particular, the choice of using a pipeline system instead of a single stylesheet was due to the fact that some actions needed to be taken before other ones could follow: for instance, inserting a capital letter after a full stop implies that the full stops are already in place. Each pipeline generates a new TEI XML file with a different level of normalization.

This proved to be a very suitable strategy for dealing with the complexity of the normalisation process. Concerning the choice of XProc pipelines, it has already been shown that small-step pipelines making use of a stateless language such as XSLT prove to be advantageous in that they reduce the complexity of computer programs, they can be easily shared with the peers and they improve the sustainability of the code (Barabucci and Schaeben 2021). Not only were the pipelines successful in generating three levels of normalisation of the texts, but they could also be applied to the transcriptions of all the witnesses edited, despite the fact that some were written in the East-Swabian dialect, and some in Bavarian. This was probably due to the geographical proximity and the contemporaneity of their production, which leads to hypothesise the possibility of creating “pools of rules” for the editing of witnesses written in specific areas and historical periods.

In my talk I will present the system at the basis of the normalisation of the texts featured in the edition of DI, giving some insights into the very development of the pipelines, both from a strictly philological and from a computational perspective. In particular, I will focus on some tricky aspects such as the cases of ambiguities and exceptions, which might represent a hurdle for the full systematisation of the editorial workflow.

## Outlook

A case of reuse of the normalisation pipelines written for Version DI of the *Devisement dou Monde* is the project for a computational edition of the medieval German versions of the *sortes* text *Prenostica Socratis Basilei* (Alonso Guardo 2015). This project is still in its early stages,<sup>5</sup> but a taste of its complexity will be given in the talk. With a prototypical example I will show how the same or slightly adapted XSLT stylesheets can be used to normalise texts different from the ones which had served as a basis for their development, as well as how they can be embedded in the editorial workflow for the creation of a fully computational edition. This will open up the discussion on the reuse of program code for different editorial projects (which parts of the code can be reused, which need to be adapted, what rules can apply to different projects and what is instead determined by each individual edition).

*Sortes* texts (in German “Losbücher”) represent an extremely interesting genre, which unfortunately has not

enjoyed much scholarly attention in the field of digital editing, although it has a long history dating back to antiquity.<sup>6</sup> They were very interactive texts which were not meant to be read linearly (Heiles 2018) and which were used to predict the future with the help of sortition mechanisms, sometimes quite extravagant ones. The computational edition of the medieval German versions of the *Prenostica Socratis Basilei* aims at establishing a new methodology for the editing of this genre, finding strategies to weave together games and textual criticism, with the goal of representing the *sortes* in their genesis and reception throughout the centuries. Specifically, the whole editing process will find expression in an actionable formal language.

## Conclusions

As can be evinced from this abstract, my contribution has a twofold purpose: on the one hand giving a short, but hopefully convincing introduction to the theoretical aspects justifying the scholarly value of a computational approach to editing and, on the other, presenting some meaningful results obtained in the praxis. In particular, the focus will mainly be on the success and challenges of the normalisation pipelines developed in collaboration with Gioele Barabucci, showing what it actually meant to translate into code the usually very analogue task of normalising medieval texts, what problems we encountered and how we solved them. Finally, as an outlook for the possible applications of the computational principle, I will briefly illustrate the work I am doing for my next project, the edition of the *Prenostica Socratis Basilei*, which was conceived from the start as a computational one. In my talk, I will show some prototypical examples concerning specific aspects of the computational workflow.

## Fußnoten

1. An example of the application of an integrated computational approach for different phases of the editorial process is the digital critical edition of the Chronicle of Matthew of Edessa (Andrews, Safaryan and Atayan 2019).
2. Although there seems to be a need for further discussions also as far as the “edition as interface” is concerned. In this respect, Andrews and Van Zundert (2018) argue that the interface is “not just an argument about the text, but also an argument about the ‘attitude’ of the editor, a window into his or her take on methodology and the digital edition itself. It is also a revelation of the technical skills available to the editor. The interface tells us something not only about the methodology but also about the import of the edition”. The authors observe, however, that many editors still underestimate the role of the interface, whose “language” is one of great complexity.
3. “Take, for example, knowledge one might have of a particular thematic concentration in a deeply familiar work of literature. In modelling one begins by privileging this knowledge, however wrong it might later turn out to be, then building a computational representation of

it, e.g. by specifying a structured vocabulary of word-forms in a text-analysis tool. In the initial stages of use, this model would be almost certain to reveal trivial errors of omission and commission. Gradually, however, through perfective iteration trivial error is replaced by meaningful surprise" (McCarty 2005).

4. The pipelines are available online on the pre-release website of the DI edition: <https://marcopolo.c-keh.uni-koeln.de>.

5. In February 2022 it received an "Initial Funding" from the Faculty of Arts and Humanities at the University of Cologne (<https://phil-fak.uni-koeln.de/en/research/research-funding/initial-funding>).

6. The roots of the *sortes* have been traced down to the Mediterranean area (Abraham 1968), but their use soon spread all over Europe. Since the Renaissance these texts have been used as entertaining games to play in a group (Urbini 2006), but this was not the case in antiquity and for the greater part of the Middle Ages. Then, they were taken seriously and could have a strong influence on people's decisions.

## Bibliographie

**Abraham, Werner.** 1968. "Studien zu einem Wahrsage-text des späten Mittelalters." *Hessische Blätter für Volkskunde* 59, 9-24.

**Alonso Guardo, Alberto.** 2015. *Prenostica Socratis Basilei: étude, édition critique et traduction*. Textes littéraires du Moyen âge 4. Paris: Classiques Garnier.

**Andrews, Tara L. and Joris J. Van Zundert.** 2018. "What Are You Trying to Say? The Interface as an Integral Element of Argument." In *Digital Scholarly Editions as Interfaces*, edited by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, Gerlinde Schneider, 3-33. Norderstedt: BoD.

**Andrews, Tara Lee, Anahit Safaryan, and Tatevik Atayan.** 2019. "Continuous Integration Systems for Critical Edition: The Chronicle of Matthew of Edessa". *Data-verseNL*. <https://doi.org/10.34894/MZ7FBI>.

**Barabucci, Gioele and Franz Fischer.** 2017. "The Formalization of Textual Criticism. Bridging the Gap between Automated Collation and Edited Critical Texts." In *Advances in Digital Scholarly Editing: Papers Presented at the Dixit Conferences in the Hague, Cologne, and Antwerp*, edited by Peter Boot, Anna Cappellotto, Wout Dillen, Franz Fischer, Aodhán Kelly, Elena Spadini, and Dirk Van Hulle, 47-53. Leiden: Sidestone Press.

**Schaeben, Marcel, and Gioele Barabucci.** 2021. 'Small-Step Pipelines Reduce the Complexity of XSLT/XPath Programs'. In *Proceedings of the 21st ACM Symposium on Document Engineering*, 1-4. Limerick Ireland: ACM. <https://doi.org/10.1145/3469096.3474922>.

**Cugliana, Elisa and Gioele Barabucci.** 2022. "Signs of the Times: Medieval Punctuation, Diplomatic Encoding and Rendition." *Journal of the Text Encoding Initiative* 14. DOI: <https://doi.org/10.4000/jtei.3715>.

**Cugliana, Elisa and Joris J. Van Zundert.** 2022. "A Computational Turn in Digital Philology". *Filologia Germanica / Germanic Philology* 14, 43-72.

**Heiles, Marco.** 2018. *Das Losbuch. Manuskriptologie Einer Textsorte Des 14. Bis 16. Jahrhunderts*. Köln: Böhlau Verlag.

**McCarty, Willard.** 2005. *Humanities Computing*. Basingstoke, New York: Palgrave Macmillan.

**Rosselli Del Turco, Roberto.** 2016. "The Battle We Forgot to Fight: Should We Make a Case for Digital Editions?" In *Digital Scholarly Editing: Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, 219-238. Cambridge: Open Book Publishers.

**Sahle, Patrick.** 2013. *Digitale Editionsformen*. 3 Voll. Schriften des Instituts für Dokumentologie und Editorik. Norderstedt: BoD.

Sahle, Patrick. 2016. "What is a scholarly digital edition (SDE)?" In *Digital Scholarly Editing. Theory, Practice and Future Perspectives*, edited by Matthew Driscoll and Elena Pierazzo 19-39. Cambridge: Open Book Publishers.

**Stella, Francesco V.** 2007. "Metodi e prospettive dell'edizione digitale di testi mediolatini." *Filologia mediolatina* XIV, 149-180.

**Urbini, Silvia.** 2006. *Il Libro delle sorti*. Modena: Franco Cosimo Panini.

**Van Zundert Joris J.** 2018. "Why the Compact Disc Was Not a Revolution and Cityfish Will Change Textual Scholarship, or What Is a Computational Edition?" *Ecdotica*, 129-56. <https://doi.org/10.7385/99283>.